IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 22 July 2022, accepted 3 September 2022, date of publication 6 September 2022, date of current version 15 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3204804

APPLIED RESEARCH

Machine Learning Approach for Forecast Analysis of Novel COVID-19 Scenarios in India

ANKIT KUMAR SRIVASTAVA[1], SAURABH MANI TRIPATHI[2], SACHIN KUMAR[3], (Senior Member, IEEE), RAJVIKRAM MADURAI ELAVARASAN[4], SIVASANKAR GANGATHARAN[4], DINESH KUMAR[1], AND LUCIAN MIHET-POPA[5], (Senior Member, IEEE)

[1]Department of Electrical Engineering, Institute of Engineering and Technology, Dr. Rammanohar Lohia Avadh University, Ayodhya, Uttar Pradesh 224001, India
[2]Power and Energy Research Centre, Department of Electrical Engineering, Kamla Nehru Institute of Technology, Sultanpur, Uttar Pradesh 228118, India
[3]Department of Electrical Engineering, Govind Ballabh Pant Institute of Engineering & Technology, Pauri, Garhwal, Uttarakhand 246194, India
[4]Department of Electrical and Electronics Engineering, Thiagarajar College of Engineering, Madurai, Tamil Nadu 625015, India
[5]Faculty of Electrical Engineering, Østfold University College, 1757 Halden, Norway

Corresponding authors: Ankit Kumar Srivastava (ankitsrivastava@rmlau.ac.in) and Lucian Mihet-Popa (lucian.mihet@hiof.no)

This work was supported by Østfold University College, Halden, Norway.

ABSTRACT The novel coronavirus (nCOV) is a new strain that needs to be hindered from spreading by taking effective preventive measures as swiftly as possible. Timely forecasting of COVID-19 cases can ultimately support in making significant decisions and planning for implementing preventive measures. In this study, three common machine learning (ML) approaches via linear regression (LR), sequential minimal optimization (SMO) regression, and M5P techniques have been discussed and implemented for forecasting novel coronavirus disease-2019 (COVID-19) pandemic scenarios. To demonstrate the forecast accuracy of the aforementioned ML approaches, a preliminary sample-study has been conducted on the first wave of the COVID-19 pandemic scenario for three different countries including the United States of America (USA), Italy, and Australia. Furthermore, the contributions of this study are extended by conducting an in-depth forecast study on COVID-19 pandemic scenarios for the first, second, and third waves in India. An accurate forecasting model has been proposed, which has been constructed on the basis of the results of the aforementioned forecasting models of COVID-19 pandemic scenarios. The findings of the research highlight that LR is a potential approach that outperforms all other forecasting models tested herein in the present COVID-19 pandemic scenario. Finally, the LR approach has been used to forecast the likely onset of the fourth wave of COVID-19 in India.

INDEX TERMS Death forecasting, linear regression (LR), M5P, machine learning (ML), novel coronavirus (nCOV), COVID-19 forecasting, SMO regression.

I. INTRODUCTION

According to real-time data, most countries around the world witnessed a rapid increase in the confirmed coronavirus disease-2019 (COVID-19) cases [1]. Human-to-human infection is evident from its rapid spreading ability, and global efforts have been concentrated on mitigating the spreading rate of the virus. The implementation of the mitigation measures affected the normal functioning of society through its consequences on travel, event cancellation, employment opportunities, food supply chain, academic, and hospital capacity. Altogether, the pandemic has imposed extraordinary pressure on the world economy, healthcare, and globalization. The healthcare systems in several countries such as Italy and Spain were overburdened by the pandemic leading to the increased loss of human lives [2]. According to several reports, COVID-19 originated in Wuhan, China [3] and it spread to almost all the countries and emerged as a pandemic within a short span [4]. COVID-19 was discovered to be closely related to severe acute respiratory syndrome (SARS) [5], however, its infection spreads faster than the infection of SARS. The severity of community transmission

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott.

of this disease can be found in WHO e-reports [6], which show that more than 469 million people across the world have been infected with COVID-19, with 6 million deaths reported as of March 21, 2022. In the absence of a vaccine or anti-drug for the novel COVID-19, governments were forced to use non-drug prevention strategies such as social distancing, use of masks, non-touch measures and sanitizing agents etc.

Further, [7] sheds light on the different technologies that have been put in place to help the healthcare systems, the government, and the general public in various ways in the fight against COVID-19. Also, [8] has explained a distinctive viewpoint to the research community and imparted greater information in relation to sustainability, while the solutions to post-COVID impacts were demonstrated for the guidance of policy and decision-makers.

A. MOTIVATION

The initial reporting of COVID-19 as a pandemic disease corresponds to the period when no major preventive measures, even in the health systems of most industrialized and developed countries were being implemented. Most countries prioritized economic performance over the prevention of this unpredictable community-transmitted disease until it became a life-threatening disease for their citizens. When the first suspect was detected in India, the new coronavirus had taken uncontrollable roots around the world.

As of 2020, India is the second-most populous country in the world, accounting for about 18% of the global population. In India, there is a wide gap between existing healthcare conditions when compared to the affluent countries. Therefore, adherence to the control measures was most likely to fail. In addition, human mobility, which is the most supportive element in COVID-19 disease transmission, was higher in India's metro cities than in any other populated city across the world due to higher population density. Drawing inspiration from China's experience in [9], the Indian government took several steps toward setting up dedicated COVID-19 isolation units in existing hospitals and quarantine centres as well as increasing the number of beds to provide quality healthcare for infected people.

Firstly, it is imperative to educate government and health experts about what to anticipate and what steps to take. Secondly, to encourage the general public to follow the preventive measures to moderate the spread of the disease, proper awareness is of the utmost [10], [11]. For most countries, it is crucial to develop an accurate model to predict the spread of COVID-19 [12], [13]. which can ultimately help in making important decisions and planning for implementing preventive measures. In statistical physics [14], investigations of pandemic systems have a long and fruitful history. There is a need for mathematical and statistical approaches for short-term forecasting of the transient rise of COVID-19 cases in order to augment the resources to deal with the pandemic. Such a timely forecast would help in better management of health care measures to restrain the pandemic.

B. OVERVIEW OF RELATED RESEARCH

Simple mathematical models that capture the fundamentals of epidemic spread can be used to fit data using a large number of parameters, and the resulting values can be used to generate accurate forecasts. In recent years, the scientific community has gathered significant justification for diverse and complex social network connection patterns [15], [16]. These are important in defining the behavior of equilibrium and non-equilibrium systems in general, as well as the spread of pandemics and the development of effective prevention measures. Digital epidemiology and the theory of epidemic processes on complex networks are the results of interdisciplinary studies at the intersection of statistical physics, network science and epidemiology, driven by the vast amounts of data documenting our health status and life style.

Various dynamic models were used to investigate and evaluate epidemiological parameters such as incubation period, transmissibility period and many others in prior pandemic outbreaks [17], [18]. Machine learning (ML)-based forecasting methods have proven to be effective in analyzing post-operative outcomes and making better decisions about future activities [19]. ML models have long been used in numerous domains, including detecting and prioritizing aversive aspects of a threat. Several studies used simple techniques to estimate the number of COVID-19 cases, assuming that government data is reliable and accurate. Auto regressive integrated moving average (ARIMA) and Holt's simple exponential method have been used in [20] for short-term forecasting of COVID-19 spread in India. For Italy, China and France, simple mean-area models and susceptible–infected–recovered–death models, the Gompertz model, the logistic model and the Bertalanffy model have been utilized [21], [22]. Researchers have applied the Gompertz model to predict the growth of tumors and many others, whereas, the logistic growth model has been used to model the outbreak of COVID-19 and predict its global spread. Similarly, the exponentially escalating model was used to forecast the final size and spread of COVID-19 in Italy, as well as the total number of confirmed COVID-19 cases in China, Italy, South Korea, Iran, and Thailand [23], [24]. Other nations, such as the United States, Iran, Slovenia, and Germany were expected to have COVID-19 instances between March 29 and April 12, 2020, according to the prediction in [25]. In addition, the number of new confirmed cases, recovery, and mortality numbers for Algeria, Australia, and Canada have been assessed [26].

Furthermore, in [27], the authors used ML and evolutionary computing methods with regression for the COVID-19 virus spread prediction and control model. In addition, [28] systematically reviewed forecasting models to identify key factors in the spread of the COVID-19 pandemic. In a study described in [29], artificial neural networks (ANNs) were used to make a real-time predictor model for COVID-19 spread. Ahmadini *et al.* [30] proposed the Kalman filter model to predict COVID-19 infections for the four most affected countries, namely the United States, India, Brazil and Russia. Zeroual *et al.* [31] applied five deep learning methods to the global forecast of daily new confirmed and recovered cases of COVID-19 based on a small volume of data. Yasminah *et al.* [32] and Dairi *et al.* [33] developed new methods to predict how COVID-19 will spread.

Five ML algorithms were used to forecast the amount of COVID-19 confirmed cases, fatalities, and cured cases for Indian states [34]. For India, [35] described the Holt-Winters technique for predicting incident cases, death cases, and active cases from September 28, 2020 to November 15, 2020, and the results were found to be comparable with the basic susceptible-infectious-recovered model. Even today, any model including real-time short-term forecasts cannot accurately confirm the natural behavior of any pandemic, natural phenomenon or tragedy. In [36], the performance of different ML approaches has been evaluated for predicting the COVID-19 outbreak for the world based on publicly accessible data on daily deaths, recovered and confirmed cases from January 22, 2020 to August 18, 2021. In the present study, three common and well-known machine learning (ML) approaches via linear regression (LR), sequential minimal optimization (SMO) regression, and M5P techniques have been discussed and implemented for forecasting novel coronavirus disease-2019 (COVID-19) pandemic scenarios. Unlike [36], this forecast study has been conducted not only for the first wave of COVID-19 scenarios in three countries (United States, Italy, and Australia) but an extensive forecast analysis has also been conducted for the first, second and third waves of COVID-19 scenarios in India.

The purpose of this research is to identify an effective model that yields maximum accuracy in forecasting COVID-19 scenarios. The findings of this study would direct future research towards more complex and accurate forecasting models for various unprecedented situations.

C. SCOPE OF THE WORK

Table 1 shows the scope of the work described in this study and highlights the works described in recently published study. In the present study, an extensive forecast analysis has been carried out for three countries (the USA, Italy, and Australia) for the first wave of COVID-19 and is extended to first, second and third waves of COVID-19 scenarios in India which is not performed in recently published studies.

Furthermore, since forecasting the upcoming wave will be highly supportive in making important decisions and planning for implementing preventive measures, the authors have also attempted to forecast the likely onset of the fourth wave of COVID-19 in India.

D. MAIN CONTRIBUTIONS

The present work has the following contributions.

- Three of the most commonly used ML approaches viz. linear regression (LR), M5P, and sequential minimal optimization (SMO) regression have been implemented for a sample study in three different countries, i.e., USA, Italy, and Australia, to predict the first wave of the COVID-19 pandemic scenario (daily new cases and deaths). In this study, a day ahead forecasting model has been considered to predict the daily confirmed COVID-19 cases and daily deaths.
- Based on the sample study conducted on the first wave of the COVID-19 pandemic scenario in three countries, the forecast accuracy of the above-mentioned ML approaches has been exemplified using several error measures such as MAPE, SMAPE, MAE, RMSE and MSE.
- The above-mentioned ML approaches have then been employed for an extensive forecast analysis of the COVID-19 scenarios for the first, second, and third waves in India.
- Based on the key findings of the forecast analyses for the first, second, and third waves of the COVID-19 pandemic scenarios in India, an accurate forecasting model has been proposed.
- Lastly, the proposed forecasting model was used to predict when the fourth wave of COVID-19 is likely to hit India.

E. ORGANIZATION OF THE STUDY

The successive contents of this research have been organized as follows—-The LR, SMO regression and M5P approaches of forecasting have been briefly explained in Section 2. Section 3 presents the results of the preliminary sample study on the first wave of the COVID-19 pandemic scenario performed for three different countries namely, the USA, Italy, and Australia. Section 4 presents the results of an in-depth forecast study of COVID-19 pandemic scenarios for the first, second, and third waves in India. Based on the proposed forecasting model, the likely onset of the fourth wave of COVID-19 in India is also hinted in Section 4. Finally, the conclusion has been drawn out in Section 5.

II. TECHNIQUES USED TO FORECAST NOVEL COVID-19

With the ever-increasing amount of data availability, ML has become an emerging technology for comprehensive data analysis over the past two decades and has become more widespread as an essential component of technological advancement. In this work, three common ML-based forecasting approaches viz. LR, SMO regression and M5P techniques have been employed to predict the COVID-19 scenarios. To predict the COVID-19 scenario in India, the authors aim to employ machine learning algorithms, which are an emerging tool nowadays and are increasingly being used in forecasting studies. All three methods are well-known and commonly used by the researchers in their recently reported literature; therefore, the authors used LR, SMO, and M5P in this study. In addition, several error measures have also been used for the assessment of the forecast accuracy of these techniques.

TABLE 1. Scope of work reported in this study *vs.* work reported in recently published study–*At-a-glance*.

