Big Data Mining and Analytics

Volume 4 | Issue 2 Article 1

2021

Analysis and Predictions of Spread, Recovery, and Death Caused by COVID-19 in India

Rajani Kumari
Department of Information Technology and Computer Application, JECRC University, Jaipur, Rajasthan 303905, India

Sandeep Kumar
CHRIST ,Bangalore, Karnataka 560029, India

Ramesh Chandra Poonia
Amity University Rajasthan, Jaipur, Rajasthan 303002, India

Vijander Singh
Manipal University Jaipur, Rajasthan 303007, India

Linesh Raja
Manipal University Jaipur, Rajasthan 303007, India

See next page for additional authors

Analysis and Predictions of Spread, Recovery, and Death Caused by COVID-19 in India

Authors

Rajani Kumari, Sandeep Kumar, Ramesh Chandra Poonia, Vijander Singh, Linesh Raja, Vaibhav Bhatnagar, and Pankaj Agarwal

BIG DATA MINING AND ANALYTICS
ISSN 2096-0654 01/07 pp65–75
Volume 4, Number 2, June 2021
DOI: 10.26599/BDMA.2020.9020013

Analysis and Predictions of Spread, Recovery, and Death Caused by COVID-19 in India

Rajani Kumari, Sandeep Kumar*, Ramesh Chandra Poonia, Vijander Singh, Linesh Raja, Vaibhav Bhatnagar, and Pankaj Agarwal

Abstract: The novel coronavirus outbreak was first reported in late December 2019 and more than 7 million people were infected with this disease and over 0.40 million worldwide lost their lives. The first case was diagnosed on 30 January 2020 in India and the figure crossed 0.24 million as of 6 June 2020. This paper presents a detailed study of recently developed forecasting models and predicts the number of confirmed, recovered, and death cases in India caused by COVID-19. The correlation coefficients and multiple linear regression applied for prediction and autocorrelation and autoregression have been used to improve the accuracy. The predicted number of cases shows a good agreement with 0.9992 R-squared score to the actual values. The finding suggests that lockdown and social distancing are two important factors that can help to suppress the increasing spread rate of COVID-19.

Key words: COVID-19; regression; correlation; machine learning; prediction

1 Introduction

The coronavirus disease spreads through getting in touch with an infected person, touching a thing or object that has the virus on its surface and then touching their mouth, eyes, ears, or nose. The first case of COVID-19 was detected in the last week of January 2020 in India and only 3 cases were diagnosed in next month. As of 6 June 2020, the total number of confirmed cases in India was 247 857 with 119 293 recovered cases and 6954 deaths. There is a need for the current situation to predict possible infected and death cases by using a computational model to arrange the necessary resources.

The virus of COVID-19 shows great resemblance with the Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS) coronavirus as investigated by pathologists[1]. It is the seventh member of the coronavirus family that can spread among humans and easily transmit human-to-human through droplets of coughs or sneezing by an infected person[1]. Major symptoms of COVID-19 are fever, cough, shortness of breath, and some patients show symptoms of diarrhea. The major problem in the case of this disease is that its symptoms generally appear after 2 to 14 days if an individual gets infected. This period is known as the incubation period and the mean incubation period is approximately 5 days[2]. An infected person may infect the number of healthy persons during the incubation period. These patients are asymptomatic and major challenge is to identify them. The rate at which one infected person transmits this disease into others is termed as transmission rate (R_o)[1]. Recent studies by leading research organizations estimated that R_o is between 1.5 and 3.5[3–5]. This rate of transmission is very high in comparison to SARS (2.0) and common flu (1.3), and thus it is very dangerous. In the early stage, the Case

• Rajani Kumari is with Department of Information Technology and Computer Application, JECRC University, Jaipur, Rajasthan 303905, India. E-mail: rajanikpoonia@gmail.com.
• Sandeep Kumar is with CHRIST (Deemed to be University), Bangalore, Karnataka 560029, India. E-mail: sandpoonia@gmail.com.
• Ramesh Chandra Poonia and Pankaj Agarwal are with Amity University Rajasthan, Jaipur, Rajasthan 303002, India. E-mail: rameshcpoonia@gmail.com; mr.pankajagarwal@gmail.com.
• Vijander Singh, Linesh Raja, and Vaibhav Bhatnagar are with Manipal University Jaipur, Rajasthan 303007, India. E-mail: vijan2005@gmail.com; lineshraja@gmail.com; vaibhav.bhatnagar15@gmail.com.
* To whom correspondence should be addressed.
Manuscript received: 2020-05-07; revised: 2020-06-14; accepted: 2020-07-28

Fatality Rate (CFR) for coronavirus was estimated at 2%[1]. While the fatality rate for SARS and MERS was approximately 10% and 30%, respectively[2]. Highest CFR was reported in France (19%), followed by Belgium (16.23%), Italy (14.4%), and UK (14.2%)[6]. Almost 5% of the total infected persons lost their lives around the world, while in India CFR was near to 2.8%. The actual CFR can only be computed based on the identification of the correct number of infected individuals. The major contributions of this research are as follows:

- This paper performs a descriptive analysis of the COVID-19 outbreak in different states of India.
- We propose a model for predicting the number of confirmed, recovered, and death cases due to COVID-19.
- The proposed model deploys multiple linear regression to analyze the existing data.
- This paper also employs the autoregression to predict the cases.

The organization of this paper is as follows: Section 2 discusses some recent studies carried out by researchers, medical practitioners, and scientists in the field of infectious disease. In Section 3 the detailed analysis is carried out from the considered dataset of COVID-19 for India. Section 4 introduces the proposed prediction model to estimate the count of infection, recovery, and death due to COVID-19. Section 5 concludes this study.

2 Recent Study on COVID-19

Many researchers involved in the study of novel coronavirus after the outbreak at Wuhan, China in late December 2019 and developed various types of models for prediction of its spread, transmission, and death caused by it. Some studies and researches are related to the development of medicine and diagnostic tool for this pandemic. Few of these recent studies discussed here.

Zhong et al.[7] developed a mathematical model for the timely prediction of the coronavirus outbreak in China. Hamzah et al.[8] developed an online platform to provide real-time information related to COVID-19 and a statistical analysis of data. Susceptible-Exposed-Infectious-Recovered (SEIR) predictive modeling was used for forecasting on daily basis. They developed their micro-services to fetch data from different sources. Morawska and Cao[9] discussed how COVID-19 spreads especially through the air.

