International Journal of Modern Research
Vol. 1, No. 1, 22-29, 2021

Analytical Review Analysis for Screening COVID-19 Disease

Pankaj Kumar Vaishnav*
Faculty of Engineering,
Pacific University, Udaipur, India
E-mail: panvas23@gmail.com

Sunil Sharma
Faculty of Engineering,
Pacific University, Udaipur, India
E-mail: sunil.sharma@pacific-it.ac.in

Prashant Sharma
Faculty of Engineering,
Pacific University, Udaipur, India
E-mail: prashant.sharma@pacific-it.ac.in

**Corresponding author:* Pankaj Kumar Vaishnav

(Received on August 5, 2021; Accepted on August 18, 2021)

Abstract
World have suffered a lot due to international pandemic caused by CORONA virus and declared as COVID-19 by world health organization (WHO) in April 2020. Subsequently various analysis have been performed to find of the cause, cure and precautionary measures against it. A lot casualty is also reported meanwhile the analysis was in progress. Here in this paper analytical analysis is performed to find out the impact of COVID-19 in terms of total cases, total recovery and total death which has been reported during this pandemic. Various Machine learning algorithms have been utilized to perform this analysis. For this analysis Indian states have been selected and all the research work is analyzed on the dataset which is available on WHO official platform. Finally Random forest and decision tree regression model have been used to perform analytical analysis.
Keywords- COVID-19; analysis; machine learning; artificial intelligence.

1. Introduction

By the end of 2019, a new type of virus was identified and named as **SARS-CoV-2** (Acaps, 2020). It was identified by the "**International Committee of Taxonomy of Viruses" (ICTV)**. It was observed that this new virus was caused by the disease named **Corona Virus disease-2019** (Bhatnagar et al., 2020). After this WHO refers this virus as **COVID-19** virus in it is every document. WHO's platform which is known as "Epidemic Intelligence from Open Sources" **(EIOS)** took up on "**ProMED**" which is a programme of the "International Society for Infectious Diseases" and is dealing with the same huddle of cases of "**Pneumonia of unknown cause**", in Wuhan city (Dey and Sinha, 2020). It was observed that **COVID-19** affects different age group of persons in different manner. It is also observed that most of the time the infected person build up meek to reasonable disease and recover without any sort of hospitalization. Among various

observations it was identified by world health organization (WHO) that this virus can have three categories of symptoms (Durairaj and Ramasamy, 2016).

1. The **Most common symptoms** are listed as fever, dry cough and tiredness.
2. The **Less common symptoms** are aches and pains, sore throat, diarrhea, conjunctivitis, headache, loss of taste or smell a rash on skin, or discoloration of fingers or toes.
3. **The Serious symptoms** are difficulty in breathing or shortness of breath, chest pain or pressure loss of speech or movement.

During two to three days when a person got infected by this virus, it was unable to detect that the person got infected or not (Hindustan Times, 2020). But after few days the patient starts feeling these above mentioned symptoms.

2. Epidemic Condition in Indian States

As the corona virus spreads globally we track the confirmed cases, deaths and recoveries with the help of tools and techniques available. Here the present epidemic condition of Indian states (India Today, 2020) (Where major impact of COVID-19 is seen) is analyzed and with the help of two available models i.e. Random forest Regression Model and Decision Tree Regression Model this analysis is presented. It has been identified that some states are in severe condition while some are in controlled condition. On this basis these two models have been presented. Here we have tabulated confirmed cases, total cured and total deaths and identified with the regression model that most severe state in terms of confirmed cases, deaths and also in terms of recoveries is Maharashtra while Telangana is least impacted in all terms (Resistrar, 2020). This analysis is considered as per state wise selection because every state has different infection rate, different environment, different precautionary measures have been take by their state government and progression over time is also different for each states. We have worn Indian COVID-19 data base accessible openly. Two major sources of the data are the Ministry of Health and Family Welfare (MoHFW), India (Shouman et al., 2012) and COVID-19 India (Shanthi et al., 2009).

3. Models Used for Analysis and Prediction

The random forest (RF) is a group classifier and consisting of numerous Decision Trees related to the means that a forest is a collection of countless trees (Unnithan, 2020). Decision Trees that are developed extremely deep repeatedly effects over fitting problem of the training dataset, resulting an elevated deviation in categorization product for a minute transform in the input statistics (Viboud et al., 2016). They are extremely responsive to their preparation data, which constructs them error-prone to the experiment dataset. The diverse Decision Trees of a Random Forecast are trained using the special elements of the training dataset (Welfare, 2020). To categorize a new illustration, the input vector of that model is essential to overtake with all Decision Tree of the forest. All Decision Tree subsequently believes a special element of that input vector and provides a categorization outcome (WHO, 2019).