Year	Author [Ref]	Research focus	Region/ Country	Forecasting		Methodology	Error Measures			
				Daily Cases	Death		MAPE	MAE	RMSE	MSE
2020	Rustam et al [26]	First Wave	Algeria; Australia; Canada	✓	✓	ES; LASSO; LR; SVM;	X	✓	✓	✓
		Second Wave	X							
		Third Wave	X							
2020	Sharma et al [37]	First Wave	Brazil; India; USA	✓	X	EVDHM-ARIMA	X	X	✓	X
		Second Wave	X							
		Third Wave	X							
2020	Singh et al [35]	First Wave	India	✓	✓	Holt-Winters models	✓	X	X	X
		Second Wave	X							
		Third Wave	X							
2021	Dash et al [38]	First Wave	Brazil; France; India;Russia; UK; USA;	✓	X	Autoregressive integrated moving average (ARIMA)	✓	✓	✓	✓
		Second Wave	X							
		Third Wave	X							
2021	Friji et al [39]	First Wave	Brazil; Italy; Russia; USA	✓	✓	Levenberg-Marquardt (LM); Broyden-Fletcher-Goldfarb-Shanno (BFGS)	✓	✓	✓	✓
		Second Wave	USA							
		Third Wave	X							
2021	Gomes et al [40]	First Wave	Brazil	✓	✓	Interval type- 2 fuzzy Kalman filter	✓	✓	✓	✓
		Second Wave	X							
		Third Wave	X							
2021	Gupta et al [34]	First Wave	India	✓	✓	Decision tree; MLR ; Neural network;Random forest (RF); Support vector machine (SVM)	X	X	X	X
		Second Wave	X							
		Third Wave	X							
2021	Gecili et al [41]	First Wave	USA; Italy	✓	✓	ARIMA; Holt ; Splines; TBATS	✓	✓	X	X
		Second Wave	X							
		Third Wave	X							
2021	Iqbal et al [42]	First Wave	Pakistan	✓	X	Long short-term memory (LSTM)	✓	X	X	X
		Second Wave	X							
		Third Wave	X							
2021	Kumari et al [43]	First Wave	India	✓	✓	Auto regression; Multiple linear regression (MLR);	X	X	✓	X
		Second Wave	X							
		Third Wave	X							
2021	Lucic, et al [44]	First Wave	USA	✓	✓	Autoregressive integrated moving average (ARIMA)	X	✓	✓	X
		Second Wave	X							
		Third Wave	X							
2022	Zhou et al [45]	First Wave	USA	✓	✓	Deep learning models; Interpretable temporal attention network (ITANet)	✓	✓	✓	X
		Second Wave	X							
		Third Wave	X							
The proposed work		**First Wave**	**Australia; India; Italy; USA**	✓	✓	**Linear regression; M5P; SMO regression**	✓	✓	✓	✓
		Second Wave	**India**							
		Third Wave	**India**							

A. LINEAR REGRESSION

The LR technique involves recognizing independent or incidental demand variables that influence and continue to influence the forecast/dependent variables, as well as expressing the forecaster's belief in the inter dependencies of all such variables as an equation or series of equations. The independent variables of any system can be internal or external. If there is only one dependent variable, the regression model will be linear; if there are multiple dependent variables, the model will be multivariate. The regression technique provides the ability to forecast not only the demand for the dependent variable, but also the forces and events that cause the dependent variable to change. The steps involved to produce forecasts for a given problem using a common LR technique have been comprehensively explained in [26] and [46].

There are two factors (x; y) involved in linear regression analysis.The equation below shows the relation between y and x which is called regression.

$$y = \beta_0 + \beta_1 x + \epsilon \tag{1}$$

or equivalently

$$E(y) = \beta_0 + \beta_1 x \tag{2}$$

where ϵ is the error term of the linear regression. Here the error term takes into account the variability between both x and y, β_0 represents the y-intercept and β_1 represents the slope.

B. SMO REGRESSION

SMO regression is a simplified technique for rapidly solving the support vector machine (SVM) quadratic programming (QP) issue without any additional matrix storage or numerical QP optimization. To achieve convergence, SMO regression decomposes the overall QP issue into QP sub-problems.

TABLE 2. Comparison of forecast accuracy of daily cases for the first wave of COVID-19 in the USA using different error measures.

S. No.	Error Measures	Methods	Duration				Average
			01-07 April 2020	8-14 April 2020	14-21 April 2020	22-28 April 2020	
1	MAPE	LR	1.44	0.67	0.79	0.32	0.81
		M5P	2.33	2.03	0.74	0.48	1.4
		SMO Regression	4.19	1.71	1	0.72	1.91
2	SMAPE	LR	1.44	0.68	0.79	0.32	0.81
		M5P	2.30	2.03	0.73	0.48	1.39
		SMO Regression	4.19	1.71	1	0.72	1.91
3	MAE	LR	3720	3080	5180	2785	3690.95
		M5P	6020	9130	4900	4020	6016.18
		SMO Regression	10800	7970	6660	6640	8007.63
4	RMSE	LR	4660	4190	6320	4320	4872.83
		M5P	7260	9380	5460	5190	6822.33
		SMO Regression	11900	9360	8240	10700	10056.35
5	MSE	LR	21800000	17500000	40000000	18600000	24477000
		M5P	52800000	88000000	29800000	26900000	49358750
		SMO Regression	142000000	87600000	67800000	115000000	103049750

TABLE 3. Daily MAPE of LR, SMO regression and M5P techniques to compare forecast accuracy of daily cases for the first wave of COVID-19 in USA.

Duration	LR	SMO Reg.	M5P	Duration	LR	SMO Reg.	M5P	Duration	LR	SMO Reg.	M5P	Duration	LR	SMO Reg.	M5P
1-Apr-20	0.25	0.76	1.78	8-Apr-20	0.13	0.24	3.27	15-Apr-20	0.77	0.1	1.28	22-Apr-20	0.01	0.02	0.96
2-Apr-20	0.38	3.14	2.27	9-Apr-20	0.56	1.11	1.93	16-Apr-20	0.1	2.16	0.05	23-Apr-20	0.08	0.08	0.99
3-Apr-20	3.28	5.34	2.17	10-Apr-20	0.85	3.67	2.17	17-Apr-20	1.13	0.32	1.06	24-Apr-20	0.9	0.22	0.21
4-Apr-20	1.54	6.92	2.26	11-Apr-20	2.07	1.62	1.07	18-Apr-20	1.82	0.83	0.36	25-Apr-20	0.98	0.98	0.01
5-Apr-20	1.74	4.55	1.45	12-Apr-20	0.47	2.87	2.07	19-Apr-20	1.02	1.88	0.72	26-Apr-20	0.06	0.06	0.83
6-Apr-20	0.46	3.86	1.67	13-Apr-20	0.53	1.57	1.71	20-Apr-20	0.37	1.32	0.88	27-Apr-20	0.09	2.71	0.02
7-Apr-20	2.39	4.75	4.73	14-Apr-20	0.11	0.91	1.99	21-Apr-20	0.29	0.37	0.82	28-Apr-20	0.16	1	0.34
Average	**1.44**	**4.19**	**2.33**	**Average**	**0.67**	**1.71**	**2.03**	**Average**	**0.79**	**1**	**0.74**	**Average**	**0.32**	**0.72**	**0.48**

Unlike previous techniques, SMO regression chooses to tackle the lowest optimization problem possible at each step. Since Lagrange multipliers must satisfy a linear equality requirement, the least feasible optimization issue for the typical SVM QP problem comprises two Lagrange multipliers. SMO regression picks two Lagrange multipliers to jointly optimize at each step, determines their ideal values, and then updates the SVM to reflect the new optimal values. Furthermore, the advantage of SMO regression is that it allows solving two Lagrange multipliers analytically. As a result, numerical QP optimization is completely avoided. Even though the method solves more optimization subproblems, each sub-issue is handled swiftly that the overall QP problem is addressed quickly. Since SMO regression does not need any additional matrix storage, even very large SVM training problems can fit into the random access memory (RAM) of a standard computer or workstation. A detailed methodology for implementing SMO regression to produce forecasts for a given problem has been described in [47]. The implementation of SMO regression has the following steps:

Step 1: Break large QP problems into a series of smallest possible QP problem. Find the most promising pair (μ_1 and μ_2).

Step 2: Solve small QP problems in promptly when compared to the QP optimization process. It is cardinal to consider that it requires memory in proportion to the smallest possible samples taken under Step 1. This enables it to handle numerous training sets, i.e., very large QP problems. Optimize μ_1 and μ_2 keeping other μ's fixed.

C. M5P TECHNIQUE

The M5P technique is a numeric prediction tool based on classification and regression analysis and is a modified version of the original M5 tree algorithm, which enables it to deal with enumerated attributes and attribute missing values. M5P is more sensitive to data segmentation and gives better results with longer data set as input. The following steps are involved in implementing the M5P technique to produce forecasts for a given problem as detailed in [48] and [49]:

Step 1: Take the input data (enumerated attributes), then convert it into binary variables and apply the algorithm to maximize standard deviation reduction (SDR).

$$SDR = \delta(C_s) - \sum\nolimits_k \frac{|Cs_k|}{|Cs|}.\delta(Cs_k) \tag{3}$$

where C_s is the set of cases, Cs_k is the k^{th} subset of cases that result from the tree splitting process, $\delta(Cs)$ is the standard deviation of C_s, and $\delta(Cs_k)$ is the standard deviation of k^{th} subset as a measure of error

Step 2: Use these binary variables to construct a tree (as the tree grows over fitting increases).

Step 3: Perform tree pruning process (which reduces the problem of over-fitting) and compensation for discontinuities.

Step 4: Carry out tree smoothing process to compensate for sharp discontinuities that occur between linear adjacent models at end nodes (leaf) of pruned tree.

Step 5: Produce tree model as the output.

D. ERROR MEASURES USED FOR THE ASSESSMENT OF FORECAST ACCURACY

In this study, several error measures such as mean absolute error (MAE), mean square error (MSE), root-mean-square error (RMSE), mean absolute percentage error (MAPE), and symmetric mean absolute percentage error (SMAPE) have also been used for the assessment of forecast accuracy of novel COVID-19 scenarios. These error measures [50], [51], [52] are mathematically defined as follows:

$$MAE = \frac{1}{K}\sum_{p=1}^{K}|(A_p - F_p)| \tag{4}$$

$$MSE = \frac{1}{K}\sum_{p=1}^{K}(A_p - F_p)^2 \tag{5}$$

$$RMSE = \sqrt{\frac{1}{K}\sum_{p=1}^{K}(A_p - F_p)^2} \tag{6}$$

$$MAPE = \frac{100}{K}\sum_{p=1}^{K}\frac{|(A_p - F_p)|}{A_p} \tag{7}$$

$$SMAPE = \frac{100}{K}\sum_{p=1}^{K}\frac{|(A_p - F_p)|}{(A_p + F_p)/2} \tag{8}$$

where A_p and F_p are the actual and forecasted values of the novel COVID-19 scenario; and K is the length of the forecast horizon.

III. FORECAST ANALYSIS OF COVID-19 SCENARIOS FOR USA, ITALY AND AUSTRALIA: A PRELIMINARY SAMPLE-STUDY

A preliminary sample study has been presented in this section to illustrate the forecast accuracy of LR, SMO regression and M5P techniques for the first wave of COVID-19 scenarios in three different countries (USA, Italy and Australia). The authors selected three countries, taking into account the following: (1) Countries should be from different continents, and (2) they should be the most or least affected. The United States and Italy were the most affected countries during the first wave of COVID-19, and Australia was the least affected. These are the only specific criteria for selecting these countries in this study. The forecast of daily cases for the first wave of COVID-19 scenario in the above-mentioned countries has been analyzed for the sample-period 01-28 April 2020. To accomplish the forecast analysis, the data has been adopted from the official website of the World Health Organization (WHO) [6]. After creating the training and testing data set, the forecast accuracy has been computed. It is worth noting that the forecasting model applied in this study does not consider the parameters such as the number of lockdowns, the number of people vaccinated, social distancing and self-isolation behavior.

Table 2 and Table 3; Table 4 and Table 5; and Table 6 and Table 7 depict the results of the preliminary sample study conducted to forecast daily cases for the first wave of COVID-19 in the USA, Italy, and Australia, respectively. These results have been obtained to exemplify the forecast accuracy of LR, SMO regression, and M5P techniques and have been grouped on a weekly basis for ease of understanding.

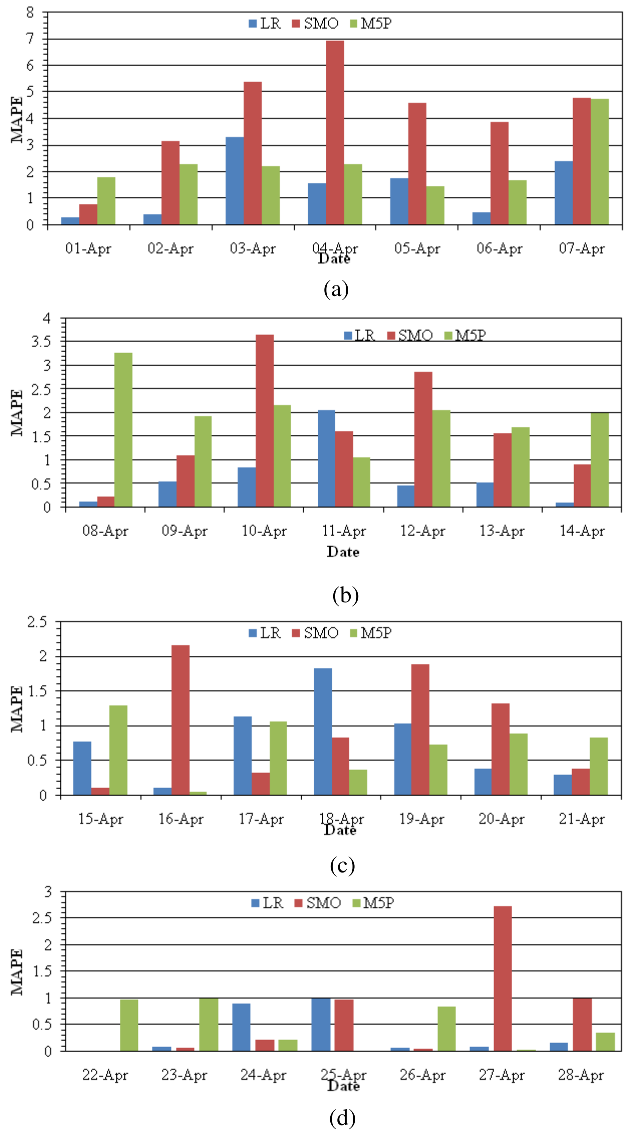

FIGURE 1. Daily MAPE of LR, SMO regression and M5P techniques to compare forecast accuracy of daily cases for the first wave of COVID-19 in the USA for the duration—(a) 01-07 April, 2020 (b) 08-14 April, 2020 (c) 15-21 April, 2020 (d) 22-28 April, 2020.

Table 2 compares the LR technique with the SMO regression and M5P techniques by evaluating error measures such as MAPE, SMAPE, RMSE, MAE and MSE for the USA. It can be observed that the average MAPE of LR is 0.81 which shows superior results relative to M5P and SMO regression. Moreover, the average SMAPE, MAE, RMSE, and MSE of LR have been evaluated as 0.81, 3690.95, 4872.83 and 24477000, respectively, depicting superior results when compared to M5P and SMO regression.

