COVID-19 epidemic: analysis and prediction

Santosini Bhutia, Bichitrananda Patra, Mitrabinda Ray
Department of Computer Science and Engineering, Siksha O Anusandhan (Deemed to be) University, Bhubaneswar, Odisha, India

Article Info

Article history:

Received Jul 31, 2021
Revised Feb 11, 2022
Accepted Feb 27, 2022

Keywords:

COVID-19
Ensemble method
Machine learning
Prediction
Regression

ABSTRACT

"Novel Coronavirus", commonly known as COVID-19 has spread nearly to the entire world. The number of impacted cases and deaths has increased significantly in each country, posing a challenge for the world's health organizations. The goal of this paper was to better comprehend and analyze the growth of the disease in India, including confirmed, recovered, fatalities, and active cases of COVID-19. Data analysis affects an organization's decision-making process with interactive visual representation. The proposed model was an ensemble model that was built using linear regression, polynomial regression, and support vector machine (SVM) regression models. The model predicted the number of confirmed cases from 30th May 2021 to 15th June 2021 based on the data available from 22 January 2020 to 29 May 2021 and improved accuracy was obtained when compared with the actual data. Forecasting the confirmed cases might assist health organizations in planning medical facilities. Following that, an appropriate machine leraning (ML) model must be found that can predict the number of new cases in the future.

Corresponding Author:

Bichitrananda Patra
Department of Computer Science and Engineering, Siksha 'O' Anusandhan (Deemed to be) University
J-15, Khandagiri Marg, Dharam Vihar, Jagamara, Bhubaneswar, Odisha 751030, India
Email: bichitranandapatra@soa.ac.in

1. INTRODUCTION

The coronavirus virus (CoV) is a particular disease virus and enhances the existing disease in the human body, making it a very dangerous virus [1], [2]. The virus causes headaches, breathability, maldigestion, and liverwort and harms animals such as cows, horses, and pigs kept, rose, and used by people, as well as various wild animals [3]. The severe acute respiratory syndrome (SARS) epidemic and the explosion of the middle east respiratory syndrome (MERS) showed in 2002-2003 the probability of the newly transmitted human, animal, and human coronavirus disease (COVID-19) in humans, and vice versa [4]. Although such cases are much rarer, they do exist. The effect of secret pneumonia on the entire world is a remarkable subject in late December 2019 [5]. On January 30, 2020, India announced the first case of coronavirus disease COVID-19 [6]. In India, 247,857 cases were reported, 119,293 of which were recovered and 6,954 were dead by 6th June 2020. After that date, new cases, which number about 10,000, continue to come to light every day. All of these details are accurately given to us on the website [7].

In Huanan Market, Wuhan, China, COVID-19 was the first case reported [8]. The main reason for this virus' spreading is the animal-to-human transmission. Yet the next COVID-19 cases were not related to the method of subjection. Therefore, it has been concluded that human-to-human transmission and the primary recurring reason for COVID-19 spread is people with viruses that are indicative. The probability that COVID-19 will be transmitted appears to be very rare before symptoms progress, although the virus cannot be prohibited from transmission. In addition, each person is advised that people who are symptomatic and asymptomatic may pass the virus, and the only way to be safe with this virus is by social distance. Rhinovirus and influenza, as well as additional wheezing bacteria, are believed to be the main reason for the virus spreading the droplets of a person's sneeze and cough [9].

There is currently no specific antiviral therapy isolated and reassuring to COVID-19. The effects of interferon (IFN) and ribavirin recombination on COVID-19 are very small. Several valuable efforts to develop new CoV protease, polymerase, and entry protein antivirals were undertaken. However, a few of them in clinical trials have proven worthwhile. The fact of their benefit in COVID-19 treatment has shown that patients who are recovered from COVID-19 can give plasma and antibodies. In addition, various vaccine schemes such as the use of disabled virus live attenuated viruses, a viral vector vaccine, subunit injection, recombinant protein, and DNA vaccines were developed [10]. Until now no effective COVID-19 injection or therapy has been provided, but the best measures are to monitor the source of infection, early diagnosis, reporting, isolating, supporting therapies, and to keep out the inherent anxiety on time [11]. Every person will benefit from the blocking of the COVID-19 virus or its inflammation from good exclusive hygiene, a formed and suitable mask, ventilation, and keeping away from crowded areas [12]. The major contribution of this paper is:

- A data-driven predictive analysis of the COVID-19 among different states in India. The analysis is done after pre-processing the data such as handling missing values and reduction of redundant data.
- Proposed a novel ensemble predictive model using linear regression, polynomial regression, and support vector machine (SVM) regression models. The model is predicting the number of confirmed cases from 30th May 2021 to 15th June 2021on the data available from 22 January 2020 to 29 May 2021 in India.

The rest of this paper is organized as some literature published in the area of analysis and prediction of COVID-19 is presented in section 2. The detailed methodology including data collection, data pre-processing and feature reduction, data visualization, and all the regression models with the proposed novel ensemble regression model are described in section 3. Thereafter results and discussion of the proposed model are presented in the section 4 followed by a conclusion in section 5.

