

COVID-19 FORECAST USING MACHINE LEARNING MODELS

Pavitra Maheshwari^{*1}, Tushar Agarwal^{*2}, Snehil Raj^{*3}, Ankit Shukla^{*4}

¹Mukesh Patel School of Technology Management and Engineering, NMIMS University-Shirpur, 425 405, India

²Mukesh Patel School of Technology Management and Engineering, NMIMS University-Shirpur, 425 405, India

³Mukesh Patel School of Technology Management and Engineering, NMIMS University-Shirpur, 425 405, India

⁴Mukesh Patel School of Technology Management and Engineering, NMIMS University-Shirpur, 425 405, India

Abstract Corona Virus unwellness 2019 (COVID-19) could be an illness caused by Severe Acute metabolism Syndrome Corona Virus 2 (SARS-CoV-2) and was initial diagnosed in China in Dec, 2019. Dr. Tedros Adhanom Ghebreyesus, World Health Organization (WHO) director-general on March eleventh declared the COVID-19 pandemic. Machine learning (ML) based mostly prediction mechanisms have established their significance to anticipate in perioperative outcomes to boost the choice creating on the long run course of actions. The ml models have long been utilized in several application domains that required the identification and prioritization of adverse factors for a threat. many predictions strategies square measure being popularly wont to handle prediction issues. This study demonstrates the potential of ml models to forecast the quantity of coming patients suffering from COVID-19 that is presently thought of as a possible threat to human beings.

In specific, normal prediction models, like linear regression (LR), least absolute shrinkage and choice operator (LASSO), support vector machine (SVM), and exponential smoothing (ES) are utilized in this study to forecast the threatening factors of COVID-19.

The results prove that the es performs best among all the used models followed by LR and LASSO that performs well in prediction the new confirmed cases, death rate similarly as recovery rate, whereas SVM performs poorly all told the prediction situations given the offered dataset.

Keywords: machine learning , covid 19, LR,LASSO

I. INTRODUCTION

This study is expounded to the unfold of novel coronavirus, conjointly called SARS-CoV-2, formally named COVID-19 by the globe Health Organization (WHO). COVID-19 is presently a awfully serious threat to human life everywhere the globe. 1st reportable in city, Hubei Province, China, has infected folks from 216 countries, and territories worldwide. Due to the inverse quantitative relation of an oversized variety of infected people and also the short amount of your time, accumulative infectious cases area unit principally taken by exponential functions because of the apace increasing values. Mortality happens principally in infected patients with a weakened system, i.e. to the old and patients with chronic diseases. the shortage of medication to treat the virus has led governments to policies that need voters to be isolated and to take care of spatial distance so as to limit the unfold the maximum amount as attainable. The virus spreads primarily through shut person to person

physical contacts, by metabolic process droplets, or by touching the contaminated surfaces. the foremost difficult side of its spread is that an individual will possess the virus for several days without showing symptoms. The causes of its unfold and

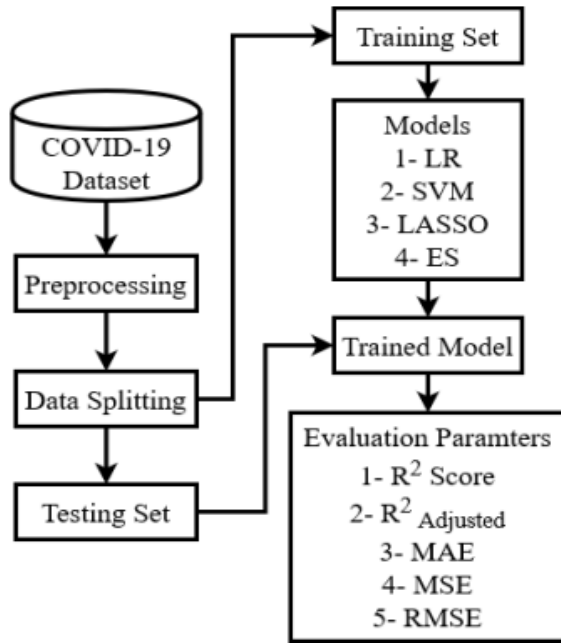
considering its danger, most the countries have declared either partial or strict lockdowns throughout the affected regions and cities. Thus, the implementation of machine learning and cloud computing can be effective in predicting wherever and once the illness can unfold or be eradicated so as to alert these communities to require applicable actions.