TABLE 4. Comparison of forecast accuracy of daily cases for the first wave of COVID-19 in Italy using different error measures.

S. No.	Error Measures	Methods	Duration				Average
			01-07 April 2020	8-14 April 2020	14-21 April 2020	22-28 April 2020	
1	MAPE	LR	0.65	0.43	0.24	0.24	0.39
		M5P	1.40	0.63	0.43	0.26	0.68
		SMO Regression	0.88	0.65	0.47	0.35	0.59
2	SMAPE	LR	0.65	0.43	0.24	0.24	0.39
		M5P	1.39	0.62	0.43	0.26	0.68
		SMO Regression	0.87	0.65	0.47	0.35	0.58
3	MAE	LR	748.71	627.57	566.64	463.00	601.48
		M5P	1650.00	902.71	737.57	502.89	948.87
		SMO Regression	1020.00	963.43	800.43	657.16	859.85
4	RMSE	LR	896.38	722.52	406.43	514.60	634.98
		M5P	1700	1100	908.43	580.26	1074.47
		SMO Regression	1190	1140	871.94	823.11	1006
5	MSE	LR	803000	522000	321000	265000.00	477855.00
		M5P	2910000	1220171	825000	337000.00	1321980.00
		SMO Regression	1410000	1293890	760000	678000.00	1035473.00

TABLE 5. Daily MAPE of LR, SMO regression and M5P techniques to compare forecast accuracy of daily cases for the first wave of COVID-19 in Italy.

Duration	LR	SMO Reg.	M5P	Duration	LR	SMO Reg.	M5P	Duration	LR	SMO Reg.	M5P	Duration	LR	SMO Reg.	M5P
1-Apr-20	0.55	1.26	2.36	8-Apr-20	0.23	0.21	1.63	15-Apr-20	0.11	0.53	0.79	22-Apr-20	0.24	0.56	0.32
2-Apr-20	1.61	0.64	1.29	9-Apr-20	0.78	1.35	0.74	16-Apr-20	0.14	0.27	0.87	23-Apr-20	0.43	0.32	0.07
3-Apr-20	0.76	1.73	1.21	10-Apr-20	0.51	0.43	0.36	17-Apr-20	0.75	0.76	0.07	24-Apr-20	0.22	0.78	0.28
4-Apr-20	0.56	1.3	1.28	11-Apr-20	0.08	0.67	0.45	18-Apr-20	0.01	0.7	0.17	25-Apr-20	0.26	0.13	0.04
5-Apr-20	0.68	0.3	0.92	12-Apr-20	0.68	0.41	0.12	19-Apr-20	0.15	0.5	0.11	26-Apr-20	0.23	0.52	0.32
6-Apr-20	0.04	0.8	1.21	13-Apr-20	0.19	1.19	0.26	20-Apr-20	0.12	0.37	0.31	27-Apr-20	0	0.06	0.27
7-Apr-20	0.34	0.11	1.51	14-Apr-20	0.52	0.31	0.81	21-Apr-20	0.37	0.15	0.7	28-Apr-20	0.3	0.05	0.5
Average	**0.65**	**0.88**	**1.4**	**Average**	**0.43**	**0.65**	**0.63**	**Average**	**0.24**	**0.47**	**0.43**	**Average**	**0.24**	**0.35**	**0.26**

TABLE 6. Comparison of forecast accuracy of daily cases for the first wave of COVID-19 in Australia using different error measures.

S. No.	Error Measures	Methods	Duration				Average
			01-07 April 2020	8-14 April 2020	14-21 April 2020	22-28 April 2020	
1	MAPE	LR	1.48	0.71	0.46	0.18	0.71
		M5P	2.80	1.61	0.65	0.24	1.33
		SMO Regression	2.78	0.96	0.54	0.56	1.21
2	SMAPE	LR	1.47	0.70	0.46	0.18	0.70
		M5P	2.76	1.59	0.65	0.24	1.31
		SMO Regression	2.74	0.96	0.54	0.56	1.20
3	MAE	LR	79.86	43.71	33.68	12.14	42.35
		M5P	149.43	98.86	42.43	16.00	76.68
		SMO Regression	149.71	59.43	35.00	37.57	70.43
4	RMSE	LR	99.38	49.84	33.68	14.40	49.32
		M5P	153.63	106.67	50.45	17.00	81.94
		SMO Regression	161.56	71.52	41.89	42.17	79.29
5	MSE	LR	9877.00	2483.70	1134.10	207.29	3425.52
		M5P	23601.00	11377.00	2544.70	289.14	9452.96
		SMO Regression	26103.00	5114.60	1754.70	1778.70	8687.75

TABLE 7. Daily MAPE of LR, SMO regression and M5P techniques to compare forecast accuracy of daily cases for the first wave of COVID-19 in Australia.

Duration	LR	SMO Reg.	M5P	Duration	LR	SMO Reg.	M5P	Duration	LR	SMO Reg.	M5P	Duration	LR	SMO Reg.	M5P
1-Apr-20	2.11	2	2.38	8-Apr-20	0.23	0.7	2.68	15-Apr-20	0.16	0.83	0.57	22-Apr-20	0.27	0.33	0.26
2-Apr-20	1.98	3.48	4.59	9-Apr-20	1.1	1.66	2.55	16-Apr-20	0.32	0.2	1.41	23-Apr-20	0.23	0.58	0.42
3-Apr-20	0.15	3.54	2.65	10-Apr-20	0.66	0.12	1.14	17-Apr-20	0.72	0.14	1.1	24-Apr-20	0.02	0.14	0.28
4-Apr-20	0.27	2.39	2.27	11-Apr-20	0.99	2	1.2	18-Apr-20	0.89	0.98	0.1	25-Apr-20	0.12	0.57	0.14
5-Apr-20	0.64	1.41	2.56	12-Apr-20	1.16	1.25	1.52	19-Apr-20	0.46	0.97	0.25	26-Apr-20	0.22	1.13	0.17
6-Apr-20	2.06	2.04	2.88	13-Apr-20	0.71	0.29	1.48	20-Apr-20	0.41	0.41	0.55	27-Apr-20	0.06	0.53	0.23
7-Apr-20	3.16	4.62	2.28	14-Apr-20	0.1	0.73	0.69	21-Apr-20	0.25	0.22	0.59	28-Apr-20	0.37	0.64	0.17
Average	**1.48**	**2.78**	**2.8**	**Average**	**0.71**	**0.96**	**1.61**	**Average**	**0.46**	**0.54**	**0.65**	**Average**	**0.18**	**0.56**	**0.24**

The MAPE is one of the most commonly used key performance indicators to measure forecast accuracy (i.e., the lower the MAPE, the higher is the forecast accuracy). However, it is interesting to note that the values of MAPE can exceed 100%, which would mean that the errors are ''much higher'' than the actual values [53]. On the other hand, setting arbitrary forecast performance targets without reference to the forecast data (e.g., MAPE<10% is excellent, MAPE<20% is good, etc.) is irrational [54]. Table 3 summarizes daily MAPE of LR, SMO regression and M5P techniques to compare the forecast accuracy of daily cases for the first wave of COVID-19 in the USA and the same has been illustrated graphically in Figure 1.

It can be observed that the average of daily MAPE values corresponding to LR, M5P and SMO regression techniques have been evaluated as 1.44, 2.33 and 4.19, respectively, for

TABLE 8. Daily MAPE of LR, SMO regression and M5P techniques to compare the accuracy of daily death forecast for the first wave of COVID-19 in the USA, Italy and Australia.

Duration	USA			Italy			Australia		
	LR	SMO Reg.	M5P	LR	SMO Reg.	M5P	LR	SMO Reg.	M5P
22-Apr-20	0.47	0.3	4.03	0.31	0.4	0.85	2.7	2.7	12.16
23-Apr-20	1.55	1.61	5.49	0.01	0.18	0.7	1.35	1.35	9.46
24-Apr-20	0.62	1.16	5.99	0.14	0.45	0.35	1.32	1.32	10.53
25-Apr-20	0.63	0.92	4.46	0	0.11	0.34	2.53	3.8	12.66
26-Apr-20	0.29	1.42	3.84	0	0.16	0.08	1.23	2.47	12.35
27-Apr-20	0.73	1.04	1.37	0.52	0.48	0.33	3.61	3.61	12.05
28-Apr-20	1.63	1.71	0.93	0.15	0.4	0.36	1.19	2.38	9.52
Average	**0.84**	**1.17**	**3.73**	**0.16**	**0.31**	**0.43**	**1.99**	**2.52**	**11.25**

TABLE 9. Comparison of the present study with other methods reported in recently published papers in terms of MAPE.

Sr. No.	Country	Reference	Methodology	Duration		
				02-08 April 2020	09-15 April 2020	16-22 April 2020
1	USA	[41]	ARIMA	10	8.7	7.6
			Holt	11.3	9.3	8.4
			Splines	13	8.1	7.1
			TBATS	9.6	10.5	8.6
		Present Study	SMO	4.11	1.69	0.99
			M5P	2.55	1.75	0.69
			LR	1.42	0.77	0.68
2	Italy	[41]	ARIMA	4.9	4.3	3.7
			Holt	12	10.7	8
			Splines	7	6	5.4
			TBATS	7.8	7.1	6
		Present Study	SMO	0.73	0.7	0.47
			M5P	1.29	0.51	0.37
			LR	0.6	0.41	0.25
3	India	[35]	Hot winter	**3 July-12 Sep 2020**		
				11.23		
		Present Study	SMO	6.83		
			M5P	6.25		
			LR	5.63		

TABLE 10. Comparison of forecast accuracy of daily cases for the first wave of COVID-19 in India using different error measures.

S. No.	Error Measures	Methods	Duration				Average
			1-7 July 2020	8-14 July 2020	14-21 July 2020	22-28 July 2020	
1	MAPE	LR	4.12	2.56	5.12	5.19	4.25
		M5P	4.20	2.56	5.12	5.18	4.27
		SMO Regression	4.52	2.98	6.18	6.06	4.94
2	SMAPE	LR	4.10	2.55	5.19	5.26	4.28
		M5P	4.20	2.55	5.19	5.21	4.29
		SMO Regression	4.53	3.03	6.37	6.25	5.05
3	MAE	LR	930.06	699.73	1880	2450	1490
		M5P	952.77	699.73	1880	2430	1492.25
		SMO Regression	1020	821.17	2280	2870	1746.77
4	RMSE	LR	1090	844.81	2180	3220	1832.85
		M5P	1170	844.81	2180	3320	1880.03
		SMO Regression	1130	973.19	2480	3640	2056.57
5	MSE	LR	1200000	714000	4740000	10300000	4247828
		M5P	1380000	714000	4740000	11000000	4470428
		SMO Regression	1280000	947000	6160000	13300000	5411748

TABLE 11. Daily MAPE of LR, SMO regression and M5P techniques to compare forecast accuracy of daily cases for the first wave of COVID-19 in India.

Duration	LR	SMO Reg.	M5P	Duration	LR	SMO Reg.	M5P	Duration	LR	SMO Reg.	M5P	Duration	LR	SMO Reg.	M5P
1-Jul-20	1.99	2.55	0.48	8-Jul-20	3.49	5.82	3.49	15-Jul-20	5.12	6.83	5.12	22-Jul-20	11.37	13.22	11.37
2-Jul-20	6.1	6.74	8.16	9-Jul-20	1.89	0.73	1.89	16-Jul-20	7.52	9.23	7.52	23-Jul-20	7.74	10.91	8.37
3-Jul-20	2.42	3.23	2.42	10-Jul-20	3.52	5.3	3.52	17-Jul-20	0.42	1.48	0.42	24-Jul-20	1.41	5.23	1.96
4-Jul-20	4.34	5.45	4.34	11-Jul-20	1.68	0.35	1.68	18-Jul-20	5.51	7.37	5.51	25-Jul-20	2.39	0.01	1.54
5-Jul-20	0.76	1.67	0.76	12-Jul-20	1.49	3.11	1.49	19-Jul-20	7.14	9.06	7.14	26-Jul-20	0.58	3	1.85
6-Jul-20	8.84	7.92	8.84	13-Jul-20	5.72	3.67	5.72	20-Jul-20	8.79	5.38	8.79	27-Jul-20	12.63	9.54	11.12
7-Jul-20	4.4	4.11	4.4	14-Jul-20	0.1	1.9	0.1	21-Jul-20	1.37	3.9	1.37	28-Jul-20	0.15	0.51	0.15
Average	**4.12**	**4.52**	**4.2**	**Average**	**2.56**	**2.98**	**2.56**	**Average**	**5.12**	**6.18**	**5.12**	**Average**	**5.18**	**6.06**	**5.19**

the duration 01–07 April 2020; 0.67, 2.03 and 1.71, respectively, for the duration 08–14 April 2020; 0.79, 0.74 and 1.0, respectively, for the duration 15–21 April 2020; and 0.32, 0.48, 0.72, respectively, for 22–28 April 2020. The daily MAPE, as summarized in Table 3 and illustrated in Figure 1, clearly indicates that the LR technique for forecasting daily cases for the first wave of COVID-19 in the USA outperforms the M5P and SMO regression.

On the other hand, Table 4 compares the LR technique with the SMO regression and M5P techniques by evaluating

TABLE 12. Comparison of forecast accuracy of daily cases for the second wave of COVID-19 in India using different error measures.

S. No.	Error Measures	Methods	Duration				Average
			01-07 May 2021	8-14 May 2021	14-21 May 2021	22-28 May 2021	
1	MAPE	LR	6.72	6.61	4.40	6.79	6.13
		M5P	9.85	11.68	4.78	7.39	8.42
		SMO Regression	7.92	7.30	5.11	8.24	7.14
2	SMAPE	LR	6.70	6.65	4.40	6.79	6.14
		M5P	9.85	11.68	4.70	7.39	8.40
		SMO Regression	7.91	7.30	5.11	8.24	7.14
3	MAE	LR	26200	23300	12000	13600	18775
		M5P	37400	41300	12900	15000	26650
		SMO Regression	30300	25600	13900	16300	21525
4	RMSE	LR	29300	29100	15400	16300	22525
		M5P	44100	48200	16400	19400	32025
		SMO Regression	33600	29100	16600	19800	24775
5	MSE	LR	857000000	846000000	238000000	266000000	551750000
		M5P	1940000000	2320000000	269000000	375000000	1226000000
		SMO Regression	1130000000	848000000	275000000	393000000	661500000

TABLE 13. Daily MAPE of LR, SMO regression and M5P techniques to compare forecast accuracy of daily cases for the second wave of COVID-19 in India.