2. RELATED WORK

Researchers, scientists, and medical professionals are executing a multitude of studies on COVID-19 to develop various types of models for the prediction of COVID-19. Some of these existing publications are addressed in this section. Yadav *et al.* [13] focused on the prediction of the transmission of the COVID-19 virus and the scenario of the spread. A novel support vector regression model is proposed instead of a simple regression line to obtain better classification accuracy and the result is compared with simple linear regression and polynomial regression model. The model predicts the spread of coronavirus, analyses the growth rate, transmission rate, number of recoveries, and the correlation between coronavirus and weather. Khanday *et al.* [14] proposed various classical and ensemble ML techniques which are used for classification after the feature engineering process to a better understanding of the viral spread of COVID-19 and the performance is measured in terms of accuracy, precision, recall, and F1 score. Rahimi *et al.* [15] designed mathematical models for COVID-19 based on susceptible, infected, and recovered (SIR) cases and susceptible, exposed, infected, quarantined, and recovered (SEIQR) cases with some parameter's settings and optimization algorithms. The optimized SIR and SEIQR models were also compared with ML models and the results demonstrate optimized SIR and SEIQR model performs better. Tiwari *et al.* [16] predicted the number of confirmed, recovered, and death cases of COVID-19 in India by using the machine learning approach that was employed by the Chinese pattern. They concluded that the growth rate will be higher in the 3rd and 4th week of April 2020 and is controlled at the end of May 2020. Tomar *et al.* [17] proposed a data-driven model using the long short-term memory (LSTM) technique and curve fitting for the forecasting of 90 days of confirmed and recovered cases of COVID-19 in India. They also analysed the impact of social distancing and lockdown on the spread of the virus. Wang *et al.* [18] proposed a hybrid predictive model based on the logistic model and FbProphet model by concluding the infection rate will be in pick by late October 2020. The logistic model is used to fit the cap of the trend of the epidemic and then fed to the FbProphet model to derive the curve and trend of the epidemic globally. Tuli *et al.* [19] proposed an improved mathematical model using machine learning and cloud computing to analyse and predict the growth of the epidemic globally. They have taken 5 sets of a global dataset of daily confirmed cases of COVID-19 to find the best fitting distribution model. Finally, they found the 5 best distributions from which Robust Weibull using an iteratively weighted approach performed better than others. They have also identified future research directions and emerging trends in their research. From these studies, it can be concluded that the COVID-19 virus is much more similar to SARS and MERS virus, and the infection rate is higher than the fatality rate. Moreover, researchers are continuously working to build the models for prediction and forecasting of the coronavirus and most of the models are built using machine learning.

3. METHODOLOGY

The current section holds the dataset description along with the methodologies used for prediction. The flowchart of the adopted methodology, which includes analysis and prediction of confirmed COVID-19 cases in India, is shown reflected in Figure 1. The COVID-19 dataset of India is collected in the first step. The data is then pre-processed and features reduced, followed by data visualization. The existing regression models including linear regression [20], [21], polynomial regression [22], [23], and SVM regression [24], [25] are implemented to predict the confirmed cases over the next 17 days. Finally, an ensemble regression model for prediction is proposed, which outperforms the existing regression models.

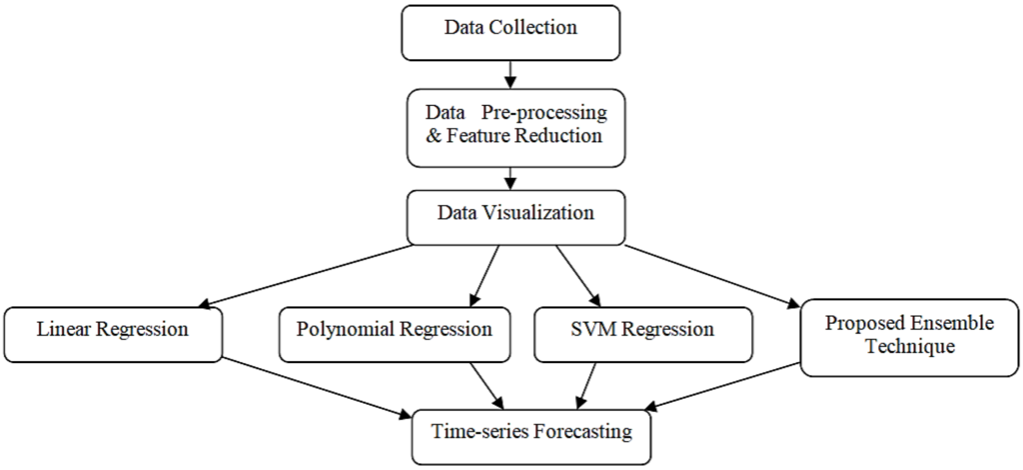

Figure 1. Flowchart of the methodology

3.1. Data collection

The dataset used in this study is collected from Johns Hopkins University Centre for Systems Science and Engineering (JHU CSSE). The dataset of COVID-19 in India is downloaded in Comma separated values (CSV) file format with 14,654 samples. The data is taken from 22 January 2020 to 29 May 2021 which consists of date, time, state/union territory, ConfirmedIndianNational, ConfirmedForeignNational, cured, deaths, and confirmed as input variables that are described in Table 1. It is also observed that the numbers of reported new cases are increasing with time.

Table 1. Description of the features of the COVID-19 dataset in India

Feature Name	Description of the Feature
Date	Date represents the observation date on which how many numbers of COVID-19 positive cases have been reported in India
Time	Time represents the time of the observed date on which COVID-19 positive cases are reported
State/Union Territory	State/Union Territory represents the name of the state or union territory in India where the COVID-19 cases were reported
ConfirmedIndianNational	ConfirmedIndianNational represents the number of COVID-19 cases that originate in India
ConfirmedForeignNational	ConfirmedForeignNational represents the number of COVID-19 cases found in India but originated from foreign countries.
Cured	Cured represents the total number of recovered cases of COVID-19 in India till the observed date
Deaths	Deaths represent the total number of deaths in COVID-19 in India till the observed date
Confirmed	Confirmed represents the total number of confirmed cases of COVID-19 in India till the observed date

3.2. Data pre-processing and feature reduction

In this study, the dataset is analysed in Jupyter Notebook with Python 3 software by importing the corresponding libraries. Data pre-processing is done by converting the Observation Date to date-time format and feature reduction is done by dropping two features which are ConfirmedIndianNational and ConfirmedForeignNational columns. Those two features are dropped because they affected the dataset only at the beginning of the virus spread. But later travel is banned and the virus spread is community transferable. Further analysis is done state-wide on the dataset of India which is one of the most populated countries in the world.