Machine Learning is evidenced itself as a distinguished field of study over the last decade by finding many terribly advanced and complex real-world issues. ML algorithms' learning is usually supported trial and error technique quite opposite of typical algorithms, which follows the programming directions ML algorithms are utilized in this space to guide the long run course of actions required in several application areas together with weather statement, illness statement, securities market statement as well as illness prognosis. numerous regression and neural network models have wide pertinency in predicting the conditions of patients within the future with a selected illness.

Mathematical modelling of accumulative infectious cases with linear, exponential regression models interprets the info a lot of accurately, as given by the R-square analysis. we have a tendency to developed polynomial coefficients victimisation {least squares method of least squares statistical technique statistical procedure} method. so as to reduce

the variance between the calculable values from the polynomial perform and also the actual values from the info set. Furtherance of polynomial's curve will predict the expected progress of the particular curve. Also, from the roots of the 2d derivatives we are going to have the chance to calculate the turning points that are called points of inflection, as these points area unit milestones in pandemic things.

I. METHODOLOGY



II. MODELLING AND ANALYSIS

A. Dataset

The main aim of this study is to demonstrate the capability of the Machine Learning models to forecast the number of upcoming new patients which are affected by this new virus known as COVID-19. The dataset used in the study has been obtained from GitHub repository made available by the John Hopkin University. The repository was first made available in 2019 and was also supported by ESRI Living Atlas Team. The repository contains daily time series summary table, number of confirmed cases of deaths and recoveries. Data sample from files are shown below: -

Table1-COVID 19 Patient Death Cases Time-Series Worldwide

Province /State	Country /Region	Lat	Long	1/22/20	1/23/20	...	3/27/20
Northern Territory	Australia	-12.46	130.84	0	0	...	0
Diamond Princess	Canada	0.000	0.000	0	0	...	1
NaN	Algeria	28.03	1.65	0	0	...	19

Table2- COVID-19 New Confirmed Cases Time-Series Worldwide



Province /State	Country /Region	Lat	Long	1/22/20	1/23/20	...	3/27/20
NaN	Afghan	33.00	65.00	0	0	...	74
Victoria	Australia	-37. 81	144. 96	0	0	...	411
NaN	Algeria	28.03	1.65	0	0	...	264

Table3- COVID-19 Recovery Cases Time-Series Worldwide

Province /State	Country /Region	Lat	Long	1/22/20	1/23/20	...	3/27/20
Colombia	Canada	49. 28	-123. 1	0	0	...	4
Victoria	Australia	-37. 81	144. 96	0	0	...	70
NaN	Algeria	28.03	1.65	0	0	...	65

B. SUPERVISED MACHINE LEARNING MODELS

Supervised learning is where you have an input variable ‘x’ and an output variable ‘Y’ and we use different algorithms to learn the mapping function from input to output i.e. $Y = f(x)$

Thus, in Supervised ML model we make prediction with an unknown input. So, we provide the learning algorithm with unknown dataset as input to train the regression model. The trained model then generates a prediction. For predictive model development the methods we can use are regression techniques and classification algorithms.

Four regression models have been used in this study -

1]Exponential Smoothing

The method uses previous data for forecasting. ES is very simple and powerful method for univariate data.

The forecast for current time is given by –

$$F_t = \alpha A_t + (1 - \alpha) F_{t-1}$$

α - smoothing cost

F_{t-1} – forecast value of previous forecast

2] Linear Regression

The predicted output is continuous and has a constant slope in linear regression. It predicts value in a continuous range and not try to classify them in different categories. Most used statistical technique in predictive analysis and find a relationship between independent and dependent variable. Determines a linear relationship between the two variables i.e. (x, y).

The equation showing how ‘y’ is related to ‘x’: -

$$y = \beta_0 + \beta_1 x + \epsilon$$

ϵ -error term

The goal is to find the best values for β_0 and β_1 to get best fit line.

3]LASSO Regression

It is a linear regression technique that uses shrinkage. Using shrinkage makes LASSO more stable and better than others and also tries to reduce the errors. Also it is very easy in terms of the number of terms it is using.

LASSO works in the motive to minimize the following-

$$\sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j|$$

4]Support Vector Machine

Used for regression as well as classification. It depends on many mathematical functions. It solves regression problems using the linear function. The linear function can be shown as –

$$f(x) = x' \beta + b$$

The goal is to make it as flat as possible which can be solved by the minimization function –

$$J(\beta) = \frac{1}{2} \beta' \beta$$

C. Evaluation parameters

We evaluate the performance of different models using –

1]R-Squared score

It shows a statistical value and shows the performance of the model. We can check the goodness-of-fit of the models after they are trained. Scores between 0-100%.