Duration	LR	SMO Reg.	M5P	Duration	LR	SMO Reg.	M5P	Duration	LR	SMO Reg.	M5P	Duration	LR	SMO Reg.	M5P
1-May-21	10.4	9.18	11.91	8-May-21	0.01	1.86	9.94	15-May-21	0.87	1.66	0.32	22-May-21	1.16	1.56	3.17
2-May-21	9.39	13.12	22.9	9-May-21	13.86	13.73	17.88	16-May-21	5.02	4.06	4.46	23-May-21	3.26	1.1	4
3-May-21	1.94	12.11	13.26	10-May-21	7.5	10.29	25.99	17-May-21	0.51	2.12	1.03	24-May-21	5.89	10.92	1.82
4-May-21	10.75	4.49	7.42	11-May-21	12.34	8.88	12.08	18-May-21	9.67	9.46	10.2	25-May-21	14.88	18.16	20.34
5-May-21	6.85	1.41	5.03	12-May-21	7.36	5.64	6.74	19-May-21	9.47	10.91	10.08	26-May-21	7.72	9.1	7.72
6-May-21	2.07	5.22	1.64	13-May-21	3.44	7.91	6.76	20-May-21	1.59	3.83	2.37	27-May-21	11.04	5.87	10.29
7-May-21	5.67	9.91	6.77	14-May-21	1.79	2.8	2.39	21-May-21	3.69	3.72	5.01	28-May-21	3.59	10.97	4.36
Average	**6.72**	**7.92**	**9.85**	**Average**	**6.61**	**7.3**	**11.68**	**Average**	**4.4**	**5.11**	**4.78**	**Average**	**6.79**	**8.24**	**7.39**

TABLE 14. Comparison of forecast accuracy of daily cases for the third wave of COVID-19 in India using different error measures.

S. No.	Error Measures	Methods	Duration				Average
			15-21 Dec 2021	22-28 Dec 2021	01-07 Jan 2022	08-14 Jan 2022	
1	MAPE	LR	5.06	10.63	12.58	12.71	10.25
		M5P	6.51	11.28	13.56	15.11	11.61
		SMO Regression	6.51	12.51	21.66	13.67	13.59
2	SMAPE	LR	4.63	11.87	13	11.95	10.36
		M5P	6.47	12.38	14.20	14.28	11.83
		SMO Regression	6.47	14.02	25.32	12.67	14.62
3	MAE	LR	294	853	9940	26000	9271.75
		M5P	425	883	10900	30900	10777.00
		SMO Regression	425	985	11100	27700	10052.50
4	RMSE	LR	536	1270	12100	28400	10576.50
		M5P	496	1190	13400	34100	12296.50
		SMO Regression	496	1380	12300	30300	11119.00
5	MSE	LR	287000	1620000	146000000	805000000	238226750
		M5P	246000	1430000	180000000	1160000000	335419000
		SMO Regression	246000	1900000	150000000	919000000	267786500

TABLE 15. Daily MAPE of LR, SMO regression and M5P techniques to compare forecast accuracy of daily cases for the third wave of COVID-19 in India.

Duration	LR	SMO Reg.	M5P	Duration	LR	SMO Reg.	M5P	Duration	LR	SMO Reg.	M5P	Duration	LR	SMO Reg.	M5P
15-Dec-21	2.53	3.17	3.17	22-Dec-21	14.94	14.51	14.51	1-Jan-22	3.39	47.74	10.18	8-Jan-22	13.82	12.08	18.79
16-Dec-21	0.27	6.2	6.2	23-Dec-21	0.02	1.28	1.28	2-Jan-22	2.48	32.01	5.33	9-Jan-22	3.19	16.09	3.19
17-Dec-21	0.56	7.72	7.72	24-Dec-21	10.61	9.97	9.97	3-Jan-22	15.55	14.92	4.53	10-Jan-22	19.52	29.66	19.52
18-Dec-21	1.43	8.9	8.9	25-Dec-21	8.36	9.96	9.96	4-Jan-22	25.08	26.86	28.95	11-Jan-22	22.63	4.8	27.64
19-Dec-21	4.77	1.33	1.33	26-Dec-21	5.26	13.32	10.3	5-Jan-22	17.63	18.51	15.18	12-Jan-22	10.6	8.16	17.95
20-Dec-21	25.6	15.94	15.94	27-Dec-21	2.57	3.65	3.56	6-Jan-22	10.43	1.62	18.16	13-Jan-22	7.04	10.68	8.57
21-Dec-21	0.26	2.28	2.28	28-Dec-21	32.64	34.91	29.36	7-Jan-22	13.53	9.97	12.61	14-Jan-22	12.2	14.22	10.13
Average	**5.06**	**6.51**	**6.51**	**Average**	**10.63**	**12.51**	**11.28**	**Average**	**12.58**	**21.66**	**13.56**	**Average**	**12.71**	**13.67**	**15.11**

error measures MAPE, SMAPE, RMSE, MAE and MSE for Italy. It can be observed from Table 4 that the average MAPE of LR is 0.39 which again demonstrated a superior result to M5P and SMO regression. Moreover, the average SMAPE, MAE, RMSE and MSE of LR have been evaluated as 0.39, 601.48, 634.98 and 477855 respectively, which depicts superior results once again when compared to M5P and SMO regression. Furthermore, Table 5 summarizes the daily MAPE of LR, SMO regression and M5P techniques to compare the forecast accuracy of daily cases for the first wave of COVID-19 in Italy and the same has been illustrated graphically in Figure 2.

It can be observed that the average daily MAPE values corresponding to LR, M5P and SMO regression techniques

TABLE 16. Daily MAPE of LR, SMO regression and M5P techniques to compare the accuracy of daily death forecast for the first, second, and third waves of COVID-19 in India.

Duration	First Wave			Duration	Second Wave			Duration	Third Wave		
	LR	SMO Reg.	M5P		LR	SMO Reg.	M5P		LR	SMO Reg.	M5P
8-Jul-20	0.51	13.57	19.49	8-May-21	7.38	6.52	12.16	22-Dec-21	23.53	29.91	23.53
9-Jul-20	5.34	17.42	21.89	9-May-21	17.79	14.07	25.92	23-Dec-21	4.34	3.36	4.34
10-Jul-20	1.88	7.4	13.63	10-May-21	7.4	7.4	6.36	24-Dec-21	6.45	12.71	6.45
11-Jul-20	2.9	0.94	5.2	11-May-21	5.42	8.71	7.77	25-Dec-21	2.08	4.26	2.08
12-Jul-20	6.18	10.97	8.83	12-May-21	4.57	4.57	8.65	26-Dec-21	15.9	26	15.9
13-Jul-20	0.71	4.06	15.83	13-May-21	8.55	7.4	9.33	27-Dec-21	14.79	5.61	14.79
14-Jul-20	8.77	4.28	2.15	14-May-21	1.12	4.28	0.07	28-Dec-21	8.19	2.32	8.19
Average	**3.76**	**8.38**	**12.43**	**Average**	**7.46**	**7.56**	**10.04**	**Average**	**10.75**	**12.02**	**10.75**

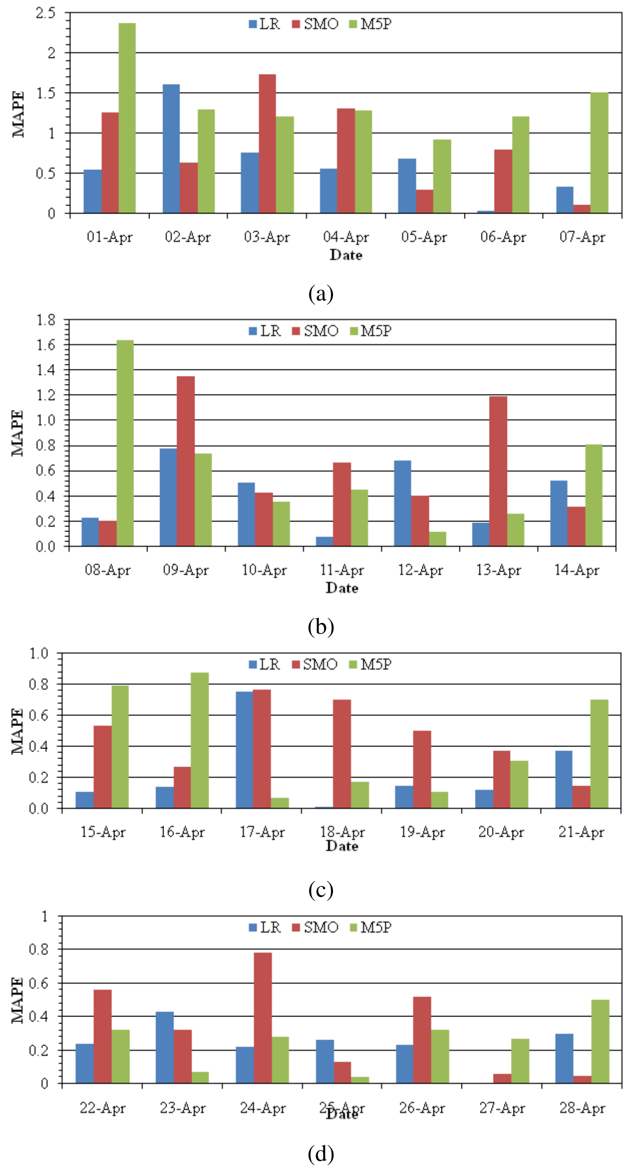

FIGURE 2. Daily MAPE of LR, SMO regression and M5P techniques to compare forecast accuracy of daily cases for the first wave of COVID-19 in Italy for the duration—(a) 01-07 April, 2020 (b) 08-14 April, 2020 (c) 15-21 April, 2020 (d) 22-28 April, 2020.

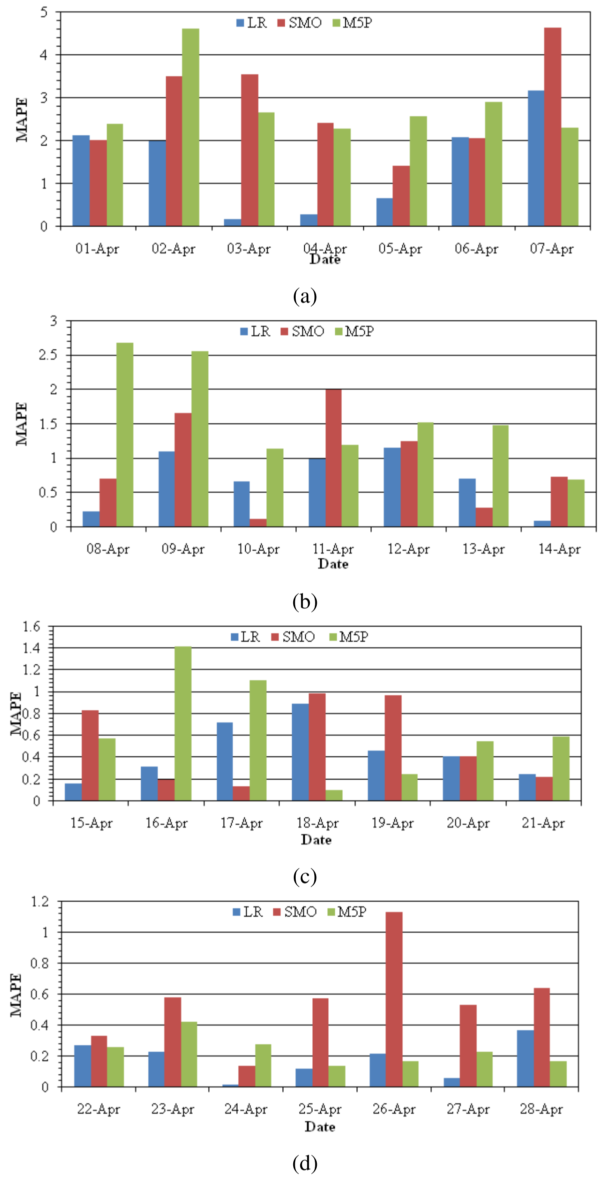

FIGURE 3. Daily MAPE of LR, SMO regression and M5P techniques to compare forecast accuracy of daily cases for the first wave of COVID-19 in Australia for the duration—(a) 01-07 April, 2020 (b) 08-14 April, 2020 (c) 15-21 April, 2020 (d) 22-28 April, 2020.

have been evaluated as 0.65, 1.4, and 0.88, respectively, for the duration 01–07 April 2020; 0.43, 0.63, 0.65, respectively, for the duration 08–14 April, 2020; 0.24, 0.43, 0.47, respectively, for the duration 15–21 April, 2020; and 0.24, 0.26, 0.35, respectively, for 22–28 April, 2020. The daily MAPE, as summarized in Table 5 and illustrated in Figure 2,

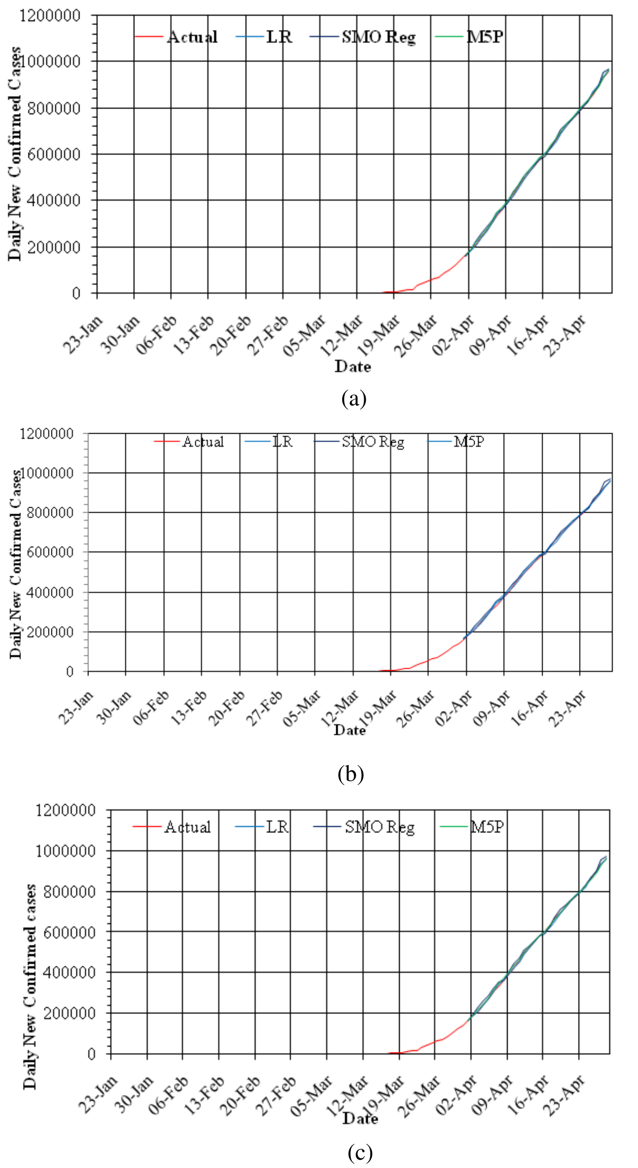

FIGURE 4. Forecast of daily cases using LR, SMO regression and M5P techniques for the first wave of COVID-19 in (a) USA (b) Italy (c) Australia.

