Wireless Personal Communications (2023) 130:1963–1991
https://doi.org/10.1007/s11277-023-10367-7

Empirical Analysis of Impact of Weather and Air Pollution Parameters on COVID-19 Spread and Control in India Using Machine Learning Algorithm

Lokesh Kumar Shrivastav[1] · Ravinder Kumar[2]

Accepted: 6 March 2023 / Published online: 16 March 2023

Abstract

The COVID-19 has affected and threatened the world health system very critically throughout the globe. In order to take preventive actions by the agencies in dealing with such a pandemic situation, it becomes very necessary to develop a system to analyze the impact of environmental parameters on the spread of this virus. Machine learning algorithms and artificial Intelligence may play an important role in the detection and analysis of the spread of COVID-19. This paper proposed a twinned gradient boosting machine (GBM) to analyze the impact of environmental parameters on the spread, recovery, and mortality rate of this virus in India. The proposed paper exploited the four weather parameters (temperature, humidity, atmospheric pressure, and wind speed) and two air pollution parameters (PM2.5 and PM10) as input to predict the infection, recovery, and mortality rate of its spread. The algorithm of the GBM model has been optimized in its four distributions for best performance by tuning its parameters. The performance of the GBM is reported as excellent (where R2 = 0.99) in training for the combined dataset comprises all three outcomes i.e. infection, recovery and mortality rates. The proposed approach achieved the best prediction results for the state, which is worst affected and highest variation in the atmospheric factors and air pollution level.

Keywords COVID-19 · Weather parameters · Air pollution · Gradient boosting machine

1 Introduction

After the first reported case of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) in Wuhan, China in December 2019, it spread exponentially covering approximately 215 countries worldwide by 28th June 2021 [1]. According to the WHO's report,

✉ Ravinder Kumar
ravinder_y@yahoo.com

Lokesh Kumar Shrivastav
lokesh.arsd@gmail.com

1 USICT, GGSIPU, New Delhi, India

2 SFET, Shri Vishwakarma Skill University, Gurugram, India

it has infected over 180,654,652 people, and 3,920,463 confirmed deaths globally by 28th June 2021. According to the report of the Ministry of Health and Family Welfare, Government of India, there is a total of 5,72,994 active cases, 29,30,9607 cured and discharged and 3,96,730 deaths by 28th June 2021 [2]. Governments made their all efforts to control the spread of COVID-19 at their level, including lockdown, social distancing measures, personal hygiene, testing, tracking, isolation, and trial of drugs already used for other diseases like malaria, HIV, tuberculosis, etc. Finally, vaccination became the main tool to control the spread of COVID-19. In India total of 32,36,63,297 vaccines are vaccinated of which 4.3% are fully vaccinated and 20% of the population are partially vaccinated upto 28th June 2021 [2].

Despite these all-available precautions, the 2nd surge in India was unexpected and affected a large percentage of the population. 2nd surge of COVID-19 spread started in 1st week of April 2021 and declined after the 1st week of June 2021. In nearly two months, the country started to struggle with inadequate of hospital beds, oxygen cylinders, essential medicines, and vaccines all around the country. On 30 April 2021, India became the first country that reported over 4,00,000 newly infected cases in a very single day (24 h). This unexpected speed of infection created a huge demand for basic essentials.

It has been observed that both spikes were reported during the particular climate conditions in India. Therefore, it becomes too necessary to study the impact of weather and atmospheric factor on the spread of COVID-19. Along with weather parameters the impact of air pollutants is equally, important to analyses its impact on COVID 19.

The initial research talks about the transmission of COVID-19 from bats to humans originating from the seafood market in Wuhan, China [1, 3–5]. However, the scientific exploration of its route of transmission is requisite. The close contact of humans increases its transmission rate rapidly, through the surface and air [6]. In some recent studies, the presence of coronavirus in the air, fecal swabs, and blood of active cases have been informed [7, 8]. The change of climate conditions provides a favorable environment to grow viruses resulting common flu. The particular climate conditions also affect the transmission rate of the pandemic by presenting emergent or hostile conditions for humans. It was confirmed in cases of past infectious diseases as well as in the case of transmission of the present situation of COVID-19 in some countries. Like the transmission rate of influenza was high at the low temperature and humidity. It is also confirmed in the case of severe acute respiratory syndrome (SARS) in July 2003 which affected by climate change [9].

As reported that COVID-19 has a similar genetic sequence to SARS, therefore, it is highly expected that its transmission rate will be affected by the change in weather parameters [10]. The effect of climate factors on the spread of COVID-19 in different countries has been established in some recent studies [11–19]. Besides, the atmospheric factor, air pollution level may also be an affecting factor of the transmission of COVID-19, as reported high-rise of COVID-19 cases in Italy [20]. The effect of concentration of nitrogen oxide on the fatality due to COVID-19 has also been reported in Italy [7]. The effect of lockdown and air pollution level at the spread rate of COVID-19 in Wuhan, China has been also reported [21], etc. Even the atmospheric factors and air pollution levels are highly correlated; the study based on their combined effect on the transmission rate of COVID-19 in India has not been reported yet. The present work tries to cover the combined effect of atmospheric factors and measures of air pollution on the spread of COVID-19 during 28th March 2020 to 20th May 2021 (Exclusively two major surges) period in India.

In the past few years, machine learning has become a very significant tool in the analysis and design of prediction models [22–24]. Many machine-learning models have been designed and applied efficiently in the analysis of COVID-19 cases [25–29]. Apart from the most famous deep learning methods, tree-based learning (extreme gradient boosting machine) was successfully applied to find the associations between microRNAs (miRNAs) and human diseases. This motivates us to design the twined gradient boosting machine (GBM) model to analyze the correlation among atmospheric factors (temperature, humidity, pressure, and wind speed), and air pollution (max and min of PM2.5 and PM10) with the infection, recovery, and death cases of COVID-19 daily in different states or places of India.

This paper proposes the following contributions:

i. The data for the period of 25th March 2020 to 20th June 2021, has been collected, and analyzed to confirm the suitability of the dataset.
ii. The analysis of the impact of atmospheric and air pollutant parameters on the spread of the disease
iii. Analysis of the impact of atmospheric and air pollutant parameters on the recovery rate of the patient
iv. Analysis of the impact of atmospheric and air pollutant parameters on the mortality rate of the disease
v. The worst affected states were analyzed and tested for spread, recovery and mortality rate of COVID-19 separately.

Rest of the paper is organized in the following manner: Sect. 2 describes the process of data collection and its analysis. Section 3 presents the proposed gradient boosting machine (GBM) approach; Experimental setup and results are presented in Sect. 4. The next Sect. 5 discusses the results, and finally Sect. 6 summarizes the critical finding and future research directions in this domain.

2 Data Collection and Analysis

The data of eight atmospheric factors (maximum and minimum temperature, maximum and minimum air pressure, maximum and minimum air humidity, and maximum and minimum wind speed) and four measures of air pollution (maximum and minimum of PM2.5 and PM10) of the 21 significant states or places of India have been collected from the Indian meteorological department (IMD) and Indian central pollution control board (CPCB) during the period of 14^{th} March 2020 to 20th May 2021 on daily basis (433 days) [30, 31]. The cases (number of infected, recovered, and death) of COVID-19 of similar states have been collected from an open-access source and information published by the ministry of health and family welfare, the government of India [32, 33]. The data of some states and union territories were not so significant for COVID-19, so it was not considered at all. The atmospheric factors, measures of air pollution, and cases of COVID-19 were used in combination for further analysis. The missing or doubtful values of the atmospheric factors, air pollution measures for some states at some days were replaced by the previous imputation technique. The variations of minimum and maximum temperature and humidity after imputation are shown in Fig. 1. The minimum and maximum of PM_{10} and $PM_{2.5}$ are shown in Fig. 2. The statistics of the dataset are presented in Table 1.

The variation in cases of COVID-19 after imputation is shown in Fig. 3. Variations of pressure wind speed are presented in the Table 1.

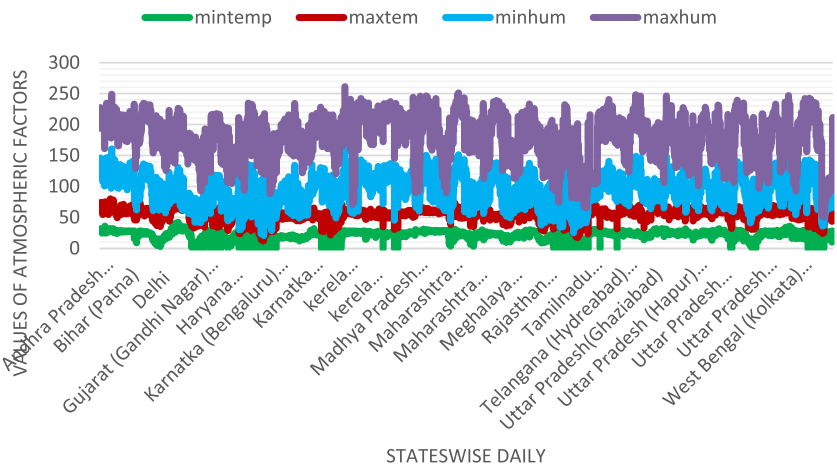

Fig. 1 The variation of the temperature (in.°C), humidity (in %)

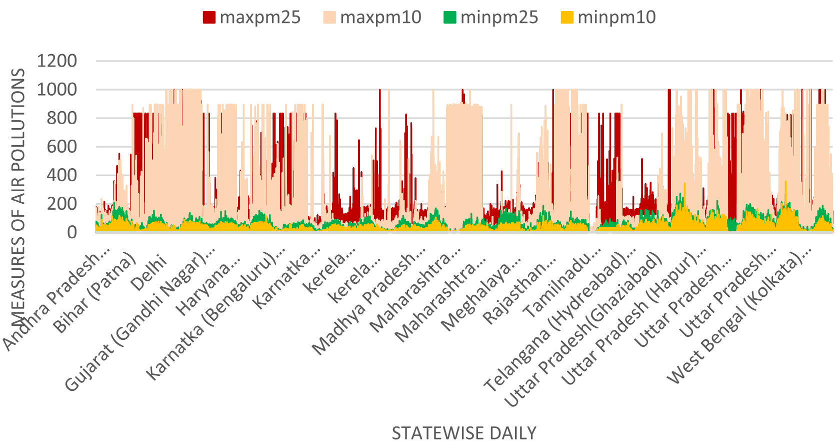

Fig. 2 The variation of PM_{25} and PM_{10} on a daily basis

Eight atmospheric parameters and four measures of air pollution were considered as input in the proposed twined GBM to analyses the correlation and forecast the infected, recovered, and death cases of COVID-19, independently. The total 9,033 instances are taken for the preprocessing that was collected between 14th March 2020 to 20th May 2021 (21 states /places × 433 days). Out of this, 5974 with 17 attributes are taken for the training and remaining 3119 with 17 attributes are taken for the testing. The performance of the proposed GBM was also evaluated by predicting the COVID-19 cases state-wise. The atmospheric factors and air pollution measures were used as input of GBM simultaneously to check their mutual influence on the cases of COVID-19. Moreover, the minimum and the maximum values of the atmospheric factors (temperature, pressure, humidity, and wind speed) used as input of GBM and GBM are suitable in the understanding of their better impact on the distribution of COVID-19 cases. Moreover, to evaluate the impact of air pollution four measures maximum and minimum PM10 and PM2.5 have been included.