The high the score the good the model is trained. It explains the percentage of variation independent variable.

$$R^2 = \frac{\text{Variance explained by model}}{\text{Total variance}}$$

2]Adjusted R-Squared Score

It is the modified version of R-square. In R-square adjusted increase in the new features can be very useful to the prediction model.

$$R^2_{adjusted} = 1 - (1 - R^2) \frac{n - 1}{n - (k + 1)}$$

3]Mean Absolute Error

It is basically the average magnitude of errors. It is average on test data between model prediction and actual data.

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|$$

4]Mean Square Error

It measures the performance of the regression models. It takes distance of data points and square them. The smaller the MSE the closer we find the line of best fit.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

5]Root Mean Square Error

Defined as standard deviation of prediction errors. Prediction error is the distance between best fit line and actual data points.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

III. APPLICATIONS

1. Medical imaging of Infected patient

The novel coronavirus 2019 is spreading globally. It results in ill effects on human health and the global economy. This novel virus is spreading globally among humans and an urgent need to detect the positive cases as soon as possible so that the spreading chain can be slowed down. Few studies have encountered changes in chest X-ray and CT images before the beginning of COVID-19 symptoms. Chest radiological imaging such as computed tomography (CT) and X-ray has

played an important role in early diagnosis and treatment of the COVID-19 pandemic. It is also found that by using radiology imaging techniques, the COVID-19 virus can be detected in the human body.

The application of machine learning-artificial intelligence (AI) techniques coupled with radiological imaging is very helpful in detecting these disease symptoms in the subjects. A new automatic model to detect COVID-19 symptoms in the patients is presented using a chest X-ray image. They used the developed automatic model to accurately diagnostics for binary and multi-class classification. It is found that as advised by the developed automatic models can be used to diagnose other chest-related diseases, including tuberculosis and pneumonia (Ozturk et al., 2020). However, there is a limitation of the developed model that it is not robust and took fewer COVID-19 X-ray images

2. Provide an intelligent platform for healthcare.

Machine learning-artificial intelligence (AI) is found as an emerging tool which shows their potential to fight against novel coronavirus disease. The existing technology with NLP, computer visions, etc. enables the computers to leverage massive data-enabled models for pattern recognition, interpretation, and prediction. It is observed that the current pandemic disease COVID-19 is spreading worldwide quickly, so there is a strong need to explore and unzip the AI to cure or diagnose the infected patients by COVID-19. It is found that China has developed AI-driven sensors that can quickly recognise individuals with a fever, even in crowds. The sensing of COVID-19 patients can be detected by using AI-based technology in Florida through facial thermal scans (V. Kumar, 2020). An Artificial Intelligence based an app, 'COVID voice detector' used to detect the infection in the human voice. Augmented reality and virtual reality technology provided the virtual environment to the patients to engage with medical professionals.

3. Prediction of future disease symptoms

It is very important to predict the symptoms of present and future diseases that may hit the humanity with the impact of deaths globally. Machine learning methods such as deep neural network and long-short term memory learning methods are used to predict the symptoms and spread of infectious diseases such as chickenpox, scarlet fever, and malaria (Chae et al., 2018). Industry 4.0 technologies are very useful in fighting against Coronavirus (Javaid et al., 2020). Similarly, machine learning methods, such as artificial intelligence or artificial neural network, can be used to predict the symptoms of future diseases such as genetic disorders etc. This prediction is created by perceptive the facial expressions or options, predict the danger of heart failure by perceptive the patterns of blood vessels on the retina at the rear of the eye, which is used to monitor a patient's posture, respiration and even sleep for early signs of unhealthiness, etc. (Kaminsky, 2019). The future health care condition of patients can be predicted using various machine learning tools by collecting data from the clinical reports, doctor's notes, and various wearable body sensors (S. Wang et al., 2016). The machine learning models are useful to predict the spread of the COVID-19 pandemic in real-time. However, the methods are very effective to forecast the spread of novel coronavirus in the future. Also, machine learning methods may be used to predict the trends in infectious pandemic diseases such as COVID-19.