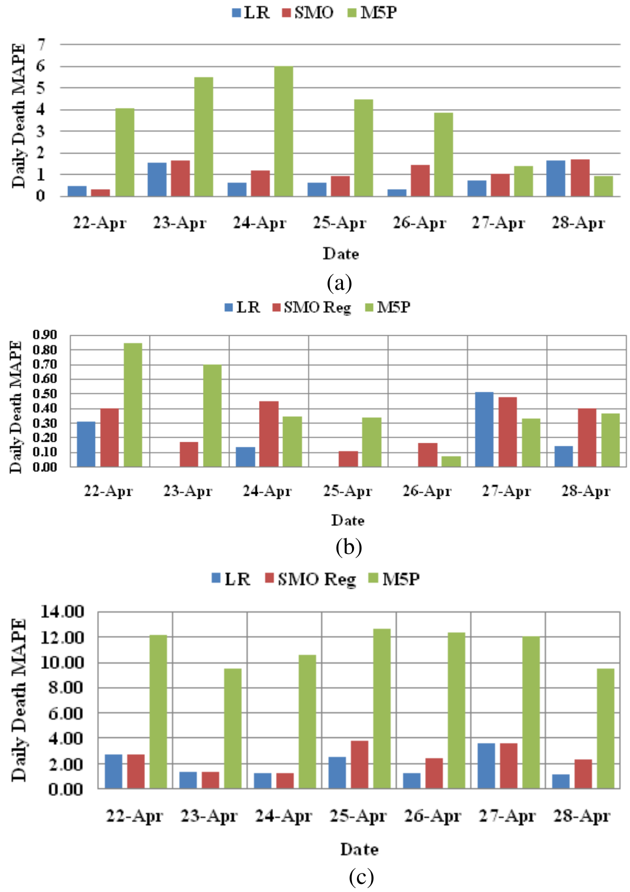

FIGURE 5. Daily MAPE for the duration of 22-28 April, 2020 of LR, SMO regression and M5P techniques to compare the accuracy of daily death forecast for the first wave of COVID-19 in (a) USA (b) Italy (c) Australia.

clearly indicates that the LR technique for forecasting daily cases for the first wave of COVID-19 in Italy outperforms the M5P and SMO regression techniques once again.

Likewise, Table 6 compares the LR technique with the SMO regression and M5P techniques by evaluating different error measures for Australia. It can be found from Table 6 that the average MAPE of LR is 0.71 which again demonstrated the ability of LR to produce superior results than M5P and SMO regression. Moreover,the average SMAPE, MAE, RMSE, and MSE of LR have been evaluated as 0.70, 42.35, 49.32 and 3425.52, respectively, which onceagain shows superior results to M5P and SMO regression. Furthermore, Table 7 summarizes daily MAPE of LR, SMO regression and M5P techniques to compare the forecast accuracy of daily cases for the first wave of COVID-19 in Australia and the same has been depicted graphically in Figure 3. It can be observed that the average of daily MAPE values corresponding to LR, M5P, and SMO regression techniques have been evaluated as 1.48, 2.8 and 2.78, respectively, for the duration 01–07 April, 2020; 0.71, 1.61, 0.96, respectively, for the duration 08–14 April, 2020; 0.46, 0.65, 0.54, respectively, for the duration 15–21 April, 2020; and 0.18, 0.24 and 0.56, respectively, for 22–28 April, 2020. The daily MAPE, as summarized in Table 7 and illustrated in Figure 3, clearly implies that the LR technique again outperforms M5P and SMO regression for forecasting daily cases for the first wave of COVID-19 in Australia. Forecast of daily cases for April 01-28, 2020 using LR, SMO regression and M5P techniques for the first wave of COVID-19 in the USA, Italy and Australia have been depicted in Figure 4. A comparison has been made with the actual data which depicts the variations between the forecasted values and the actual values. Table 8 summarizes daily MAPE for the duration range of Apr 22-28, 2020 of LR, SMO regression and M5P techniques to compare the accuracy of daily death forecast for the first wave of COVID-19 in the USA, Italy and Australia and the same has been depicted graphically in Figure 5. The average

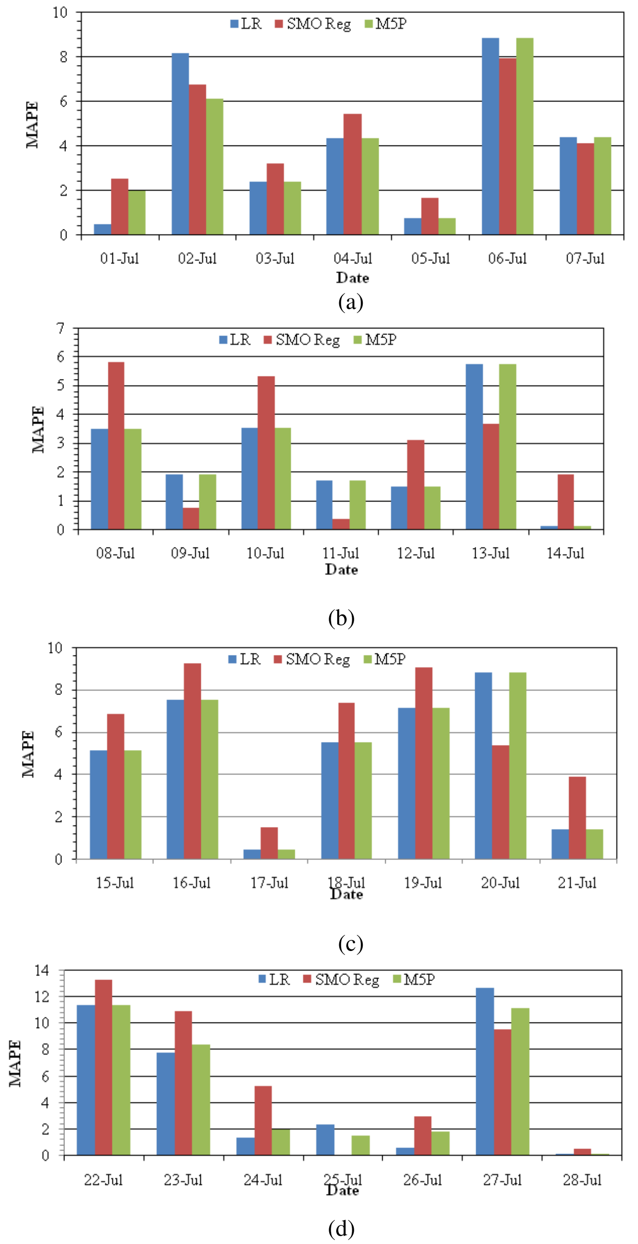

FIGURE 6. Daily MAPE of LR, SMO regression and M5P techniques to compare forecast accuracy of daily cases for the first wave of COVID-19 in India for the duration–(a) 01-07 July, 2020 (b) 08-14 July, 2020 (c) 15-21 July, 2020 (d) 22-28 July, 2020.

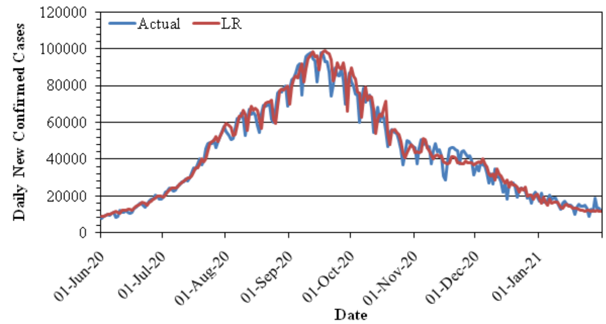

FIGURE 7. Forecast of daily cases for the duration from June 01, 2020 to Jan 31, 2021 for the first wave of COVID-19 in India.

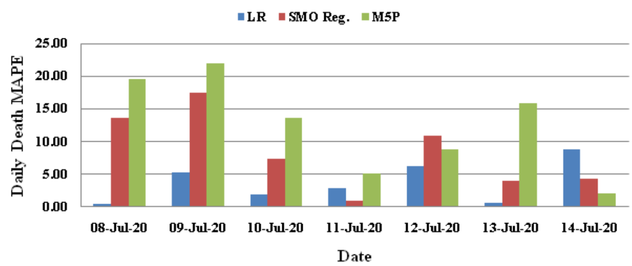

FIGURE 8. Daily MAPE for the duration 08-14 July, 2020 of LR, SMO regression and M5P techniques to compare the accuracy of daily death forecast for the first wave of COVID-19 in India.

daily MAPE values corresponding to LR, M5P, and SMO regression techniques have been evaluated as 0.84, 3.73, and 1.17 respectively, for the USA; 0.16, 0.43, 0.31 respectively for Italy; and 1.99, 11.25 and 2.52, respectively, for Australia. This again suggests that the LR technique outperforms M5P and SMO regression for death forecast for the first wave of COVID-19 in the USA, Italy and Australia. The authors have also compared the MAPE values of the several other methods reported in [35] and [41] with the MAPE values of the forecast models presented in this paper (viz. M5P, SMO, and LR), considering similar data sets. Table 9 compares the work presented with other methods reported in recently published papers.

IV. EXTENSIVE FORECAST ANALYSIS OF COVID-19 SCENARIOS FOR INDIA

The various error measures are summarized in Table 2—Table 8 indicate that all three approaches viz. LR, SMO regression and M5P employed in the preliminary sample study presented in the previous section have acceptable forecast accuracy and the LR technique outperforms M5P and SMO regression techniques. Although the LR technique outperformed during the first wave in three different countries, using the LR technique alone would not be sufficient for extensive forecast analysis in the Indian scenario since ML algorithms rely heavily on quality data to learn future trends and build better performing forecasting models.This prompted the authors to continue their extensive forecast analysis on the first, second and third waves of the COVID-19 pandemic scenarios in India using all three techniques.

A. FORECAST ANALYSIS OF COVID-19 SCENARIO FOR FIRST WAVE IN INDIA

In this sub-section, the authors have conducted a complete forecast study on the first wave of the COVID-19 pandemic scenario in India. Table 10 compares the LR technique with SMO regression and M5P techniques by evaluating error measures MAPE, SMAPE, RMSE, MAE and MSE for forecasting daily cases for the first wave of COVID-19 in India. Table 10 shows that the average MAPE of LR

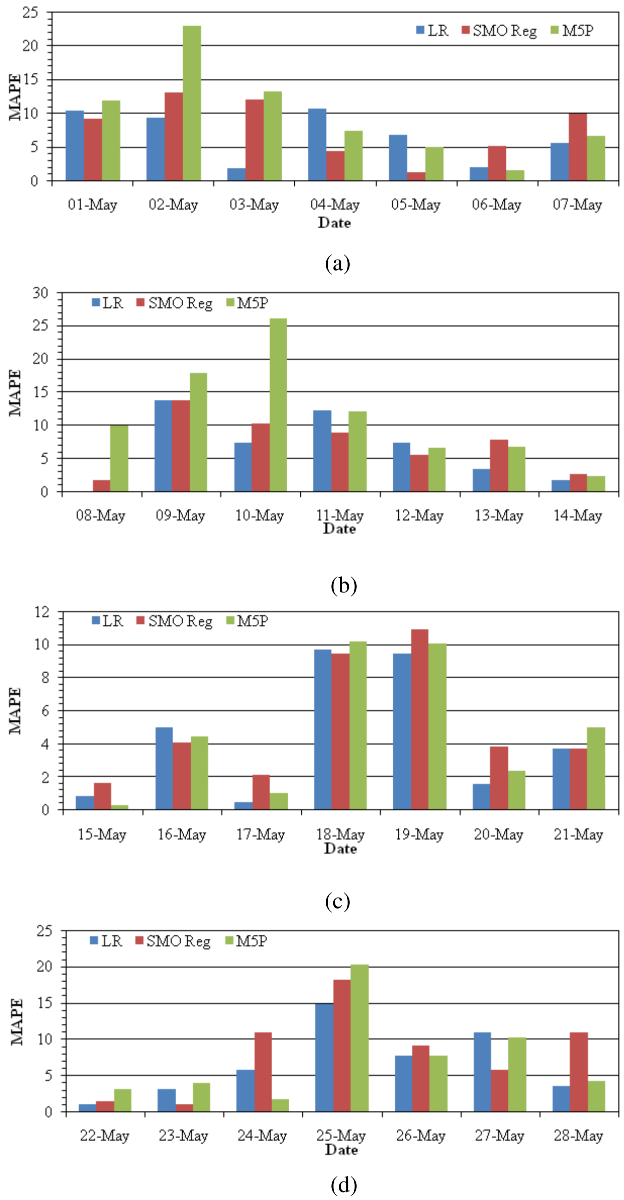

FIGURE 9. Daily MAPE of LR, SMO regression and M5P techniques to compare forecast accuracy of daily cases for the second wave of COVID-19 in India for the duration—(a) 01-07 May, 2021 (b) 08-14 May, 2021 (c) 15-21 May, 2021 (d) 22-28 May, 2021.