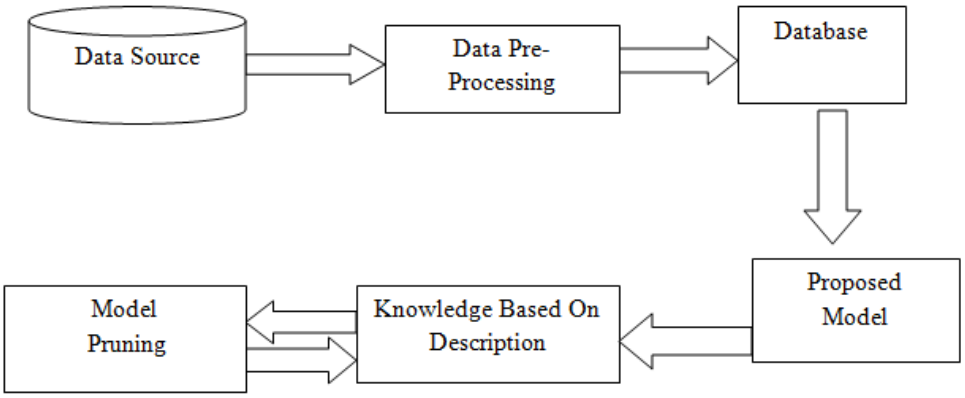


Figure 1: Models used for analysis and prediction.

The forest after that selects the categorization of having the majority 'votes' for distinct categorization outcome or the standard of all trees in the forest for numeric categorization outcome. Prediction in terms of confirmed cases (Wiki, 2020), total cured and total deaths is analyzed and presented with both the models (Zhu, 2020).

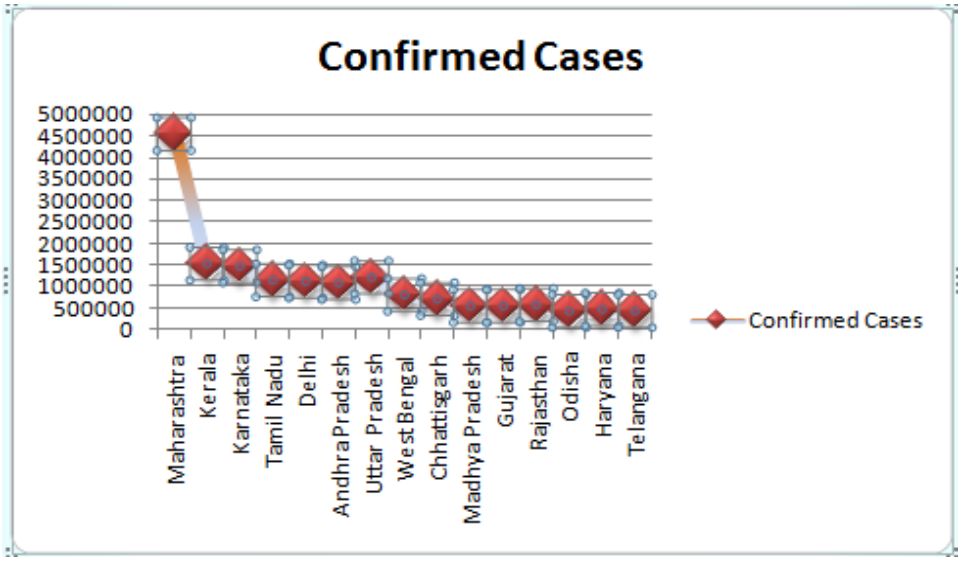
4. Results and Analysis

State wise analysis is selected and is tabulated for all these above mentioned cases. According to Table 1, Maharashtra is severe impacted while Telangana is least impacted with COVID-19.