Li et al.[10] investigated the genetic evolution of the virus that is responsible for COVID-19. This study identified that novel coronavirus has a genetic similarity with coronavirus derived from rhinolophus sinicus, paradoxurus hermaphroditus, paguma larvata, aselliscus stoliczkanus, and civet, while homology analysis shows that it has close resemblance with bat coronavirus.

Ma et al.[11] analyzed the effect of humidity and changes in temperature on COVID-19 patients, but the study was limited to Wuhan city only. This study established a correlation with variation in temperature and humidity on daily death due to the virus. Singh et al.[12] studied and compared SARS, MERS, and COVID-19 viruses based on transmission cycle, etiology, genetics, hosts, diagnosis, reproductive rates, laboratory diagnosis, clinical features, and radiological features. Pal et al.[13] illustrated the classification of ribonucleic acid group of viruses and origin of severe acute respiratory syndrome coronavirus along with virion structure and genetic characteristics of COVID-19. Dutheil et al.[14] investigated the role of COVID-19 for decreasing air pollution as most of the industries are shut down and traffic is also significantly low.

Vellingiri et al.[15] discussed the cause of infection, symptoms, and the structure of the virus in detail and compared it with common flu, SARS, and MERS at various parameters. They also discussed ongoing treatment to the infected people and suggested some Indian plants for medical use. Henry and Lippi[16] suggested that Extracorporeal Membrane Oxygenation (ECMO) is one of the options for survival therapy to COVID-19 patients. Some limitations of ECMO were also discussed here. Lai et al.[17] also discussed major symptoms and ongoing methods of cure for COVID-19. They raised some unresolved issues, like the presence of SARS-CoV-2 in patient stool and the efficiency of disinfection agents used for sanitization.

Ghosal et al.[18] developed a model to predict week wise death in India due to COVID-19. They used linear regression and multiple regression for prediction and deployed autoregression to enhance the prediction capability of the proposed model. The projected model is based on data analysis of 15 highly infected countries.

Liang[19] compared the spread characteristics of novel coronavirus with characteristics of SARS and MERS. A new mathematical model was proposed to identify the symptoms of coronavirus diseases. Nicola et al.[20] suggested that veterinary medicine may be helpful in the cure of COVID-19.

Lee et al.[21] discussed the importance of Computed Tomography (CT) images in the diagnosis of COVID-19 infected individuals. As technology growing, there are many applications and tools being produced that utilize

various algorithms. In many fields, computer-assisted tools are being designed and employed successfully. The efficient use of such computer-aided systems in medicine is no exception. Medical images are very useful for the doctor, and their detailing can have a decisive influence on the correctness of the diagnosis. One of the branches of the healthcare system that tightly works with images is the radiology. There have been generated several datasets containing CT scans, Magnetic Resonance Imaging (MRI), etc., to detect novel coronavirus pneumonia diseases. Pan et al.[22] illustrated changes in the lung of COVID-19 patients during the recovery process with the help of CT images. Singh et al.[23] classified COVID-19 patients based on their chest CT images. For this classification they implemented a convolutional neural network based on differential evolution algorithm. Jaiswal et al.[24] deployed DenseNet201 for classification. The proposed approaches achieved a higher rate of accuracy and precision.

Singh et al.[25] analyzed time series data and predicted the registered, deceased, and death numbers per reported case (mortality rate) based on COVID-19's world health data for the world population. This study concluded that COVID-19's regular mortality is positively correlated with the number of confirmed cases. It may also be dependent upon the population's dietary routine and robustness of the immune system. This study suggested that an emergency can awaken before the proper vaccine is invented. Some critical issues were measured by several researchers, considering individual countries, provinces, and derived some conclusions. Bhatnagar et al.[26] presented a detailed analysis of the COVID-19 pandemic with the help of a boxplot and Q plot.

Ivanov et al.[27] analyzed and predicted the effect of the ongoing pandemic on global supply chains. They also performed a simulation-based analysis in the case of supply chains and the impact of COVID-19 on supply chains along with associated risks. Hou et al.[28] performed SEIR model analysis to examine the effectiveness of quarantine especially for Wuhan city and developed a new variant of the SEIR model. They concluded that quarantine and isolation are two powerful and unique tools to reduce the risk of infection. Roosa et al.[29] developed a system for forecasting the COVID-19 in real-time in China for a specific time period. Tuli et al.[30] employed the latest technologies, like machine learning and cloud computing, for predicting the growth rate of COVID-19 pandemic with the help of the Weibull model.

Xu et al.[31] explained the pathological characteristics of COVID-19 and compared them with SARS and MERS. These pathological features are highly similar to SARS and MERS. This study provided some recommendations to physicians so that they can timely plan a therapeutic strategy for the patient. Kucharski et al.[32] developed a mathematical model and analyzed four datasets. This study revealed that the transmission rate is between 1.6 to 2.6. Here they classified patients into four different classes: susceptible, exposed (but not yet infectious), infectious, and removed (i.e., isolated, recovered, or otherwise no longer infectious). Yuvaraj et al.[33] used a deep neural network for the analysis of interactions of protein-ligand for SARS-CoV-2 against selective drugs. Some studies focused on psychological health of farmers engaged in the business of poultry[34].

Researchers are also working on test procedures and trying to reduce testing time. In this sequence, Assad et al.[35] suggested that sample pooling is the best option to reduce the testing time that leads to reduce fatality but with a limitation of 10% positive cases. If positive cases are very low, binary elimination algorithms are the better option.

These studies revealed that symptoms of COVID-19 are similar to SARS and MERS. COVID-19 is more infectious but has a low fatality rate. The virus' root cause is still unclear, and virologists are actively working to establish its antidote. However, physicians are continuously trying to cure patients by using antiviral therapy, antibiotic treatment, systematic corticosteroids, etc. Table 1 summarises a few of the recently developed prediction and forecasting models. Most of the models are based on the SEIR model and its extended version, like symptomatic infectious, asymptomatic infectious, quarantined, hospitalized, recovered, dead model (SEI_DI_UQHRD)[36–38]. Machine learning and deep learning models are also used for prediction and forecasting[39–41].