4. Disease and patient behavioural analysis

The patient's behaviour and the analysis of the current and future pandemic disease behaviour are very important. Machine learning-based artificial intelligence focuses on developing mathematical models to analyse the COVID-19 pandemic disease using national data (Punn et al., 2020). Machine learning and deep learning models are used to understand the daily exponential behaviour by using real-time information from the Johns Hopkins dashboard. A comparative analysis of machine learning and soft computing models to predict the COVID-19 outbreak is proposed (Ardabili et al., 2020). It is found that the two machine learning models, such as multi-layered perceptron, and adaptive network-based fuzzy inference system, showed promising results towards the model of the behaviour of the COVID-19 outbreak from nation-to-nation. In this view, it is imperative to predict the behaviour of the spreading the virus so that the early actions may be planned to stop the spread of the COVID-19 disease. Machine learning algorithms play an important role in epidemic analysis and forecasting in the availability of massive epidemic data.

5. Clinical trials

The machine-learning-based risk predictive analytical approach is employed (Schaaf et al., 2020) to modify pharmaceutical firms to create higher predictions and perform new clinical trials. The current developed approach with help of machine learning on the cases of Covid 19 will be proved to be very helpful for the next generation platform transformation of clinical operations management. The key challenges in clinical trials are: the patients and the health care staff may come to the exposure of the infectious or pandemic disease during the trial operations, clinics refuse to having a part in the trials, and the patients may falling out of trials, etc. During clinical trials, it is very important to regularly review the COVID-19 status, trends, and predictions (Jaly et al., 2020). However, it's vital to require the recommendation from the epidemiologists from time to time, so the danger are often avoided.

IV. RESULTS AND DISCUSSION

This examination endeavours to build up a framework for the future estimating of the quantity of cases influenced by COVID-19 utilizing AI techniques. The dataset utilized for the investigation contains data about the everyday reports of the number of recently tainted cases, the number of recoveries, and the number of deaths because of COVID-19 around the world. As the passing rate and affirmed cases are expanding step by step which is a disturbing circumstance for the world isn't notable. This investigation is an endeavour to estimate the quantity of individuals that can be influenced regarding new contaminated cases and deaths including the number of recoveries for the forthcoming 10 days. Two AI models LASSO, and ES have been utilized to anticipate the quantity of recently contaminated cases, the number of deaths, and the number of recoveries.

A. DEATH RATE FUTURE FORECASTING

The investigation performs forecasts on death rate and as per results ES performs better among every one of the models, LASSO perform well and accomplish practically a similar R2 score. The outcomes are appeared in Table 1.

TABLE 1. Models performance on future forecasting for death rate.

Model	R ² Score	R ² Adjusted	MSE
LASSO	0.85	0.81	3244066.79
ES	0.98	0.97	662228.72

Figures 1 and 2 show the performance of LASSO and ES respectively. The Graphs show the death rate will increase in the upcoming days. The figure 14 shows the models' prediction is correct.

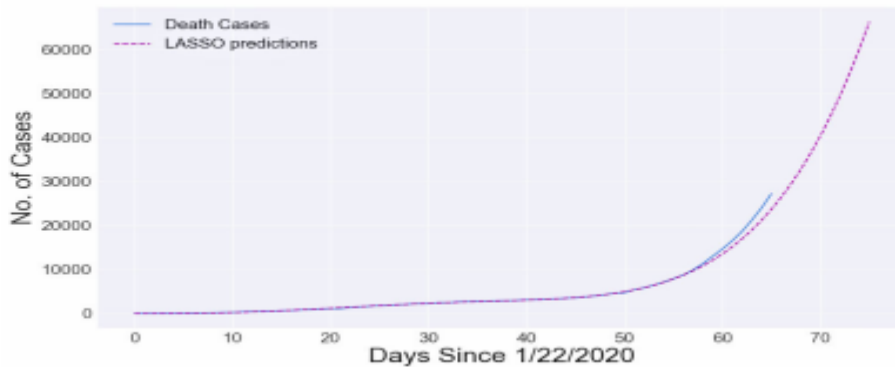


FIGURE 1. Death prediction by LASSO for the upcoming 10 days.

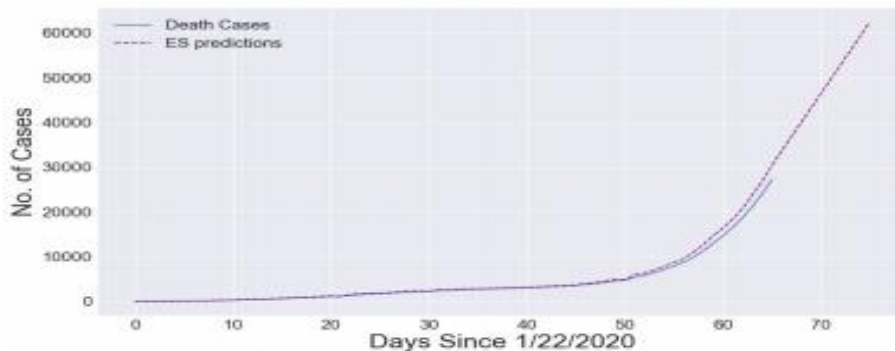


FIGURE 2. Death prediction by ES for the upcoming 10 days.