Table 1 Basic statistics of atmospheric and air pollution dataset

Atmospheric parameters, pollution measures, and COVID-cases	Statistical measures								
	Minimum	Maximum	Range	First quartile	Third quartile	Median	Mean	Standard deviation	Skewness
Minimum temperature	0	42	0 to 42	18.50	26.40	23.70	21.45	7.43	−1.31
Maximum temperature	11.20	48.50	11.20 to 48.50 44	31	35.80	33.30	33.39	4.29	−0.18
Minimum humidity	0.10	121.50	0.10–121.50	21.70	55	37.50	38.43	20.14	0.14
Maximum humidity	11	114.70	11–114.70	80.20	99	93	85.97	18.25	−1.77
Minimum Pressure	0	1014.20	0–1014.20	745	765.3	741	732.80	224.83	−1.94
Maximum Pressure	3	3174	708–1160	745	1012	1005.5	898.1	198.83	−1.40
Minimum wind speed	0	48	0–48	.10	.30	.37	.20	2.05	20.16
Maximum wind speed	0.30	999.90	0.30 –999.90	2.50	5.90	3.90	14.91	59.63	1.23
MIN PM 2.5	0.1	274	0.1–274	12	67	38	46.4	41.76	1.24
MAX PM 2.5	0.3	999	0.3–999	141	261	171	245.3	201.17	1.24
MIN PM 10	0.10	359	0.10–359	9	46	24	46	28.99	1.68
MAX PM 10	3	999	3–999	68	303	115	238.1	263.77	1.60
Infected	0	68,631	0–68,631	191	3957	1157	4079	8422.27	4.06
Recovered	0	99,651	0–99,651	153	3513	986	3553	7854.13	4.90
Mortality	0	1409	0–1409	2	41	12	45.19	100.61	4.90

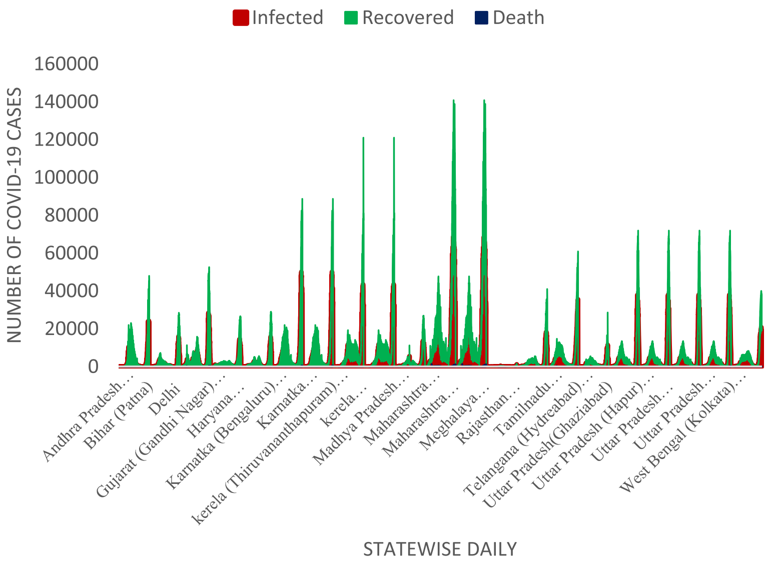

Fig. 3 The variation in COVID-19 cases in India on a daily basis

3 Gradient Boosting Machine (GBM) Approach

The gradient boosting machine (GBM) is an efficient method in regression analysis since it selects the adaptive characteristics of the dataset in the analysis. The optimal values of the predicted variables are obtained in several iterations by using the values of the dependent variable of the previous iteration and average weights. The GBM approach is implemented using the H2O package in R [33]. The basic steps of the GBM approach are described as follows [34]:

Step-I: *For* $k=1, 2\ldots K$ $\{f_{k0}=0\}$

Step-II: *For* $m=1, 2, 3 \ldots M$

$$\left\{ p_k(x) = \frac{e^{f_k(x)}}{\sum_{l=1}^{K} e^{f_l(x)}} k = 1,\ 2\ \ldots\ K\right.$$

Step-III: For $k=1, 2\ldots K$

$$\left\{ r_{ikm} = y_{ik} - p_k\left(x_i\right),\ i = 1,\ 2, \ldots, N\right.$$

Fitting regression tree to the targets r_{ikm}, i = 1, 2… N to obtain the terminal regions $R_{jim}, j = 1, 2, \ldots J_m$

$$\gamma_{jkm} = \frac{K-1}{K}\left(\frac{\sum_{x_i} \in R_{jkm}\left(r_{ikm}\right)}{\sum_{\sum_{x_i} \in R_{jkm}} \left|r_{ikm}\right|\left(1 - \left|r_{ikm}\right|\right)}\right), j = 1, 2, \ldots, J_m$$

$$f_{km}(x) = f_{k,m-1} + \sum_{j=1}^{J_m} r_{ikm} I\left(x \in R_{jkm}\right) \Big\}$$

$f_k(x) = f_{kM}(x)$, where $k = 1, 2 \ldots K\}$

The additional classifier can support to further enhancing the performance metrics of the GBM without disturbing its overall speed. Such a combination reduces the process of parameter tuning by providing a parallelizable and distributable feature. Furthermore, it can result in optimal accuracy in big data analysis.

4 Analysis of Experimental Results

4.1 Statistical Analysis of the COVID-19 Dataset

Table 2 summarizes the statistical analysis using ANOVA method of the complete dataset (atmospheric factors, measures of air pollution, and cases of COVID-19). Results indicate that eight atmospheric factors, four pollution measures, and three significant parameters of COVID-19 are significant for further prediction modeling. Specifically, P-value is less than 0.05 indicates the confirmation in contrast to the null hypothesis for each of the dependent and independent variables. The F value represents the ratio of the variation between sample means and variation within the sample. Hence, a large value of F indicates a higher value of variation between sample means than within the sample. It also indicates that the null hypothesis is wrong (Table 2).

Table 2 Statistical analysis of the complete dataset using ANOVA methods

Atmospheric/air pollution COVID-19 metrics	Statistical analysis using ANOVA methods		
	DF	F Value	*P* value
Minimum temperature	20 & 9072	396.60	$P = < 2 \times 10^{-16}$ (less than 0.05)
Maximum temperature	20 & 9072	127.00	$P = < 2 \times 10^{-16}$ (less than 0.05)
Minimum humidity	20 & 9072	284.00	$P = < 2 \times 10^{-16}$ (less than 0.05)
Maximum humidity	20 & 9072	120.30	$P = < 2 \times 10^{-16}$ (less than 0.05)
Minimum pressure	20 & 9072	455.20	$P = < 2 \times 10^{-16}$ (less than 0.05)
Maximum pressure	20 & 9072	107.4	$P = < 2 \times 10^{-16}$ (less than 0.05)
Minimum wind speed	20 & 9072	30.87	$P = < 2 \times 10^{-16}$ (less than 0.05)
Maximum wind speed	20 & 9072	79.76	$P = < 2 \times 10^{-16}$ (less than 0.05)
Minimum PM2.5	20 & 9072	165.00	$P = < 2 \times 10^{-16}$ (less than 0.05)
Maximum PM2.5	20 & 9072	280.50	$P = < 2 \times 10^{-16}$ (less than 0.05)
Minimum PM10	20 & 9072	222.30	$P = < 2 \times 10^{-16}$ (less than 0.05)
Maximum PM10	20 & 9072	251.30	$P = < 2 \times 10^{-16}$ (less than 0.05)
Infected cases of COVID-19	20 & 9072	74.31	$P = < 1.2 \times 10^{-9}$ (less than 0.05)
Mortality cases of COVID-19	20 & 9072	164.10	$P = < 2 \times 10^{-16}$ (less than 0.05)
Recovery cases of COVID-19	20 & 9072	69.16	$P = < 2 \times 10^{-16}$ (less than 0.05)

4.2 Experimental Setup

The GBM models was trained with learning rate = 0.01, sample rate = 0.8 the number of trees = 10,000, and folds = 10 on Intel(R) Core (TM) i7-8565U CPU @ 1.80 GHz 1.99 GHz with 8 GB RAM to get the optimal performance.

4.3 Gradient Boosting Machine Model Analysis Results

The optimal GBM model was obtained after tuning the parameters of distribution functions, including the learning rate, the number of trees, folds, etc. Four result-oriented distribution functions were used in GBM, including Poisson, Gaussian, Tweedie, and Gamma out of seven compared distributions (excluding Huber, Laplace, and Quantile). The performance of the twinned GBM model using four different distribution functions is summarized in Table 3 (the rest distribution is discarded). The performance measures, including the goodness-fit-measures (R2), root mean square error (RMSE), mean residual deviance (MRD), and mean average error (MAE) were used to evaluate the efficiency of the GBM. In the training, the optimal prediction performance of the GBM was achieved with the Poisson distribution (R2 = 0.99) in all the three metrics of COVID-19 as infected, recovered, and mortality cases as shown in Table 3. The performance metrics of the GBM model in the forecast of the COVID-19 cases of the test dataset are demonstrated in Figs. 4, 5, and 6, respectively. Figure 4 exhibits a detailed performance analysis of different distribution functions of GBM to forecast the infected, recovered, and mortality cases of COVID-19, respectively for the combined dataset of different states/places of India.

All seven worst-affected states (Maharashtra, Delhi, Karnataka, Kerala, Madhya Pradesh, Uttar Pradesh, and West Bengal) data were tested with the twinned GBM with the four most result-oriented distributions as Poisson, Gaussian, Tweedie, and Gamma distributions. Five performance parameters were used as R2, MSE, RMSE, MAE, and

Table 3 Performance metrics of twined GBM in training with combined dataset of India

COVID parameters	Applied distributions	Performance metrics of India using twined GBM Model				
		R^2	MSE	RMSE	MAE	MRD
Infection	Poisson	0.99	697,073.2	834.90	549.45	− 68,306.81
	Gaussian	0.97	2,334,416	1527.28	958.07	2,334,416
	Tweedie	0.96	1,474,786	1214.40	600.14	9.32
	Gamma	0.85	9,943,415	3153.31	1239.84	15.59
Recovery	Poisson	0.99	508,361.5	712.99	777.06	− 58,347.96
	Gaussian	0.98	1,549,178	1244.66	794.12	1,549,178
	Tweedie	0.97	1,107,026	1052.15	519.56	9.63
	Gamma	0.81	10,711,357	3272.82	1291.17	15.05
Mortality	Poisson	0.99	72.24	8.49	5.62	− 352.88
	Gaussian	0.97	214.41	14.64	9.19	214.41
	Tweedie	0.98	133.50	11.55	5.94	1.20
	Gamma	0.85	1459.253	38.20	15.07	5.64

Performance metrics: R^2 (R squared), RMSE (root mean square error), MSE (mean square error), MRD (mean residual deviance) and MAE (mean average error)

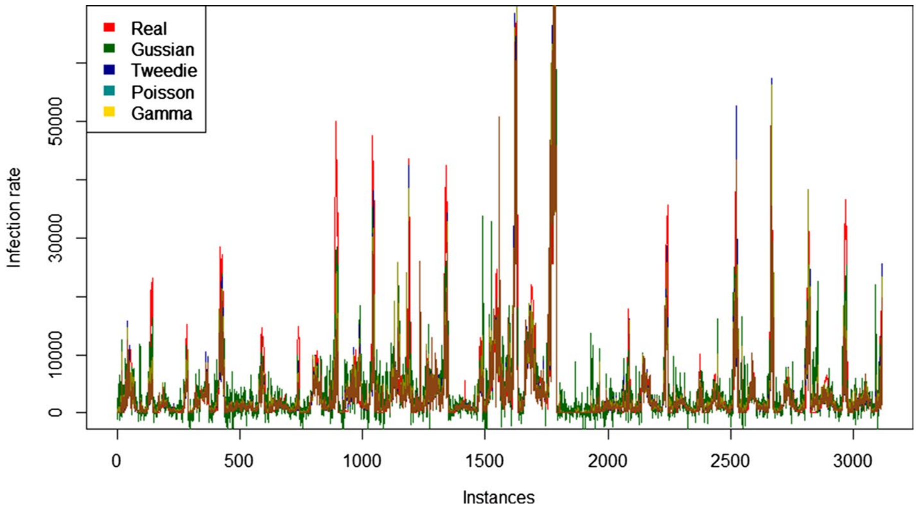

Fig. 4 Correlative capability of twinned GBM in the training for infected cases of COVID-19 in terms of the combined dataset

