ORIGINAL ARTICLE

Time Series Analysis Using Different Forecast Methods and Case Fatality Rate for Covid-19 Pandemic

Atanu Bhattacharjee[1,2] | Gajendra K. Vishwakarma[3] (ORCID) |
Namrata Gajare[1] | Neha Singh[3]

[1]Section of Biostatistics, Centre for Cancer Epidemiology, Tata Memorial Center, Navi Mumbai, India

[2]Homi Bhabha National Institute, Mumbai, India

[3]Department of Mathematics & Computing, Indian Institute of Technology Dhanbad, Dhanbad, India

Correspondence
Gajendra K. Vishwakarma, Department of Mathematics & Computing, Indian Institute of Technology Dhanbad, Dhanbad-826004, India.
Email: vishwagk@rediffmail.com

Abstract

This study presents forecasting methods using time series analysis for confirmed cases, the number of deaths and recovery cases, and individual vaccination status in different states of India. It aims to forecast the confirmed cases and mortality rate and develop an artificial intelligence method and different statistical methodologies that can help predict the future of Covid-19 cases. Various forecasting methods in time series analysis such as ARIMA, Holt's trend, naive, simple exponential smoothing, TBATS, and MAPE are extended for the study. It also involved the case fatality rate for the number of deaths and confirmed cases for respective states in India. This study includes the forecast values for the number of positive cases, cured patients, mortality rate, and case fatality rate for Covid-19 cases. Among all forecast methods involved in this study, the naive and simple exponential smoothing method shows an increased number of positive instances and cured patients.

KEYWORDS
case fatality rateARIMA, forecasting, Holt, MAPE, simple exponential smoothing, TBATS, time series analysis

JEL CLASSIFICATION
C32, C35, I81

1 | INTRODUCTION

An epidemic is a disease occurrence surpassing that expected in a community or a region. When an epidemic affects a large population or occurs worldwide, it is called a pandemic (Park, 2007). Over the centuries, the world has been facing menaces in the form of pandemics. Shinde et al. (2020) stated that the repercussions or after-effects have always had an enormous impact globally. Epidemics and pandemics substantially increase mortality and morbidity, and cause remarkable social, political, and economic disruption (Madhav et al., 2018). Despite the lack of literature, it can be assumed that the likelihood of these conditions has increased over time because of increased urbanization, global travel, and greater exploitation of the natural environment. In human history, according to the World Health Organization (WHO), the worst registered epidemics are smallpox (1896–1980) (Fenner et al., 1988), Spanish flu (1918–1919), malaria (since 1880), cholera (1817–1824), and acquired immunodeficiency syndrome (AIDS) (since 1981). In recent years, a few large-scale outbreaks have occurred: severe acute respiratory syndrome (SARS), H1N1 influenza, Ebola virus epidemic, and H5N1 influenza (Gostin et al., 2016).

Covid-19, caused by the coronavirus SARS-CoV-2, began in late December 2019 and is spreaded devastatingly worldwide (Li et al., 2020). It has substantially spread among 221 countries and infected more than 166 million people, with more than 3 million deaths to date. Jin et al. (2020) concluded that Covid-19 is a virus-induced pneumonia on the basis of blood tests, clinical diagnosis, and chest radiology. According to the Centers for Disease Control and Prevention (CDC), Covid-19 is spread from human to human via air or by touching surfaces or things harboring virus particles. The incubation period is a minimum of 14 days (Wang et al., 2020), and it affects this incubation period. Recent studies have indicated that the incubation period and median age of suspected cases is 3 days and 47 years, respectively (Guan et al., 2020).

Almost 30% of nations had no response plans for the spread of this pandemic (Nilima et al., 2021). Some countries took necessary measures for prevention, and specific programs were implemented, including sanitation, social distancing, mask-wearing, and proper hygienic conditions. These measures are still in force to halt the progression of this spread (Flaxman et al., 2020; Lai et al., 2020). However, owing to the high mortality rate, it was widely accepted that an effective and safe vaccine was required (Sultana et al., 2020). On the other hand, false-positive diagnoses may result in unnecessary wastage of resources, such as testing kits, hospital beds, and protective gear, and increased pressure on healthcare workers (Burstyn et al., 2020).

A vaccine is defined as a biological preparation that provides acquired immunity to a specific infectious disease (Guimarães et al., 2015). According to the Ministry of Health and Family Welfare, the Covid-19 vaccine in India was launched on January 16, 2021. As of February 18, 2021, seven different vaccines were rolled out globally across three platforms, as stated by the WHO. More than 186 million doses of vaccine have been administered in India, and more than 41 million people are now fully vaccinated (i.e., 3% of the total population). Even after a speedy immunization program against Covid-19, there is a significant variation in the pattern of this epidemic (Gecili et al., 2021). According to the credible report by WHO as well as public health reports, the suspected cases have increased globally. An increase in the number of recoveries indicates that the situation is stabilizing globally, and hence, modeling or forecasting the measures of Covid-19 is vital (Zarrin et al., 2019).