B. NEW INFECTED CONFIRM CASES' FUTURE FORECASTING

The new confirmed cases of COVID-19 increase day by day Table 2 shows the forecasting results of the models used in this study.

TABLE 2. Models performance on future forecasting for new infected confirm cases.

Model	R2 Score	R ² Adjusted	MSE
LASSO	0.98	0.97	234489560.99
ES	0.98	0.97	283201302.2

The graphs in figure 4 and 5 show the predictions of the model.

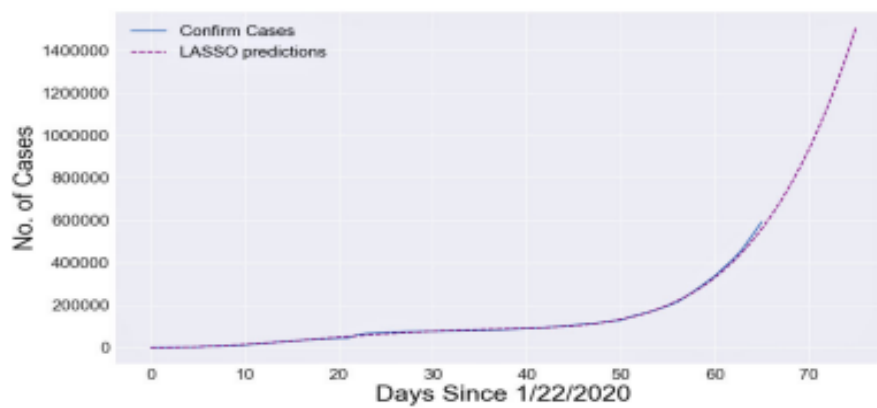


FIGURE 4. New infected confirm cases prediction by LASSO for the upcoming 10 days.

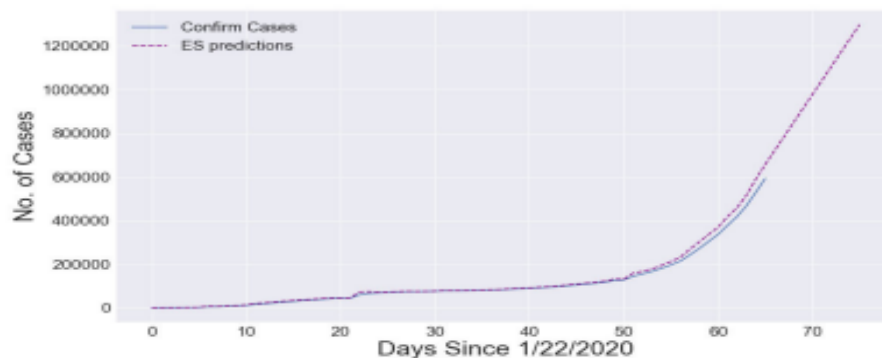


FIGURE 5. New infected confirm cases prediction by ES for the upcoming 10 days.

C. RECOVERY RATE FUTURE FORECASTING

In recovery rate future estimating the ES again performs better among the wide range of various models. The expectation patterns for the coming days are appeared in Figures 6 and 7.

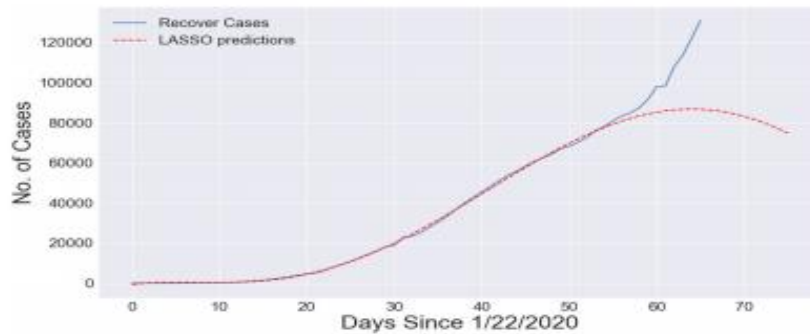


FIGURE 6. Recovery rate prediction by LASSO for the upcoming 10 days.