3 Analysis of COVID-19 Data

Analysis of the COVID-19 dataset for coronavirus disease is performed on the basis of reported cases (confirmed, recovered, and death) in India. We have collected data from 29 February 2020 to 6 June 2020 (hereafter it is termed as $Week_1$ to $Week_{15}$) of some states in India that are worst hit by this virus. The dataset is taken from www.kaggle.com[54] and shown in Table 2. Attributes which are considered in this dataset

Table 1 Comparative study of recent prediction and forecasting models for COVID-19.

Author(s)	Activity performed	Methodology used	Strength	Drawback
Ghosal et al.[18]	Prediction	Linear regression analysis	Statistical analysis results prove its reliability.	Results are over-estimated and significance of predictor is very low.
Singh et al.[23]	Prediction	Gaussian mixture model	Predicted values with 95% confidence intervals	Predicted end dates are not true.
Sarkar et al.[36]	Modeling and forecasting	$SARII_qS_q$ model	Considered six components: Susceptible (S), Asymptomatic (A), Recovered (R), Infected (I), Isolated Infected (Iq), and Quarantined Susceptible (Sq)	Accuracy and reliability of this model depend on the assumption of parameter values and initial population size.
Nabi[37]	Forecasting	Developed a new model (SEI_DI_UQHRD)	Developed a reliable model using trust-region-reflective algorithm	Accuracy and reliability of this model depend on the assumption of parameter values.
Kanagarathinam and Sekar[38]	Prediction	SEIR model	Maximum-likelihood and bootstrap strategy are used to analyze the R_0 and re-sampling, respectively.	Data considered from 2 March 2020 to 2 April 2020 only.
Arora et al.[39]	Prediction and analysis	Deep learning	Deep LSTM, convolutional LSTM, and bi-directional LSTM are deployed for accurate prediction.	Tested for 15 days only
Wang et al.[40]	Prediction	Logistic model and machine learning technique	Fbprophet model deployed for forecasting in various countries	Predicted number of infected persons only
Sujath et al.[41]	Forecasting	Machine learning	Deployed multilayered perceptron, linear regression, and vector autoregression for better results	Results may be improved by deep learning.
Rafiq et al.[42]	Evaluation and prediction	Predictive error minimization-based approach	Performed prediction for one month	Used different models for different states
Sahoo and Sapra[43]	Prediction	Mathematical model developed for predication	Predicted peak time and end time of the pandemic, this model also analyzed the effect of lockdown.	Only 10 days data used for testing purpose
Goswami et al.[44]	Prediction and analysis of meteorological factors	Generalized additive model, Sen's slope, Man-Kendall test, and Verhulst (logistic) population model	Considered the impact of temperature and humidity	Results are not consistent throughout the study area.
Salgotra et al.[45]	Forecasting	Gene expression programming	Model is highly reliable	Considered small size data (till 24 March 2020)
Tomar and Gupta[46]	Prediction	LSTM and curve-fitting	Predicted the effect of social distancing for 30 days and lockdown also analyzed	The error rate is very high.
Gupta et al.[47]	Predicted risk based on weather conditions	Daily temperature and relative humidity-weighted against cases	Analyzed data of the USA and India	Considered only weather conditions that less significant
Mandal et al.[48]	Prediction	Mathematical model	Analyzed reproduction number and sensitivity analysis to decide the preventive measure	Data were taken till 27 March 2020 only.
Chakraborty and Ghosh[49]	Forecasts and risk assessment	Hybrid of wavelet-based forecasting and ARIMA model	A risk assessment performed using regression tree	Forecasts for ten days only
Tiwari et al.[50]	Prediction	Machine learning	Predicted results with higher accuracy	Proposed model used data till 3 April 2020 only
Maheshwari et al.[51]	Forecasting	ARIMA model	R statistical package used for forecasting for the next 76 days	Achieved accuracy (93.695%, 86.96%, 87.94%, and 90.91% for confirmed, recovered, death, and death rate, respectively) is significantly low.
Bhattacharjee et al.[52]	Prediction	New mathematical model proposed	Cases load rate based on cumulative confirmed cases and the recovery rate are used for prediction.	Study is restricted till 24 April 2020.
Sree[53]	Prediction	Cellular automata classifier	Hybrid non-linear cellular automata deployed for prediction	Achieved accuracy (78.8%) is significantly low.

are mainly week wise confirmed cases in concerning states. After collecting the required data, they are refined and analyzed.

Tables 3 and 4 illustrate the mathematical description of considered dataset and correlation among those data, respectively. In Table 3, the notations: Count, Mean,

Table 2 Dataset from $Week_1$ to $Week_{15}$ including confirmed cases in different states of India. CH: Chandigarh, KR: Karnataka, MP: Madhya Pradesh, MH: Maharashtra, and TL: Telengana.

Week	Kerala	Gujarat	CH	Delhi	KR	Ladakh	MP	MH	TL
1	3	2	1	1	1	1	1	1	1
2	8	7	2	4	3	2	2	2	1
3	22	13	3	7	6	3	3	32	3
4	52	18	5	29	26	12	4	67	22
5	182	58	8	49	76	13	30	186	66
6	306	122	18	503	144	14	165	490	269
7	364	308	18	903	214	15	443	1574	504
8	396	1272	21	1707	371	18	1355	3323	791
9	453	2848	37	2501	482	19	1974	7029	988
10	499	4948	49	4068	611	32	2838	11 823	1062
11	505	7404	59	6261	750	42	3433	19 101	1143
12	587	9931	66	8895	1056	43	4595	29 100	1454
13	794	13 268	172	12 319	1743	44	6170	44 582	1761
14	1208	15 953	289	17 415	2728	74	7672	62 357	2378
15	1807	18 584	301	25 004	4329	90	8762	77 793	3147

Table 3 Mathematical description of datasets.