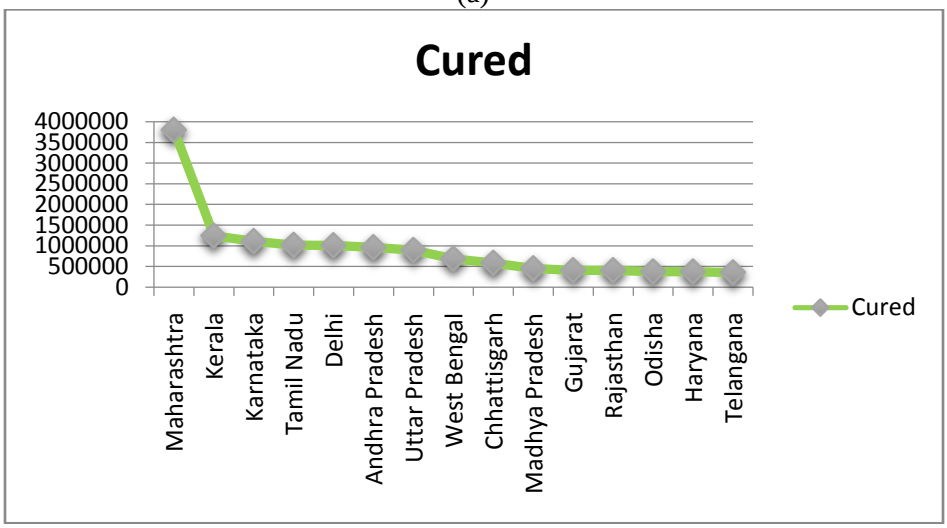
Table 1. State wise analysis of confirmed cases, total cure and total death due to COVID-19.

S.No.	State/UT	Confirmed Cases	Cured	Deaths
1	Maharashtra	4539553	3799266	67985
2	Kerala	1533984	1244301	5259
3	Karnataka	1474846	1110025	15306
4	Tamil Nadu	1148064	1021575	13933
5	Delhi	1122286	1008537	15772
6	Andhra Pradesh	1084336	962250	7928
7	Uttar Pradesh	1217952	896477	12238
8	West Bengal	810955	689466	11248
9	Chhattisgarh	713706	587484	8312
10	Madhya Pradesh	550927	453331	5519
11	Gujarat	553172	408368	7010
12	Rajasthan	580846	407243	4084
13	Odessa	435513	380400	2029
14	Haryana	474145	376852	4118
15	Telangana	435606	355618	2261

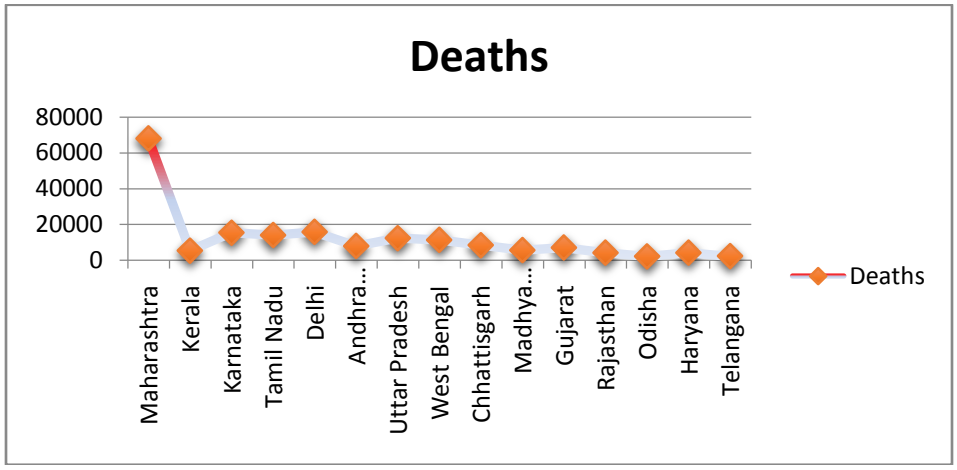
With the help of this dataset we have obtained plots for confirmed cases, total cured and total death in these states. These are plotted in below Figure 1.



(a)



(b)



(c)

Figure 2: (a) Confirmed cases (b) Confirmed cured patients (c) Confirmed deaths, in selected states of India.

On the basis of this analysis a combined plot is drawn and presented in Figure 2. It indicates the variation of all selected parameters in identified states.