This study considers the frequency of cumulative confirmed cases, recovered cases, and deaths related to Covid-19 in India. To assess the progression of Covid-19, we propose a few models to forecast the frequency of Covid-19 cases. Predictive statistical models play a significant role in understanding the pattern of disease. Forecasts are likely to be beneficial because they help us identify the best- and worst-case scenarios. Forecasting the number of cases and deaths of Covid-19 can help health authorities, and can aid in predicting virus transmission. It can also facilitate the development of appropriate preventive methods and strategies to halt the spread of the virus beforehand. It is necessary to develop specific techniques that will help us deal with this pandemic. Many researchers have used mathematical and statistical modeling to study pandemics. The early development of the epidemic follows an exponential growth in cases. It is confirmed that recent declines in cases are likely due to an under-ascertainment of cases with recent onset and delayed identification and reporting, rather than a true turning point in incidence (Li et al., 2020).

The need for transparency in reporting testing policies, with clear reporting of the denominators used to calculate case-fatality rates, is clearly expressed by Onder et al. (2020). It is also mentioned that age, sex, and comorbid

clinical status of affected persons differ when comparing COVID-19 cases and mortality rates between different countries and regions. Forecasting the outcomes of a pandemic is a challenging task with massive potential value to decisionmakers and policymakers. However, they adopted a simple time series model that shows a good level of accuracy and uncertainty while more data are integrated. These forecasts are useful to explore disease burden in the future and help policymakers (Petropoulos et al., 2022). Santosh (2020) expressed and well documented the importance of artificial intelligence (AI)-driven tools to control coronavirus outbreaks.

Owing to the advances in analytics, a new technique has been developed to study geographical data using spatial data representation. Various applications like Corona, Madrid, Data wrapper, etc. are created using spatial modeling to manage the consequences more efficiently to collect and analyze the data with regard to location.

Similarly, another approach to this pandemic is social distancing. Authorities are using social distancing strategies to reduce the transmission of the virus. Spatial analysis is quite helpful in determining areas at risk from overcrowding to ensure efficient rehabilitation of those suspected individuals. In addition, spatial analysis contributes to effective data visualization. Maps help as an intuitive way to guide citizens in relation to Covid-19. Variables regarding the pandemic situation are largely unknown, and very few are related to spatial dimensions, which relate to the geographical phenomenon. The study of Covid-19 includes: (i) various cross variables that resemble spatial analysis, (ii) decision making on geographical conditions, and (iii) predictive modeling.

The focus can be kept on adverse medical conditions, risk of medical formalities, or high healthcare costs. Hence, geospatial modeling has become a vital statistical tool for this global pandemic.

The present study reveals many efforts to address Covid-19 through big data and artificial intelligence. Several other disease outbreaks have been studied using AI, which plays a vital role in overcoming Covid-19. This technique is being used to identify particular diseases in the form of clusters, to monitor suspected cases, mortality rate, disease trends, and much more. It is hoped that AI can be applied in the fight against Covid-19 (Rahmatizadeh et al., 2020). AI can be engaged in forecasting the spread of the virus and providing information about the more vulnerable sites, helping to predict morbidity and mortality. AI can also be used to predict Covid-19 recovery or mortality rate, provide daily updates, storage, and trend analysis, and chart vaccination within regions.

2 | METHODOLOGY

2.1 | Data used

Owing to the high peak in the number of Covid-19 cases in India, all its states were chosen as the target population for this study. In essence, three datasets, collected from the official website, were examined in this study:

1. Coviddata: this dataset comprises cases that tested positive and negative for Covid-19 in different states of India for a duration of 453 days (April 1, 2020, to June 27, 2021).
2. cdata: this dataset contains the number of deaths from and confirmed cases of Covid-19 in all states of India from January 16, 2021, to July 7, 2021 (173 days).
3. c19data: this dataset shows the vaccination status of individuals along with the number of deaths from and confirmed cases of Covid-19 from January 16, 2021, to July 7, 2021 (173 days).

2.2 | Developed R package

R packages are an ideal way to package and distribute R code and data for reuse by others. They contain a variety of resources, including R code, datasets, and documentation functions, as well as other supporting files (Cryer & Chan, 2008). Since the pandemic has been ongoing for over a year now, there is an urge to make predictions

regarding the situation and to study it with newly emerging Covid-19 data. Thus, it is vital to develop an R package that will help researchers to analyze Covid-19 datasets.

A package was developed in R studio to analyze the Covid-19 dataset named 'CovidTSA'. This package has been built up with multiple functions, namely 'tsoutput', 'cfrfn', 'mapefn', and 'forebayesjags'. These functions provide forecast value using various methods such as Arima, HoltT, naive, simple exponential smoothing, Tbats, and multi-layer perceptron. In addition, it provides accurate error values and case fatality rate. This package can be downloaded and installed in R Studio and help to predict the parameters regarding Covid-19.

2.3 | Forecasting techniques used

In this study, forecasting using time series analysis is used, which involved methods including ARIMA, Holt's trend method, naive, TBATS, simple exponential smoothing, and neural network multilayer perceptron (Montgomery et al., 2015). Case fatality rate for the number of deaths and prediction accuracy error MAPE was also studied.

A time series is a series of data points measured at consistent time intervals. This time interval may be hourly, daily, weekly, monthly, quarterly, yearly, etc. In a time series, each data point depends on the previous data points. Time-series patterns may be classified as follows: (1) trend, a long-term increase or decrease in the data; (2) seasonal, when the time series is of a fixed and known frequency; (3) cyclic, when the time series is not of a fixed frequency.

Forecasting is defined as making predictions about the future on the basis of past and present data and analyzing trends. Time-series forecasting uses a model to predict future values on the basis of previously observed values (Box et al., 2015).