Week	Count	Mean	Std	Min	25%	50%	75%	Max
1	23	2.13	2.88	1	1	1	1.5	14
2	23	4.22	4.75	1	1	2	4.5	18
3	23	13.74	17.29	1	3	7	21	80
4	23	38.04	43.83	3	10	26	50.5	190
5	22	104.5	126.74	6	22.5	57.5	163	485
6	18	254.44	239.93	14	75.5	171	401.25	911
7	9	482.56	490.04	15	214	364	504	1574
8	9	1028.22	1049.76	18	371	791	1355	3323
9	9	1814.56	2220.68	19	453	988	2501	7029
10	9	2881.11	3806.50	32	499	1062	4068	11 823
11	9	4299.78	6188.59	42	505	1143	6261	19 101
12	9	6191.89	9382.13	43	587	1454	8895	29 100
13	9	8983.67	14 291.02	44	794	1761	12 319	44 582
14	9	12 230.44	19 916.79	74	1208	2728	15 953	62 357
15	9	15 535.22	24 897.31	90	1807	4329	18 584	77 793

Std, Min, Max, 25%, 50%, and 75% are used to denote the number of non-null values, mean of values, the standard deviation of the values, minimum value, maximum value, first quartile, second quartile, and third quartile, respectively. These same notations also used in Table 5 for the statistical description of the considered dataset. The objective of this analysis is to find the correlation between $Week_1$ to $Week_{15}$ for all confirmed cases in different states. Through this analysis, it is observed that there is a strong correlation between the complete datasets. Table 4 represents the correlation analysis which determines the relationship among $Week_1$ to $Week_{15}$ data. According to descriptive analysis concerning spread rate of coronavirus disease in different states, it is observed that in the first four to five weeks the spread rate of this virus is very less in India, but after that spread rate is very high in some of the states in India due to social gathering by a single source (refer to Table 3). It is observed from Figs. 1 and 2 that till $Week_4$ the spread rate of confirmed cases is very low and $Week_5$ onwards spread rate is very high. Figure 1 illustrates the confirmed cases in considered states and shows that exponential growth in confirmed cases occurs after the fourth week. Similarly, Figs. 2 and 3 also illustrate an exponential growth pattern for confirmed cases in considered states and it indicates that in the near future situation it will be very tough if it is not controlled.

Table 4 Correlation analysis of determining relationship between datasets.

Week	Week														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.00														
2	0.91	1.00													
3	0.45	0.40	1.00												
4	0.37	0.30	0.92	1.00											
5	0.53	0.40	0.92	0.95	1.00										
6	0.06	0.11	0.60	0.75	0.64	1.00									
7	−0.15	−0.11	0.70	0.74	0.62	0.89	1.00								
8	−0.11	0.13	−0.07	0.00	−0.19	0.45	0.28	1.00							
9	−0.16	0.09	−0.15	−0.06	−0.25	0.43	0.23	0.98	1.00						
10	−0.16	0.10	−0.18	−0.10	−0.28	0.39	0.19	0.97	1.00	1.00					
11	−0.15	0.12	−0.18	−0.10	−0.28	0.39	0.18	0.96	0.99	1.00	1.00				
12	−0.14	0.13	−0.17	−0.09	−0.27	0.41	0.20	0.95	0.99	1.00	1.00	1.00			
13	−0.14	0.14	−0.17	−0.07	−0.25	0.43	0.21	0.95	0.99	0.99	1.00	1.00	1.00		
14	−0.12	0.16	−0.15	−0.05	−0.24	0.45	0.22	0.94	0.98	0.98	0.99	1.00	1.00	1.00	
15	−0.09	0.20	−0.13	−0.05	−0.22	0.45	0.22	0.95	0.98	0.98	0.99	0.99	1.00	1.00	1.00

Table 5 Statistical description of datasets.

Dataset	Count	Mean	Std	Min	25%	50%	75%	Max
Confirmed cases	70.00	69 595.39	70 086.76	1024.00	12 599.00	41 102.00	110 639.75	246 622.00
Recovered cases	70.00	27 765.33	34 180.32	95.00	1516.00	11 297.00	44 643.75	118 695.00
Death cases	70.00	2087.83	1981.72	27.00	415.75	1357.00	3401.00	6946.00

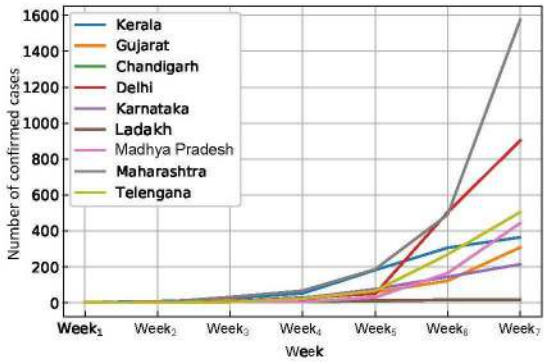

Fig. 1 Confirmed cases of COVID-19 in India.

4 Proposed Model for COVID-19 Predictions

Prediction using the proposed model is performed for data from 20 March 2020 to 6 June 2020. This date range is different from the date range considered for analysis, because initially the number of cases is very low and the use of that data may affect the accuracy of the model, data after 20 March 2020 are considered when the number of COVID-19 patients are significantly higher. Data were collected in the CSV file (from www.kaggle.com[54]) and imported in Jupyter notebook through anaconda navigator and analyzed with Python 3.7.6 software. Attributes that were considered in this dataset are mainly confirmed, recovered, and death cases. Figure 3 shows the graphical representation of the dataset. It is assumed that the coronavirus-infected persons are available in India and they come into contact with other healthy persons. Since it is an infectious disease, it is going to spread into others also. Consequently, the number of cases is growing rapidly.

During the development of the model, the collected data were analyzed by using functions in Python Software. For understanding the dataset properly, a statistical description was performed on the complete dataset. The description of statistical data is shown in Table 5.

The proposed model is summarised in Fig. 4. It is important to discover and compute the degree of variables in the dataset and this information is helpful for better preparation of dataset to meet the expectations of machine learning algorithms. A recovery strategy and correlation analysis are performed on data using Python Software. It reveals a statistical summary of confirmed, recovered, and death cases and also finds a strong relationship among current data. The consequential correlation analysis is shown in Table 6.