3.3. Data visualization

The graphical representation of the growth of COVID-19 disease across the country is shown in Figure 2 and it is an exponential growth in the reported cases. It is observed that the spread of the virus is very less in the first two months of the beginning of the outbreak, but later the spread is very high. Figures 3 to 5 show the total number of confirmed cases, recovered, and deaths respectively from the different states and union territories in India. Maharashtra, Karnataka, and Kerala are the three most affected states in India. The objective of the study is to understand and visualize the outbreak of COVID-19 disease. Figure 6 gives the detailed information of confirmed cases, cured, deaths, active cases, death rate, and cured rate among different states and union territories in India which is sorted according to the confirmed cases. The more the number of cases the higher the intensity of the colour.

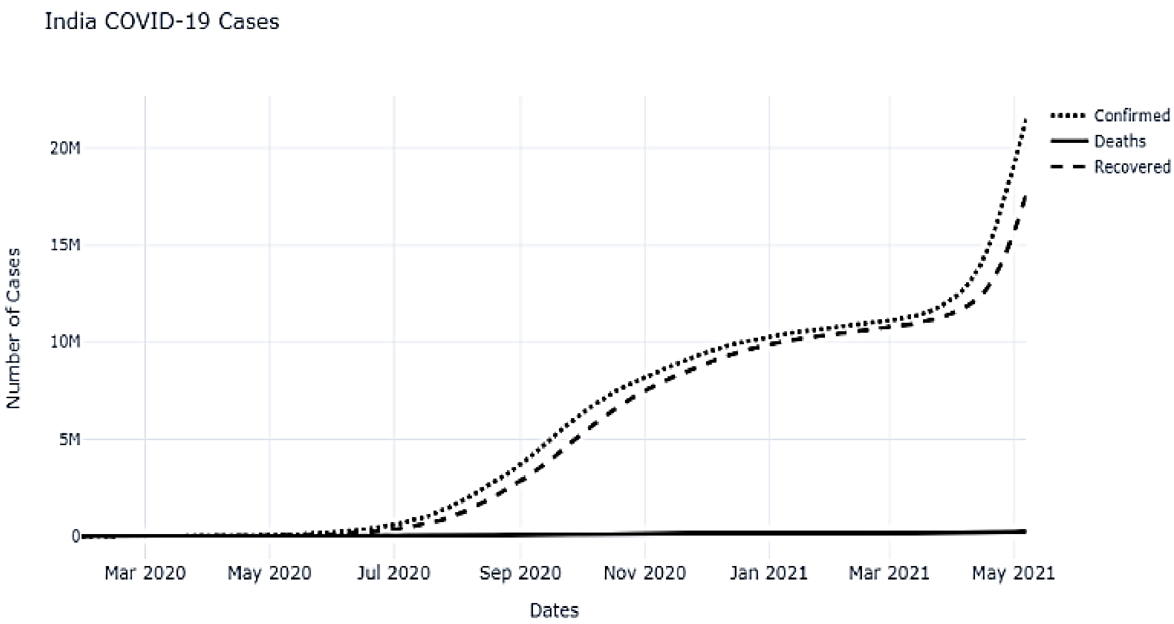

Figure 2. Spread of COVID-19 across the country

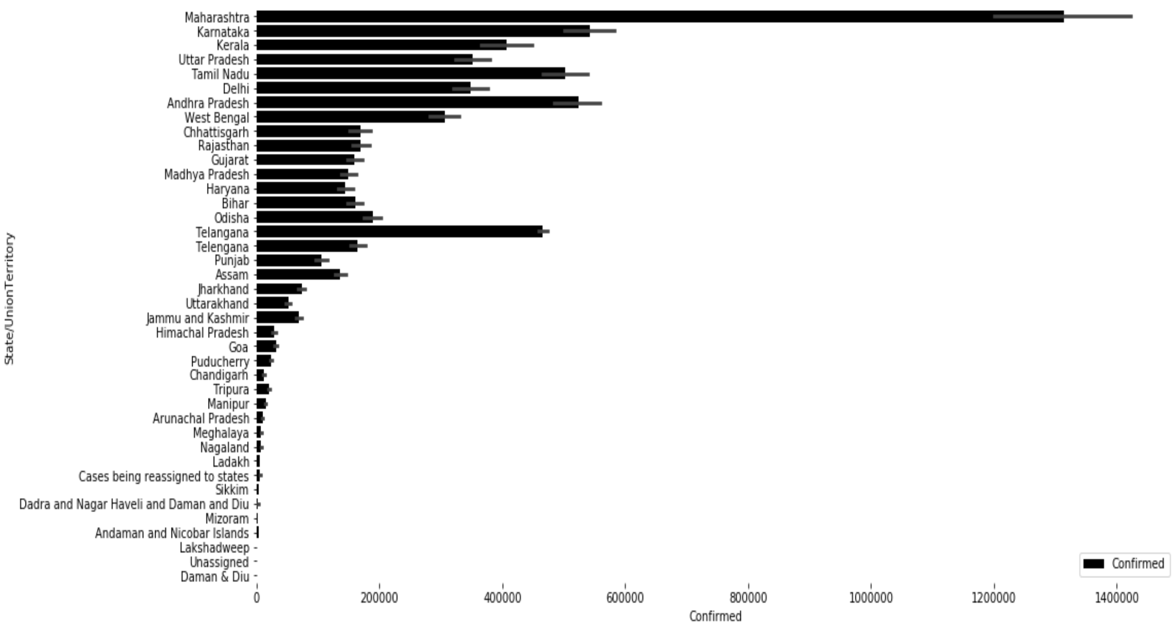

Figure 3. Confirmed cases of COVID-19

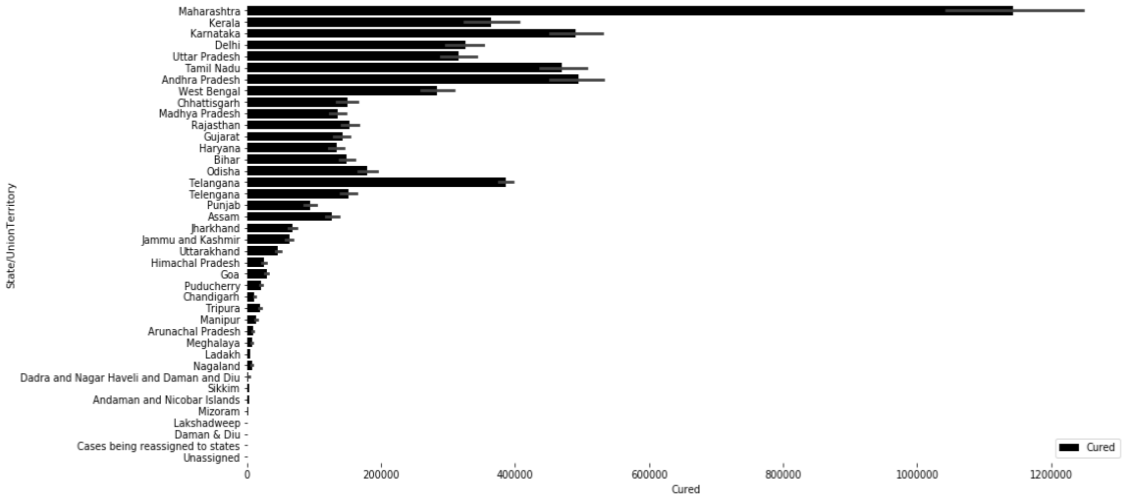

Figure 4. Recovered cases of COVID-19

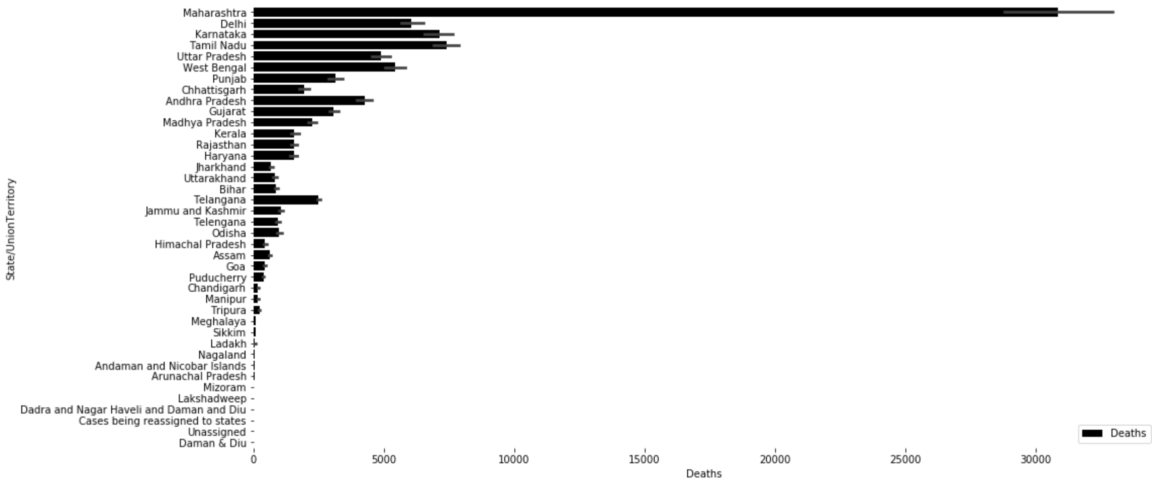

Figure 5. Deaths cases of COVID-19

	State/UnionTerritory	Confirmed	Deaths	Cured	Active	Death Rate (per 100)	Cure Rate (per 100)
22	Maharashtra	4942736	73515	4227940	641281	1.490000	85.540000
17	Karnataka	1790104	17212	1255797	517095	0.960000	70.150000
18	Kerala	1786396	5628	1389515	391253	0.320000	77.780000
37	Uttar Pradesh	1425916	14501	1151571	259844	1.020000	80.760000
32	Tamil Nadu	1297500	14974	1151058	131468	1.150000	88.710000
10	Delhi	1273035	18398	1164008	90629	1.450000	91.440000
1	Andhra Pradesh	1228186	8446	1037411	182329	0.690000	84.470000
39	West Bengal	935066	11964	800328	122774	1.280000	85.590000
7	Chhattisgarh	816489	9950	675294	131245	1.220000	82.710000
30	Rajasthan	702568	5182	499376	198010	0.740000	71.080000
12	Gujarat	645972	8035	490412	147525	1.240000	75.920000
21	Madhya Pradesh	637406	6160	542632	88614	0.970000	85.130000