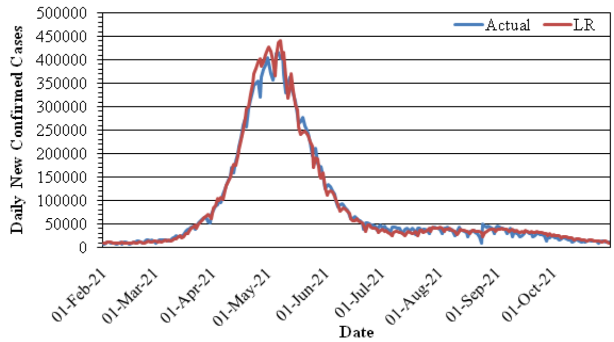

FIGURE 10. Forecast of daily cases for the duration 01 Feb, 2021 to 31 Oct, 2021 for the second wave of COVID-19 in India.

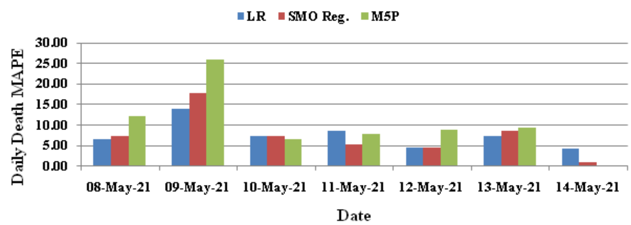

FIGURE 11. Daily MAPE for the duration 08-14 May, 2021 of LR, SMO regression and M5P techniques to compare the accuracy of daily death forecast for the second wave of COVID-19 in India.

is 4.25, indicating that LR outperformed M5P and SMO regression (with MAPE of 4.27 and 4.94 respectively). Moreover, the average SMAPE, MAE, RMSE, and MSE of LR have been evaluated as 4.28, 1490, 1832.85, and 4247828, respectively. Hence, LR has higher forecasting performance when compared to M5P and SMO regression. As stated in the previous section, MAPE is one of the most commonly used key performance indicators to measure forecast accuracy. Therefore, Table 11 has been prepared to summarize the daily MAPE of LR, SMO regression and M5P techniques to compare the forecast accuracy of daily cases for the first wave of COVID-19 in India and the same has been graphically represented in Figure 6.It can be observed that the average of daily MAPE values corresponding to LR, M5P and SMO regression techniques have been evaluated as 4.12, 4.20 and 4.52, respectively, for the duration 01–07 July 2020; 2.56, 2.56 and 2.98 respectively, for the duration 08–14 July 2020; 5.12, 5.12 and 6.18 respectively for the duration 15–21 July 2020; and 5.18, 5.19, 6.06 respectively for 22–28 July 2020. The daily MAPE values, as summarized in Table 11 and shown in Figure 6, clearly indicate that the LR technique for forecasting daily cases for the first wave of COVID-19 in India outperforms the M5P and SMO regression.

A forecast of daily cases for the duration from 01 June 2020 to 31 Jan 2021, using the LR technique for the first wave of COVID-19 in India has been depicted in Figure 7. A comparison has been made with the actual data which clearly shows how closely the forecasted values match the actual data. On the other hand, the first part of Table 16 summarizes daily MAPE for the duration 08-14 July 2020 of LR, SMO regression and M5P techniques to compare the accuracy of daily death forecast for the first wave of COVID-19 in India and the same has been depicted graphically in Figure 8. The average of daily MAPE values corresponding to LR, M5P and SMO regression techniques have been evaluated as 3.76, 12.43 and 8.38 respectively, for the first wave of COVID-19 in India. It can be concluded that the LR techniques provide a better forecasting outcome when compared to M5P and SMO regression techniques for death forecast for the first wave of COVID-19 in India.

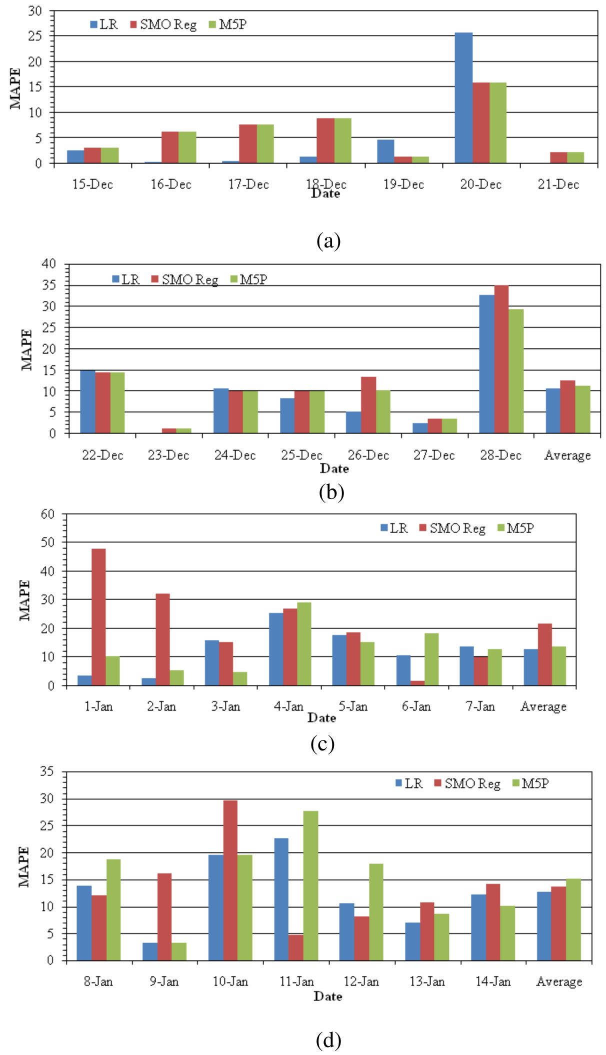

FIGURE 12. **Daily MAPE of LR, SMO regression and M5P techniques to compare forecast accuracy of daily cases for the third wave of COVID-19 in India for the duration—(a) 15-21 Dec, 2021 (b) 22-28 Dec, 2022 (c) 01-07 Jan, 2022 (d) 08-14 Jan, 2022.**

B. FORECAST ANALYSIS OF COVID-19 SCENARIO FOR SECOND WAVE IN INDIA

In this sub-section, the authors have conducted a complete forecast study on the second wave of the COVID-19 pandemic scenario in India. Table 12 compares the LR technique with SMO regression and M5P techniques by evaluating error measures MAPE, SMAPE, RMSE, MAE and MSE for forecasting daily cases for the second wave of COVID-19 in India. Table 12 shows that the average MAPE of LR is 6.13, indicating that LR supersedes M5P and SMO regression (with MAPE of 8.42 and 7.14 respectively). Moreover, the average SMAPE, MAE, RMSE, and MSE of LR have been evaluated as 6.14, 18775, 22525, and 551750000, respectively; hence, confirming superior results relative to M5P and SMO regression.

Table 13 has been prepared to summarize daily MAPE of LR, SMO regression and M5P techniques to compare the forecast accuracy of daily cases for the second wave

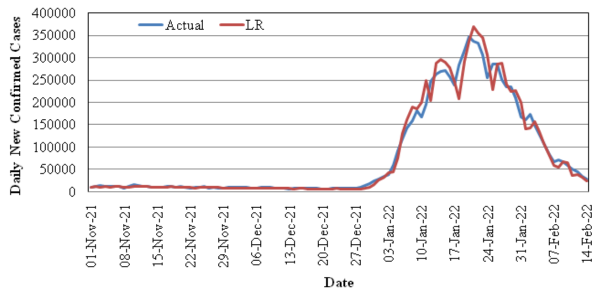

FIGURE 13. **Forecast of daily cases for the duration from 01 Nov, 2021 to 14 Feb, 2022 for the third wave of COVID-19 in India.**

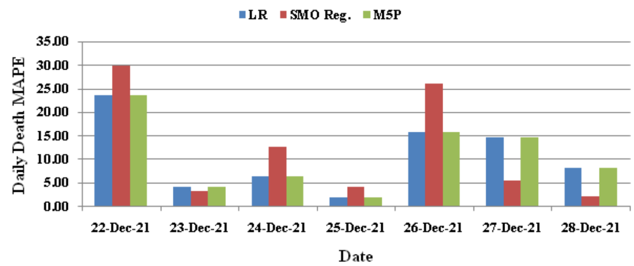

FIGURE 14. **Daily MAPE for the duration from 22 Dec, to 28 Dec, 2021 of LR, SMO regression and M5P techniques to compare the accuracy of daily death forecast for the third wave of COVID-19 in India.**

of COVID-19 in India and the same has been illustrated graphically in Figure 9. It can be observed that the average of daily MAPE values corresponding to LR, M5P, and SMO regression techniques have been evaluated as 6.72, 9.85 and 7.92, respectively, for the duration 01–07 May 2021; 6.61, 11.68 and 7.30, respectively, for the duration 08–14 May 2021; 4.40, 4.78 and 5.11, respectively, for the duration 15–21 May 2021; and 6.79, 7.39 and 8.24, respectively, for 22–28 May 2021. The daily MAPE, as summarized in Table 13 and shown in Figure 9, clearly indicates that the LR technique for forecasting daily cases for the second wave of COVID-19 in India outmatch the M5P and SMO regression.

A forecast of daily cases for the duration from 01 Feb 2021 to 31 Oct 2021 using the LR technique for the second wave of COVID-19 in India has been depicted in Figure 10. A comparison has been made with the actual data which clearly shows that the forecasted dataset matches with the actual dataset. On the other hand, the mid-part of Table 16 summarizes daily MAPE for the duration 08-14 May 2021 of LR, SMO regression and M5P techniques to compare the accuracy of the daily death forecast for the second wave of COVID-19 in India and the same has been depicted graphically in Figure 11. The average daily MAPE values corresponding to LR, M5P, and SMO regression techniques have been evaluated as 7.46, 10.04 and 7.56, respectively, for the second wave of COVID-19 in India. It can be concluded that the LR techniques again provide a better forecasting outcome when compared to M5P and SMO regression techniques for death forecast for the second wave of COVID-19 in India.

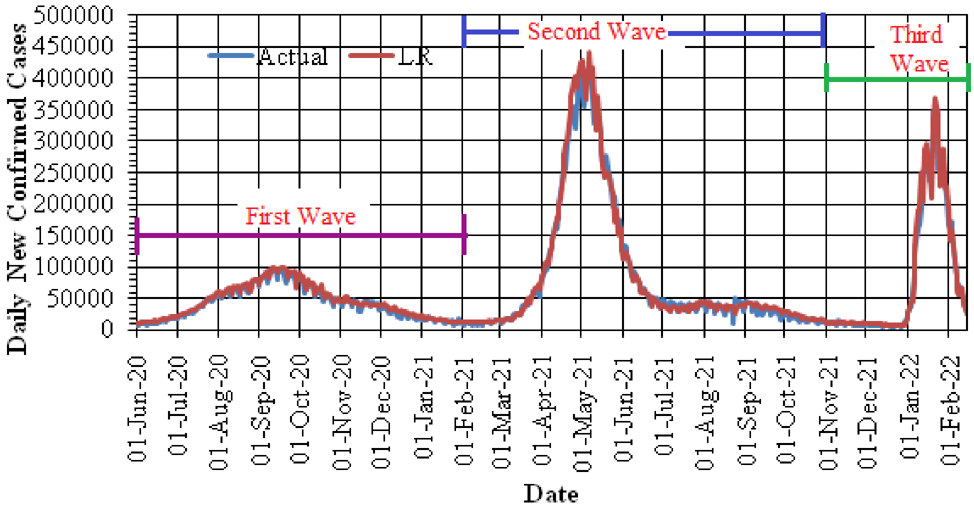

FIGURE 15. Comprehensive illustration of COVID-19 forecast of daily new cases for first, second, and third waves in India from 01 June 2020 to 14 February 2022.

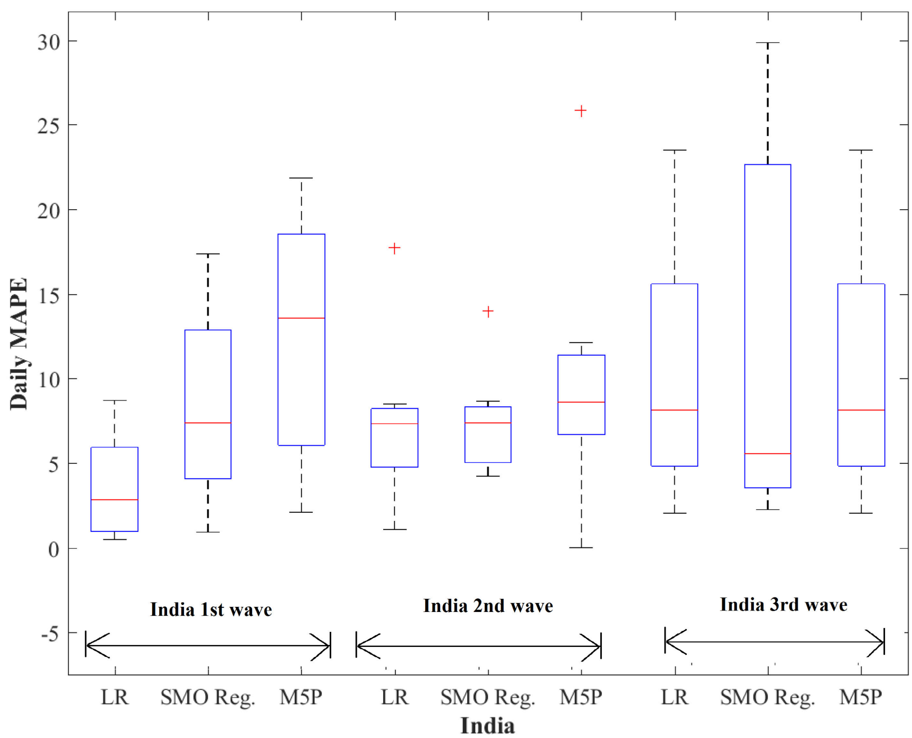

FIGURE 16. Box plot of daily MAPE of new cases during First, Second and Third waves of COVID-19 in India.

C. FORECAST ANALYSIS OF COVID-19 SCENARIO FOR THIRD WAVE IN INDIA

In this sub-section, the authors have conducted a complete forecast analysis on the third wave of the COVID-19 pandemic scenario in India. Table 14 compares the LR technique with SMO regression and M5P techniques by evaluating different error measures for forecasting daily cases for the third wave of COVID-19 in India. Table14 indicates that the average MAPE of LR is 10.25, indicating that LR outmatches M5P and SMO regression. Furthermore, the average SMAPE, MAE, RMSE and MSE of LR have been evaluated as 10.36, 9271.75, 10576.50 and 238226750, respectively, therefore, reaffirm superior results to M5P and SMO regression.