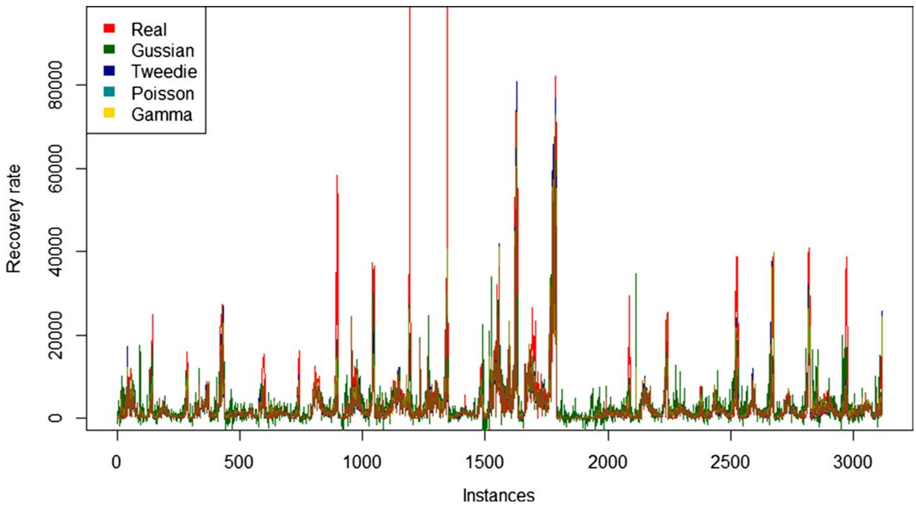

Fig. 5 Correlative capability of twinned GBM in the training for recovery cases of COVID-19 in terms of the combined dataset

MRD to find the proper correlation and efficiency of the individual model. The test performance of seven states of India was summarized and presented in Tables 4, 5, 6, 7, 8, 9 and 10 and Figs. 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 and 21 respectively below:

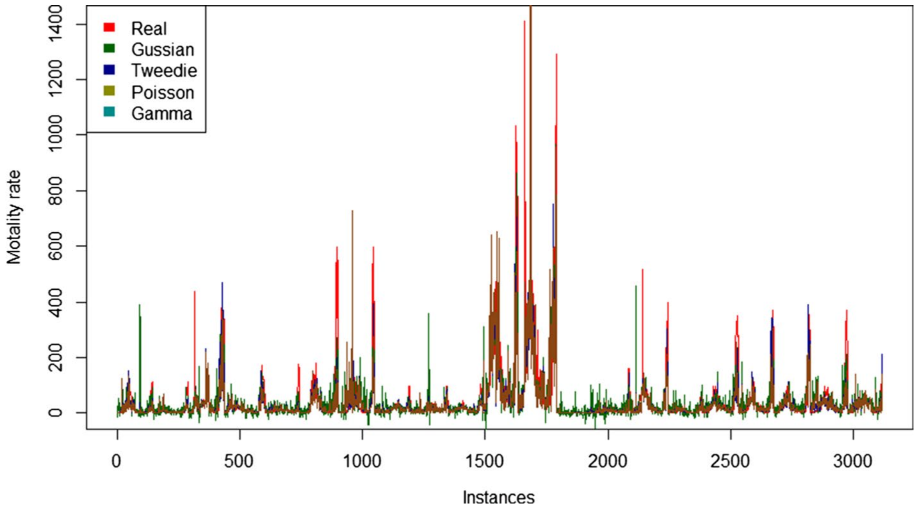

Fig. 6 Correlative capability of twinned GBM in the training for mortality cases of COVID-19 in terms of the combined dataset

Table 4 Performance metrics of twinned GBM in the forecast of infected, recovered and mortality cases of COVID-19 in Maharashtra

COVID parameters	Applied distributions	Performance metrics of Maharashtra using twinned GBM Model				
		R^2	MSE	RMSE	MAE	MRD
Infection	Poisson	0.90	26,641,098	5161.50	2150.96	−230,002.7
	Gaussian	0.90	27,409,054	5235.36	275,374	27,409,054
	Tweedie	0.88	32,401,221	5692.20	2506.71	23.05
	Gamma	0.78	61,476,424	7840.69	3989.27	19.70
Recovery	Poisson	0.87	35,233,383	5935.77	2083.57	−208,900.6
	Gaussian	0.89	29,515,814	5432.84	2635.31	29,515,814
	Tweedie	0.85	40,477,651	6362.20	2427.32	28.22
	Gamma	0.71	76,864,938	8767.26	4242.95	19.32
Mortality	Poisson	0.84	7480.54	86.49	31.87	−1857.321
	Gaussian	0.88	5770.93	75.96	36.84	5770.931
	Tweedie	0.83	8169.448	90.38	37.05	3.43
	Gamma	0.65	17,140.35	130.92	72.00	11.78

Performance metrics: R^2 (R squared), RMSE (root mean square error), MSE (mean square error), MRD (mean residual deviance) and MAE (mean average error)

5 Discussion of Results

Tree-based machine learning approaches have high accuracy in the analysis of small and big datasets in previous research studies [35, 36]. In the case of analysis of the disease data, the GBM was used to predict the association of miRNAs [35]. Besides, the improved

Table 5 Performance metrics of twinned GBM in the prediction of COVID-19 in Delhi

COVID parameters	Applied distributions	Performance metrics of Delhi using twinned GBM Model				
		R^2	MSE	RMSE	MAE	MRD
Infection	Poisson	0.75	7,099,035	2664.40	920.95	−50,832.85
	Gaussian	0.78	7,424,884	2724.864	902.08	13.80
	Tweedie	0.74	6,258,603	2501.72	1232.27	6,258,603
	Gamma	0.69	8,748,782	2957.83	1119.90	16.29
Recovery	Poisson	0.70	7,416,954	2723.40	936.06	−48,065.37
	Gaussian	0.78	5,590,525	2364.42	1092.77	5,590,525
	Tweedie	0.73	6,680,122	2584.59	893.76	19.05
	Gamma	0.67	8,247,308	2871.81	1164.22	15.79
Mortality	Poisson	0.77	1687.84	41.08	13.41	−381.93
	Gaussian	0.83	1238.01	35.18	15.62	1238.01
	Tweedie	0.75	1809.62	42.53	13.84	13.06
	Gamma	0.59	3026.10	55.01	22.80	626.64

Performance metrics: R^2 (R squared), RMSE (root mean square error), MSE (mean square error), MRD (mean residual deviance) and MAE (mean average error)

Table 6 Performance metrics of twinned GBM in the prediction of infected, recovered and mortality cases of COVID-19 in Karnataka

COVID parameters	Applied distributions	Performance metrics of Karnataka using twinned GBM Model				
		R^2	MSE	RMSE	MAE	MRD
Infection	Poisson	0.79	19,862,527	4456.73	1500.29	−91,129.5
	Gaussian	0.84	14,337,580	3786.5	1849.86	14,337,580
	Tweedie	0.74	24,453,918	4945.09	1655.20	25.38
	Gamma	0.54	43,640,970	6606.13	2505.40	16.71
Recovery	Poisson	0.55	24,694,898	4969.39	1380.51	−66,065.49
	Gaussian	0.63	19,921,899	4463.39	164,133	19,921,899
	Tweedie	0.50	27,565,253	5250	1483.41	29.55
	Gamma	0.31	37,742,917	6143.52	2137.15	16.08
Mortality	Poisson	0.64	3241.45	56.93	17.11	−416.7302
	Gaussian	0.71	2674.02	51.71	20.25	2674.02
	Tweedie	0.60	3604.25	60.03	18.61	3.29
	Gamma	0.39	5582.60	74.71	27.88	7.68

Performance metrics: R^2 (R squared), RMSE (root mean square error), MSE (mean square error), MRD (mean residual deviance) and MAE (mean average error)

performance of the GBM in the predictive modeling of the pandemic has been discussed [35]. This is the reason for selecting the GBM model in the prediction of the COVID-19 cases in India using the atmospheric factors and pollution levels. Due to a large geographical area, there is a huge variation in atmospheric factors (Fig. 1 and Table 1) in different states of India. Besides, the pollution levels also vary in different states, which is obvious

Table 7 Performance metrics of twinned GBM in the prediction of infected, recovered and mortality cases of COVID-19 in Kerala

COVID parameters	Applied distributions	Performance metrics of Kerala using twinned GBM Model				
		R^2	MSE	RMSE	MAE	MRD
Infection	Poisson	0.76	15,856,941	3982.07	1761.23	−85,513.48
	Gaussian	0.76	16,218,005	4027.15	2402.33	16,218,005
	Tweedie	0.74	17,305,083	4159.93	1832.27	41.30
	Gamma	0.59	28,000,496	5251.54	2513.33	17.51
Recovery	Poisson	0.47	34,757,267	5895.52	1563.41	70,499.11
	Gaussian	0.56	28,294,945	5319.29	1872.39	28,294,945
	Tweedie	0.43	36,155,911	6212.97	1591.96	71.48
	Gamma	0.19	46,725,269	6835.58	2331.723	19.08
Mortality	Poisson	0.59	129.47	11.37	6.47	−65.85
	Gaussian	0.37	198.53	14.09	9.96	198.53
	Tweedie	0.58	130.45	11.42	5.65	3.09
	Gamma	0.46	171.18	13.08	6.64	6.76

Performance metrics: R^2 (R squared), RMSE (root mean square error), MSE (mean square error), MRD (mean residual deviance) and MAE (mean average error)

Table 8 Performance metrics of twinned GBM in prediction of infected, recovered and mortality cases of COVID-19 in Madhya Pradesh

COVID parameters	Applied distributions	Performance metrics of Madhya Pradesh using twinned GBM Model				
		R^2	MSE	RMSE	MAE	MRD
Infection	Poisson	0.87	1,099,131	1048.39	543.74	−25,225.9
	Gaussian	0.80	1,736,249	1317.66	836.93	1,736,249
	Tweedie	0.86	1,230,043	1109.07	498.90	8.37
	Gamma	0.75	2,195,843	1481.84	642.10	27.65
Recovery	Poisson	0.88	936,599.5	967.78	557.09	−22,263.02
	Gaussian	0.81	1,460,947	1208.69	756.27	1,460,947
	Tweedie	0.85	1,142,666	1068.95	433.96	5.58
	Gamma	0.67	2,522,093	1588.11	661.08	12.80
Mortality	Poisson	0.84	77.58	8.80	5.38	−78.62
	Gaussian	0.74	127.49	11.29	7.36	127.49
	Tweedie	0.84	75.86	8.70	4.94	1.54
	Gamma	0.65	171.64	13.10	7.27	6.21

Performance metrics: R^2 (R squared), RMSE (root mean square error), MSE (mean square error), MRD (mean residual deviance) and MAE (mean average error)

from the variation of minimum and maximum PM_{10} and $PM_{2.5}$ (Fig. 2 and Table 1). The basic statistics in Table 1 and Fig. 3 demonstrates the variation in the cases of COVID-19 in different states of India. The basic statistics on the atmospheric factors, pollution measures, and cases of COVID-19 suggest their unequal distribution.

