

# COVID-19 Future Forecasting Using Supervised Machine Learning Models

Mr D.Purushothaman MCA.,M.E.,<sup>[1]</sup> , G K Jayaprakash<sup>[2]</sup>

<sup>[1]</sup> Asst. Professor, Department of Computer Applications

<sup>[2]</sup> Student, Department of Computer Applications

<sup>[1],[2]</sup> Chadalawada Ramanamma Engineering College (Autonomous)

## ABSTRACT

Machine learning (ML) based forecasting mechanisms have proved their significance to anticipate in peri-operative outcomes to improve the decision making on the future course of actions. The ML models have long been used in many application domains which needed the identification and prioritization of adverse factors for a threat. Several prediction methods are being popularly used to handle forecasting problems. This study demonstrates the capability of ML models to forecast the number of upcoming patients affected by COVID-19 which is presently considered as a potential threat to mankind. In particular, four standard forecasting models, such as linear regression (LR), least absolute shrinkage and selection operator (LASSO), support vector machine (SVM), and exponential smoothing (ES) have been used in this study to forecast the threatening factors of COVID-19. Three types of predictions are made by each of the models, such as the number of newly infected cases, the number of deaths, and the number of recoveries in the next 10 days. The results produced by the study proves it a promising mechanism to use these methods for the current scenario of the COVID-19 pandemic. The results prove that the ES performs best among all the used models followed by LR and LASSO which performs well in forecasting the new confirmed cases, death rate as well as recovery rate, while SVM performs poorly in all the prediction scenarios given the available dataset.

**Keywords:** - Machine Learning, Covid19, Forecasting, Supervised Learning.

## I. INTRODUCTION

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 corned cases and study is also focused on the forecast of COVID-19 outbreak and early response. Computational methods for machine learning (ML) have shown their meaning for the projection of potential results for informed decisions. Machine learning algorithms have been applied for a long time in many applications requiring the detection of adverse risk factors. This study shows the ability to predict the number of individuals who are affected by the COVID-19 as a potential threat to human beings by ML modeling.

In this existing analysis, the risk factors of COVID-19 were exponential smoothing (ES). The Lower Absolute Reductor and Selection Operator, (LASSo), Vector Assistance (SVM), four normal potential forecasts, such as Linear Regression (LR)). [Each of these machine-learning models has three distinct kinds of predictions: the number of newly infected COVID 19 people, mortality rates and the recovered COVID-19 estimates in the next 10 days. These approaches are better used in the latest COVID-19 situation, as shown by the findings of the analysis. The LR, that is effective in predicting new cases of corona, death numbers and recovery. In the existing work, the system is low effective due to lack of supervised learning model. This system is less performance due to lack of the finding number of death cases.

## II. RELATEDWORKS

S V Subramanian and K S James proposed Use of the Demographic and Health Survey framework as a population surveillance strategy for COVID-19. In the absence of universal testing, a random-sample-based population surveillance framework is urgently needed. they propose using the well-established Demographic and Health Survey (DHS) framework as a solution to ascertain the true prevalence of COVID-19. Layering a COVID-19-

focused data-collection effort onto the NFHS infrastructure would keep operational costs low, with the major expense being laboratory costs for testing samples.

They estimated the minimum required sample of individuals who would need to be tested under three scenarios of anticipated COVID-19 prevalence in the population. Under a scenario of 0-5% prevalence, They would only need a sample of some people to be tested. the minimum sample size needed to reliably estimate the true prevalence would be smaller and, therefore, fewer resources would be needed. Stuart A Kinner, Kathryn Snow proposed Prisons and custodial settings are part of a comprehensive response to COVID-19. Prisons are epicenters for infectious diseases because of the higher background prevalence of infection, the higher levels of risk factors for infection, the unavoidable close contact in often overcrowded, poorly ventilated, and unsanitary facilities, and the poor access to health-care services relative to that in community settings. Infections can be transmitted between prisoners, staff and visitors, between prisons through transfers and staff cross-deployment, and to and from the community. As such, prisons and other custodial settings are an integral part of the public health response to coronavirus disease 2019. As such, COVID-19 outbreaks in custodial settings are of importance for public health, for at least two reasons: first, that explosive outbreaks in these settings have the potential to overwhelm prison health-care services and place additional demands on overburdened specialist facilities in the community; and second, that, with an estimated 30 million people released from custody each year globally, prisons are a vector for community transmission that will disproportionately impact marginalized communities.