An auto-regressive integrated moving average (ARIMA) model is a type of statistical model for analyzing and forecasting time-series data. It explicitly caters to a suite of standard structures in time-series data and, as such, provides a simple yet powerful method for making skilful time-series forecasts (Ord, 2004). ARIMA corresponds to a linear regression equation written as

$$y_t' = c + \emptyset_1 \ y_{t-1}' + \dots + \emptyset_p \ y_{t-p}' + \theta_1 \in_{t-1} + \dots + \theta_q \in_{t-q} + \in_t, \tag{1}$$

where y_t' is the differenced series. The predictors include both lagged values of y_t' and lagged errors. It depends on the following parameters:

i. p (lag order): number of lag observations included in the model
ii. d (degree of differencing): number of times the raw observations are differenced
iii. q (order of moving average): the size of the moving average window

ARIMA can be used in cases where the data are stationary, do not contain any anomaly, and are univariate. Cases of nonstationarity can be eliminated by applying the difference once or twice.

Holt (1957) extended simple exponential smoothing (SES) to allow the forecasting of data with a trend. This is also known as linear exponential smoothing. It is a popular smoothing model with three separate equations that work together to generate a final forecast. The first is a basic smoothing equation that directly adjusts the last smoothed value to the last period's trend. The trend itself is updated over time through the second equation, where the trend is expressed as the difference between the last two smoothed values. Finally, the third equation is used to generate the final forecast. This method is also called double exponential smoothing or trend-enhanced exponential smoothing. This method involves a forecast equation and two smoothing equations (one for the level and one for the trend):

Forecast equation:

For level

$$l_t = \alpha v_t + (1 - \alpha)(l_{t-1} + b_{t-1}) . \tag{2}$$

For trend

$$b_t = \beta l_{t-1} + (1 - \beta)b_{t-1} . \tag{3}$$

For seasonal

$$s_t = \gamma s_{t-1} + (1 - \gamma)(y_t/l_t) , \tag{4}$$

where s is length of the seasonal period, for $0 \le \alpha \le 1$, $0 \le \beta \le 1$ and $0 \le \gamma \le 1$, l_t is an estimate of the level of the series at the time t; b_t is an estimate of the trend (slope) of the time series at the time t; α is a smoothing parameter for the level, $0 \le \alpha \le 1$; and β is a smoothing parameter for the trend, $0 \le \beta \le 1$.

As with simple exponential smoothing (SES), the level equation here shows that l_t is a weighted average of observation y_t and the one-step-ahead training forecast for time t, here given by $(l_{t-1} + b_{t-1})$. The trend equation shows that b_t is the weighted average of the estimated trend at time t based on $(l_t - l_{t-1})$, the previous estimate of the trend.

The forecast function is no longer flat but trending. The h-step-ahead forecast is equal to the last estimated level plus h times the last estimated trend value. Hence, the forecasts are a linear function of h.

Naive forecasting is the technique in which the last period's sales are used for the next period's forecast without prediction or adjusting factors. Forecasts produced using a naive approach are equal to the final observed value. This method works quite well for economic and financial time series, which often have patterns that are difficult to predict with reliability and accuracy. If the time series is believed to have seasonality, the seasonal naive approach may be more appropriate where the forecasts are equal to the values from the last season. In this approach, the predictions of all future values are equal to the mean of the past data. This approach can be used with any sort of data where past data are available. In time-series notation:

$$\widehat{y}_{T+1} = (y_1 + y_2 + ... + y_T)/T , \tag{5}$$

where $(y_1, y_2, ..., y_T)$ is past data.

Although the time-series notation has been used here, the average approach can also be used for cross-sectional data (when we are predicting unobserved values; values that are not included in the dataset). Then, the prediction for unobserved values is the average of the observed values.

De Livera et al. (2011) developed an alternative approach by using a combination of Fourier terms with an exponential smoothing state-space model and a Box–Cox transformation, in a completely automated manner. TBATS is a forecasting method to model time-series data. The main aim of this method is to forecast time series with complex seasonal patterns using exponential smoothing. This forecasting method is capable of modelling time series with multiple seasonalities. It constitutes the following elements:

T: trigonometric terms for seasonality,
B: Box–Cox transformations for heterogeneity,
A: ARIMA errors for short-term dynamics,
T: trend,
S: seasonal (including multiple and non-integer periods)

The TBATS model can deal with complex seasonalities (e.g., non-integer seasonality, non-nested seasonality, and large-period seasonality) with no seasonality constraints, making it possible to create detailed, long-term forecasts.

A multilayer perceptron (MLP), first introduced by Rosenblatt (1958), is a feed-forward artificial neural network (ANN) model, consisting of multiple layers. As aforementioned, MLP is a feed-forward neural network, which means that data are transmitted from the input layer to the output layer in the forward direction. Connections between layers are assigned weights. The weight of a connection specifies its importance. This concept is the backbone of an MLP's learning process.

Simple exponential smoothing (SES) is a time-series forecasting method for univariate data without trend or seasonality. It requires a single parameter, called alpha (α), the smoothing factor or smoothing coefficient. It is a powerful forecasting method that may be an alternative to the famous Box–Jenkins ARIMA family of methods. This parameter controls the rate at which the influence of the observations at prior time steps decays exponentially. Alpha is often set to a value between 0 and 1. Large values mean that the model pays attention mainly to the most recent past observations. In contrast, smaller values mean more of the history is taken into account when making a prediction. SES works on weighted averages, i.e., the average of the previous level and the current observation. The enormous weights are associated with the recent observations, and the most negligible consequences are the oldest observations.