Table 15 has been prepared to summarize daily MAPE of LR, SMO regression and M5P techniques to compare the forecast accuracy of daily cases for the third wave of COVID-19 in India and the same has been depicted

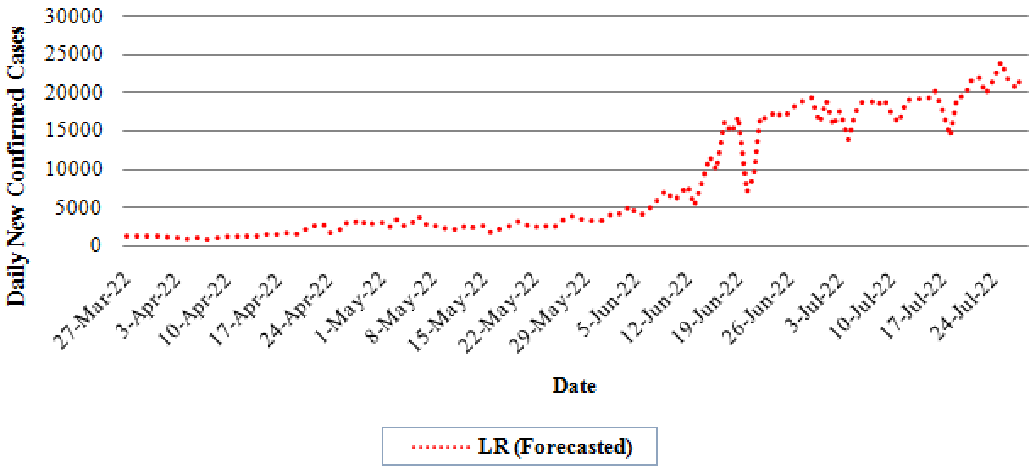

FIGURE 17. Forecasting of likely onset of the fourth wave of COVID-19 in India.

graphically in Figure 12. It can be observed that the average daily MAPE values corresponding to LR, M5P, and SMO regression techniques have been evaluated as 5.06, 6.51 and 6.51, respectively, for 15–21 December 2021; 10.63, 11.28 and 12.51, respectively, for 22–28 December 2021; 12.58, 13.56 and 21.66, respectively, for 01–07 January, 2022; and 12.71, 15.11, 13.67, respectively, for 08–14 January, 2022. The daily MAPE, as summarized in Table 15 and shown in Figure 12, clearly indicates that the LR technique for forecasting daily cases for the third wave of COVID-19 in India outmatch the M5P and SMO regression.

Forecast of daily cases for the duration from 01 Nov 2021, to 14 Feb 2022 using the LR technique for the third wave of COVID-19 in India has been plotted in Figure 13.

A comparison has been made with the actual data which indicates that the forecasted values match the actual dataset. On the other hand, the last part of Table 16 summarizes the daily MAPE of LR, SMO regression and M5P techniques for the duration 22-28 December 2021 to compare the accuracy of the daily death forecast for the third wave of COVID-19 in India and the same has been illustrated graphically in 14. The average daily MAPE values corresponding to LR, M5P, and SMO regression techniques have been evaluated as 10.75, 10.75 and 12.02 respectively, for the third wave of COVID-19 in India. It can be concluded that the LR technique once more yields better forecasting outcome when compared to M5P and SMO regression techniques for death forecast for the third wave of COVID-19 in India.

D. CLOSING REMARKS ON ALL THREE WAVES OF COVID-19 IN INDIA AND FORECASTING OF LIKELY ONSET OF THE FOURTH WAVE

Figure 15 comprehensively illustrates the results of forecast analysis for daily new cases of COVID-19 using the LR technique for the first, second and third waves in India for the duration from 01 June 2020 to 14 Feb 2022. The comparison has been made with the actual dataset of daily cases of COVID-19 obtained for all three waves in India. In this study, the authors used bar charts to represent MAPE values that allow a visual check of the accuracy of the forecast and the rationality of the calculations. In addition, the authors have also included the box plots showing the distribution of MAPE values for the prediction of daily COVID-19 cases during the first, second, and third waves of COVID-19 in India. Figure 16 depicts the box plot of daily MAPE of new cases during first, second and third waves of COVID-19 in India.

On analyzing the combined plot of the daily cases of COVID-19 for all three waves in India, it is evident that the duration of the first wave of COVID-19 in India was longer than that of the second wave. However, the number of daily new cases of COVID-19 was the lowest for the first wave compared to the second and third waves. On the other hand, the duration of the third wave was the shortest among the three waves of COVID-19 in India. Nevertheless, the number of daily new cases of COVID-19 for the third wave in India was slightly lower than for the second wave. It is a matter of fact that the second wave of COVID-19 infected people more severely than the first and third waves of COVID-19 in India. However, the people of India are fortunate that the Indian government took action against COVID-19 before it could get worse, which was a concern for many experts, given India's large population.

Forecasting the likely onset of the fourth wave will be of great help in making important decisions and planning for the implementation of preventive measures. Therefore, based on the extensive analysis conducted for the first, second and third waves of COVID-19 in the Indian scenario, the LR technique alone would be sufficient to forecast the likely onset of the fourth wave of COVID-19 in India. The forecast result using the LR technique for daily new cases for the period 27 March, 2022 to 28 July, 2022 is shown in Figure 17 which shows the upswing in daily new cases after May 2022. Looking at the rapidly increasing daily new cases during June-July 2022, it seems that India is likely to witness a fourth wave of COVID-19 in the coming days if preventive measures are not taken.

V. CONCLUSION

The ML approach has proven to be a potential strategy to forecast the current COVID-19 pandemic scenarios. Three commonly used ML approaches, namely LR, M5P and SMO regression techniques were used in this study for forecast analysis of the COVID-19 pandemic scenarios.

- A preliminary sample study was conducted using LR, M5P and SMO regression techniques, to forecast COVID-19 pandemic scenarios for the three countries (USA, Italy and Australia) from different subcontinents for the first wave of COVID-19.
- Forecast results obtained for three countries in the preliminary sample study showed satisfactory forecast accuracy.
- Based on a preliminary sample study conducted for three countries from different subcontinents, it has been established that the LR technique outperformed M5P and SMO regression in forecasting the COVID-19 pandemic scenarios for the first wave.
- In addition, the comprehensive analysis conducted for the first, second and third waves of COVID-19 in India also established that the forecast accuracy of LR was better than that of M5P and SMO regression.
- This way, the LR technique can be suggested as the most suitable model for forecasting COVID-19 pandemic scenarios in India.
- Accordingly, the LR technique has been used to forecast the likely onset of the fourth wave of COVID-19 in India from the perspective of better-advanced decision-making about the future course of action
- As forecasted in this study, it seems that India may witness a fourth wave of COVID-19 in the coming days if preventive measures are not taken. The Government of India should keep an eye on the increase in daily new cases and take appropriate steps as needed to prevent its spread.
- The people of India are also advised not to be negligent and follow the instructions given by the Government of India from time to time to fight against the possible fourth wave of COVID-19.

The authors believe that the findings of this paper will certainly inspire further studies to develop more accurate forecasting models of the COVID-19 pandemic scenarios.

REFERENCES

[1] E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track COVID-19 in real time," *Lancet Infectious Diseases*, vol. 20, no. 5, pp. 533–534, 2020.
[2] A. Remuzzi and G. Remuzzi, "COVID-19 and Italy: What next?" *Lancet*, vol. 395, no. 10231, pp. 1225–1228, Apr. 2020.
[3] Q. Li *et al.*, "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia," *New England J. Med.*, vol. 382, no. 13, pp. 1199–1207, Mar. 2020.
[4] S. Ai, G. Zhu, F. Tian, H. Li, Y. Gao, Y. Wu, Q. Liu, and H. Lin, "Population movement, city closure and spatial transmission of the 2019-nCoV infection in China," *MedRxiv*, 2020.
[5] J. Riou and C. L. Althaus, "Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020," *Eurosurveillance*, vol. 25, no. 4, Jan. 2020, Art. no. 2000058.
[6] *Coronavirus Disease (COVID-19) Weekly Epidemiological Update and Weekly Operational Update*. World Health Organization (WHO). Accessed: Mar. 18, 2022. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-re%ports
[7] R. M. Elavarasan and R. Pugazhendhi, "Restructured society and environment: A review on potential technological strategies to control the COVID-19 pandemic," *Sci. Total Environ.*, vol. 725, Jul. 2020, Art. no. 138858.
[8] R. M. Elavarasan, R. Pugazhendhi, G. M. Shafiullah, N. M. Kumar, M. T. Arif, T. Jamal, S. S. Chopra, and J. Dyduch, "Impacts of COVID-19 on sustainable development goals and effective approaches to maneuver them in the post-pandemic environment," *Environ. Sci. Pollut. Res.*, vol. 29, no. 23, pp. 33957–33987, May 2022.
[9] Y. Steinbuch, *China Opens Second New Hospital for Coronavirus Patients*. New York, NY, USA: New York Post, 2020.
[10] G. Ippolito, D. S. Hui, F. Ntoumi, M. Maeurer, and A. Zumla, "Toning down the 2019-nCoV media hype—And restoring hope," *Lancet Respiratory Med.*, vol. 8, no. 3, pp. 230–231, Mar. 2020.
[11] B. McCloskey, A. Zumla, G. Ippolito, L. Blumberg, P. Arbon, A. Cicero, T. Endericks, P. L. Lim, and M. Borodina, "Mass gathering events and reducing further global spread of COVID-19: A political and public health dilemma," *Lancet*, vol. 395, no. 10230, pp. 1096–1099, Apr. 2020.
[12] Y. Liu, A. A. Gayle, A. Wilder-Smith, and J. Rocklöv, "The reproductive number of COVID-19 is higher compared to SARS coronavirus," *J. Travel Med.*, vol. 27, no. 2, Mar. 2020, Art. no. taaa021.
[13] A. Lai, A. Bergna, C. Acciarri, M. Galli, and G. Zehender, "Early phylogenetic estimate of the effective reproduction number of SARS-CoV-2," *J. Med. Virol.*, vol. 92, no. 6, pp. 675–679, Jun. 2020.
[14] Z. Wang, C. T. Bauch, S. Bhattacharyya, A. d'Onofrio, P. Manfredi, M. Perc, N. Perra, M. Salathé, and D. Zhao, "Statistical physics of vaccination," *Phys. Rep.*, vol. 664, pp. 1–113, Dec. 2016.
[15] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, "Multilayer networks," *J. Complex Netw.*, vol. 2, no. 3, pp. 203–271, 2014.
[16] L. Lü, D. Chen, X.-L. Ren, Q.-M. Zhang, Y.-C. Zhang, and T. Zhou, "Vital nodes identification in complex networks," *Phys. Rep.*, vol. 650, pp. 1–63, Sep. 2016.
[17] G. Chowell, D. Hincapie-Palacio, J. Ospina, B. Pell, A. Tariq, S. Dahal, S. Moghadas, A. Smirnova, L. Simonsen, and C. Viboud, "Using phenomenological models to characterize transmissibility and forecast patterns and final burden of Zika epidemics," *PLoS Currents*, vol. 8, May 2016.
[18] B. Pell, Y. Kuang, C. Viboud, and G. Chowell, "Using phenomenological models for forecasting the 2015 Ebola challenge," *Epidemics*, vol. 22, pp. 62–70, Mar. 2018.
[19] G. Bontempi, S. B. Taieb, and Y.-A. L. Borgne, "Machine learning strategies for time series forecasting," in *European Business Intelligence Summer School*. Cham, Switzerland: Springer, 2012, pp. 62–77.
[20] N. Poonia and S. Azad, "Short-term forecasts of COVID-19 spread across Indian states until 29 May 2020 under the worst-case scenario," *Sci. Rep.*, 2020. [Online]. Available: https://arxiv.org/abs/2004.13538
[21] D. Fanelli and F. Piazza, "Analysis and forecast of COVID-19 spreading in China, Italy and France," *Chaos, Solitons Fractals*, vol. 134, May 2020, Art. no. 109761.
[22] Y. Hao, T. Xu, H. Hu, P. Wang, and Y. Bai, "Prediction and analysis of corona virus disease 2019," *PLoS ONE*, vol. 15, no. 10, Oct. 2020, Art. no. e0239960.
[23] T. Dehesh, H. Mardani-Fard, and P. Dehesh, "Forecasting of COVID-19 confirmed cases in different countries with ARIMA models," *MedRxiv*, 2020.
[24] G. Perone, "An ARIMA model to forecast the spread and the final size of COVID-2019 epidemic in Italy," in *HEDG-Health Econometrics and Data Group Working Paper Series*. Heslington, U.K.: Univ. York, 2020.
[25] M. Perc, N. G. Miksic, M. Slavinec, and A. Stozer, "Forecasting COVID-19," *Frontiers Phys.*, vol. 8, p. 127, Apr. 2020.
[26] F. Rustam, A. A. Reshi, A. Mehmood, S. Ullah, B.-W. On, W. Aslam, and G. S. Choi, "COVID-19 future forecasting using supervised machine learning models," *IEEE Access*, vol. 8, pp. 101489–101499, 2020.
[27] J. Musulin, S. Baressi Šegota, D. Štifanić, I. Lorencin, N. Anđelić, T. Šušteršič, A. Blagojević, N. Filipović, T. Cabov, and E. Markova-Car, "Application of artificial intelligence-based regression methods in the problem of COVID-19 spread prediction: A systematic review," *Int. J. Environ. Res. Public Health*, vol. 18, no. 8, p. 4287, Apr. 2021.
[28] I. Rahimi, F. Chen, and A. H. Gandomi, "A review on COVID-19 forecasting models," *Neural Comput. Appl.*, 2021, doi: 10.1007/s00521-020-05626-8.