N Oliver proposed Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. This Research describes how mobile phone data can support government and public health policymaking throughout the COVID-19 pandemic lifecycle, providing increased situational awareness, more accurate predictions, impact assessment of the policies and cause-and-effect inferences. It identifies key gaps and reasons why this kind of data is only scarcely used, although their value has been proven by previous work in the context of other infectious diseases, such as malaria, dengue, ebola or the H1N1 flu. It presents ways to overcome these gaps and key recommendations for urgent action, most notably the establishment of mixed expert groups on national and regional level, and the inclusion and support of governments and public authorities early on. The manuscript is authored by a group of experienced data scientists,

epidemiologists, demographers and representatives of mobile network operators who jointly put their work at the service of the global effort to combat the COVID-19 pandemic.

### **III. PROPOSED SYSTEM ARCHITECTURE**

This study aims to provide an early forecast model for the spread of novel corona virus, also known as SARS-CoV-2, officially named as COVID-19 by the World Health Organization (WHO). COVID-19 is presently a very serious threat to human life all over the world. At the end of 2019, the virus was first identified in a city of China called Wuhan, when a large number of people developed symptoms like pneumonia. It has a diverse effect on the human body, including severe acute respiratory syndrome and multi-organ failure which can ultimately lead to death in a very short duration. Hundreds of thousands of people are affected by this pandemic throughout the world with thousands of deaths every coming day. Thousands of new people are reported to be positive every day from countries across the world. The virus spreads primarily through close person to person physical contacts, by respiratory droplets, or by touching the contaminated surfaces. 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. ES performs best when the time-series dataset has very limited entries. Different ML algorithms seem to perform better in different class predictions.

Most of the ML algorithms require an ample amount of data to predict the future, as the size of the dataset increases the model performances improve. ML model based forecasting can be very useful for decision-makers to contain pandemics like COVID-19. In this proposed system, there are two modules. They are: Covid 19 Server And User. In the Covid-19 server module, the Covid19 Server has to login by

using valid user name and password. After login successful he can do some operations such as View All Student Academic Details, Find COVID19 Forecasting Prediction, View COVID19 Forecasting By Date Wise, View All Registered Users, View Death Cases Prediction Results, View Recovery Cases Prediction Results, Extract All Deaths Prediction, Extract All Recovery Prediction, Extract All New Case Prediction, Extract All New Case

Prediction. In the remote user module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like Browse Covid19 Data Sets, View All Covid19 Data Sets Details, Search Covid19 Forecasting In Data Set.

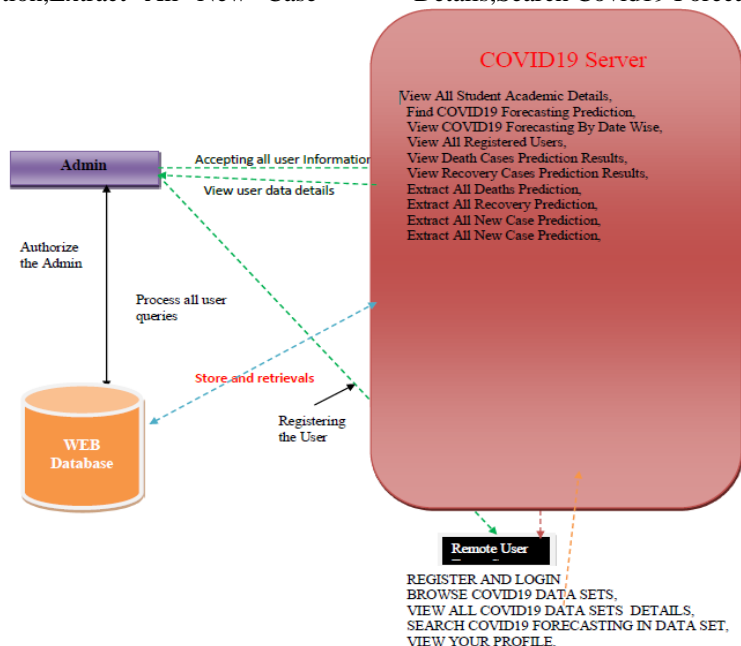


Fig.1 Proposed system architecture

#### IV. RESULTS AND DISCUSSION

The output screens obtained after running and executing the system are shown from Fig.2 to Fig.10