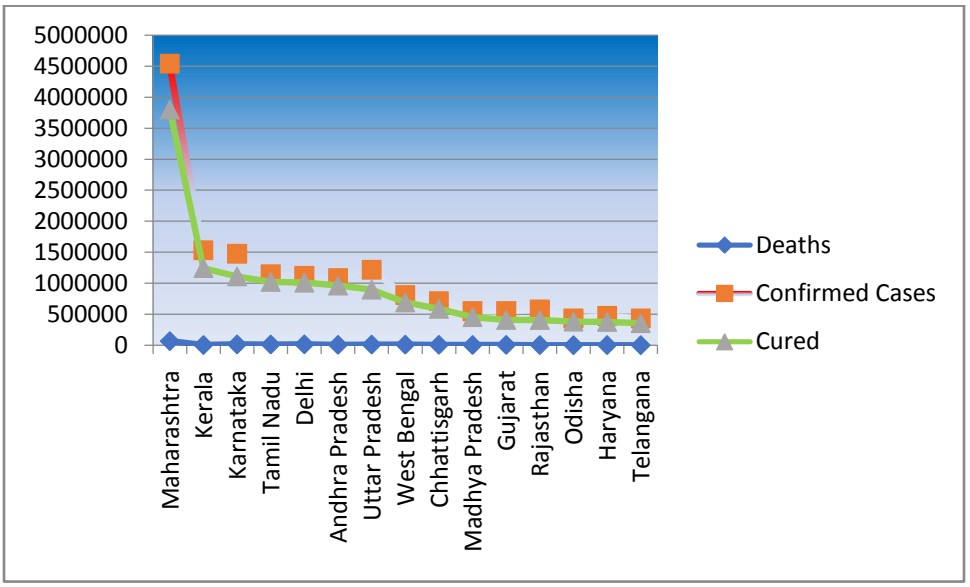


Figure 3: Comparison of all cases in selected states of India.

On the basis of above mentioned dataset two regression models have been used to identify the accuracy and prediction of the obtained outcomes. And subsequently is has been observed that Decision tree regression model provides 70% accuracy prediction while the Random Forest regression model provides 76% accuracy with prediction. To calculate accuracy and prediction following formula have been used for the identified models as shown in Figures 3-4 and Table 2.

$$\text{Accuracy} = \frac{\text{Number of Correct Prediction}}{\text{Total No.of Prediction}} \quad (1)$$

$$\text{Gain (T,X)} = \text{Entropy(T)} — \text{Entropy(T,X)} \quad (2)$$

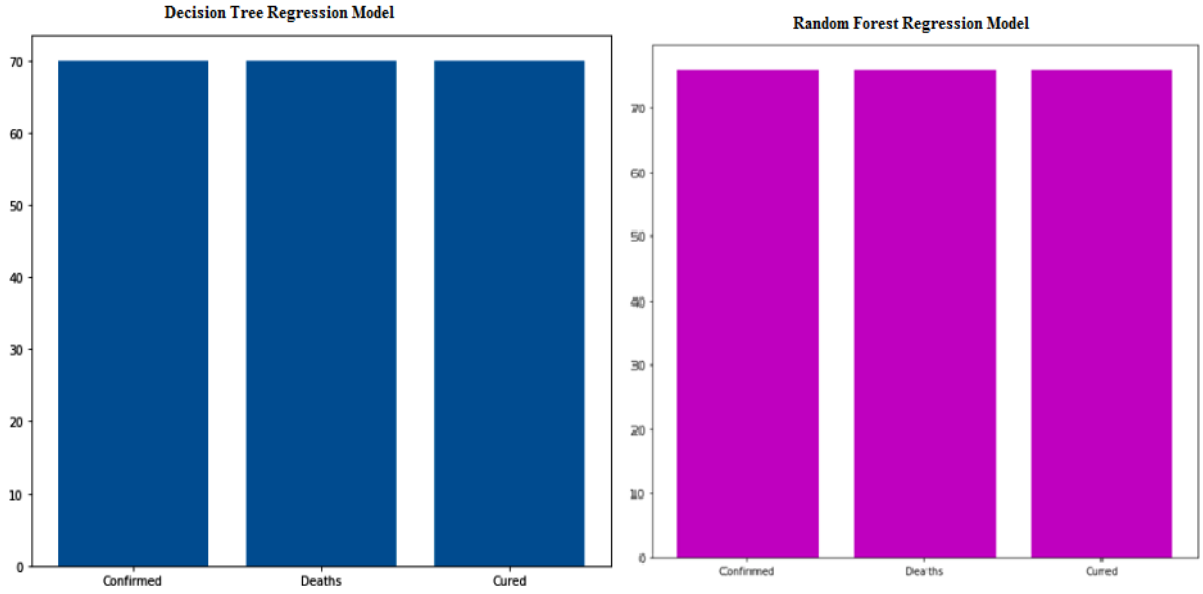
T = target variable

X = Feature to be split on

Entropy (T,X) = The entropy calculated after the data is split on feature X

Table 2: Accuracy prediction of both regression models.