13	Haryana	573815	5137	452836	115842	0.900000	78.920000
4	Bihar	553803	3077	435574	115152	0.560000	78.650000
27	Odisha	500162	2121	423257	74784	0.420000	84.620000
33	Telangana	481640	2625	405164	73851	0.550000	84.120000
34	Telengana	443360	2312	362160	78888	0.520000	81.690000
29	Punjab	416350	9979	339803	66568	2.400000	81.610000
3	Assam	277687	1531	242980	33176	0.550000	87.500000
16	Jharkhand	270089	3479	205977	60633	1.290000	76.260000
38	Uttarakhand	220351	3293	154147	62911	1.490000	69.960000
15	Jammu and Kashmir	201511	2562	157283	41666	1.270000	78.050000

Figure 6. State/union territory-wide analysis of COVID-19 in India

3.4. Proposed ensemble model

Ensemble learning is the process of integrating several individual models to improve the model's accuracy and performance. Algorithm-1 explains the working principles of the proposed ensembled model. The proposed ensemble model is performed for the available data from 22 January 2020 to 29 May 2021. It is assumed that the coronavirus infected person can transmit the virus to another person directly or indirectly as it is a transmitted disease and consequently the number of reported cases is growing rapidly. Hence, we are developing an ensemble machine learning model based on Linear Regression, Polynomial Regression, and SVM Regression. The data is split into 80% and 20% for training and testing purposes respectively. While developing the model the dataset was analyzed using the functions available in Python. Table 2 shows the methods, packages, and parameters used by these models for the prediction of confirmed cases of COVID-19 in India in the Python programming language. The linear regression, polynomial regression, and SVM regression models are generated for the forecasting of confirmed cases over the next 17 days in India. Finally, an ensemble model is proposed by taking weighted average predictions. This ensemble model combines the prediction from each statistical model with a weighted value proportionally and the predicted values are collected.