Table 9 Performance metrics of twinned GBM in the prediction of infection, recovery, and mortality cases of COVID-19 in Uttar Pradesh

COVID parameters	Applied distributions	Performance metrics of Uttar Pradesh using twinned GBM Model				
		R^2	MSE	RMSE	MAE	MRD
Infection	Poisson	0.79	10,577,070	3552.24	1381.75	−61,904.25
	Gaussian	0.80	10,406,203	3225.86	1743.42	10,406,203
	Tweedie	0.75	13,076,517	3616.147	1399.83	30.04
	Gamma	0.67	17,071,418	4131.75	1747.28	16.68
Recovery	Poisson	0.88	936,599.5	967.78	557.09	−22,263.02
	Gaussian	0.81	1,460,947	1208.69	756.27	1,460,947
	Tweedie	0.85	1,142,666	1068.95	433.96	8.58
	Gamma	0.67	2,522,093	1588.11	12.80	661.08
Mortality	Poisson	0.73	1301.39	36.07	14.88	−289.60
	Gaussian	0.72	1375.24	37.08	19.75	1375.24
	Tweedie	0.69	1499.57	38.72	15.54	3.69
	Gamma	0.51	2423.18	49.22	22.37	8.29

Performance metrics: R^2 (R squared), RMSE (root mean square error), MSE (mean square error), MRD (mean residual deviance) and MAE (mean average error)

Table 10 Performance of twinned GBM in the prediction of infestation, recovery, and mortality cases of COVID-19 in West Bengal

COVID parameters	Applied distributions	Performance metrics of West Bengal using twinned GBM Model				
		R^2	MSE	RMSE	MAE	MRD
Infection	Poisson	0.79	4,051,807	2012.91	883.78	−42,267.37
	Gaussian	0.78	4,271,421	2066.74	856.38	26.78
	Tweedie	0.64	6,969,710	2640.02	1329.80	6,969,710
	Gamma	0.68	6,107,182	2471.27	1055.63	10.37
Recovery	Poisson	0.80	2,969,163	1723.12	756.79	−36,596.96
	Gaussian	0.72	4,218,383	2053.87	1106.53	4,218,383
	Tweedie	0.78	3,331,449	1825.22	737.33	26.88
	Gamma	0.60	6,129,715	2475.82	1145.73	16.34
Mortality	Poisson	0.78	219.41	14.81	9.05	−181.25
	Gaussian	0.63	367.59	19.17	12.32	367.59
	Tweedie	0.75	244.71	15.64	9.27	5.54
	Gamma	0.57	424.39	20.60	13.14	8.94

Performance metrics: R^2 (R squared), RMSE (root mean square error), MSE (mean square error), MRD (mean residual deviance) and MAE (mean average error)

The training performance results of twinned GBM for infected cases on the combined dataset of significant states of India provide $R^2=0.99$, and RMSE = 834.90 with Poisson distribution, $R^2=0.97$, and RMSE = 1527.28 with Gaussian distribution, $R^2=0.96$,

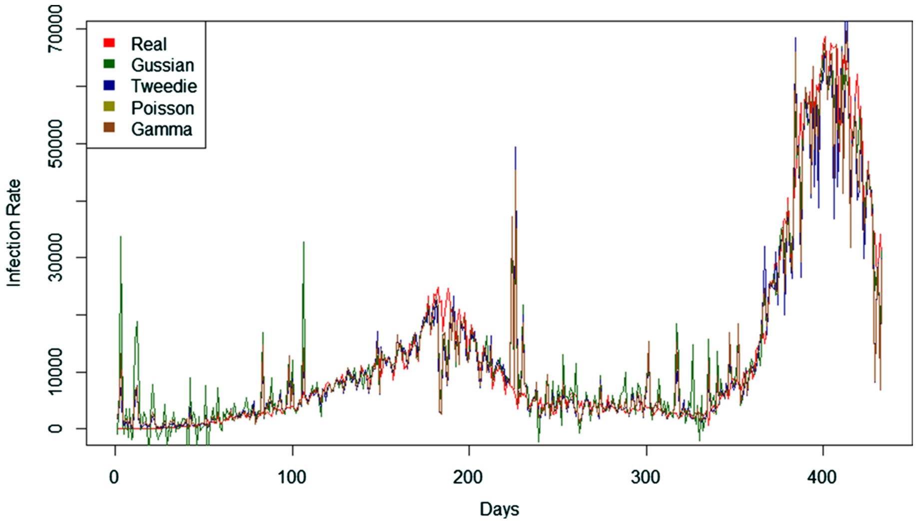

Fig. 7 Correlative capability of twined GBM to forecast the infection rate of COVID-19 in Maharashtra

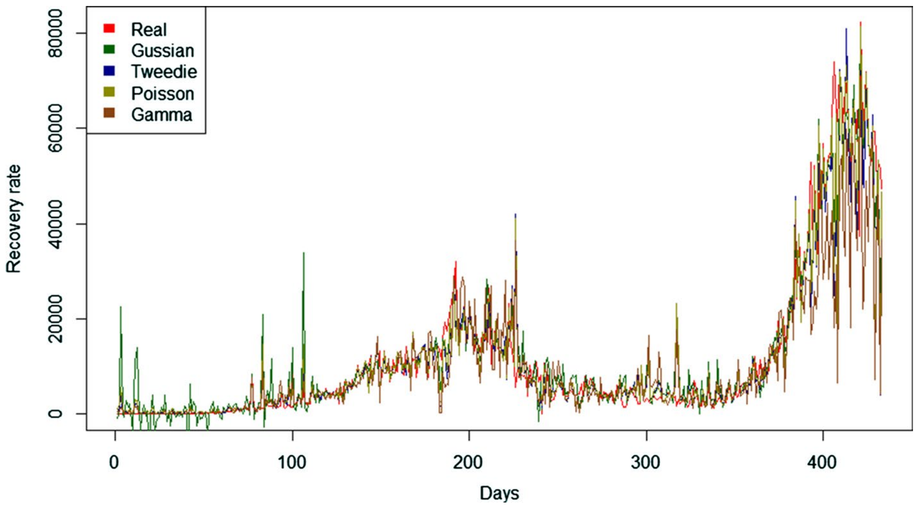

Fig. 8 Correlative capability of twined GBM to forecast the recovery rate of COVID-19 in Maharashtra

and RMSE = 1214.40 with Tweedie distribution and $R^2 = 0.85$ and RMSE = 1239.84 with Gamma distributions. The training performance results of twinned GBM for recovered cases on the combined dataset of significant states of India provide $R^2 = 0.99$, and RMSE = 712.99 with Poisson distribution, $R^2 = 0.98$, and RMSE = 1244.66 with Gaussian distribution, $R^2 = 0.97$, and RMSE = 1052.15 with Tweedie distribution and $R^2 = 0.81$ and RMSE = 3272.82 with Gamma distributions. The training performance results of twinned GBM for mortality case on the combined dataset of significant states of India provides $R^2 = 0.99$, and RMSE = 8.49 with Poisson distribution, $R^2 = 0.97$, and RMSE = 14.64

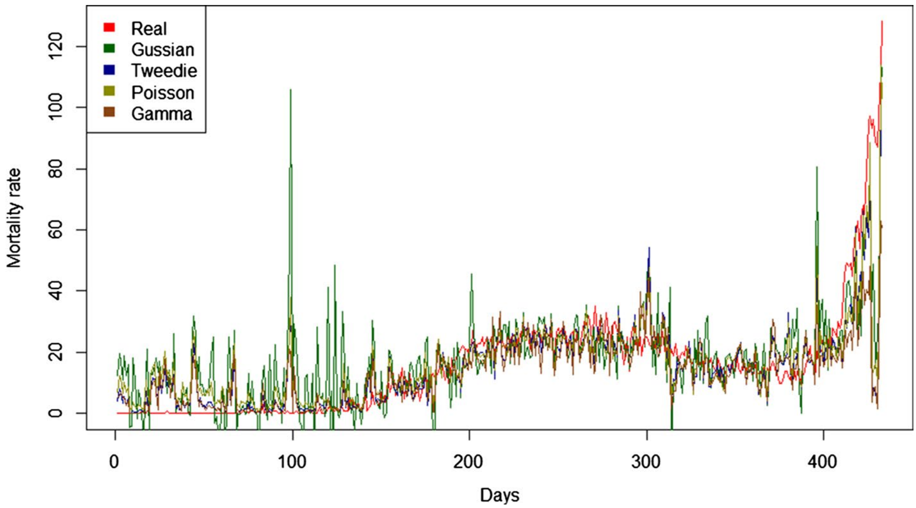

Fig. 9 Correlative capability of twined GBM to forecast the mortality rate of COVID-19 in Maharashtra

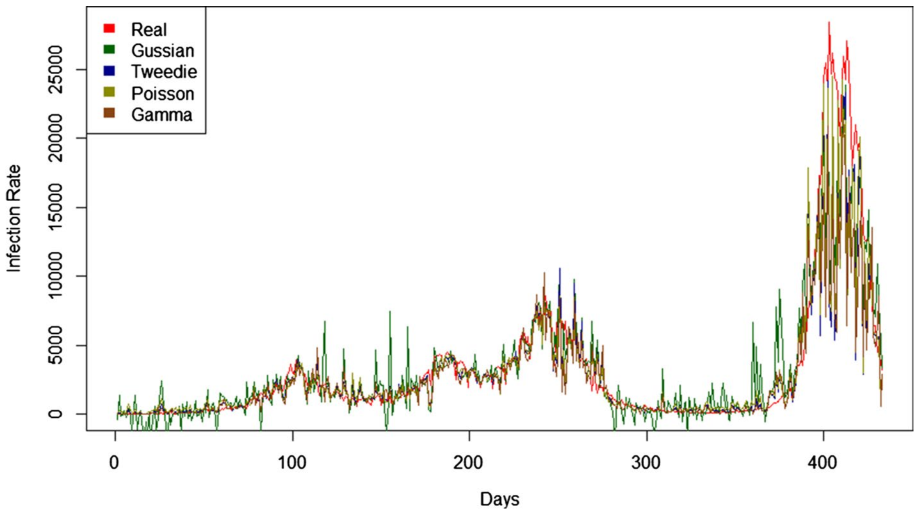

Fig. 10 Correlative capability of twined GBM to forecast the infection rate of COVID-19 in Delhi

with Gaussian distribution, $R^2=0.98$ and RMSE = 11.55 with Tweedie distribution and $R^2=0.85$ and RMSE = 38.20 with Gamma distributions. The complete performance result for infected, recovery, and mortality cases are presented in Table 3, Figs. 4, 5, and 6 respectively. The performance results of the twined GBM with all four selected four distributions (Poisson, Gaussian, Tweedie, and Gamma) are quite good and quite better it assures that there is a close correlation among the atmospheric factor, air pollutants, and COVID-19 parameters and the study may move for the further processing.

Now the trained model has applied the dataset to the seven largely affected states of India to explore the deeper analysis and correlation for testing. At first, one of the worst

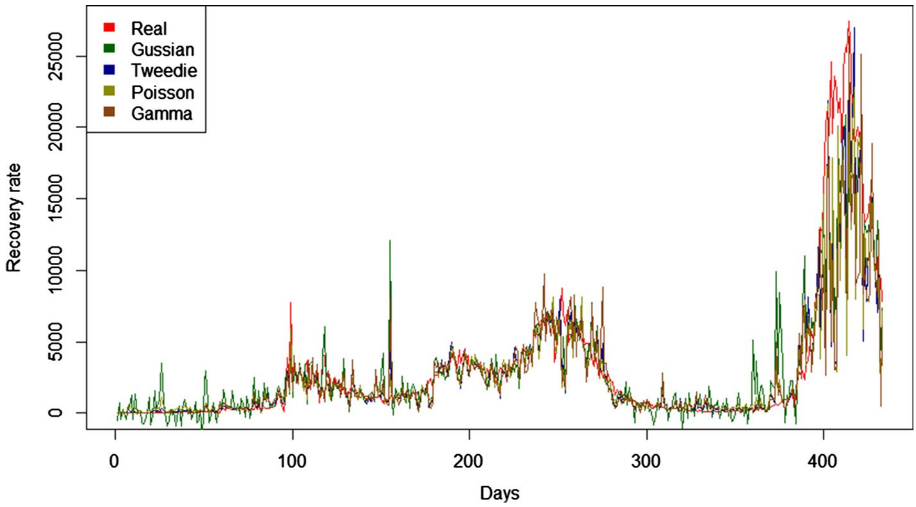

Fig. 11 Correlative capability of twined GBM to forecast the recovery rate of COVID-19 in Delhi