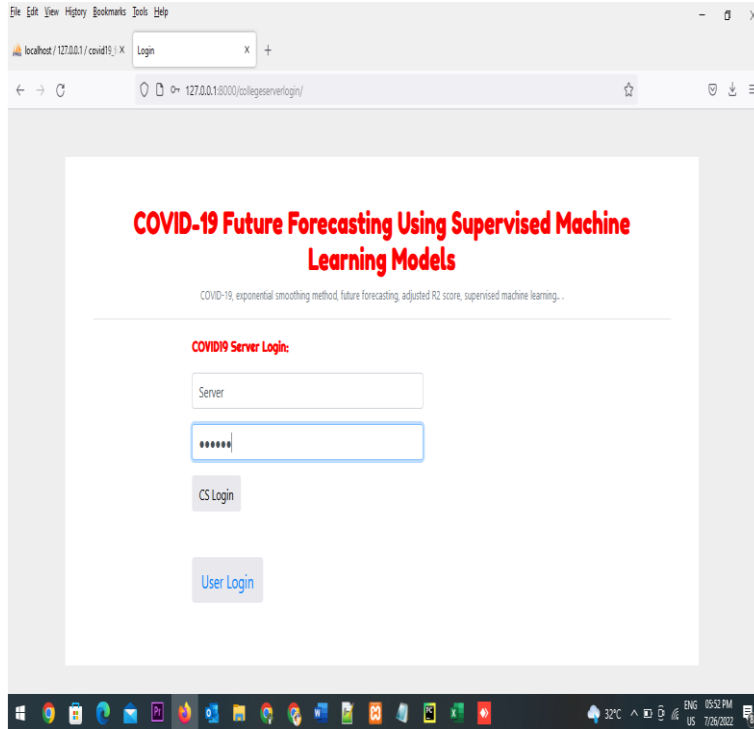


Fig.2 Server Login

The screenshot shows a web browser window with the address bar displaying 'localhost / 127.0.0.1 / covid19 / Add\_Data\_Sets/'. The main content area displays a table with the following data:

Province or State	Country or Region	Latitude	Longitude	From Date	To Date	Number Of Days	New Cases	Death Cases	Recovery Cases	Ongoing Treatment Cases
Colombia	Canada	49.28	-123.1	22/4/2020	30/4/2020	8	400	100	200	100
Victoria	Australia	-37.8	144.96	22/4/2020	30/4/2020	8	600	50	300	250
Las Anges	USA	34.05	118.24	14/5/2020	30/5/2020	16	5000	300	1200	3500
New York	USA	40.71	74.01	14/5/2020	30/5/2020	16	12000	2500	2000	7500
California	USA	36.77	119.41	14/5/2020	30/5/2020	16	6000	1500	1200	3300
Chennai	India	49.28	-123.1	22/4/2020	30/4/2020	8	10000	1000	3000	6000
Delhi	India	-37.8	144.96	22/4/2020	30/4/2020	8	7000	1000	2000	4000
Mumbay	India	34.05	118.24	14/5/2020	30/5/2020	16	25000	3000	4000	18000
Calcutta	India	40.71	74.01	14/5/2020	30/5/2020	16	18000	1000	4000	13000
Bangalore	India	36.77	119.41	14/5/2020	30/5/2020	16	8000	1000	2000	5000
Colombia	Canada	49.28	-123.1	22/5/2020	30/5/2020	8	18000	4000	6000	8000
Victoria	Australia	-37.8	144.96	22/6/2020	30/6/2020	8	5000	1000	2000	2000
Las Anges	USA	34.05	118.24	14/6/2020	30/6/2020	16	8000	1000	3000	4000
New York	USA	40.71	74.01	14/7/2020	30/7/2020	16	13000	2500	2000	8500
California	USA	36.77	119.41	14/7/2020	30/7/2020	16	16000	1500	1200	13300
Colombia	Canada	49.28	-123.1	22/8/2020	30/8/2020	8	5000	1000	2000	1000
Victoria	Australia	-37.8	144.96	22/8/2020	30/8/2020	8	6000	500	2000	3500
Las Anges	USA	34.05	118.24	14/8/2020	30/8/2020	16	15000	2000	3000	10000
New York	USA	40.71	74.01	14/8/2020	30/8/2020	16	9000	1000	2000	6000
California	USA	36.77	119.41	14/9/2020	30/9/2020	16	18500	1000	500	17000