Table 2. Description of machine learning models

Model	Method	Required Package	Tuning Parameter
Linear Regression	LinearRegression()	LinearRegression	Normalize=True
Polynomial Regression	PolynomialFeatures()	PolynomialFeatures	Degree=8
SVM Regression	SVR()	SVR	C=1,degree=6,kernel='poly', epsilon=0.01

Algorithm-1 for proposed ensemble model

```
Start
Load the dataset with a CSV file. The data is taken from 22 January 2020 to 29 May 2021.
for i in range(1,18):
    new_date.append(datewise_india.index[-1]+timedelta(days=i))
    new_prediction_lr.append(lin_reg.predict(np.array(datewise_india["DaysSince"].max()+i).
    reshape(-1,1))[0][0])
    new_prediction_poly.append(poly.predict(np.array(datewise_india["DaysSince"].max()+i).r
    eshape(-1,1))[0][0])
    new_prediction_svm.append(svm.predict(np.array(datewise_india["Days
    Since"].max()+i).reshape(-1,1))[0][0])
The forecast data is obtained from 30th May 2021 to 15th June 2021.
The result is displaying in
    model_predictions=pd.DataFrame(zip(new_date,new_prediction_lr,new_prediction_poly,new_p
    rediction_svm), columns=["Dates","Linear Regression Prediction","Polynonmial Regression
    Prediction","SVM Regression Prediction"])
Proposed ensemble model_predictions
     final_pred =[]
    for i in range(0,17):
        x = np.sum(prediction_lr[i]*0.5+prediction_poly[i]*0.07+prediction_svm[i]*0.6)
        final_pred.append(x)
End
```

4. RESULT ANALYSIS AND COMPARISON OF ACTUAL AND PREDICTED CASES

The pandemic has had a significant impact on world health and the economy. In this study, a total of 14,653 number samples of the COVID-19 dataset of India are taken into consideration. The forecast models are built using linear regression, polynomial regression, and SVM regression. The graphical representations of the models are shown in Figures 7 to 9 respectively. The training data for confirmed cases is represented by the gray line, while the best fit line for the corresponding model is represented by the black dot line. The formulation of root means square error (RMSE) values evaluated after the prediction of confirmed cases, recovered, and deaths of the models and is presented in Table 3. RMSE is the prediction error that tells us how concentrated the data is around the best fit line. It is commonly used for forecasting and regression analysis to verify the experimental result. The lower the value of RMSE the better is the model. In this study, it is found that the SVM Regression model gives a lower RMSE value for the confirmed cases. Hence, the better the model is as compared to the other two models. But for the number of recovered cases and deaths polynomial regression model gives little better results compared to the SVM regression model.

$$RMSE = \sqrt{\overline{(F - O)^2}} \quad (1)$$

Where F=forecasts or expected values and O=observed values.

COVID-19 analysis is carried out in India from 22 January 2020 to 29 May 2021. An ensemble model is proposed in this paper, which used a weighted average of predictions. This novel ensemble model proportionally combines each statistical model's forecast with a weighted value. The proposed models will forecast over the next 17 days that is from 30th May 2021 to 15th June 2021 confirmed cases in India which are presented in Table 4. When we compare the predicted data to the actual data, the ensemble regression model has a higher percentage of accuracy than the other regression models provided in Table 5. These predicted values are compared to actual data, and the model's accuracy is assessed using (2). The ensemble model outperformed the statistical prediction models, according to the findings.

$$\%Acc = 100 - \left(\frac{Actual\ Value - Predicted\ Value}{Actual\ Value} * 100\right) \quad (2)$$