The mean absolute percentage error (MAPE), also known as mean absolute percentage deviation (MAPD), measures the prediction accuracy of a forecasting method in statistics. It usually expresses the accuracy as a ratio defined by the formula:

$$MAPE = \frac{100}{N} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|, \tag{6}$$

where A_t is the actual value and F_t is the forecast value. Their difference is divided by the actual value (A_t). The absolute value of this ratio is summed for every forecasted point in time and divided by the number of fitted points (n).

Case fatality rate (CFR)—sometimes called case fatality risk or case fatality ratio—is the proportion of deaths from a particular disease compared with the total number of people diagnosed with the disease for a particular period. CFR is conventionally expressed as a percentage and represents a measure of disease severity. CFR is most often used for diseases with discrete, limited time courses, such as outbreaks of acute infections. A CFR can only be considered final when all cases have been resolved (either died or recovered). The case fatality rate is typically used as a measure of disease severity. It is often used for prognosis (predicting disease course or outcome), where comparatively high rates are indicative of relatively poor results. It also can be used to evaluate the effect of new treatments, with measures decreasing as treatments improve. Case fatality rates are not constant; they can vary between populations and over time, depending on the interplay between the causative agent of a disease, the host, and the environment, available treatments, and the quality of patient care. Case fatality rate is calculated using the following formula:

$$CRF = \frac{\text{Number of deaths}}{\text{Confirmed cases}} \times 100. \tag{7}$$

3 | RESULTS

A time series is said to be stationary if its statistical properties do not vary with time. Stationarity tests allow verifying whether a series is fixed or not. We need to implement the augmented Dickey–Fuller (ADF) test to check stationarity. If the p-value is less than the significance level, then we can say that data are stationary. We imposed the ADF test on variables, including the Covid-19 data. Below is the result of the ADF test for the number of deaths due to Covid-19:

To test H0: There is no stationarity in the data. Against H1: There is stationarity.

Then ADF statistic is -19, and $p = 0.01$.

As the *p*-value is lower than the level of significance, we reject the null hypothesis.

3.1 | Methods of forecasts

In this study, various methods of forecast, using time series analysis, are operated using the 'CovidTSA' package in RStudio. Here we obtained the forecast values for the specified days for any particular state using the function 'tsoutput' forecast methods, ARIMA, Holt's trend, naive, simple exponential smoothing, TBATS, and neutral network as multilayer perceptron (MLP). These figures give the forecast values of positive cases for the next 25 days within the state of 'Manipur' using various forecast methods. We can see that the forecast values using the naive method and simple exponential smoothing method are mostly similar. From the figures below, we can predict that the number of Covid-19 positive cases will increase in the upcoming days in the state of Manipur on December 15, 2021.

3.2 | Prediction accuracy error

Every method deserves to be presented with a prediction error. This allows the fluctuation of the method's estimate from the actual values. It is required to submit the prediction errors for each proposed method and compare them. Table 2 below shows us the mean absolute error obtained using various forecast methods. These errors using different forecast methods are obtained using the function 'mapefn' developed using the R package.

3.3 | Case fatality rate

This study evaluated data showing the number of deaths and confirmed cases due to Covid-19 for a particular period (January 16, 2021, to July 7, 2021). Here, we calculated CFR in one million population by using the function 'cfrfn' developed in the R package for Andhra Pradesh with a lag of 7 days and the upper and lower confidence limit. The
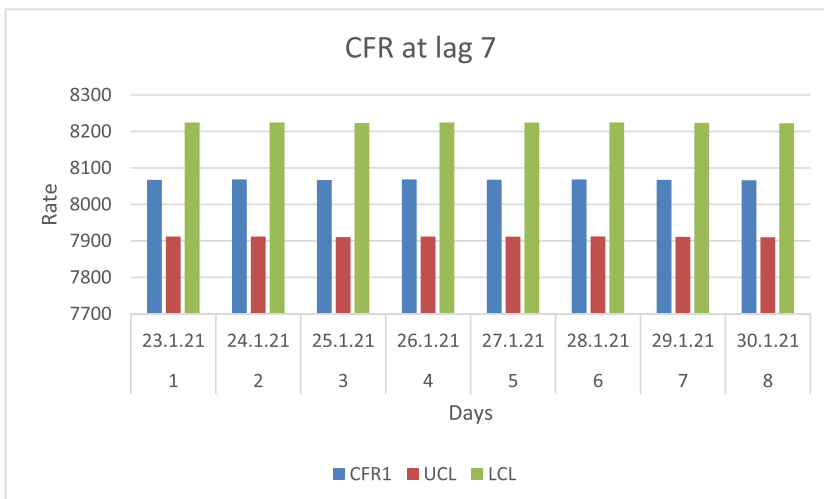

FIGURE 1 CFR for the state of 'Andhra Pradesh'

definition of the case fatality rate is presented in Equation (7). Figure 1 shows that the CFR is almost constant for the lag of 7 days. Even the confidence interval remains more or less the same.