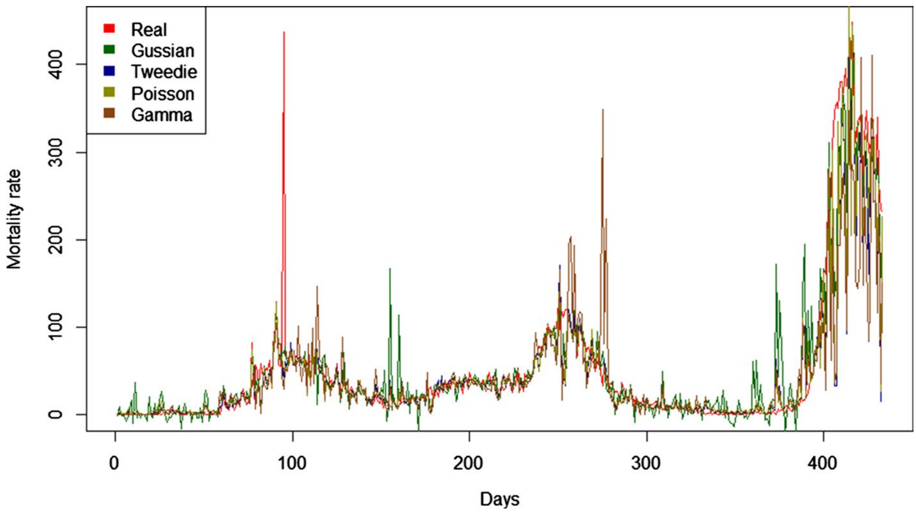

Fig. 12 Correlative capability of twined GBM to forecast the mortality rate of COVID-19 in Delhi

affected Maharashtra is taken for testing. Surprisingly the performance result of the infected case provides a very convincing correlation as $R^2=0.90$, and RMSE = 5161.50 with Poisson distribution, $R^2=0.90$, and RMSE = 5235.36 with Gaussian distribution, $R^2=0.88$, and RMSE = 5692.20 with Tweedie distribution and $R^2=0.78$ and RMSE = 7840.69 with Gamma distributions. In the case of recovery, it also approves the hypothesis with $R^2=0.87$, and RMSE = 5935.77 with Poisson distribution, $R^2=0.89$, and RMSE = 5432.84 with Gaussian distribution, $R^2=0.85$ and RMSE = 6362.20 with Tweedie distribution and $R^2=0.71$ and RMSE = 8767.26 with Gamma distributions. In the case of mortality, the performance results are also in the same hypothesis line as $R^2=0.84$, and RMSE = 86.49

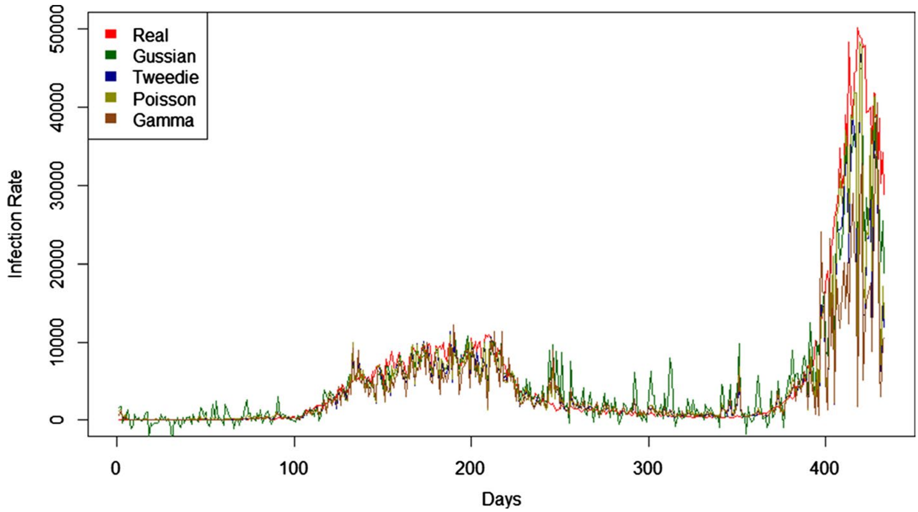

Fig. 13 Correlative capability of twined GBM to forecast the infection rate of COVID-19 in Karnataka

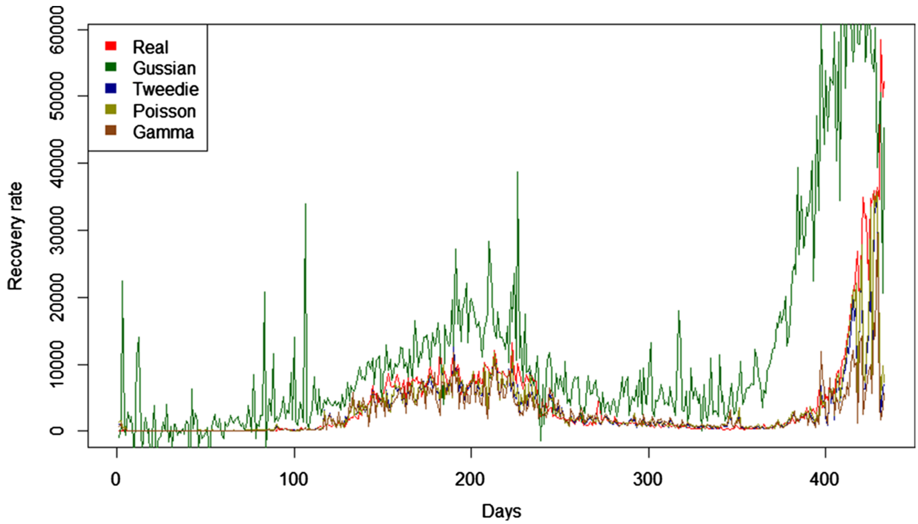

Fig. 14 Correlative capability of twined GBM to forecast the recovery rate of COVID-19 in Karnataka

with Poisson distribution, $R^2=0.88$, and RMSE = 75.96 with Gaussian distribution, $R^2=0.83$ and RMSE = 90.38 with Tweedie distribution and $R^2=0.65$ and RMSE = 130.92 with Gamma distributions. The complete performance result for Maharashtra is already shown in Table 4, Figs. 7, 8, and 9 respectively.

Secondly, the model is tested for the largely affected state of Delhi. The performance result of this testing is $R^2=0.75$, and RMSE = 2664.40 with Poisson distribution, $R^2=0.78$, and RMSE = 2724.86 with Gaussian distribution, $R^2=0.74$, and RMSE = 2501.72 with Tweedie distribution and $R^2=0.69$ and RMSE = 2957.83 with Gamma distributions. In the

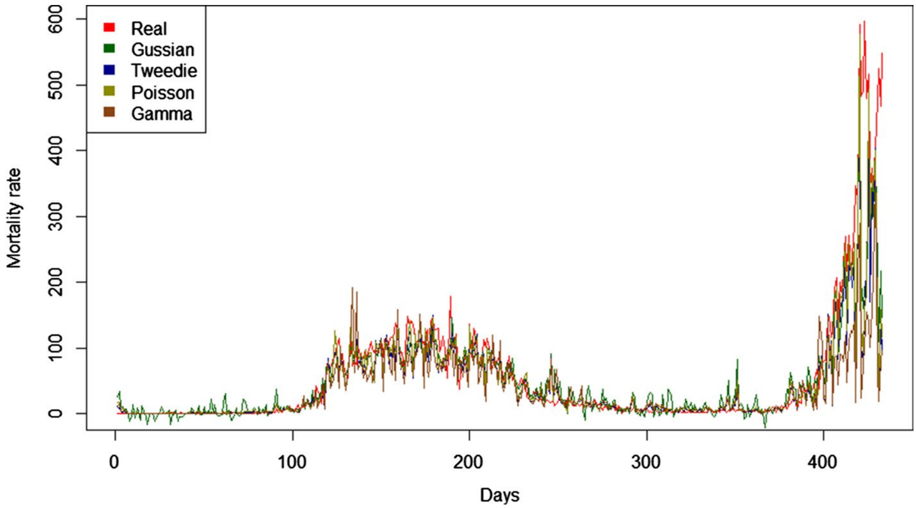

Fig. 15 Correlative capability of twined GBM to forecast the mortality rate of COVID-19 in Karnataka

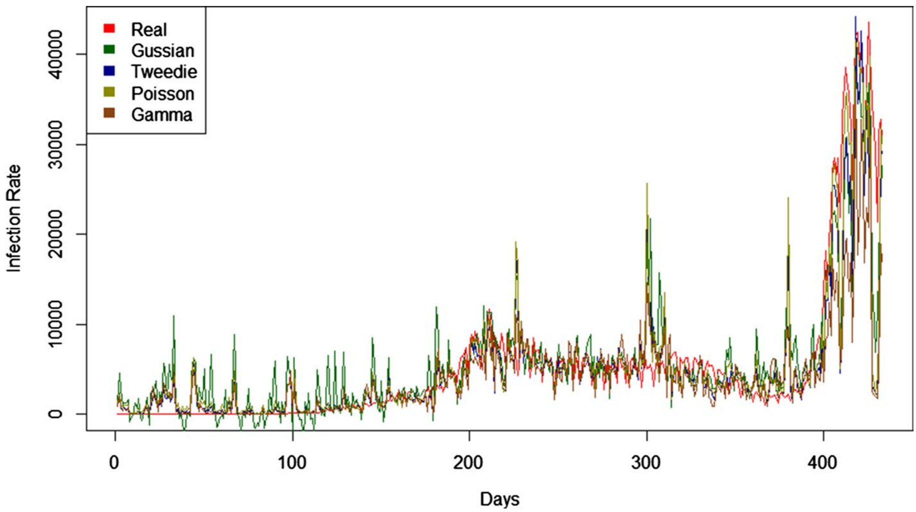

Fig. 16 Correlative capability of twined GBM to forecast infection rate of COVID-19 in Kerala

case of recovery, it also approves the hypothesis with $R^2=0.88$, and RMSE = 5935.77 with Poisson distribution, $R^2=0.81$, and RMSE = 5432.84 with Gaussian distribution, $R^2=0.85$ and RMSE = 6362.20 with Tweedie distribution and R2 = 0.67 and RMSE = 8767.26 with Gamma distributions. In the case of mortality, the performance results are also in the same hypothesis line as $R^2=0.73$, and RMSE = 36.07 with Poisson distribution, $R^2=0.72$, and RMSE = 37.08 with Gaussian distribution, $R^2=0.69$ and RMSE = 38.72 with Tweedie distribution and $R^2=0.51$ and RMSE = 49.22 with Gamma distributions. The complete performance result for Maharashtra is already shown in Table 5, Figs. 10, 11, and 12 respectively.

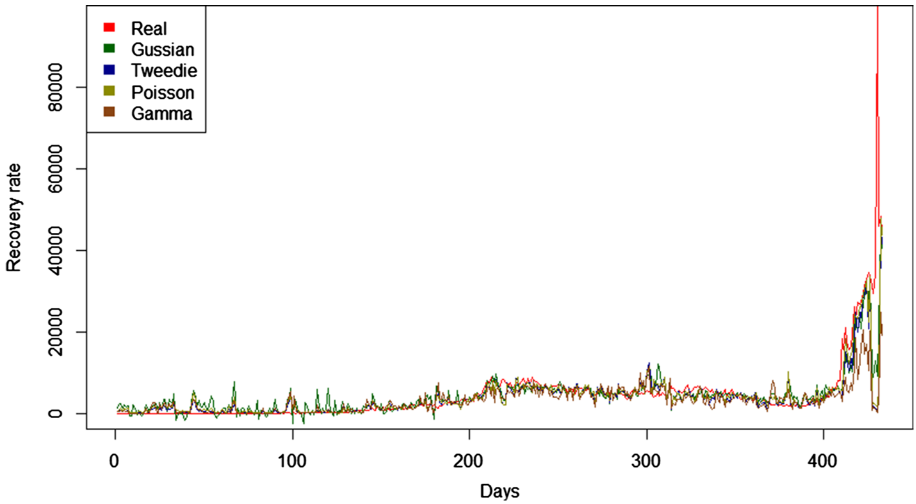

Fig. 17 Correlative capability of twined GBM to forecast recovery rate of COVID-19 in Kerala

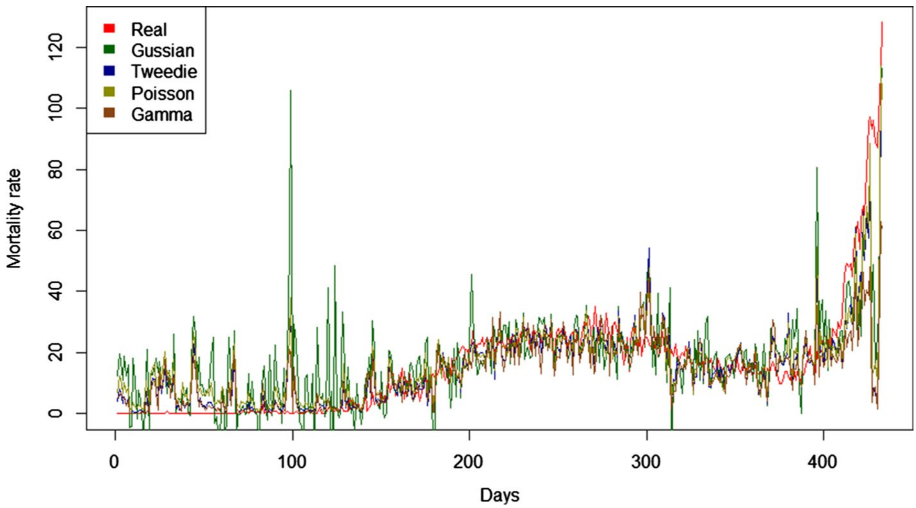

Fig. 18 Correlative capability of twined GBM to forecast mortality rate of COVID-19 in Kerala