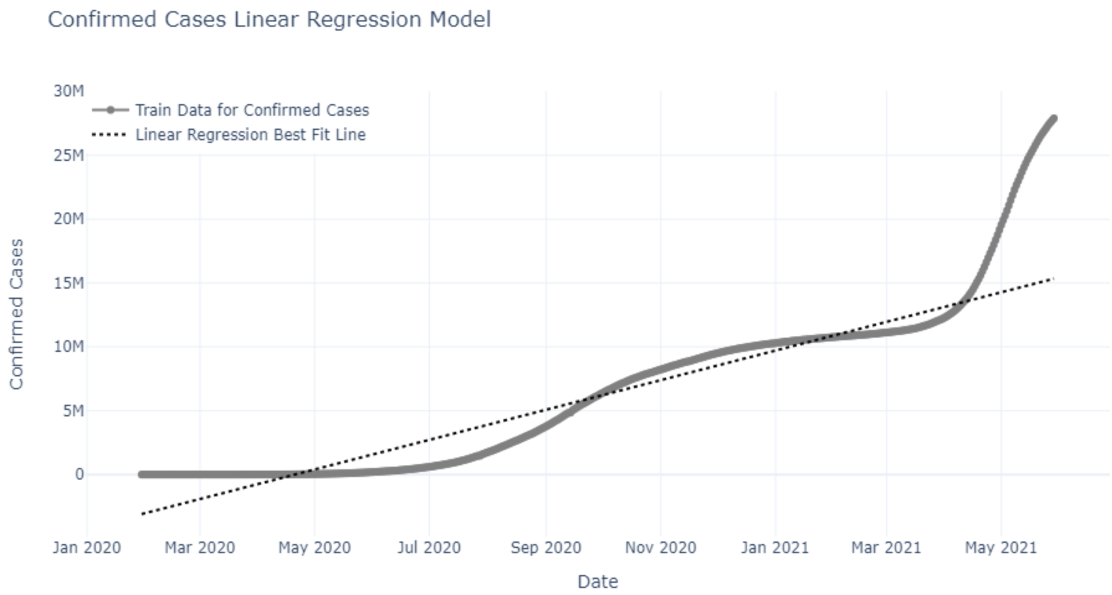

Figure 7. Linear regression model for confirmed cases

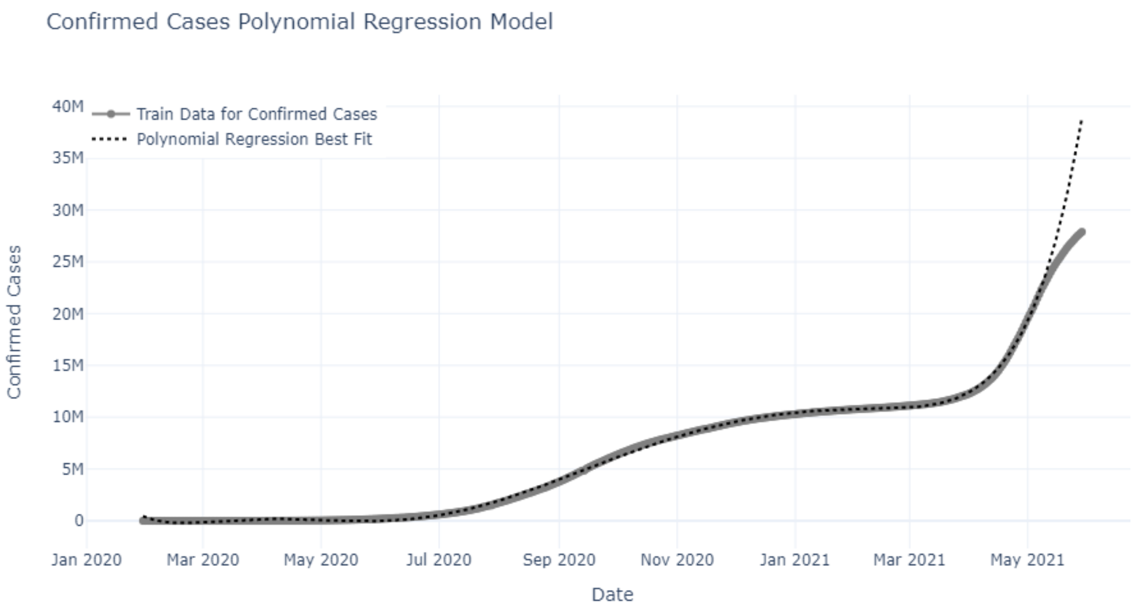

Figure 8. Polynomial regression model for confirmed cases

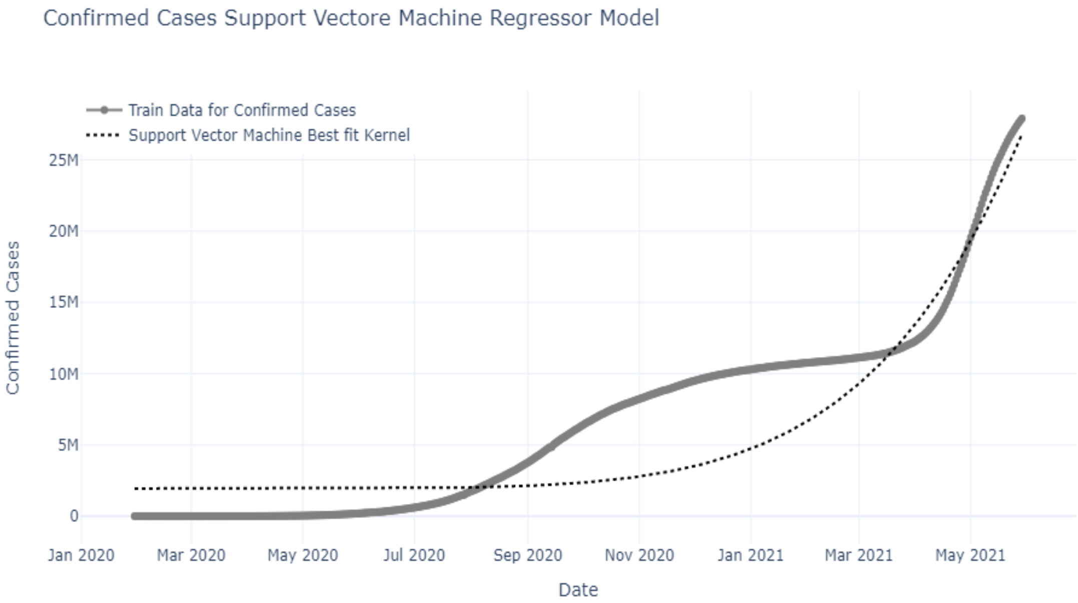

Figure 9. SVM regression model for confirmed cases

Table 3. RMSE of the models in millions

	Linear Regression	Polynomial Regression	SVM Regression
Confirmed	10.24×10^6	5.12×10^6	1.60×10^6
Recovered	8.14×10^6	0.77×10^6	0.85×10^6
Death	0.07×10^6	0.01×10^6	0.02×10^6