Parameter	Decision Tree		Random Forest	
	Training Dataset	Test Dataset	Training Dataset	Test Dataset
Dataset	70%	30%	70%	30%
Accuracy Prediction	70%		76%	



(a) (b)

Figure 4: Comparison of selected parameters with (a) Decision Tree regression model (b) Random Forest regression models.

5. Conclusion

State wise selection for the analysis and screening of COVID-19 is done with two regression models. Total Confirmed cases, Total cured and Total death have been identifies in the selected states and analysis is presented in above mentioned figures. Finally it is concluded that Decision tree regression model produces 70% accuracy while the Random Forest regression model produces 76% accuracy with prediction. On the bases of this proposed model if in near

future if third wave or next variant of COVID-19 will be observed in India then we are able to predict confirm cases, death rate and cured cases by prediction model.

Conflict of Interest
The author confirms that there is no conflict of interest to declare for this publication.

Acknowledgments
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors would like to thank the editor and anonymous reviewers for their comments that help improve the quality of this work.

References

Acaps (2020). COVID19 Government Measures Dataset. acaps. [2020-04-09]. https://www.acaps.org/covid19-government-measures-dataset

Bhatnagar, T., Murhekar, M. V., Soneja, M., Gupta, N., Giri, S., Wig, N., & Gangakhedkar, R. (2020). Lopinavir/ritonavir combination therapy amongst symptomatic coronavirus disease 2019 patients in India: Protocol for restricted public health emergency use. *The Indian journal of medical research*, *151*(2-3), 184.

Dey S., & Sinha S. (2020). Covid-19: govt bans all international flights to India from March 22 to 29. *The Times of India*. 2020. Mar 20, [2020-03-28]. https://timesofindia.indiatimes.com/india/govt-bans-all-international-flights-to-india-from-march-22-to-29/articleshow/74720923.cms

Durairaj, M., & Ramasamy, N. (2016). A comparison of the perceptive approaches for preprocessing the data set for predicting fertility success rate. *Int. J. Control Theory Appl*, *9*(27), 255-260.

Hindustan Times (2020). Coronavirus update: has Covid-19 entered Stage 3? Experts, government disagree. *Hindustan Times.* 2020. [2020-04-01]. https://www.hindustantimes.com/india-news/has-covid-19-entered-stage-3-experts-government-disagree/story-u22337reY9uO1ZSeHPIUiK.html

India Today (2020). Is India entering stage 3 of covid-19 outbreak? *India Today*. 2020. Mar 30, [2020-04-01]. https://www.indiatoday.in/programme/to-the-point/video/is-india-entering-stage-3-of-covid-19-outbreak-1661544-2020-03-30

National Herald (2020). Stage 3 begins, we are running out of time: Covid-19 task force. *National Herald*. 2020. Mar 28, [2020-04-01]. https://www.nationalheraldindia.com/flick-past/stage-3-begins-we-are-running-out-of-time-covid-19-task-force

Resistrar (2020). Area and population. *Office of the Registrar General & Census Commissioner, India*. [2020-03-28]. https://censusindia.gov.in/Census_And_You/area_and_population.aspx

Shanthi, D., Sahoo, G., & Saravanan, N. (2009). Designing an artificial neural network model for the prediction of thrombo-embolic stroke. *International Journals of Biometric and Bioinformatics (IJBB)*, *3*(1), 10-18.

Shouman, M., Turner, T., & Stocker, R. (2012). Integrating decision tree and k-means clustering with different initial centroid selection methods in the diagnosis of heart disease patients. In *Proceedings of the International Conference on Data Science (ICDATA)* (p. 1). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).

Unnithan, P. S. G. (2020). Kerala reports first confirmed coronavirus case in India. *India Today*.

Viboud, C., Simonsen, L., & Chowell, G. (2016). A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks. *Epidemics*, *15*, 27-37.

Welfare (2020). *Ministry of Health and Family Welfare.* [2020-04-21]. https://www.mohfw.gov.in/

WHO (2019). Novel corona virus (2019-nCoV) situation report - 1. *World Health Organization*. 2020. Jan 21, [20200405] https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf

Wiki (2020). Coronavirus pandemic in India. Wikipedia. [2020-04-21]. https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_India#Statistics

Zhu, W., Xie, K., Lu, H., Xu, L., Zhou, S., & Fang, S. (2020). Initial clinical features of suspected coronavirus disease 2019 in two emergency departments outside of Hubei, China. *Journal of medical virology*, *92*(9), 1525-1532.