Third, the trained model has applied the testing dataset of the significant state of Karnataka. The performance result of this testing is as $R^2=0.79$, and RMSE = 4456.73 with Poisson distribution, $R^2=0.84$ and RMSE = 3786.50 with Gaussian distribution, $R^2=0.74$ and RMSE = 4945.09 with Tweedie distribution and $R^2=0.54$ and RMSE = 6606.13 with Gamma distributions. In the case of recovery, it also approves the hypothesis with $R^2=0.55$, and RMSE = 4969.39 with Poisson distribution, $R^2=0.63$, and RMSE = 4463.39 with Gaussian distribution, $R^2=0.50$, and RMSE = 5250 with Tweedie distribution and $R^2=0.31$ and RMSE = 6143.52 with Gamma distributions. In the case of mortality, the performance results are also in the same hypothesis line as $R^2=0.64$, and RMSE = 56.93

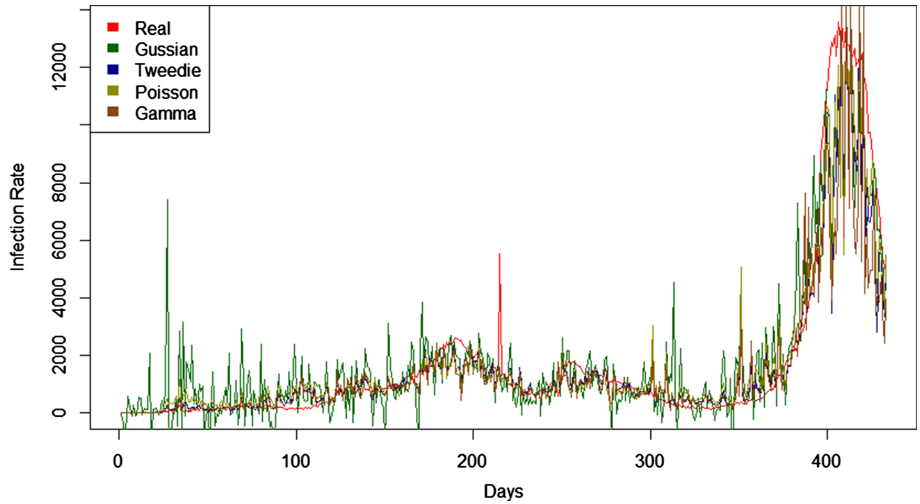

Fig. 19 Correlative capability of twined GBM to forecast infection rate of COVID-19 in Madhya Pradesh

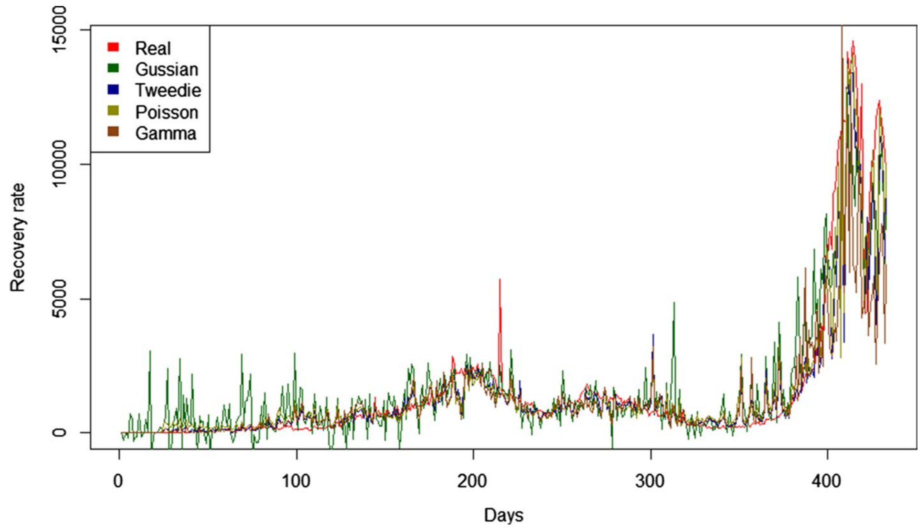

Fig. 20 Correlative capability of twined GBM to forecast recovery rate of COVID-19 in Madhya Pradesh

with Poisson distribution, $R^2=0.71$, and RMSE = 51.71 with Gaussian distribution, $R^2=0.60$, and RMSE = 60.03 with Tweedie distribution and $R^2=0.39$ and RMSE = 74.71 with Gamma distributions. The complete performance result for Karnatka is already shown in Table 6, Figs. 13, 14, and 15 respectively.

Fourth, the trained model has applied the testing dataset of the significant state of Kerala. The performance result of this testing is as $R^2=0.76$, and RMSE = 3982.07 with Poisson distribution, $R^2=0.76$ and RMSE = 4027.15 with Gaussian distribution, $R^2=0.74$, and RMSE = 4159.93 with Tweedie distribution and $R^2=0.59$ and RMSE = 5251.54

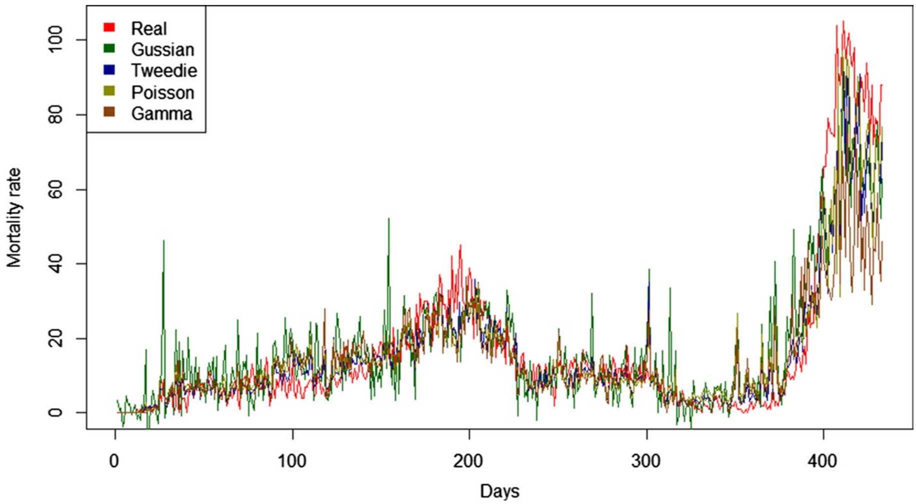

Fig. 21 Correlative capability of twined GBM to forecast mortality rate of COVID-19 in Madhya Pradesh

with Gamma distributions. In the case of recovery, it also approves the hypothesis with $R^2 = 0.47$, and RMSE = 5895.52 with Poisson distribution, $R^2 = 0.56$, and RMSE = 5319.29 with Gaussian distribution, $R^2 = 0.43$, and RMSE = 6212.97 with Tweedie distribution and $R^2 = 0.19$ and RMSE = 6835.58 with Gamma distributions. In the case of mortality, the performance results are also in the same hypothesis line as $R^2 = 0.59$, and RMSE = 11.37 with Poisson distribution, $R^2 = 0.37$, and RMSE = 14.09 with Gaussian distribution, $R^2 = 0.58$, and RMSE = 11.42with Tweedie distribution and $R^2 = 0.46$ and RMSE = 13.08 with Gamma distributions. The complete performance result for Kerala is already shown in Table 7, Figs. 16, 17, and 18 respectively.

Fifth, the trained model has applied the testing dataset of the significant state of Madhya Pradesh. The performance result of this testing is R2 = 0.87, and RMSE = 1048.39 with Poisson distribution, R2 = 0.80, and RMSE = 1317.66 with Gaussian distribution, R2 = 0.86, and RMSE = 1109.07 with Tweedie distribution and R2 = 0.59 and RMSE = 1481.84 with Gamma distributions. In the case of recovery, it also approves the hypothesis with R2 = 0.88, and RMSE = 5895.52 with Poisson distribution, R2 = 0.81, and RMSE = 5319.29 with Gaussian distribution, R2 = 0.85, and RMSE = 6212.97 with Tweedie distribution and R2 = 0.67 and RMSE = 6835.58 with Gamma distributions. In the case of mortality, the performance results are also in the same hypothesis line as R2 = 0.84, and RMSE = 8.80 with Poisson distribution, R2 = 0.74, and RMSE = 11.29 with Gaussian distribution, R2 = 0.84, and RMSE = 8.70with Tweedie distribution and R2 = 0.65 and RMSE = 13.10 with Gamma distributions. The complete performance result for Madhya Pradesh is already shown in Table 8, Figs. 19, 20, and 21 respectively.

Sixth, the trained model has applied the testing dataset of the significant state of Uttar Pradesh. The performance result of this testing is as R2 = 0.79, and RMSE = 3552.24 with Poisson distribution, R2 = 0.80, and RMSE = 3225.86 with Gaussian distribution, R2 = 0.75, and RMSE = 3616.14 with Tweedie distribution and R2 = 0.67 and RMSE = 4131.75with Gamma distributions. In the case of recovery, it also approves the hypothesis with R2 = 0.88, and RMSE = 967.78 with Poisson distribution, R2 = 0.81, and RMSE = 1208.69 with Gaussian distribution, R2 = 0.85, and RMSE = 1068.95 with

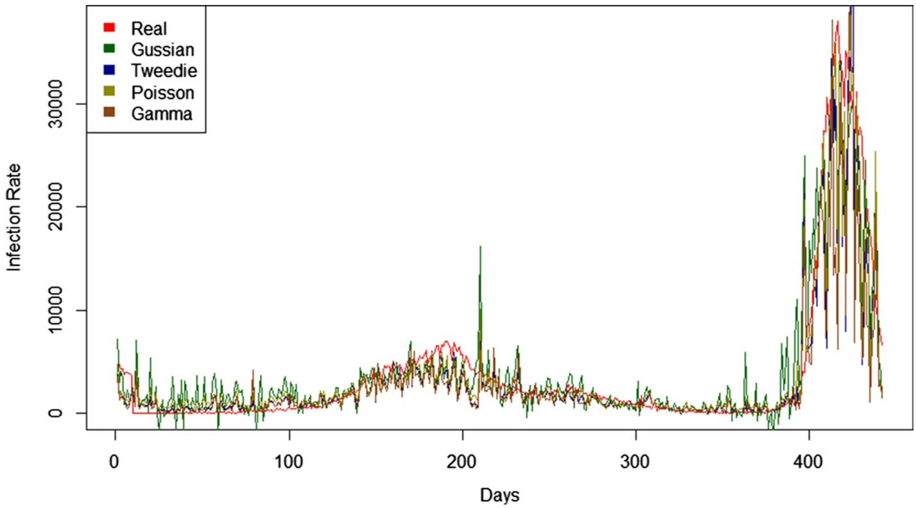

Fig. 22 Correlative capability of twined GBM to forecast infection rate of COVID-19 in Uttar Pradesh

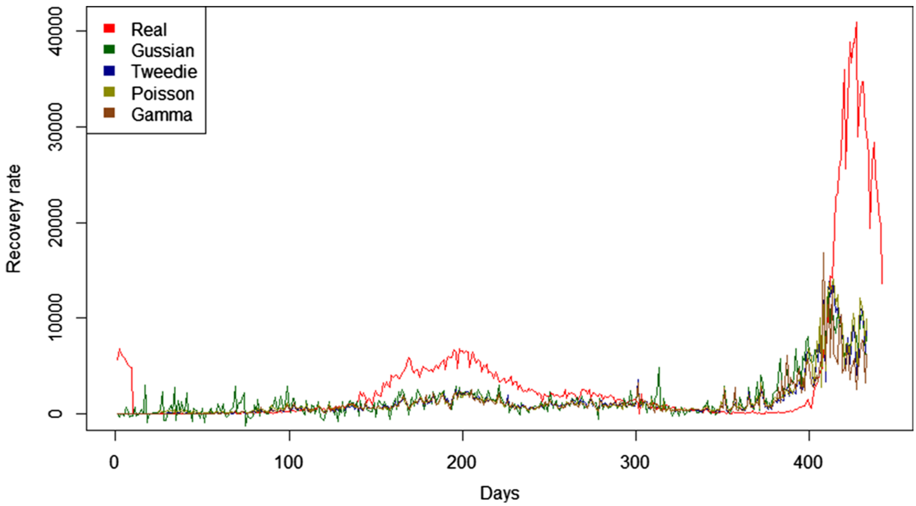

Fig. 23 Correlative capability of twined GBM to forecast recovery rate of COVID-19 in Uttar Pradesh

Tweedie distribution and R2 = 0.67 and RMSE = 1588.11 with Gamma distributions. In the case of mortality, the performance results are also in the same hypothesis line as R2 = 0.73, and RMSE = 36.07 with Poisson distribution, R2 = 0.72, and RMSE = 37.08 with Gaussian distribution, R2 = 0.69 and RMSE = 38.72 with Tweedie distribution and R2 = 0.51 and RMSE = 49.22 with Gamma distributions. The complete performance result for Uttar Pradesh is already shown in Table 9, Figs. 22, 23, and Figs. 24 respectively.