Multiple regression analysis is used for predicting

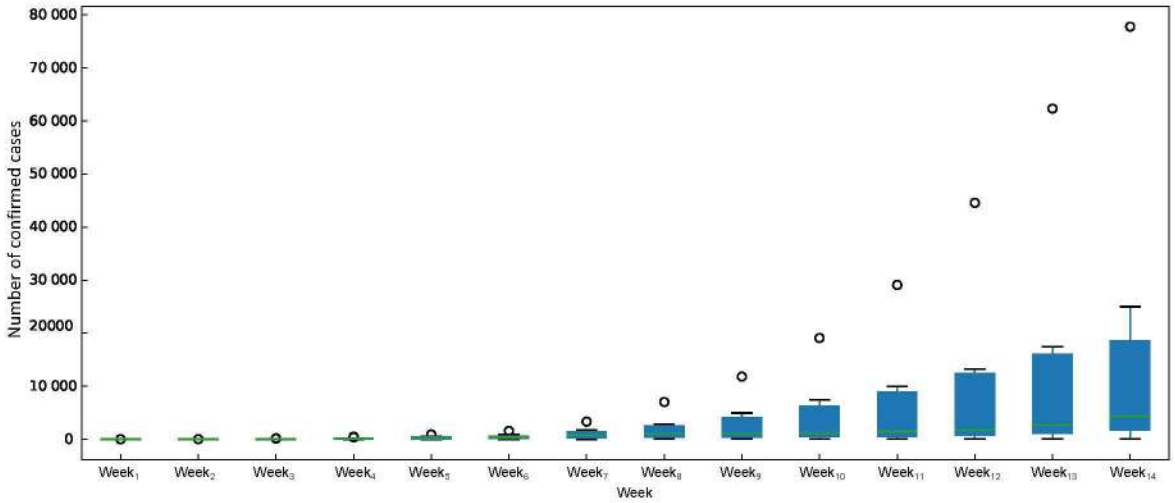

Fig. 2 Boxplot for confirmed cases of COVID-19 in India.

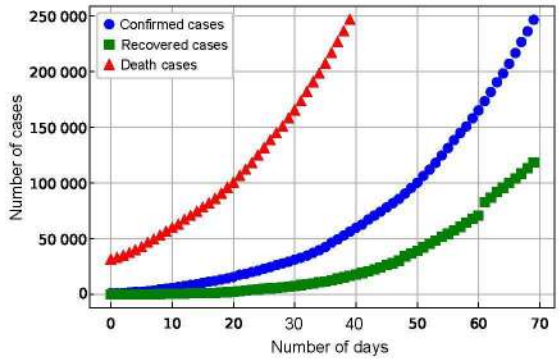

Fig. 3 Confirmed, recovered, and death cases in India.

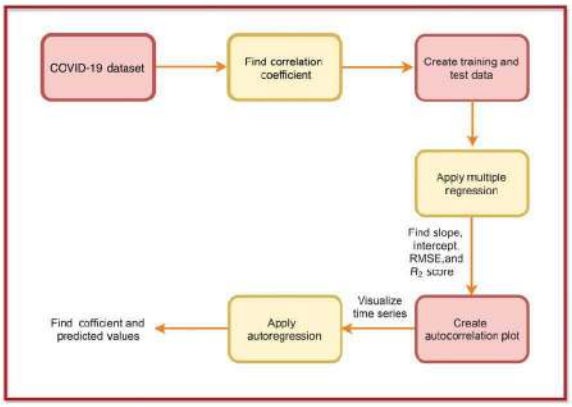

Fig. 4 Proposed model for predictions of confirmed, recovered, and death cases of COVID-19 in India.

Table 6 Correlation analysis of determining relationship between datasets.

	Confirmed cases	Recovered cases	Death cases
Confirmed cases	1		
Recovered cases	0.993 734 20	1	
Death cases	0.998 139 25	0.985 844 39	1

death cases with the help of confirmed and recovered cases. This regression technique has more than one variables to predict the output. It is helpful in predicting a target variable using more than one independent variables. For predictive analysis, the multiple linear regression techniques are used to explain the relationship between two independent variables (confirmed and recovered cases) and one dependent variable (death cases). Here the dataset is divided into training and test datasets as 70% for training and 30% for testing. This model has a very strong predictive capacity after training with the dataset and found Root Mean Square Error (RMSE) as 3085.4305 and R_2 score as 0.9992. The summary of output for multiple linear regression analysis is shown in Table 7.

Decision tree learning techniques are used to continuously split training and test data according to a certain parameter. It is a widely accepted supervised learning approach that split our dataset based on conditions. It is equally useful for regression and classification. This approach assigns the most feasible class for each record for classification. At the time of testing with a different set of input values, it is observed that predicted output is very close to actual values.

During the analysis of the complete dataset, it is

Table 7 Summary of output for multiple linear regression analysis.

Parameter	Value
Slope	[0.0418 −0.0280]
Intercept	−43.5073
RMSE	3085.4305
R_2 score	0.9992

observed that the model can use regression against itself and also able to use the autocorrelation plot to check the randomness within the data. Figure 5 shows the autocorrelation plot. Now create an autoregression model that uses observation from the previous steps as input. The time-series model is used to predict the values at the next time step. Results prove that the forecasted range of time series is accurate. Now fit the model using the existing dataset and find the lag and coefficients. Based on the lag value, a separate analysis is performed for confirmed, recovered, and death cases. It is observed from Table 8 that the testing of existing data is very close to the actual dataset and predicted values are also very relevant to the existing dataset.

5 Conclusion and Future Scope

This study discussed the spread of COVID-19 in different states of India and proposed a model for predicting the number of confirmed, recovered, and death cases. Multiple linear regression and autoregression were used to predict the possible number of cases in the future. The predicted confirmed cases of India for the next 30 days are recorded in Table 8. The predicted values and actual values are together in good agreement (see Fig. 6). This prediction may be helpful in resource management, like health services, and timely action may be taken with prior preparation to reduce the loss of human life.

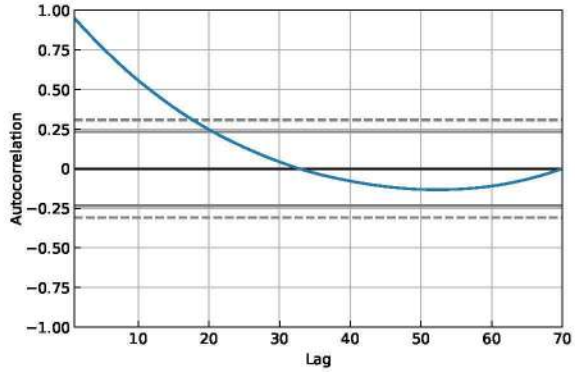

Fig. 5 Autocorrelation plot.