4 | INTERPRETATION

As far as prediction is concerned, a forecast using time series analysis can be helpful. Figures 2 and 3 indicate the autocorrelation and partial autocorrelation (for the positive cases), which shows that the data are stationary. The forecast uses various methods, including ARIMA, Holt's trend, naive, TBATS, simple exponential smoothing, and multilayer perceptron (MLP). The forecast values for the next 25 days using the methods mentioned above are presented in Table 1 for the state of Manipur. This indicates that the forecast values using the naive method and SES are the same. We can infer that the number of positive Covid-19 cases is reaching a high peak (Figure 4).
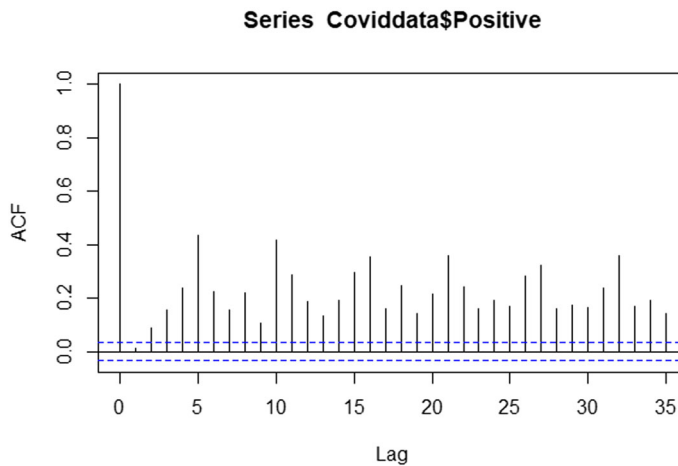

FIGURE 2 Autocorrelation showing stationarity for Covid-19-positive cases
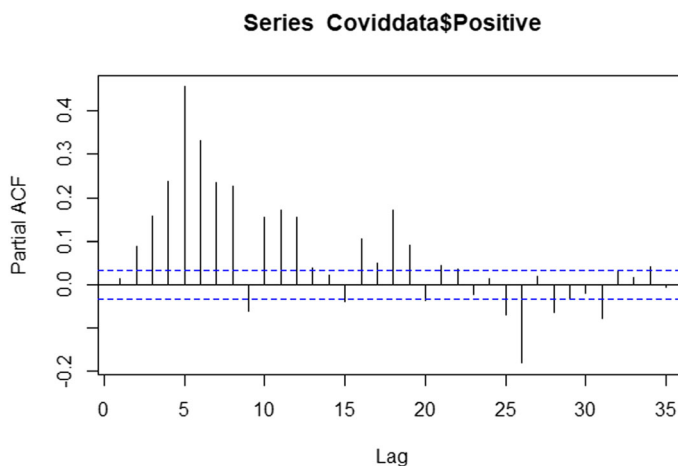

FIGURE 3 Partial autocorrelation showing stationarity for Covid-19-positive cases

TABLE 1 Different methods of the forecast for positive cases for the next 25 days in the state of 'Manipur'

	Arima	HoltT	Naive	SES	Tbats	Mlpnnet
1	1620863	1620855	1609516	1609515	1619760	1622979
2	1632210	1632194	1609516	1609515	1630830	1635487
3	1643557	1643533	1609516	1609515	1641933	1648012
4	1654904	1654872	1609516	1609515	1653068	1659697
5	1666251	1666211	1609516	1609515	1664236	1670994
6	1677598	1677550	1609516	1609515	1675435	1681733
7	1688945	1688889	1609516	1609515	1686667	1692037
8	1700292	1700228	1609516	1609515	1697932	1702768
9	1711639	1711567	1609516	1609515	1709228	1713659
10	1722987	1722906	1609516	1609515	1720556	1724042
11	1734334	1734245	1609516	1609515	1731917	1734878
12	1745681	1745584	1609516	1609515	1743310	1745044
13	1757028	1756923	1609516	1609515	1754735	1756013
14	1768375	1768262	1609516	1609515	1766192	1767044
15	1779722	1779601	1609516	1609515	1777680	1777419
16	1791069	1790940	1609516	1609515	1789201	1788151
17	1802416	1802279	1609516	1609515	1800754	1798843
18	1813763	1813618	1609516	1609515	1812339	1809180
19	1825110	1824957	1609516	1609515	1823955	1819310
20	1836457	1836296	1609516	1609515	1835604	1830044
21	1847804	1847635	1609516	1609515	1847284	1840846
22	1859151	1858974	1609516	1609515	1858996	1851645
23	1870498	1870313	1609516	1609515	1870740	1862438
24	1881845	1881652	1609516	1609515	1882516	1872738
25	1893192	1892991	1609516	1609515	1894323	1882842
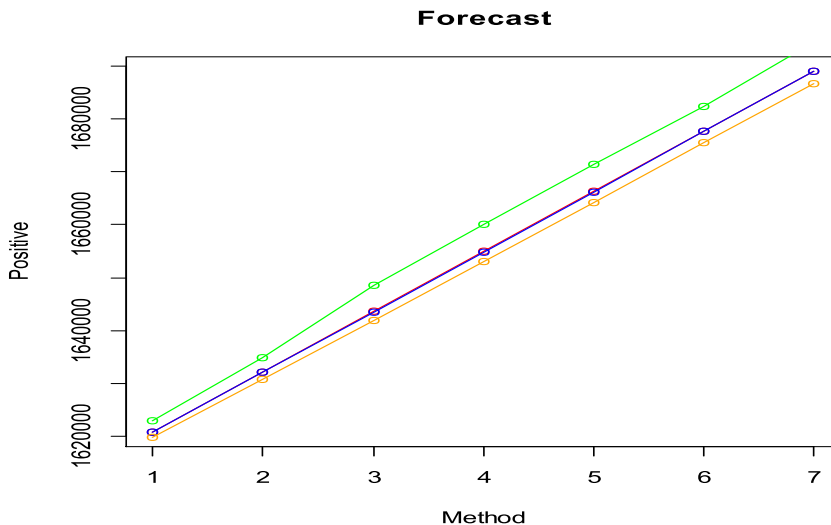
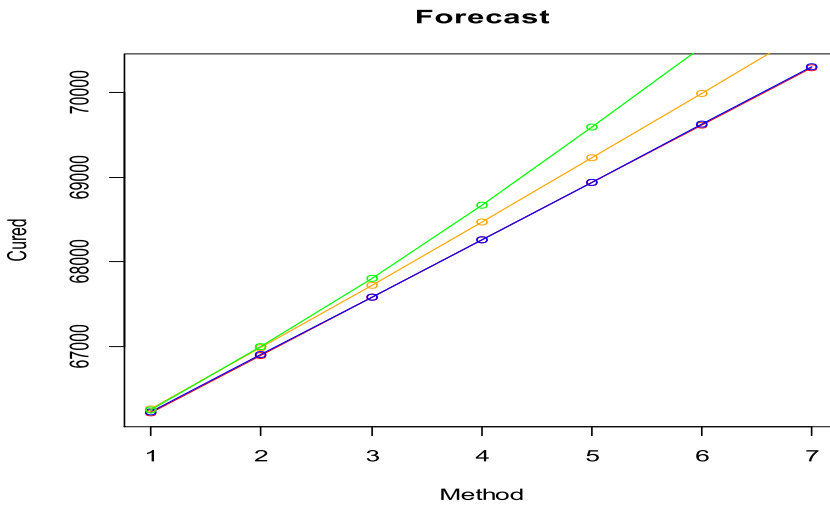**FIGURE 4** Forecast for the next 7 days (Covid-19-positive cases)