Seventh, the trained model has applied the testing dataset of the significant state of West Bengal. The performance result of this testing is as $R^2 = 0.79$, and RMSE = 2012.91 with Poisson distribution, $R^2 = 0.78$ and RMSE = 2066.74 with

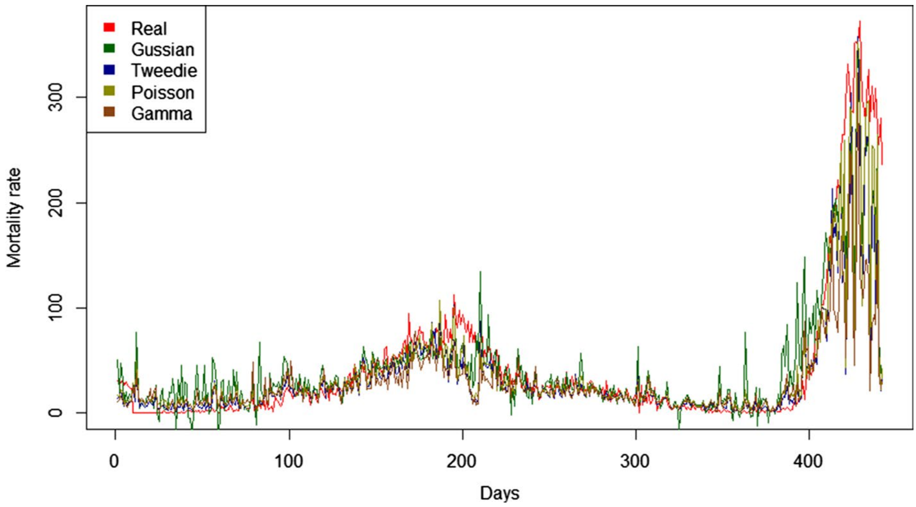

Fig. 24 Correlative capability of twined GBM to forecast mortality rate of COVID-19 in Uttar Pradesh

Gaussian distribution, $R^2=0.64$ and RMSE = 2640.02 with Tweedie distribution and $R^2=0.68$ and RMSE = 247,127 with Gamma distributions. In the case of recovery, it also approves the hypothesis with $R^2=0.80$, and RMSE = 1723.12 with Poisson distribution, $R^2=0.72$, and RMSE = 2053.87 with Gaussian distribution, $R^2=0.78$, and RMSE = 1825.22 with Tweedie distribution and $R^2=0.60$ and RMSE = 2475.82 with Gamma distributions. In the case of mortality, the performance results are also in the same hypothesis line as $R^2=0.78$, and RMSE = 14.81 with Poisson distribution, $R^2=0.63$, and RMSE = 19.17 with Gaussian distribution, $R^2=0.75$ and RMSE = 15.64 with Tweedie distribution and $R^2=0.57$ and RMSE = 20.60 with Gamma distributions. The complete performance result for West Bengal is already shown in Table 10, Figs. 25, 26, and 27 respectively.

The above-discussed performance parameter and the rest of the parameters are demonstrated in Tables 4, 5, 6, 7, 8, 9 and 10 and Figs. 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26 and 27 suggests that the Maharashtra had an ideal atmosphere for infection, recovery, and mortality with $R^2=0.99$ in all three with the Poisson distribution. The testing model on Delhi is not so much performing on infection and recovery rate but it supports the mortality rate. The maximum performance was given by Gaussian distribution with $R^2=0.78$ for the infection rate, R2 = 0.78 for recovery rate and $R^2=0.83$ for the mortality rate. $R^2=0.84$ for an infection rate for the Karnataka state, recovery provides $R^2=0.63$ and mortality $R^2=0.71$ by Gaussian distribution. Kerala infection rate $R^2=0.71$ and recovery rate $R^2=0.56$ provided by Gaussian distribution and mortality rate $R^2=0.59$ by Poisson distribution does not support; it might lack non-arability/missing of the correct atmospheric or pollution dataset. Madhya Pradesh, maximum infection rate, recovery rate, and mortality rate R2 = 0.87, $R^2=0.88$, and $R^2=0.84$ respectively by Poisson distribution. Uttar Pradesh, maximum infection rate, recovery rate, and mortality rate $R^2=0.80$, $R^2=0.88$, and $R^2=0.73$ respectively by Poisson distribution. West Bengal, maximum infection rate, recovery rate, and mortality rate $R^2=0.79$, $R^2=0.80$, and $R^2=0.78$ respectively by Poisson distribution.

The COVID parameter according to the testing performance conclusion:

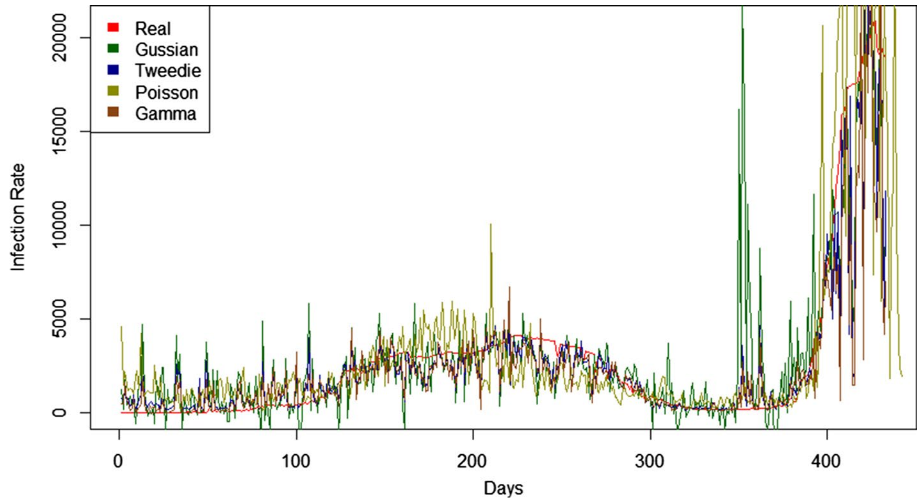

Fig. 25 Correlative capability of twined GBM to forecast infection rate of COVID-19 in West Bengal

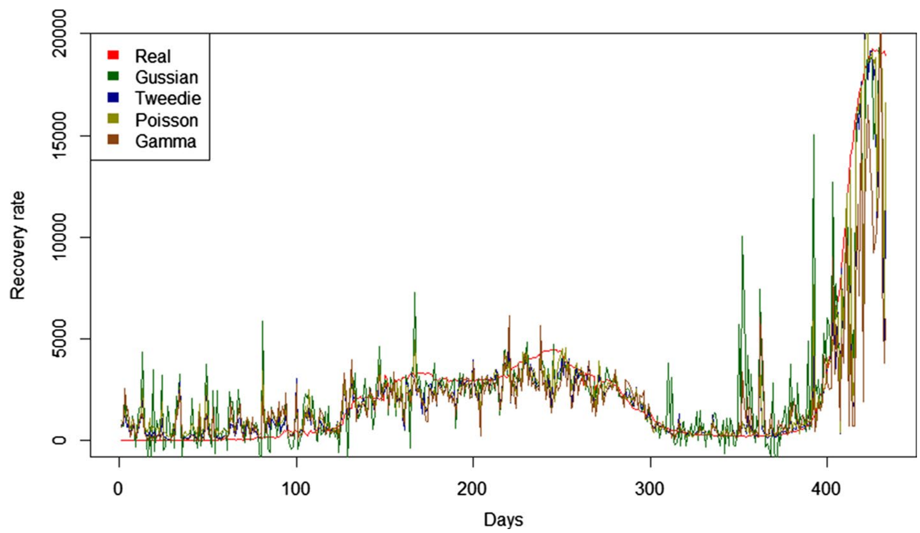

Fig. 26 Correlative capability of twined GBM to forecast recovery rate of COVID-19 in West Bengal

Infection Rate: Maharashtra > Madhya Pradesh > Uttar Pradesh > West Bengal > Karnataka > Delhi > Kerala.

Recovery Rate: Maharashtra > Madhya Pradesh > Uttar Pradesh > West Bengal > Karnataka > Kerala.

Mortality Rate: Maharashtra > Madhya Pradesh > Delhi > West Bengal > Uttar Pradesh > Delhi > Karnataka > Kerala.

The adverse effect of weather parameters like temperature and humidity on the cases of COVID-19 has been reported in some of the recently published research, like high

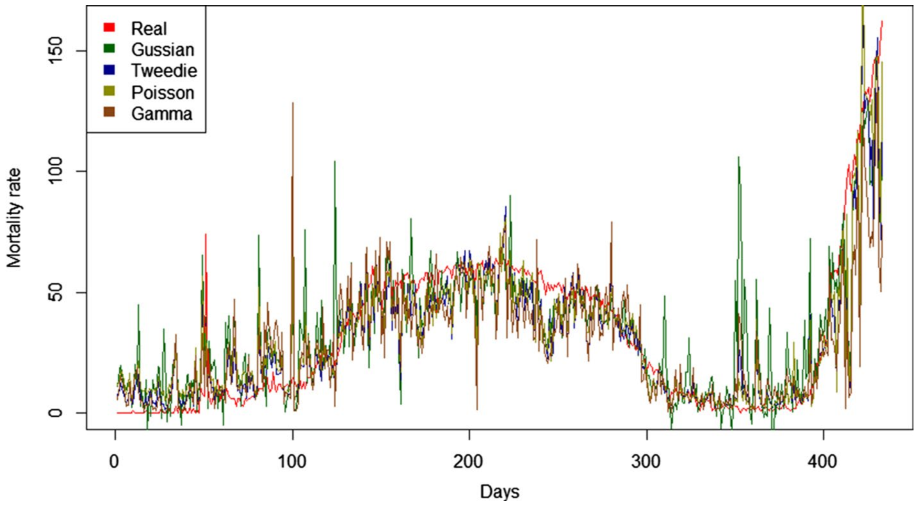

Fig. 27 Correlative capability of twined GBM to forecast mortality rate of COVID-19 in West Bengal

spread rate at low temperature and humidity in Iran [11]; low spread rate at high humidity and temperature in China [16]; and low spread rate of high average humidity and temperature [15]. The impact of additional atmospheric factors like air pressure and wind speed are not been properly noticed in any recent studies. A positive correlation between air pollution and the cases of COVID-19 has been established in some studies, like air pollution and spread rate in Italy and China [7, 20, 21]. Moreover, the atmospheric factors and the air pollution levels are also related; therefore, the present study explored their combined effect (rate of spread) of COVID-19 in major states/places of India using the twinned GBM model. It was noticed that the states having lower mean temperature, humidity, and air pollution as Uttarakhand, Arunachal Pradesh, Himachal Pradesh, Sikkim, Mizoram, etc. have a smaller number of infected, and mortality cases and a higher number of recovered cases than other states/places with high mean temperature, humidity, and air pollution as Maharashtra, Delhi, Karnataka, Kerala, and Madhya Pradesh, etc. However, in some states, it is still difficult to understand the correlation between the spread rate of COVID-19, atmospheric factors, and air pollution measures. The collected data and the analysis outcomes of the different distribution of GBM suggest a significant correlation between the spread rate of COVID-19, atmospheric factors, and air pollution measures in most of the states of India. Besides, the high population density of some of the states and activities of people towards the government regulations, movement of migrant workers, social gatherings, etc. during the lockdown period are also some factors responsible for the spread of COVID-19.