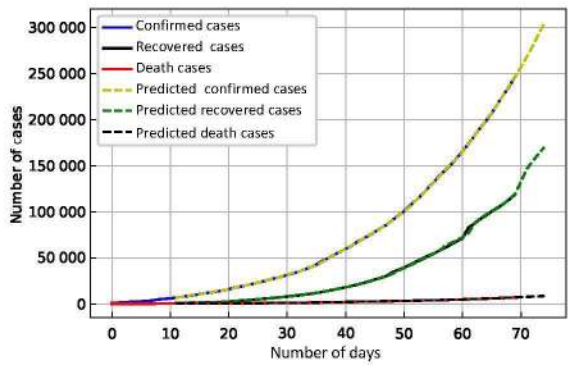

Fig. 6 Actual and predicted: confirmed, recovered, and death cases in India.

Table 8 Prediction for the next 30 days (7 June 2020 to 6 July 2020).

Date	Predicted confirmed cases	Predicted recovered cases	Predicted death cases
07-Jun-20	256 972	132 398	7230
08-Jun-20	268 021	146 535	7529
09-Jun-20	279 935	154 647	7834
10-Jun-20	292 393	162 747	8156
11-Jun-20	304 886	169 880	8500
12-Jun-20	318 185	175 762	8851
13-Jun-20	331 964	180 150	9215
14-Jun-20	346 329	185 645	9589
15-Jun-20	361 364	193 564	9971
16-Jun-20	377 129	210 844	10 373
17-Jun-20	393 235	236 021	10 789
18-Jun-20	410 083	256 198	11 223
19-Jun-20	427 679	271 631	11 676
20-Jun-20	445 982	285 667	12 142
21-Jun-20	465 005	295 537	12 626
22-Jun-20	484 898	300 267	13 128
23-Jun-20	505 477	303 552	13 648
24-Jun-20	526 898	311 531	14 191
25-Jun-20	549 246	332 608	14 753
26-Jun-20	572 521	370 832	15 337
27-Jun-20	596 690	412 622	15 943
28-Jun-20	621 891	447 115	16 571
29-Jun-20	648 083	476 246	17 223
30-Jun-20	675 324	497 141	17 899
01-Jul-20	703 696	504 343	18 602
02-Jul-20	733 247	501 616	19 331
03-Jul-20	763 961	501 854	20 088
04-Jul-20	795 938	521 520	20 873
05-Jul-20	829 215	572 506	21 688
06-Jul-20	863 836	645 885	22 534

The proposed model may be extended to predict the end of this pandemic in a particular region. Total causality and total economic losses may be predicted with the help of this model.

References

[1] World Healthy Organization, Statement on the second meeting of the international health regulations (2005) emergency committee regarding the outbreak of novel coronavirus (2019-nCoV), https://www.who.int/news/item/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov), 2019.

[2] M. Xie and Q. Chen, Insight into 2019 novel coronavirus–

An updated intrim review and lessons from SARS-CoV and MERS-CoV, *International Journal of Infectious Diseases*, vol. 94, pp. 119–124, 2020.

[3] N. Imai, A. Cori, I. Dorigatti, M. Baguelin, C. A. Donnelly, S. Riley, and N. M. Ferguson, Report 3: Transmissibility of 2019-nCOV, https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-2019-nCoV-transmissibility.pdf (2020), 2020.

[4] M. Majumder and K. D. Mandl, Early transmissibility assessment of a novel coronavirus in Wuhan, China, https://papers.ssrn.com/abstract=3524675, 2020.

[5] J. M. Read, J. R. Bridgen, D. A. Cummings, A. Ho, and C. P. Jewell, Novel coronavirus 2019-nCoV: Early estimation of epidemiological parameters and epidemic predictions, doi: 10.1101/2020.01.23.20018549.

[6] COVID-19 coronavirus pandemic, https://www.worldometers.info/coronavirus, 2020.

[7] L. Zhong, L. Mu, J. Li, J. Wang, Z. Yin, and D. Liu, Early prediction of the 2019 novel coronavirus outbreak in the mainland China based on simple mathematical model, *IEEE Access*, vol. 8, pp. 51 761–51 769, 2020.

[8] F. A. B. Hamzah, C. H. Lau, H. Nazri, D. V. Ligot, G. Lee, C. L. Tan, M. K. B. M. Shaib, U. H. B. Zaidon, A. B. Abdullah, M. H. Chung, et al., Coronatracker: Worldwide COVID-19 outbreak data analysis and prediction, *Bull World Health Organ*, http:// dx.doi.org/10.2471/BLT.20.255695.

[9] L. Morawska and J. Cao, Airborne transmission of SARS-CoV-2: The world should face the reality, *Environment International*, vol. 139, p. 105730, 2020.

[10] C. Li, Y. Yang, and L. Ren, Genetic evolution analysis of 2019 novel coronavirus and coronavirus from other species, *Infection, Genetics and Evolution*, vol. 82, p. 104285, 2020.

[11] Y. Ma, Y. Zhao, J. Liu, X. He, B. Wang, S. Fu, J. Yan, J. Niu, J. Zhou, and B. Luo, Effects of temperature variation and humidity on the death of COVID-19 in Wuhan, China, *Science of the Total Environment*, vol. 724, p. 138226, 2020.

[12] A. Singh, A. Shaikh, R. Singh, and A. K. Singh, COVID-19: From bench to bed side, *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 14, no. 4, pp. 277–281, 2020.

[13] M. Pal, G. Berhanu, C. Desalegn, and V. Kandi, Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2): An update, *Cureus*, vol. 12, no. 3, pp. 1–13, 2020.

[14] F. Dutheil, J. S. Baker, and V. Navel, COVID-19 as a factor influencing air pollution? *Environmental Pollution*, vol. 263, p.114466, 2020.

[15] B. Vellingiri, K. Jayaramayya, M. Iyer, A. Narayanasamy, V. Govindasamy, B. Giridharan, S. Ganesan, A. Venugopal, D. Venkatesan, H. Ganesan, et al., COVID-19: A promising cure for the global panic, *Science of the Total Environment*, vol. 725, p. 138277, 2020.

[16] B. M. Henry and G. Lippi, Poor survival with extracorporeal membrane oxygenation in acute respiratory distress syndrome (ARDS) due to coronavirus disease 2019 (COVID-19): Pooled analysis of early reports, *Journal of Critical Care*, vol. 58, pp. 27–28, 2020.

[17] C.-C. Lai, T.-P. Shih, W.-C. Ko, H.-J. Tang, and P.-R. Hsueh, Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): The epidemic and the challenges, *International Journal of Antimicrobial Agents*, vol. 55, no. 3, p. 105924, 2020.