FIGURE 5 Forecast for the next 7 days (cured Covid-19 cases)
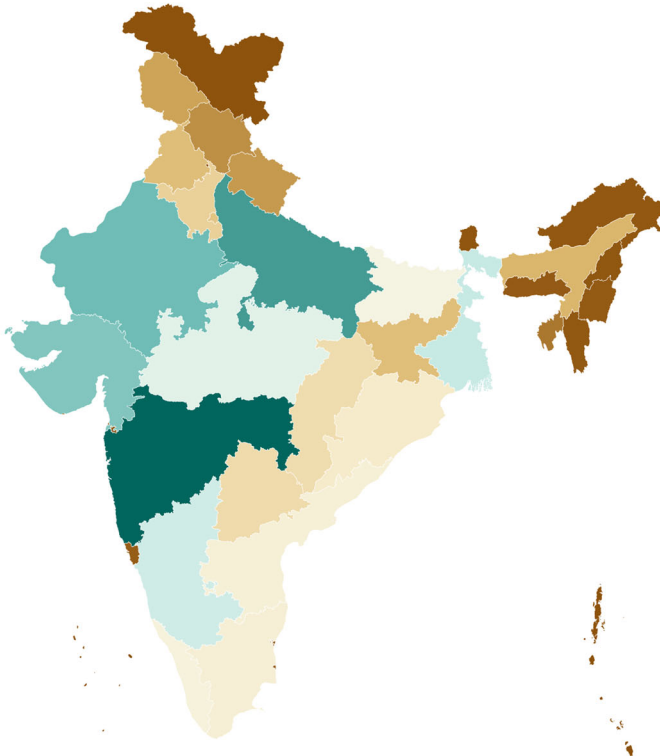
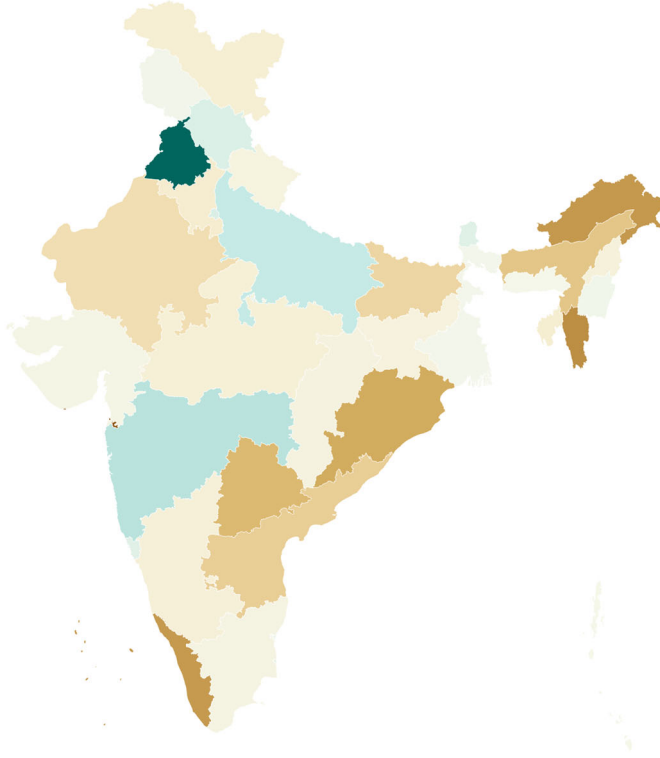**FIGURE 6** Vaccination status of India