Table 4. Predicted values of the models

Sl no.	Date	Actual Data	Linear Regression	Polynomial Regression	SVM Regression	Ensemble Technique
1	30-05-2021	28047000	15375267	40035027	27103831	26752384
2	31-05-2021	28173883	15413275	41177458	27416098	27038718
3	01-06-2021	28307035	15451283	42357143	27731587	27329594
4	02-06-2021	28441079	15489290	43575080	28050325	27625096
5	03-06-2021	28573503	15527298	44832289	28372339	27925313
6	04-06-2021	28693957	15565306	46129813	28697656	28230334
7	05-06-2021	28808445	15603313	47468717	29026303	28540249
8	06-06-2021	28909654	15641321	48850089	29358306	28855150
9	07-06-2021	28995458	15679329	50275040	29693694	29175134
10	08-06-2021	29088245	15717336	51744706	30032493	29500294
11	09-06-2021	29182128	15755344	53260244	30374732	29830729
12	10-06-2021	29273977	15793352	54822838	30720439	30166538
13	11-06-2021	29358551	15831359	56433696	31069641	30507823
14	12-06-2021	29439076	15869367	58094052	31422367	30854687
15	13-06-2021	29510077	15907375	59805163	31778645	31207236
16	14-06-2021	29570085	15945382	61568316	32138503	31565575
17	15-06-2021	29632302	15983390	63384821	32501971	31929815

Table 5. The Accuracy of the models

Date	% Of Accuracy of Linear Regression	% Of Accuracy of polynomial Regression	% Of Accuracy of SVM Regression	% Of Accuracy of the ensemble method	Date	% Of Accuracy of Linear Regression	% Of Accuracy of polynomial Regression	% Of Accuracy of SVM Regression	% Of Accuracy of the ensemble method
30-05-2021	54.18	57.25	96.63	95.39	07-06-2021	54.07	26.61	97.59	99.38
31-05-2021	54.70	53.84	97.31	95.98	08-06-2021	54.03	22.11	96.75	98.58
01-06-2021	54.58	50.36	97.96	96.55	09-06-2021	53.98	17.49	95.91	97.77
02-06-2021	54.46	46.78	98.62	97.14	10-06-2021	53.95	15.36	95.07	96.95
03-06-2021	54.34	43.09	99.29	97.74	11-06-2021	53.92	12.72	94.17	96.08
04-06-2021	54.24	39.23	99.98	98.39	12-06-2021	53.90	10.66	92.31	95.19
05-06-2021	54.16	35.22	99.24	99.07	13-06-2021	53.90	9.68	92.31	94.24
06-06-2021	54.10	31.02	98.44	99.82	14-06-2021	53.92	8.68	91.31	93.25
07-06-2021	54.07	26.61	97.59	99.38	15-06-2021	53.93	9.68	90.31	92.24

A reliable method, regression analysis is to determine the relationship between independent variable date (x) and dependent variable confirmed cases (y). Linear regression, Polynomial regression, and SVM regression are some of the known regression methods that are easy to implement and relatively efficient. These approaches are used to provide predictions for confirmed cases of COVID-19. Thereafter an ensemble model is proposed which outperformed the existing models as the predicted data is very close to the actual data.

5. CONCLUSION AND FUTURE SCOPE

In this study, the spread of COVID-19 in different Indian states is discussed, and an ensemble model employing linear regression, polynomial regression, SVM regression is proposed and experimentally verified for forecasting confirmed COVID-19 cases in India. All of the models are assessed for correctness here. When the models are compared, it is observed that the ensemble model provides more accurate predicted values for time series data forecasting than the other models. According to the finding, substantially more COVID-19 restriction requirements are needed to control the spread of the disease. The prediction could aid in healthcare decision-making, and proactive measures could be made to decrease human life loss. The proposed ensemble model can be extended for the prediction of recovery and fatalities in a certain location.

REFERENCES

[1] S. Kooraki, M. Hosseiny, L. Myers, and A. Gholamrezanezhad, "Coronavirus (COVID-19) outbreak: what the department of radiology should know," *Journal of the American College of Radiology*, vol. 17, no. 4, pp. 447–451, Apr. 2020, doi: 10.1016/j.jacr.2020.02.008.

[2] M. Smith and F. Alvarez, "Identifying mortality factors from machine learning using shapley values – a case of COVID19," *Expert Systems with Applications*, vol. 176, Aug. 2021, Art. no. 114832, doi: 10.1016/j.eswa.2021.114832.

[3] O. P. Singh, M. Vallejo, I. M. El-Badawy, A. Aysha, J. Madhanagopal, and A. A. Mohd Faudzi, "Classification of SARS-CoV-2 and non-SARS-CoV-2 using machine learning algorithms," *Computers in Biology and Medicine*, vol. 136, Sep. 2021, doi: 10.1016/j.compbiomed.2021.104650.

[4] S. Rajendran and P. Jayagopal, "Accessing Covid19 epidemic outbreak in Tamilnadu and the impact of lockdown through epidemiological models and dynamic systems," *Measurement*, vol. 169, Feb. 2021, doi: 10.1016/j.measurement.2020.108432.

[5] N. S. Punn, S. K. Sonbhadra, and S. Agarwal, "COVID-19 epidemic analysis using machine learning and deep learning algorithms," *medRxiv*, Apr. 2020, doi: 10.1101/2020.04.08.20057679.

[6] C. Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *The Lancet*, vol. 395, no. 10223, pp. 497–506, Feb. 2020, doi: 10.1016/S0140-6736(20)30183-5.

[7] S. M. Saleem, R. Quansar, I. Haq, and S. M. Salim Khan, "Imposing COVID-19 lockdown and reported dog bite cases: an experience from a tertiary antirabies center of North India," *Journal of Veterinary Behavior*, vol. 42, pp. 48–52, Mar. 2021, doi: 10.1016/j.jveb.2021.01.007.

[8] D. S. Hui *et al.*, "The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health — the latest 2019 novel coronavirus outbreak in Wuhan, China," *International Journal of Infectious Diseases*, vol. 91, pp. 264–266, Feb. 2020, doi: 10.1016/j.ijid.2020.01.009.

[9] C. Menni *et al.*, "Loss of smell and taste in combination with other symptoms is a strong predictor of COVID-19 infection," *Nature Medicine*, Apr. 2020, [Online]. Available: https://doi.org/10.1101%2F2020.04.05.20048421.