[18] S. Ghosal, S. Sengupta, M. Majumder, and B. Sinha, Prediction of the number of deaths in India due to SARS-CoV-2 at 5–6 weeks, *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 14, no. 4, pp. 311–315, 2020.

[19] K. Liang, Mathematical model of infection kinetics and its analysis for COVID-19, SARS and MERS, *Infection, Genetics and Evolution*, vol. 82, p. 104306, 2020.

[20] D. Nicola, M. Vito, J. S. Linda, and B. Canio, COVID-19 from veterinary medicine and one health perspectives: What animal coronaviruses have taught us? *Research in Veterinary Science*, vol. 131, pp. 21–23, 2020.

[21] E. Y. Lee, M.-Y. Ng, and P.-L. Khong, COVID-19 pneumonia: What has CT taught us? *The Lancet Infectious Diseases*, vol. 20, no. 4, pp. 384–385, 2020.

[22] F. Pan, T. Ye, P. Sun, S. Gui, B. Liang, L. Li, D. Zheng, J. Wang, R. L. Hesketh, L. Yang, et al., Time course of lung changes on chest CT during recovery from 2019 novel coronavirus (COVID-19) pneumonia, *Radiology*, vol. 295, no. 3, pp. 715–721, 2020.

[23] D. Singh, V. Kumar, and M. Kaur, Classification of COVID-19 patients from chest CT images using multi-objective differential evolution-based convolutional neural networks, *European Journal of Clinical Microbiology & Infectious Diseases*, vol. 39, pp. 1379–1389, 2020.

[24] A. Jaiswal, N. Gianchandani, D. Singh, V. Kumar, and M. Kaur, Classification of the COVID-19 infected patients using densenet201-based deep transfer learning, *Journal of Biomolecular Structure and Dynamics*, doi:10.1080/07391102.2020.1788642.

[25] D. Singh, V. Kumar, V. Yadav, and M. Kaur, Deep convolutional neural networks-based classification model for COVID-19 infected patients using chest X-ray images, *International Journal of Pattern Recognition and Artificial Intelligence*, doi:10.1142/S0218001421510046.

[26] V. Bhatnagar, R. C. Poonia, P. Nagar, S. Kumar, V. Singh, L. Raja, and P. Dass, Descriptive analysis of COVID-19 patients in the context of India, *Journal of Interdisciplinary Mathematics*, doi:10.1080/09720502.2020.1761635.

[27] D. Ivanov, Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2) case, *Transportation Research Part E: Logistics and Transportation Review*, vol. 136, p. 101 922, 2020.

[28] C. Hou, J. Chen, Y. Zhou, L. Hua, J. Yuan, S. He, Y. Guo, S. Zhang, Q. Jia, C. Zhao, et al., The effectiveness of quarantine of Wuhan city against the corona virus disease 2019 (COVID-19): A wellmixed SEIR model analysis, *Journal of Medical Virology*, vol. 92, pp. 841–848, 2020.

[29] K. Roosa, Y. Lee, R. Luo, A. Kirpich, R. Rothenberg, J. Hyman, P. Yan, and G. Chowell, Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020, *Infectious Disease Modelling*, doi:10.1016/j.idm.2020.02.002.

[30] 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, p. 100 222, 2020.

[31] Z. Xu, L. Shi, Y. Wang, J. Zhang, L. Huang, C. Zhang, S. Liu, P. Zhao, H. Liu, L. Zhu, et al., Pathological findings of COVID-19 associated with acute respiratory distress syndrome, *The Lancet Respiratory Medicine*, vol. 8, no. 4, pp. 420–422, 2020.

[32] A. J. Kucharski, T. W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, R. M. Eggo, F. Sun, M. Jit, J. D. Munday, et al., Early dynamics of transmission and control of COVID-19: A mathematical modelling study, *The Lancet Infectious Diseases*, vol. 20, no. 5, pp. 553–558, 2020.

[33] N. Yuvaraj, K. Srihari, S. Chandragandhi, R. A. Raja, G. Dhiman, and A. Kaur, Analysis of protein-ligand interactions of SARS-CoV-2 against selective drug using deep neural networks, *Big Data Mining and Analytics*, doi: 10.26599/BDMA.2020.9020007.

[34] G. Dhiman, The effects of coronavirus (COVID-19) on the psychological health of indian poultry farmers, *Coronaviruses*, doi:10.2174/2666796701999200617160755.

[35] A. Assad, M. A. Wani, and K. Deep, A comprehensive strategy to lower number of COVID-19 tests, *SSRN Electronic Journal*, doi: 10.2139/ssrn.3578240.

[36] K. Sarkar, S. Khajanchi, and J. J. Nieto, Modeling and forecasting the COVID-19 pandemic in India, *Chaos, Solitons & Fractals*, vol. 139, p. 110049, 2020.

[37] K. N. Nabi, Forecasting COVID-19 pandemic: A data-driven analysis, *Chaos, Solitons & Fractals*, vol. 139, p. 110046, 2020.

[38] K. Kanagarathinam and K. Sekar, Estimation of the reproduction number and early prediction of the COVID-19 outbreak in India using a statistical computing approach, *Epidemiology and Health*, vol. 42, p. e2020028, 2020.

[39] P. Arora, H. Kumar, and B. K. Panigrahi, Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India, *Chaos, Solitons & Fractals*, vol. 139, p. 110017, 2020.

[40] 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 & Fractals*, vol. 139, p. 110 058, 2020.

[41] R. Sujath, J. M. Chatterjee, and A. E. Hassanien, A machine learning forecasting model for COVID-19 pandemic in India, *Stochastic Environmental Research and Risk Assessment*, vol. 34, pp. 959–972, 2020.

[42] D. Rafiq, S. A. Suhail, and M. A. Bazaz, Evaluation and prediction of COVID-19 in India: A case study of worst hit states, *Chaos, Solitons & Fractals*, vol. 139, p. 110 014, 2020.

[43] B. K. Sahoo and B. K. Sapra, A data driven epidemic model to analyze the lockdown effect and predict the course of COVID-19 progress in India, *Chaos, Solitons & Fractals*, vol. 139, p. 110034, 2020.

[44] K. Goswami, S. Bharali, and J. Hazarika, Projections for COVID-19 pandemic in India and effect of temperature and humidity, *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 14, no. 5, pp. 801–805, 2020.