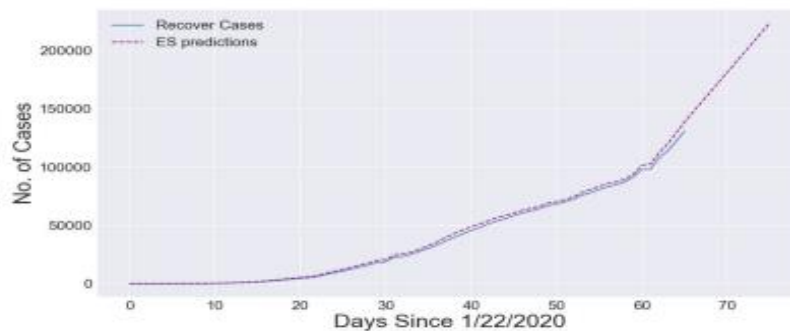


FIGURE 7. Recovery rate prediction by ES for the upcoming 10 days.

The execution consequences of learning models are appeared in Table 3 below:

TABLE 3. Models performance on future forecasting for recovery rate.

Model	R ² Score	R ² Adjusted	MSE
LASSO	0.29	0.08	1462144344.82
ES	0.99	0.99	5970634.07

In any case, looking at the current recuperation statistics with our models' expectations, the ES forecast is following the patterns which are near the real circumstance. Also, some more investigation has been performed following 5 days of trials on the refreshed dataset and some significant measurements have been found as demonstrated in Figure 8 and 9. The Figure 10 and 11 show that our model is working well.



FIGURE 8. Mortality rate after 5 days of this study prediction.



FIGURE 9. Recovery rate after 5 days of this study prediction.

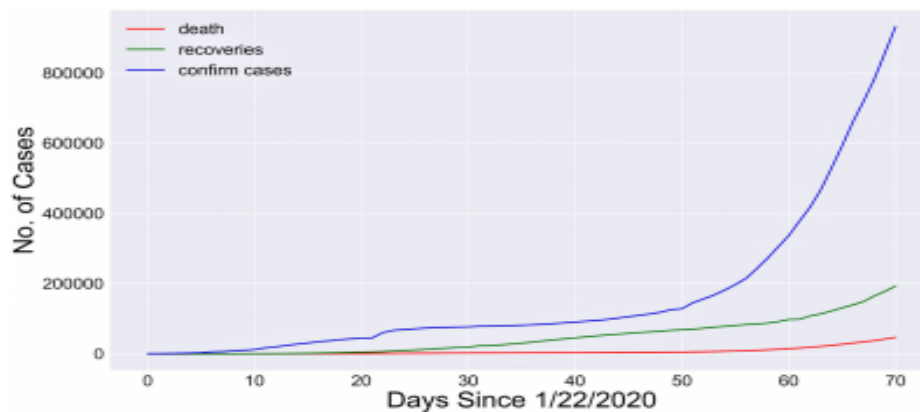


FIGURE 10. Comparison between death rate, recovery rate and confirm case rate after 5 days of this study prediction.

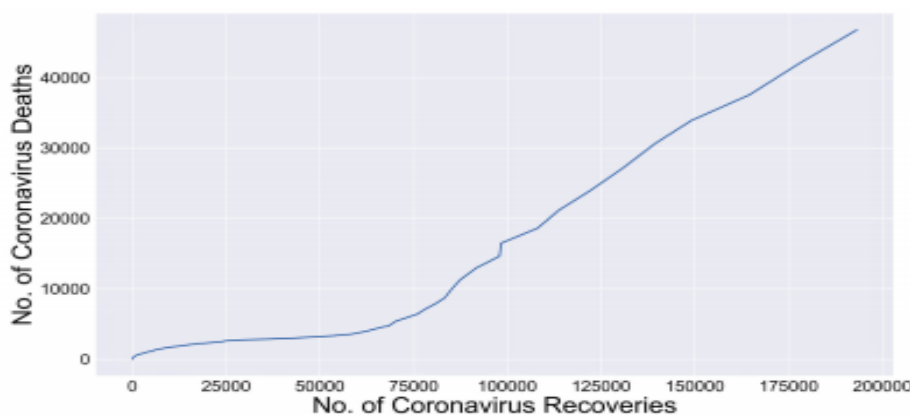


FIGURE 11. Ratio between recovery rate and death rate after 5 days of this study prediction.

D. Model Performances With 10-15 Days Prediction Intervals

As demonstrated in the past segments, ES performed best in all three cases, for example, passing rate determining, the quantity of new confirmed cases determining, and recuperation rate anticipating. Considering the best exhibition given by ES model in all the three anticipating cases among every one of the four models, the model has been utilized for

additional investigation with stretch forecast [7]. Figure 12 presents the model presentation on the passing rate, recuperation rate, and new confirmed cases with 15 days span period.