Maharashtra, Delhi, Kerala, Karnataka, Madhya Pradesh, Uttar Pradesh, and West Bengal are worst affected states than other states of India. The predicted numbers of infected cases in Maharashtra, Madhya Pradesh, and Uttar Pradesh by different distribution of GBM are equal to their exact values for most of the day (Figs. 19, 20, 21, 22, 23 and 24). Therefore, Maharashtra was the ideal place for the spread and mortality. The missing information on the atmospheric factors, air pollution measures, and cases of COVID-19 in the duration of data collection may be one of the reasons for the average and poor forecast metrics of the different distribution of GBM for some states.

6 Conclusions and Future Research Scope

This paper presents a correlation between the atmospheric factors, air pollution measures, and infection, recovery, and mortality rate of COVID-19 in the significant states/places of India. The paper proposed a twin GBM model to capture the deep and intrinsic nature of the different datasets. The experimental results confirms that the improved GBM model is proficient enough to determine the correlation among atmospheric parameters, air pollution measures, and COVID-19 impact (infection, recovery, and mortality rate) in the aggregate dataset of different states/places of India. The enhanced performance metrics (R^2 and different errors mechanism) of the improved GBM establish a convinced connotation of transmission rates of COVID-19 with air pollution measures and atmospheric factors. Particularly in some states like Maharashtra, Delhi, Karnataka, Kerala, Madhya Pradesh, Uttar Pradesh, and West Bengal where maximum number of COVID-19 cases have been reported, the air pollution measures and atmospheric factors have a significant role in the spread of the pandemic. Future research will focus on improving the state-wise prediction efficiency of COVID-19 cases by considering more parameters of the weather and atmospheric pollutants.

Funding No funding is available for this work.

Data Availability Data may be provided on individual requests.

Declarations

Conflict of interest The author declares no conflict of interest directly or indirectly related to the work submitted for publication.

Informed Consent Informed consent was obtained from all individual participants included in the study.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. World Health Organization. (2019). *Coronavirus disease (COVID-19) Pandemic*. Retrieved from https://www.who.int/emergencies/diseases/novel-coronavirus-2019
2. Ministry of Health and Family Welfare, *Government of India*. Retrieved from https://www.mohfw.gov.in/
3. Adhikari, S. P., Meng, S., Wu, Y. J., Mao, Y. P., Ye, R. X., Wang, Q. Z., Sun, C., Sylvia, S., Rozelle, S., Raat, H., & Zhou, H. (2020). Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (COVID-19) during the early outbreak period: A scoping review. *Infectious Diseases of Poverty, 9*(1), 1–12. https://doi.org/10.1186/s40249-020-00646-x
4. Singhal, T. (2020). A review of coronavirus disease-2019 (COVID-19). *The Indian Journal of Pediatrics*. https://doi.org/10.1007/s12098-020-03263-6
5. Xu, X., Chen, P., Wang, J., Feng, J., Zhou, H., Li, X., Zhong, W., & Hao, P. (2020). Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. *Science China Life Sciences, 63*(3), 457–460. https://doi.org/10.1007/s11427-020-1637-5
6. Guo, Y. R., Cao, Q. D., Hong, Z. S., Tan, Y. Y., Chen, S. D., Jin, H. J., Tan, K. S., Wang, D. Y., & Yan, Y. (2020). The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19)

outbreak–an update on the status. *Military Medical Research, 7*(1), 1–10. https://doi.org/10.1186/s40779-020-00240-0
7. Ogen, Y. (2020). Assessing nitrogen dioxide (NO2) levels as a contributing factor to the coronavirus (COVID-19) fatality rate. *Science of the Total Environment, 726*, 138605. https://doi.org/10.1016/j.scitotenv.2020.138605
8. Zhang, W., Du, R. H., Li, B., Zheng, X. S., Yang, X. L., Hu, B., Wang, Y. Y., Xiao, G. F., Yan, B., Shi, Z. L., & Zhou, P. (2020). Molecular and serological investigation of 2019-nCoV infected patients: Implication of multiple shedding routes. *Emerging Microbes & Infections, 9*(1), 386–389. https://doi.org/10.1080/22221751.2020.1729071
9. Lowen, A. C., Mubareka, S., Steel, J., & Palese, P. (2007). Influenza virus transmission is dependent on relative humidity and temperature. *PLoS Pathog, 3*(10), e151. https://doi.org/10.1371/journal.ppat.0030151
10. Lin, K., Fong, D. Y. T., Zhu, B., & Karlberg, J. (2006). Environmental factors on the SARS epidemic: Air temperature, passage of time and multiplicative effect of hospital infection. *Epidemiology & Infection, 134*(2), 223–230. https://doi.org/10.1017/S0950268805005054
11. Ahmadi, M., Sharifi, A., Dorosti, S., Ghoushchi, S. J., & Ghanbari, N. (2020). Investigation of effective climatology parameters on COVID-19 outbreak in Iran. *Science of The Total Environment, 729*, 138705. https://doi.org/10.1016/j.scitotenv.2020.138705
12. Ma, Y., Zhao, Y., Liu, J., He, X., Wang, B., Fu, S., Yan, J., Niu, J., Zhou, J., & Luo, B. (2020). Effects of temperature variation and humidity on the death of COVID-19 in Wuhan, China. *Science of The Total Environment, 724*, 138226. https://doi.org/10.1016/j.scitotenv.2020.138226
13. Mecenas, P., Bastos, R., Vallinoto, A., & Normando, D. (2020). Effects of temperature and humidity on the spread of COVID-19: A systematic review. *MedRxiv*. https://doi.org/10.1101/2020.04.14.20064923
14. Oliveiros, B., Caramelo, L., Ferreira, N. C., & Caramelo, F. (2020). Role of temperature and humidity in the modulation of the doubling time of COVID-19 cases. *MedRxiv*. https://doi.org/10.1101/2020.03.05.20031872
15. Qi, H., Xiao, S., Shi, R., Ward, M. P., Chen, Y., Tu, W., Su, Q., Wang, W., Wang, X., & Zhang, Z. (2020). 'COVID-19 transmission in Mainland China is associated with temperature and humidity: A time-series analysis. *Science of The Total Environment, 728*, 138778. https://doi.org/10.1016/j.scitotenv.2020.138778
16. Wang, M., Jiang, A., Gong, L., Luo, L., Guo, W., Li, C., Zheng, J., Li, C., Yang, B., Zeng, J., & Chen, Y. (2020). Temperature significant change COVID-19 Transmission in 429 cities. *MedRxiv*. https://doi.org/10.1101/2020.02.22.20025791
17. Zhu, Y., & Xie, J. (2020). Association between ambient temperature and COVID-19 infection in 122 cities from China. *Science of The Total Environment, 724*, 138201. https://doi.org/10.1016/j.scitotenv.2020.138201
18. Gupta, A., Banerjee, S., & Das, S. (2020). Significance of geographical factors to the COVID-19 outbreak in India. *Modeling Earth Systems and Environment, 6*, 2645–2653. https://doi.org/10.1007/s40808-020-00838-2
19. Baldasano, J. M. (2020). COVID-19 lockdown effects on air quality by NO2 in the cities of Barcelona and Madrid (Spain). *Science of the Total Environment, 741*, 140353. https://doi.org/10.1016/j.scitotenv.2020.140353
20. Conticini, E., Frediani, B., & Caro, D. (2020). Can atmospheric pollution be considered a co-factor in extremely high level of SARS-CoV-2 lethality in Northern Italy? *Environmental Pollution, 261*, 114465. https://doi.org/10.1016/j.envpol.2020.114465
21. Han, Y., Lam, J.C., Li, V.O., Guo, P., Zhang, Q., Wang, A., Crowcroft, J., Wang, S., Fu, J., Gilani, Z., & Downey, J. (2020). The Effects of Outdoor Air Pollution Concentrations and Lockdowns on COVID-19 Infections in Wuhan and Other Provincial Capitals in China. *Preprints*, 2020030364. https://doi.org/10.20944/preprints202003.0364.v1.
22. Jha, S. K., Pan, Z., Elahi, E., & Patel, N. (2019). A comprehensive search for expert classification methods in disease diagnosis and prediction. *Expert Systems, 36*(1), e12343. https://doi.org/10.1111/exsy.12343
23. Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology, 2*(4), 230–243. https://doi.org/10.1136/svn-2017-000101
24. Ramesh, A. N., Kambhampati, C., Monson, J. R., & Drew, P. J. (2004). Artificial intelligence in medicine. *Annals of the Royal College of Surgeons of England, 86*(5), 334–338. https://doi.org/10.1308/147870804290
25. Allam, Z. and Jones, D.S. (2020). On the coronavirus (COVID-19) outbreak and the smart city network: Universal data sharing standards coupled with artificial intelligence (AI) to benefit urban health monitoring and management', In *Healthcare* (Vol. 8, No. 1, p. 46). Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/healthcare8010046.

26. Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X., Kong, B., Bai, J., Lu, Y., Fang, Z., Song, Q., & Cao, K. (2020). Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct. *Radiology*. https://doi.org/10.1148/radiol.2020200905
27. McCall, B. (2020). COVID-19 and artificial intelligence: Protecting health-care workers and curbing the spread. *The Lancet Digital Health, 2*(4), e166–e167. https://doi.org/10.1016/S2589-7500(20)30054-6
28. Pham, Q. V., Nguyen, D. C., Hwang, W. J., & Pathirana, P. N. (2020). Artificial Intelligence (AI) and big data for coronavirus (COVID-19) pandemic: A survey on the state-of-the-arts. *IEEE Access, 8*, 130820–130839. https://doi.org/10.20944/preprints202004.0383.v1
29. Rao, A. S. S., & Vazquez, J. A. (2020). Identification of COVID-19 can be quicker through artificial intelligence framework using a mobile phone-based survey in the populations when cities/towns are under quarantine. *Infection Control & Hospital Epidemiology*. https://doi.org/10.1017/ice.2020.61
30. India Metrological Department, Ministry of Earth Sciences. Government of India. https://mausam.imd.gov.in/
31. Central Pollution Control Board, Ministry of Environment, Government of India. https://cpcb.nic.in/
32. A volunteer-driven crowdsourced effort to track the coronavirus in India. https://www.covid19india.org/
33. The H_2O.ai Team (2015). *h2o: R Interface for H_2O*, R package version 3.1.0.99999. http://www.h2o.ai
34. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.
35. Chen, X., Huang, L., Xie, D., & Zhao, Q. (2018). EGBMMDA: Extreme gradient boosting machine for MiRNA-disease association prediction. *Cell Death & Disease, 9*(1), 1–16. https://doi.org/10.1038/s41419-017-0003-x
36. Geurts, P., Irrthum, A., & Wehenkel, L. (2009). Supervised learning with decision tree-based methods in computational and systems biology. *Molecular Biosystems, 5*(12), 1593–1605. https://doi.org/10.1039/B907946G

Lokesh Kumar Shrivastav is an Assistant Professor in the Department of Computer Science, Atma Ram Sanatan Dharma College, University of Delhi, India. He received his B. Sc. in Physics, MCA, and M. Tech in CSE respectively. He is pursuing his Ph. D from Guru Gobind Singh Indraprastha University, Delhi, India. His research interest includes mathematical modeling of infectious disease dynamics and prediction, machine Learning, data analytics etc.

Ravinder Kumar has completed his M tech. CSE from GJUS & T Hisar and Ph.D. from GGSIPU Delhi India in the domain of Biometrics. He is presently Associate Professor in CSE/IT at Shri Vishwakarma Skill University, Palwal, India. His research interest includes Image processing, Biometrics, Artificial Neural Network, Privacy Preservation and Data Analytics.