[29] M. Wieczorek, J. Siłka, D. Połap, M. Woźniak, and R. Damaševičius, "Real-time neural network based predictor for cov19 virus spread," *PLoS ONE*, vol. 15, no. 12, Dec. 2020, Art. no. e0243189.
[30] A. A. H. Ahmadini, M. Naeem, M. Aamir, R. Dewan, S. S. A. Alshqaq, and W. K. Mashwani, "Analysis and forecast of the number of deaths, recovered cases, and confirmed cases from COVID-19 for the top four affected countries using Kalman filter," *Frontiers Phys.*, vol. 9, p. 629320, Aug. 2021.
[31] A. Zeroual, F. Harrou, A. Dairi, and Y. Sun, "Deep learning methods for forecasting COVID-19 time-series data: A comparative study," *Chaos, Solitons Fractals*, vol. 140, Nov. 2020, Art. no. 110121.
[32] Y. Alali, F. Harrou, and Y. Sun, "A proficient approach to forecast COVID-19 spread via optimized dynamic machine learning models," *Sci. Rep.*, vol. 12, no. 1, pp. 1–20, Dec. 2022.
[33] A. Dairi, F. Harrou, Y. Sun, and S. Khadraoui, "A hybrid deep learning method with attention for COVID-19 spread forecasting," *Artif. Intell. Strategies Analyzing COVID-19 Pneumonia Lung Imag.*, vol. 1, p. 5, Apr. 2022.
[34] V. K. Gupta, A. Gupta, D. Kumar, and A. Sardana, "Prediction of COVID-19 confirmed, death, and cured cases in India using random forest model," *Big Data Mining Anal.*, vol. 4, no. 2, pp. 116–123, Jun. 2021.
[35] R. K. Singh, M. Drews, M. D. L. Sen, M. Kumar, S. S. Singh, A. K. Pandey, P. K. Srivastava, M. Dobriyal, M. Rani, P. Kumari, and P. Kumar, "Short-term statistical forecasts of COVID-19 infections in India," *IEEE Access*, vol. 8, pp. 186932–186938, 2020.
[36] M. Naeem, J. Yu, M. Aamir, S. A. Khan, O. Adeleye, and Z. Khan, "Comparative analysis of machine learning approaches to analyze and predict the COVID-19 outbreak," *PeerJ Comput. Sci.*, vol. 7, p. e746, Dec. 2021.
[37] R. R. Sharma, M. Kumar, S. Maheshwari, and K. P. Ray, "EVDHM-ARIMA-based time series forecasting model and its application for COVID-19 cases," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021.
[38] S. Dash, C. Chakraborty, S. K. Giri, S. K. Pani, and J. Frnda, "BIFM: Big-data driven intelligent forecasting model for COVID-19," *IEEE Access*, vol. 9, pp. 97505–97517, 2021.
[39] H. Friji, R. Hamadi, H. Ghazzai, H. Besbes, and Y. Massoud, "A generalized mechanistic model for assessing and forecasting the spread of the COVID-19 pandemic," *IEEE Access*, vol. 9, pp. 13266–13285, 2021.
[40] D. C. D. S. Gomes and G. L. D. O. Serra, "Machine learning model for computational tracking and forecasting the COVID-19 dynamic propagation," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 3, pp. 615–622, Mar. 2021.
[41] E. Gecili, A. Ziady, and R. D. Szczesniak, "Forecasting COVID-19 confirmed cases, deaths and recoveries: Revisiting established time series modeling through novel applications for the USA and Italy," *PLoS ONE*, vol. 16, no. 1, Jan. 2021, Art. no. e0244173.
[42] M. Iqbal, F. Al-Obeidat, F. Maqbool, S. Razzaq, S. Anwar, A. Tubaishat, M. S. Khan, and B. Shah, "COVID-19 patient count prediction using LSTM," *IEEE Trans. Computat. Social Syst.*, vol. 8, no. 4, pp. 974–981, Aug. 2021.
[43] R. Kumari, S. Kumar, R. C. Poonia, V. Singh, L. Raja, V. Bhatnagar, and P. Agarwal, "Analysis and predictions of spread, recovery, and death caused by COVID-19 in India," *Big Data Mining Anal.*, vol. 4, no. 2, pp. 65–75, Jun. 2021.
[44] M. Lucic, H. Ghazzai, C. Lipizzi, and Y. Massoud, "Integrating county-level socioeconomic data for COVID-19 forecasting in the United States," *IEEE Open J. Eng. Med. Biol.*, vol. 2, pp. 235–248, 2021.
[45] B. Zhou, G. Yang, Z. Shi, and S. Ma, "Interpretable temporal attention network for COVID-19 forecasting," *Appl. Soft Comput.*, vol. 120, May 2022, Art. no. 108691.
[46] V. Bianco, O. Manca, and S. Nardini, "Electricity consumption forecasting in Italy using linear regression models," *Energy*, vol. 34, no. 9, pp. 1413–1421, Sep. 2009.
[47] A. K. Srivastava, A. S. Pandey, R. M. Elavarasan, U. Subramaniam, S. Mekhilef, and L. Mihet-Popa, "A novel hybrid feature selection method for day-ahead electricity price forecasting," *Energies*, vol. 14, no. 24, p. 8455, Dec. 2021.
[48] A. K. Srivastava, "Short term load forecasting using regression trees: Random forest, bagging and M5P," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 2, pp. 1898–1902, Apr. 2020.
[49] A. Akgündoğdu, I. Öz, and C. P. Uzunoğlu, "Signal quality based power output prediction of a real distribution transformer station using M5P model tree," *Electric Power Syst. Res.*, vol. 177, Dec. 2019, Art. no. 106003.
[50] P. Xu, M. Aamir, A. Shabri, M. Ishaq, A. Aslam, and L. Li, "A new approach for reconstruction of IMFs of decomposition and ensemble model for forecasting crude oil prices," *Math. Problems Eng.*, vol. 2020, pp. 1–23, Oct. 2020.
[51] W. Gao, M. Aamir, A. B. Shabri, R. Dewan, and A. Aslam, "Forecasting crude oil price using Kalman filter based on the reconstruction of modes of decomposition ensemble model," *IEEE Access*, vol. 7, pp. 149908–149925, 2019.
[52] J. Devaraj, R. M. Elavarasan, R. Pugazhendhi, G. M. Shafiullah, S. Ganesan, A. K. Jeysree, I. A. Khan, and E. Hossain, "Forecasting of COVID-19 cases using deep learning models: Is it reliable and practically significant?" *Results Phys.*, vol. 21, Feb. 2021, Art. no. 103817.
[53] D. A. Swanson, "On the relationship among values of the same summary measure of error when used across multiple characteristics at the same point in time: An examination of MALPE and MAPE," *Rev. Econ. Finance*, vol. 5, no. 1, pp. 1–14, 2015.
[54] Michael Gilliland," *The Business Forecasting Deal: Exposing Myths, Eliminating Bad Practices, Providing Practical Solutions*. Hoboken, NJ, USA: Wiley, 2010.

ANKIT KUMAR SRIVASTAVA received the B.Tech. degree in electrical engineering from V. B. S. Purvanchal University, Jaunpur, Uttar Pradesh, India, the M.Tech. degree in power electronics and drives from the Kamla Nehru Institute of Technology, Sultanpur, Uttar Pradesh, and the Ph.D. degree from Dr. A. P. J. Abdul Kalam Technical University, Lucknow, Uttar Pradesh. He is currently working as an Assistant Professor with the Department of Electrical Engineering, Institute of Engineering and Technology, Dr. Rammanohar Lohia Avadh University, Ayodhya, India. He has published several papers in reputed international journals and various national/international conferences. His current research interests include the area of short-term load and price forecasting, AI applications in power systems, computational intelligence, and renewable energy.

SAURABH MANI TRIPATHI received the B.Tech. degree in electrical and electronics engineering from the G. L. A. Institute of Technology and Management, Mathura, India, the M.Tech. degree in electrical engineering with specialization in power electronics and drives from the M. M. M. Engineering College, Gorakhpur, India, and the Ph.D. degree in electrical engineering from Dr. A. P. J. Abdul Kalam Technical University, Lucknow, India. He is currently working as an Associate Professor in electrical engineering with the Kamla Nehru Institute of Technology, Sultanpur, India. He was a recipient of the prestigious "IEI Young Engineers Award (2018–2019)" presented by The Institution of Engineers, India. He was also a recipient of the "Best Teachers Award 2020" presented by Dr. A. P. J. Abdul Kalam Technical University. He achieved HIL Specialist Certification from Typhoon HIL Inc., Switzerland. At the age of 23 years, he has authored his first book titled *A Course in Modern Control Systems* (Laxmi Publications, New Delhi). Subsequently, he has authored two other books in the domain of electrical engineering namely *Analysis of Basic Systems* (Laxmi Publications, New Delhi) and *Modern Control Systems: An Introduction* (Jones & Bartlett Learning, USA). He served as an Editor for the book titled *Energy Conversion Systems: An Overview* (Nova Science Publisher, USA). He has published several research papers in reputed international journals and various national/international conferences. He is serving as a reviewer for various research journals of international repute. His research interests include renewable energy systems, electrical drives, real-time simulation, and hardware-in-the-loop testing. He is also an Advisory Board Member (Engineering) of Heliyon (A *Cell Press* journal). He is a member of several professional societies, such as Institution of Engineers (India), International Association of Engineers, and World Academy of Science, Engineering and Technology. He is a Life Member of the Indian Society for Technical Education.

SACHIN KUMAR (Senior Member, IEEE) was born in Dehradun, Uttarakhand, India. He received the B.Tech. degree in electrical engineering from Harcourt Butler Technological University (formerly Harcourt Butler Technological Institute), Kanpur, Uttar Pradesh, India, in 2008, the M.Tech. degree with a specialization in power system engineering from the Indian Institute of Technology Kharagpur, West Bengal, India, in 2010, and the Ph.D. degree in electrical engineering from the Indian Institute of Technology (BHU), Varanasi, Uttar Pradesh, in 2021. He is currently working as an Assistant Professor with the Department of Electrical Engineering, Govind Ballabh Pant Institute of Engineering and Technology, Ghurdauri, Pauri-Garhwal, Uttarakhand. He has published over 20 research articles in the indexed international journals and book chapters in prestigious handbooks. He is a Reviewer of IEEE Access, *IET Renewable Power Generation*, *International Transactions on Electrical Energy Systems* (Wiley) and River Publishers, Denmark. His research interests include reliability engineering, power system reliability, renewable energy systems, and green energy conversion systems.

RAJVIKRAM MADURAI ELAVARASAN received the B.E. degree in electrical and electronics engineering from Anna University, Chennai, India, and the M.E. degree in power system engineering from the Thiagarajar College of Engineering, Madurai, India. He held an associate technical operations position with the IBM Global Technology Services Division. He was as an Assistant Professor with the Department of Electrical and Electronics, Sri Venkateswara College of Engineering, Sriperumbudur, Chennai. He was also a Visiting Scholar with the Clean and Resilient Energy Systems (CARES) Laboratory, Texas A&M University, Galveston, TX, USA. He is currently holding various positions in academia and industry, which are as follows, such as an Adjunct Faculty with the Department of Electrical and Electronics Engineering, Thiagarajar College of Engineering, and a Subject Matter Expert with the Research and Development Unit (Power and Energy), Nestlives Pvt Ltd., Chennai. He has a Google citations of more than 1000. His research interests include renewable energy and smart grids, demand side management, multi criteria decision analysis, thermal energy storage, cyber-physical power systems, microgrids, power systems operation and control, and artificial intelligence-based control techniques. He was the Department Topper and a Gold Medalist in his master's degree. He has been a reviewer of more than 40 reputed SCIE journals.

SIVASANKAR GANGATHARAN received the B.E. degree in electrical & electronics engineering from Madurai Kamaraj University, Madurai, India, in 2002, the M.E. degree (Hons.) in power electronics and drives from Anna University, Chennai, India, in 2006, and the Ph.D. degree from Anna University. He was the Best Outgoing Student with the Department of Electrical and Electronics Engineering. From 2006 to 2008, he was an Assistant Professor with VIT University. Since 2008, he has been with the Faculty of Electrical and Electronics Engineering, Thiagarajar College of Engineering, Madurai, where he is currently working as an Assistant Professor. He has published papers in international journals, and international and national conferences. His research interests include distributed generation, power quality, and power electronics and drives. He is also a member of the IEEE Society.

DINESH KUMAR received the B.Tech. degree in electrical engineering from Uttar Pradesh Technical University, Lucknow, Uttar Pradesh, India, and the M.Tech. degree in power electronics and drives from the Kamla Nehru Institute of Technology, Sultanpur, Uttar Pradesh. He is currently working as an Assistant Professor with the Department of Electrical Engineering, Institute of Engineering and Technology, Dr. Rammanohar Lohia Avadh University, Ayodhya. His current research interests include power electronics, renewable energy, and control systems.

LUCIAN MIHET-POPA (Senior Member, IEEE) was born in 1969. He received the bachelor's degree in electrical engineering, the master's degree in electric drives and power electronics, and the Ph.D. and Habilitation degrees in electrical engineering from the Politehnica University of Timisoara, Romania, in 1999, 2000, 2002, and 2015, respectively. Since 2016, he has been working as a Full Professor in smart energy technology with the Østfold University College, Norway. From 1999 to 2016, he was with the Politehnica University of Timisoara. He has also worked as a Research Scientist with the Technical University of Denmark, from 2011 to 2014, and also with Aalborg University, Denmark, from 2000 to 2002. He was a Postdoctoral Researcher with the University of Siegen, Germany, in 2004. He is also the Head of the Research Laboratory ''Intelligent Control of Energy Conversion and Storage Systems'' and is one of the Coordinator of the master's program in ''Green Energy Technology'' with the Faculty of Information Technology, Engineering and Economics. He has published more than 180 papers in national and international journals and conference proceedings, and 15 books. He has been granted more than 20 international grants/projects, such as FP7, EEA, and Horizon 2020. He has been awarded more than ten national research grants. His research interests include modeling, simulation, control, and testing of energy conversion systems, distributed energy resources (DER) components and systems, distributed battery storage systems-DBSS, energy efficiency in smart buildings, and smart grids. He has served as a scientific and technical program committee member for many IEEE conferences. He was invited to join the energy and automotive committees as an Expert by the President and the Honorary President of Atomium European Institute, working in close cooperation with—and under the umbrella—of the EC and EU Parliament, and was Chairing Energy Section of AI4People 2020. Since 2017, he has been a Guest Editor for six Special Issues of the *Energies* and *Applied Sciences* (MDPI), *Maljesi*, and the *Advances in Meteorology* Journal.