To begin with, every one of the models have been prepared from the dataset of 22 Jan 2020 to 16 Feb 2020, and expectations were made for the impending 10 days from 16/02/2020. Since the information accessible in this dataset was of just 26 days.

In the third stretch next 15 days were added to the dataset. ES in this span while performing great shows some deviation as demonstrated in the diagrams of Figure 12, from the real information arrangement as a result of a abrupt ascent taking all things together the three cases in this period. In the fourth Interval information of 10 additional days have been added expanding the size of the preparation set to 66, in this span all the models can be viewed as improved significantly and making the general outcomes close to the real circumstance.

Be that as it may, ES outflanks every one of the models in the forecast of every one of the three cases. The expectation results have been contrasted and the genuine information reports of these particular day spans. The expectations results given by these models have been discovered nearer to the real reports. The span subtleties have been gathered and given in Table 11. To proceed and expand further the extent of the of this examination in anticipating. A similar system has been applied to further estimate the quantity of confirmed cases, deaths, and recovery up to 6 Apr 2020. Figure 13 presents the plots of confirmed cases, deaths, and recuperation on the first four sheets followed by the plot of real circumstance accumulated from the genuine information reports of the inspecting time of the investigation in the fifth sheet. The outcomes in the diagrams demonstrate that the ML models utilized in this examination befit the determining task making the route towards the ease of use of the examination and future exploration of the comparative nature.

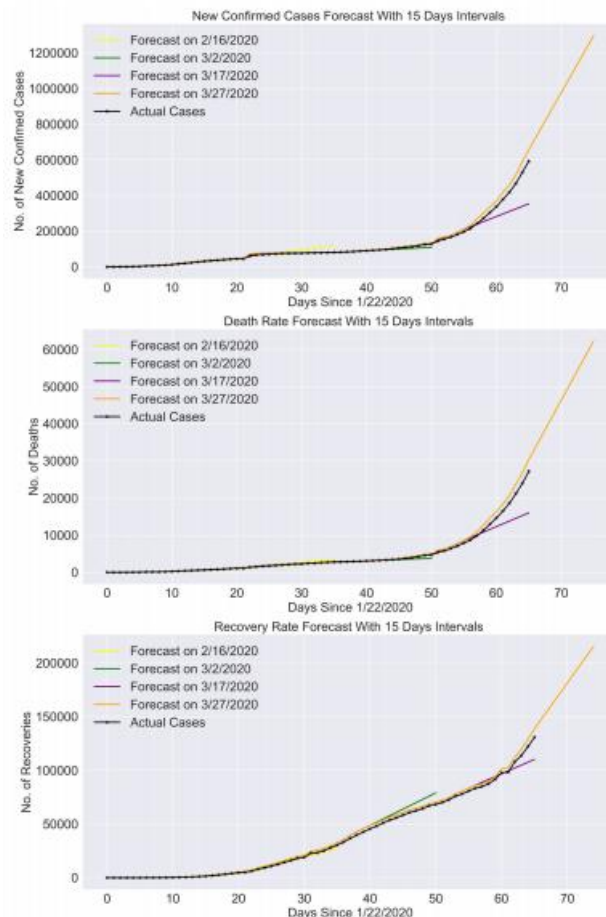


FIGURE 12. ES performances on death rate, recovery rate and new confirmed case with 10-15 days intervals.