[10] B. Patra, S. Bhutia, and N. Panda, "Machine learning techniques for cancer risk prediction," *TEST Engineering and Management*, vol. 83, pp. 7414–7420, 2020.

[11] K. Bhanu Prakash and M. A. Dorai Rangaswamy, "Content extraction studies using neural network and attribute generation," *Indian Journal of Science and Technology*, vol. 9, no. 22, pp. 1–10, Jun. 2016, doi: 10.17485/ijst/2016/v9i22/95165.

[12] A. Remuzzi and G. Remuzzi, "COVID-19 and Italy: what next?," *The Lancet*, vol. 395, no. 10231, pp. 1225–1228, Apr. 2020, doi: 10.1016/S0140-6736(20)30627-9.

[13] M. Yadav, M. Perumal, and M. Srinivas, "Analysis on novel coronavirus (COVID-19) using machine learning methods," *Chaos, Solitons and Fractals*, vol. 139, Oct. 2020, Art. no. 110050, doi: 10.1016/j.chaos.2020.110050.

[14] A. M. U. D. Khanday, S. T. Rabani, Q. R. Khan, N. Rouf, and M. Mohi Ud Din, "Machine learning based approaches for detecting COVID-19 using clinical text data," *International Journal of Information Technology*, vol. 12, no. 3, pp. 731–739, Sep. 2020, doi: 10.1007/s41870-020-00495-9.

[15] I. Rahimi, A. H. Gandomi, P. G. Asteris, and F. Chen, "Analysis and prediction of COVID-19 using SIR, SEIQR, and machine learning models: Australia, Italy, and UK aases," *Information*, vol. 12, no. 3, Mar. 2021, doi: 10.3390/info12030109.

[16] S. Tiwari, S. Kumar, and K. Guleria, "Outbreak trends of Coronavirus Disease–2019 in India: a prediction," *Disaster Medicine and Public Health Preparedness*, vol. 14, no. 5, pp. e33–e38, Oct. 2020, doi: 10.1017/dmp.2020.115.

[17] A. Tomar and N. Gupta, "Prediction for the spread of COVID-19 in India and effectiveness of preventive measures," *Science of The Total Environment*, vol. 728, Aug. 2020, doi: 10.1016/j.scitotenv.2020.138762.

[18] P. Wang, X. Zheng, J. Li, and B. Zhu, "Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics," *Chaos, Solitons and Fractals*, vol. 139, p. 110058, Oct. 2020, doi: 10.1016/j.chaos.2020.110058.

[19] S. Tuli, S. Tuli, R. Tuli, and S. S. Gill, "Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing," *Internet of Things*, vol. 11, Sep. 2020, doi: 10.1016/j.iot.2020.100222.

[20] R. O. Ogundokun, A. F. Lukman, G. B. M. Kibria, J. B. Awotunde, and B. B. Aladeitan, "Predictive modelling of COVID-19 confirmed cases in Nigeria," *Infectious Disease Modelling*, vol. 5, pp. 543–548, 2020, doi: 10.1016/j.idm.2020.08.003.

[21] S. Rath, A. Tripathy, and A. R. Tripathy, "Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model," *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, vol. 14, no. 5, pp. 1467–1474, Sep. 2020, doi: 10.1016/j.dsx.2020.07.045.

[22] A. Jaglan, D. Trehan, M. Megha, and P. Singhal, "COVID-19 trend analysis using machine learning techniques," *International*

Journal of Scientific and Engineering Research, vol. 11, no. 12, pp. 1162–1167, Dec. 2020, doi: 10.14299/ijser.2020.12.08.

[23] A. A. Suleiman, A. Suleiman, U. A. Abdullahi, and S. A. Suleiman, "Estimation of the case fatality rate of COVID-19 epidemiological data in Nigeria using statistical regression analysis," *Biosafety and Health*, vol. 3, no. 1, pp. 4–7, Feb. 2021, doi: 10.1016/j.bsheal.2020.09.003.

[24] B. Patra and S. S. Bisoyi, "CFSES optimization feature selection with neural network classification for microarray data analysis," in *2018 2nd International Conference on Data Science and Business Analytics (ICDSBA)*, Sep. 2018, pp. 45–50, doi: 10.1109/ICDSBA.2018.00016.

[25] B. Patra, L. Jena, S. Bhutia, and S. Nayak, "Evolutionary hybrid feature selection for cancer diagnosis," in *Smart Innovation, Systems and Technologies*, vol. 153, Springer Singapore, 2021, pp. 279–287.

BIOGRAPHIES OF AUTHORS

Mrs. Santosini Bhutia is currently working as a Research Scholar in Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India. She received her Bachelor of Technology (B. Tech.) in Information Technology, Master of Technology (M. Tech.) in Computer Science and Engineering, and pursuing her Ph. D. in Computer Science and Engineering. She can be contacted at email: santosinibhutia@gmail.com.

Dr. Bichitrananda Patra is an Associate Professor in the Department of Computer Science Engineering at the Institute of Technical Education and Research, Siksha 'O' Anusandhan (Deemed to be University). Bhubaneswar, Odisha, India. He received his M.Tech. in Computer Science from the Utkal University, Bhubaneswar, and Ph.D. in Computer Science from Berhampur University, Orissa, India. He has published more research papers in international and national journals, conferences, and book chapters in different books and also has membership in different professional bodies like ISTE and CSI. He can be contacted at email: bichitranandapatra@soa.ac.in.

Dr. Mitrabinda Ray is currently working as an Associate Prof. in the Department of Computer Science and Engineering in Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India. She holds her Master of Technology (M. Tech.) and Ph.D. in Computer Science and Engineering from the National Institute of Technology, Rourkela, Odisha, India. She is having more than 15 years of teaching and research experience. She has published more than 25 international Journals. Her research areas of interest include software testing, software reliability analysis. She can be contacted at email: mitrabindaray@soa.ac.in.igure.