FIGURE 7 Case fatality rate of India for Covid-19

TABLE 2 Prediction accuracy by MAPE, uMAPE, and vMAPE

A	MAPE	uMAPE	vMAPE
Arima	1.2136756	1.151604	0.06685406
HoltT	0.9249319	0.875306	0.05135845
Naive	5.1440383	4.950895	0.26923399
SES	5.1440794	4.950935	0.26923600
Tbat	1.5542915	1.471946	0.08613727
Mlpnnet	2.7783078	2.650826	0.15030495

On the other hand, the forecast graph during the second wave (when vaccination was in progress) (Figure 5) shows us that the number of cured cases is also rising. In other words, the vital advantage is that, even though the number of positive cases is increasing, so is the number of cured cases. This implies the recovery rate is higher.

The nationwide total vaccination coverage and case fatality rates are shown in Figures 6 and 7. They show the overall vaccination coverage, indicating that states with high vaccination coverage are also affected by a high of case fatality rate. Because of the rapid spread of the virus, case fatality was the major concern. However, the

TABLE 3 Case fatality rate at lag 7 along with UCL and LCL

Cfr value at lag 7	Date	cfr1	UCL	LCL
1	23/01/2021	8068.104	7911.750	8224.457
2	24/01/2021	8068.194	7911.850	8224.539
3	25/01/2021	8066.728	7910.412	8223.044
4	26/01/2021	8068.248	7911.924	8224.571
5	27/01/2021	8067.747	7911.443	8224.050
6	28/01/2021	8068.428	7912.134	8224.723
7	29/01/2021	8067.163	7910.893	8223.433
8	30/01/2021	8065.917	7909.671	8222.163

government's emphasis on controlling the infection through fast vaccination coverage in those states shows the high vaccination coverage.

According to several surveys reviewed by Gneiting (2011), MAPE is the most commonly used measure for assessing forecasts in organizations. Table 2 presents the prediction error calculated for various methods of the forecast. Table 3 presents CFR for the population (in one million) of a particular state in India with a lag of 7 days. Estimating the CFR in response to the Covid-19 pandemic is a high priority. Its interpretation should be made carefully as we are in the middle of an outbreak. Therefore, bias should be considered before making conclusions from the observed data.

5 | DISCUSSION

The Covid-19 global pandemic has had many negative impacts in India (Pringle et al., 2020). It is essential to study this pandemic, as it has major ramifications for the country's social, economic, and political status and has caused widespread disruption in the Indian economy. According to the Ministry of Statistics, India's growth in the year 2020 was reduced by 3.1% (Das & Patnaik, 2020). The pandemic has resulted in high costs to the economy and challenges with regard to meeting economic goals, employment, and work efficiency.

In this study, we concentrated on different states of India since a large number of Covid-19 victims are noted each day. The period covered by this study spanned from the start of Covid-19 (first wave) to the recession of the second wave.

Here we evaluated various methods of forecast to assess whether there is a rise in positive cases, number of patients cured, and mortality rate. In addition, we studied case fatality rate as well as individual vaccination status for all states in India. Since this situation has been ongoing for about a year now, these forecast values will help us to make predictions regarding Covid-19.

The following question arises: Can pandemic forecasts be improved with future research? Improvements are feasible by introducing new methods that combine estimates from various sources, such as epidemiological models, time-series modeling, etc. Thus, these models can be enhanced by studying aggregated levels of data at the country and local levels, in an age-specific manner using hierarchical structures.

6 | CONCLUSION

Forecasting the pandemic is a very complex task. Many prediction problems involve a time component. These problems are left aside since the time component makes it difficult to make predictions. Time-series analysis should

introduce time-to-time changes in the data, and forecasting helps to derive future strategies for decision making. The various forecast methods utilized in this study will be helpful for further predictions. Such forecasts are used to monitor the progression of the pandemic and help policymakers to decide the plan of action by decreasing the impact, and to put forward measures that relax the mitigating interventions. Mistakes may be made in the way Covid-19 is handled, but there is much to learn from experience, which will help humanity prepare for future pandemics.

ORCID

Gajendra K. Vishwakarma https://orcid.org/0000-0002-2804-4334

REFERENCES

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control*. John Wiley & Sons.

Burstyn, I., Goldstein, N. D., & Gustafson, P. (2020). Towards reduction in bias in epidemic curves due to outcome misclassification through Bayesian analysis of time-series of laboratory test results: Case study of COVID-19 in Alberta, Canada and Philadelphia, USA. *BMC Med Res Methodol*, *20*(1), 1–10. https://doi.org/10.1186/s12874-020-01037-4

Cryer, J. D., & Chan, K. S. (2008). *Time series analysis: with applications in R* (Vol. 2). Springer. https://doi.org/10.1007/978-0-387-75959-3

Das, K. K., & Patnaik, S. (2020). The impact of COVID-19. In Indian economy—an empirical study. *IJEET*, *10*(3), 194–202.