[45] R. Salgotra, M. Gandomi, and A. H. Gandomi, Time series analysis and forecast of the COVID-19 pandemic in India using genetic programming, *Chaos, Solitons & Fractals*, vol. 138, p. 109 945, 2020.

[46] 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, p. 138 762, 2020.

[47] S. Gupta, G. S. Raghuwanshi, and A. Chanda, Effect of weather on COVID-19 spread in the us: A prediction model for India in 2020, *Science of the Total Environment*, vol. 728, p. 138 860, 2020.

[48] M. Mandal, S. Jana, S. K. Nandi, A. Khatua, S. Adak, and T. Kar, A model-based study on the dynamics of COVID-19: Prediction and control, *Chaos, Solitons & Fractals*, vol. 136, p. 109 889, 2020.

[49] T. Chakraborty and I. Ghosh, Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis, *Chaos, Solitons & Fractals*, vol. 135, p. 109 850, 2020.

[50] S. Tiwari, S. Kumar, and K. Guleria, Outbreak trends of coronavirus disease-2019 in India: A prediction, *Disaster Medicine and Public Health Preparedness*, doi:10.1017/dmp.2020.115.

[51] H. Maheshwari, D. Yadav, U. Chandra, and D. S. Rai, Forecasting epidemic spread of COVID-19 in India using ARIMA model and effectiveness of lockdown, *Advances in Mathematics: Scientific Journal*, vol. 9, no. 6, pp. 3417–3430, 2020.

[52] A. Bhattacharjee, M. Kumar, and K. K. Patel, When COVID-19 will decline in India? Prediction by combination of recovery and case load rate, *Clinical Epidemiology and Global Health*, doi:10.1016/j.cegh.2020.06.004.

[53] P. K. Sree, A novel cellular automata classifier for COVID-19 trend prediction, *Journal of Health Sciences*, vol. 10, no. 1, pp. 1–5, 2019.

[54] COVID-19 open research dataset challenge (CORD-19), https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge, 2020.

Sandeep Kumar received the PhD degree in computer science & engineering from Jagannath University, Jaipur in 2015, the master of technology degree from RTU, Kota, India in 2011, and the bachelor of engineering degree from Engineering College, Kota, India in 2005. He is currently an assistant professor at CHRIST (Deemed to be University), Bangalore, India. He was an assistant professor at ACEIT, Jaipur from 2008 to 2011, and an assistant professor at Jagannath University, Jaipur from 2011 to 2017. He was the head of computer science at Jagannath University from 2013 to 2017. He edited special issue for many journals, including *IJGUC*, *IJIIDS*, *IJARGE*, *IJESD*, *JIM*, *JDMSC*, *JSMS*, and *JIOS*. He has published more than fifty articles in well-known SCI/SCOPUS indexed international journals and conferences, and attended

several national and international conferences and workshops. He has authored/edited four books in the area of computer science. His research interests include nature inspired algorithms, swarm intelligence, soft computing, and computational intelligence.

Rajani Kumari received the PhD degree in the field of soft computing from Jagannath University, Jaipur in 2015. Currently she is working at Department of Information Technology, JECRC University, India. She has published more than 20 research papers in refereed journals and international conferences. She works as a guest editor in *Int. Journal of Intelligent and Database System*. Her research interests include fuzzy logic, swarm intelligence, and evolutionary computing.

Ramesh Chandra Poonia received the PhD degree in computer science from Apaji Institute of Mathematics & Applied Computer Technology, Banasthali University, Banasthali, India in 2013. He is a postdoctoral fellow at Cyber-Physical Systems Laboratory (CPS Lab), Department of Information and Communications Technology (ICT) and Natural Sciences, Norwegian University of Science and Technology (NTNU), Alesund, Norway. He is currently an associate professor at Amity Institute of Information Technology, Amity University Rajasthan, Jaipur, India. He is the chief editor of *TARU Journal of Sustainable Technologies and Computing* (*JSTC*) and the associate editor of the *Journal of Sustainable Computing*: *Informatics and Systems*, Elsevier. His research interests include sustainable technologies, cyber-physical systems, internet of things, and network protocol evaluation.

Vijander Singh received the PhD degree from Banasthali University, Banasthali, India in 2017. He is working as an assistant professor at Manipal University, Jaipur, India. He has published 25 research papers in indexed journals and several book chapters for international publishers. He has authored 2 books and handled/handling journals of international repute as guest editor. He is an associate editor of *TARU Journal of Sustainable Technologies and Computing* (*JSTC*). He has organized several international conferences, Faculty Development Programs (FDPs), and Workshops as core team member of the organizing committee. His research area includes machine learning, deep learning, and precision agriculture and networking.

Linesh Raja received the PhD degree in computer science from Jaipur National University, India in 2015. Before that he received the master and bachelor degrees in computer application from Birla Institute of Technology, Mesra, Ranchi, India in 2009 and 2006, respectively. He is currently working as an assistant professor at Manipal University, Jaipur, India. He has published several research papers in the field of wireless communication, mobile networks security, and internet of things in various reputed national and international journals. He has chaired various sessions of international conferences. He has edited the handbook of *Research on Smart Farming Technologies for Sustainable Development*, IGI Global. At the same time he is also acting as a guest editor of various reputed journal publishing houses, such as Taylor & Francis, Inderscience and Bentham Science.

Vaibhav Bhatnagar received the PhD degree in information technology from Amity University Rajasthan, Jaipur in 2019. He is currently working as an assistant professor at Manipal University, Jaipur, India. He is the Gold Medalist Bachelor of Computer Application (BCA) and Silver Medalist Master of Computer Application (MCA). He has published more than 15 papers indexed by Scopus and Web of Science. His specialization is data science and internet of things.

Pankaj Agarwal received the MS degree in mechanical engineering specialization from Jagannath University, Jaipur, India in 2013 and the BEng degree in mechanical engineering from University of Rajasthan, Jaipur, India in 2007. He is an assistant professor at Amity University Rajasthan, India. He has published more than 20 research articles in national and international journals, as well as in conferences and 2 book chapters for international publishers. He has handled/handling journals of international repute as guest editor. He has organized several international conferences, FDPs, and Workshops as a core team member of the organizing committee. His research interests are optimization, composite materials, simulation and modelling, and soft computing.