Fig.3 Dataset details

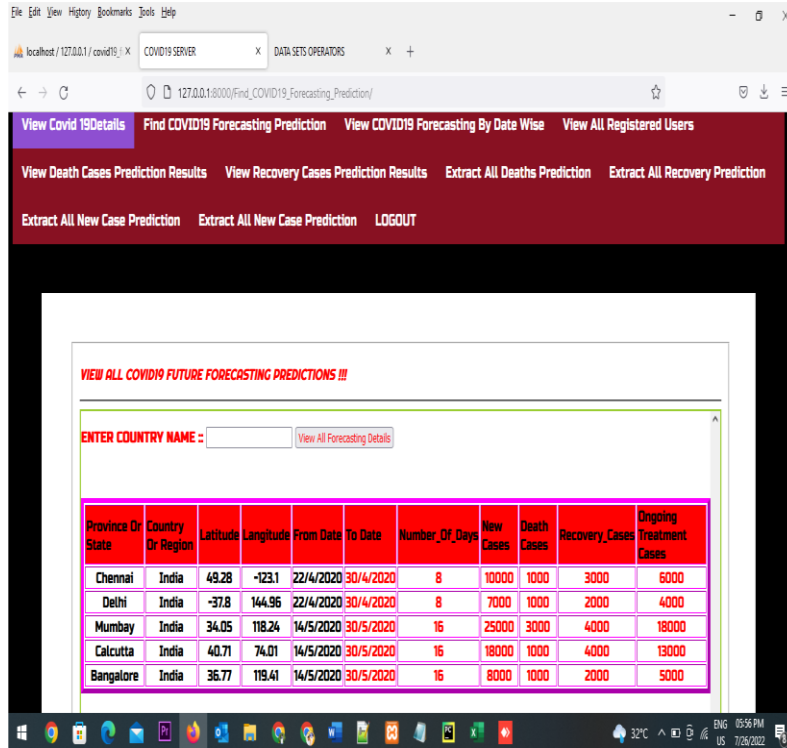


Fig.4 Predicting Covid 19 Future Forecasting

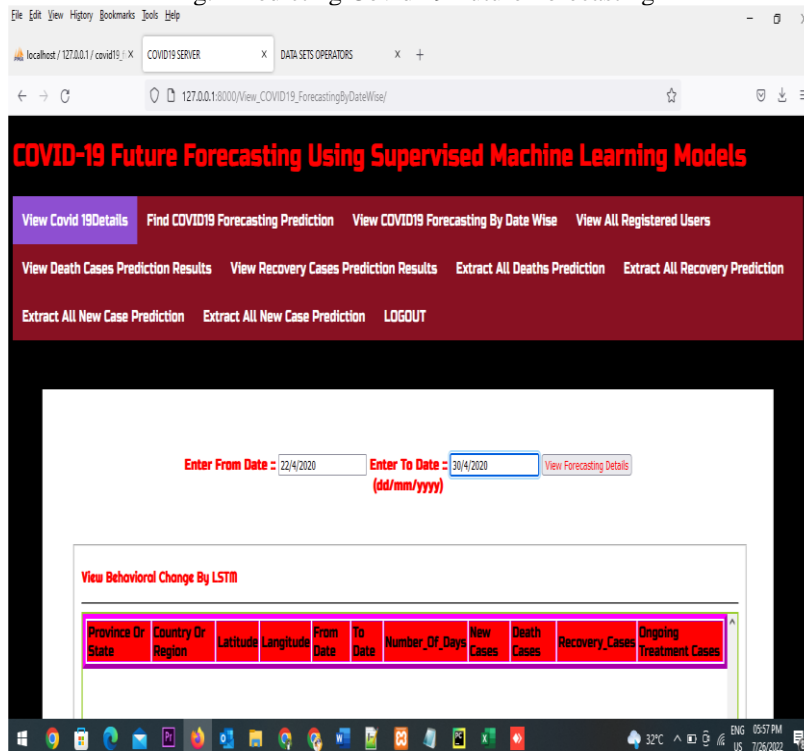


Fig.5 COVID-19 prediction by date wise

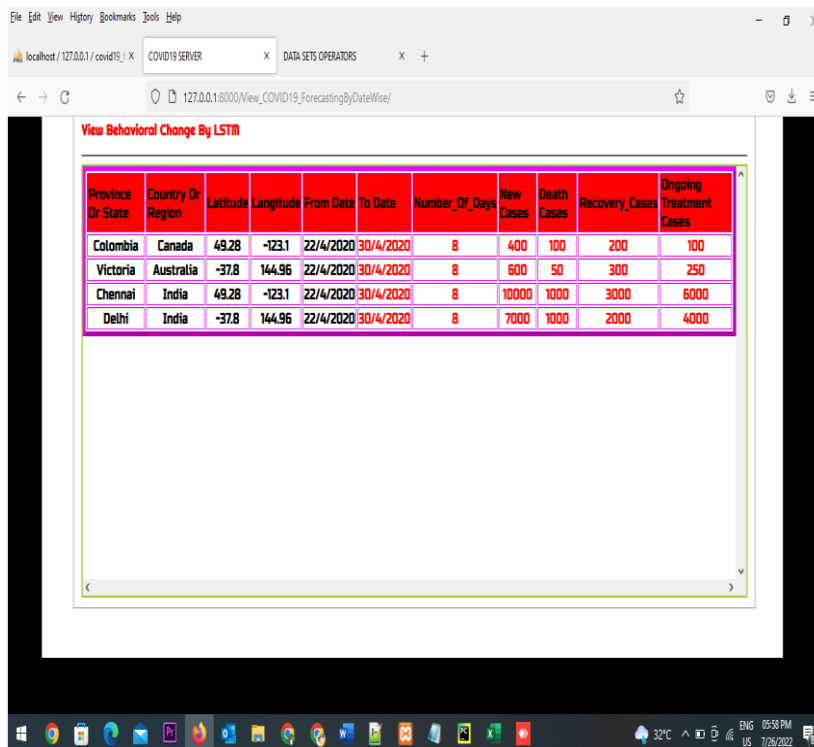


Fig.6 Forecasting details

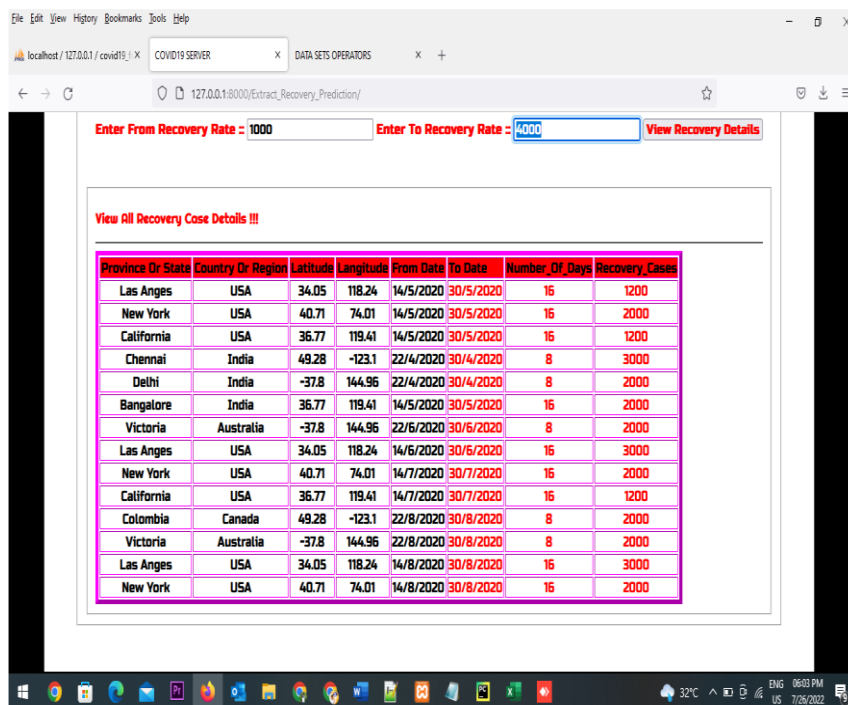


Fig.7 Recovery details

## V. FUTURE SCOPE AND CONCLUSION

In particular, four standard forecasting models, such as linear regression (LR), least absolute shrinkage and selection operator (LASSO), support

vector machine (SVM), and exponential smoothing (ES) have been used in this study to forecast the threatening factors of COVID-19. Three types of predictions are made by each of the models, such as the number of newly infected cases, the number of deaths, and the number of recoveries in the next 10

days. The results produced by the study proves it a promising mechanism to use these methods for the current scenario of the COVID-19 pandemic. The results prove that the ES performs best among all the used models followed by LR and LASSO which performs well in forecasting the new confirmed cases, death rate as well as recovery rate

## REFERENCES

- [1] Chen J., L. Wu, J. Zhang, L. Zhang, D. Gong, Y. Zhao, S. Hu, Y. Wang, X. Hu, B. Zheng et al., “Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study,” medRxiv, 2020.
- [2] “Novel coronavirus (2019-ncov). world health organization.” <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>., Accessed in March, 2020.
- [3] “Summary of probable sars cases with onset of illness from 1 november 2002 to 31 july 2003. world health organization.” <https://www.who.int/csr/sars/country/table20040421/en/>., Published April 21, 2004.
- [4] D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong et al., “Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in wuhan, china,” *Jama*, 2020.
- [5] National health commission of china. ( March, 2020). [Online]. Available: <http://www.chinadaily.com.cn/m/chinahealth/index.html>
- [6] Canada broadcast cooperation. (2020, March). [Online]. Available: <https://www.cbc.ca/player/play/1709650499517>
- [7] Bioworld. (2020, March). [Online]. Available: <https://www.bioworld.com/articles/433530-china-uses-ai-in-medical-imaging-to-speed-up-covid-19-diagnosis>