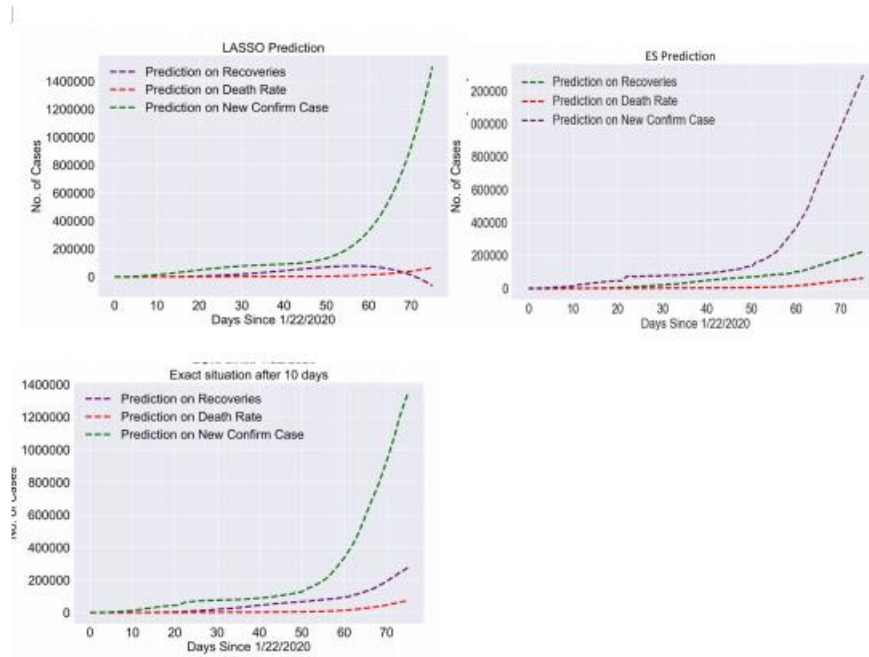


FIGURE 13. All models predictions form 1/22/2020 to 4/6/2020 and real situations form 1/22/2020 to 4/6/2020.

TABLE . Models performance on future forecasting for recovery rate.

Interval	Dataset Size (Number of Days)	Dates (From 22 Jan 2020) To	LASSO Performance	ES Performance
1.	26	16 Feb 2020	Very poor	Best
2.	41	2 Mar	Very poor	Best
3.	56	2020 17 Mar 2020	Poor	Best
4.	66	27 Mar 2020	Better	Best

V. CONCLUSION

The COVID-19 has become a global crisis. Some research and government predicted that this virus will affect a large population of the world. In this study we use a ML based predictive system that helps us to find the reach of this virus. The result of the study states that ES gives the best result in the current forecasting domain that is followed by LR. LASSO also shows good result to some extent for predicting the death rate and confirm cases. SVM shows poor results in all the cases. Overall, we can predict that the model shows correct predictions with respect to the current scenario and the upcoming situations. This study will be used more and more in the future and we plan to use the daily updated dataset and predict the outcome with more accurate result. Also use more new ML methods that help in giving accurate forecasting. Real time forecasting will be more used in the future.

VII. REFERENCES

- Assaf, Dan, et al. "Utilization of machine-learning models to accurately predict the risk for critical COVID-19." *Internal and emergency medicine* 15.8 (2020): 1435-1443.
- S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and machine learning forecasting methods: Concerns and ways forward," *PLoS ONE*, vol. 13, no. 3, Mar. 2018, Art. no. e0194889.
- P. Lapuerta, S. P. Azen, and L. Labree, "Use of neural networks in predicting the risk of coronary artery disease," *Comput. Biomed. Res.*, vol. 28, no. 1, pp. 38–52, Feb. 1995.
- K. M. Anderson, P. M. Odell, P. W. Wilson, and W. B. Kannel, "Cardiovascular disease risk profiles," *Amer. heart J.*, vol. 121, no. 1, pp. 293–298, 1991.



DOI: [10.17148/IARJSET.2021.8462](https://doi.org/10.17148/IARJSET.2021.8462)

- H. Asri, H. Mousannif, H. A. Moatassime, and T. Noel, "Using machine learning algorithms for breast cancer risk prediction and diagnosis," *Procedia Comput. Sci.*, vol. 83, pp. 1064–1069, Jan. 2016.
- S. Baran and D. Nemoda, "Censored and shifted gamma distribution based EMOS model for probabilistic quantitative precipitation forecasting," *Environmetrics*, vol. 27, no. 5, pp. 280–292, Aug. 2016.

BIOGRAPHY

Snehil Raj is a 3rd year undergraduate student pursuing Bachelor of Technology with majors in Information Technology. Having done the courses and curriculum pertaining to cloud engineering, he has done research and authored a paper on cloud computing and has come up with a generalized study of covid-19 forecast using machine learning models.

Pavitra Maheshwari is a 3rd year student of Bachelor of Technology with majors in Information Technology. Having done the courses and curriculum pertaining to Machine Learning he has done research and has co-authored the paper on covid-19 forecast using machine learning models

Tushar Agarwal is a 3rd year student of Bachelor of Technology with majors in Information Technology. Having done the courses and curriculum pertaining to cloud engineering, he has done research and has co-authored the paper on covid-19 forecast using machine learning models

Ankit Shukla is a 3rd year student of Bachelor of Technology with majors in Information Technology. Having done the courses and curriculum pertaining to cloud engineering, he has done research and has co-authored the paper on cloud covid-19 forecast using machine learning models