De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *J Am Stat Assoc*, *106*(496), 1513–1527. https://doi.org/10.1198/jasa.2011.tm09771

Fenner, F., Henderson, D. A., Arita, I., Jezek, Z., & Ladnyi, I. D. (1988). *Smallpox and its eradication* (Vol. 6, pp. 1–1421). World Health Organization.

Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., Whittaker, C., Zhu, H., Berah, T., Eaton, J. W., Monod, M., Imperial College COVID-19 Response Team, Ghani, A. C., Donnelly, C. A., Riley, S., Vollmer, M. A. C., Ferguson, N. M., Okell, L. C., & Bhatt, S. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, *584*(7820), 257–261. https://doi.org/10.1038/s41586-020-2405-7

Gecili, E., Ziady, A., & Szczesniak, R. D. (2021). Forecasting COVID-19 confirmed cases, deaths and recoveries: Revisiting established time series modeling through novel applications for the USA and Italy. *PLoS ONE*, *16*(1), e0244173. https://doi.org/10.1371/journal.pone.0244173

Gneiting, T. (2011). Making and evaluating point forecasts. *J Am Stat Assoc*, *106*(494), 746–762. https://doi.org/10.1198/jasa.2011.r10138

Gostin, L. O., Tomori, O., Wibulpolprasert, S., Jha, A. K., Frenk, J., Moon, S., Phumaphi, J., Piot, P., Stocking, B., Dzau, V. J., & Leung, G. M. (2016). Toward a common secure future: Four global commissions in the wake of Ebola. *PLoS Med*, *13*(5), e1002042. https://doi.org/10.1371/journal.pmed.1002042

Guan, W. J., Ni, Z. Y., Hu, Y., Liang, W. H., Ou, C. Q., He, J. X., Liu, L., Shan, H., Lei, C. L., Hui, D. S. C., du, B., Li, L. J., Zeng, G., Yuen, K. Y., Chen, R. C., Tang, C. L., Wang, T., Chen, P. Y., Xiang, J., ... China Medical Treatment Expert Group for Covid-19. (2020). Clinical characteristics of 2019 novel coronavirus infection in China. *N Engl J Med*, *382*, 1708–1720. https://doi.org/10.1056/NEJMoa2002032

Guimarães, L. E., Baker, B., Perricone, C., & Shoenfeld, Y. (2015). Vaccines, adjuvants and autoimmunity. *Pharmacol Res*, *100*, 190–209. https://doi.org/10.1016/j.phrs.2015.08.003

Holt, C. E. (1957). *Forecasting seasonals and trends by exponentially weighted averages (ONR memorandum no. 52)*. Carnegie Institute of Technology.

Jin, Y., Yang, H., Ji, W., Wu, W., Chen, S., Zhang, W., & Duan, G. (2020). Virology, epidemiology, pathogenesis, and control of COVID-19. *Viruses*, *12*(4), 372. https://doi.org/10.3390/v12040372

Lai, S., Ruktanonchai, N. W., Zhou, L., Prosper, O., Luo, W., Floyd, J. R., Wesolowski, A., Santillana, M., Zhang, C., du, X., Yu, H., & Tatem, A. J. (2020). Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature*, *585*(7825), 410–413. https://doi.org/10.1038/s41586-020-2293-x

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K. S. M., Lau, E. H. Y., Wong, J. Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., ... Feng, Z. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *N Engl J Med*, *382*, 1199–1207. https://doi.org/10.1056/NEJMoa2001316

Madhav, N., Oppenheim, B., Gallivan, M., Mulembakani, P., Rubin, E., & Wolfe, N. (2018). Pandemics: risks, impacts, and mitigation. Chapter 17. In *Disease control priorities: Improving health and reducing poverty* (3rd ed.) (pp. 315–345). Department of Global Health, University of Washington. https://doi.org/10.1596/978-1-4648-0527-1_ch17

Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). *Introduction to time series analysis and forecasting*. John Wiley & Sons.

Nilima, N., Kaushik, S., Tiwary, B., & Pandey, P. K. (2021). Psycho-social factors associated with the nationwide lockdown in India during COVID-19 pandemic. *Clin Epidemiology Glob Health*, *9*, 47–52. https://doi.org/10.1016/j.cegh.2020.06.010

Onder, G., Rezza, G., & Brusaferro, S. (2020). Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. *J Am Med Assoc*, *323*(18), 1775–1776. https://doi.org/10.1001/jama.2020.4683

Ord, K. (2004). Charles Holt's report on exponentially weighted moving averages: An introduction and appreciation. *Int J Forecast*, *20*(1), 1–3. https://doi.org/10.1016/j.ijforecast.2003.09.016

Park, K. (2007). *Park's textbook of preventive and social medicine India* (p. 741). Banarsidas Bhanot Publishers.

Petropoulos, F., Makridakis, S., & Stylianou, N. (2022). COVID-19: Forecasting confirmed cases and deaths with a simple time series model. *Int J Forecast*, *38*(2), 439–452. https://doi.org/10.1016/j.ijforecast.2020.11.010

Pringle, J. C., Leikauskas, J., Ransom-Kelley, S., Webster, B., Santos, S., Fox, H., Marcoux, S., Kelso, P., & Kwit, N. (2020). COVID-19 in a correctional facility employee following multiple brief exposures to persons with COVID-19—Vermont, July–august 2020. *Morb Mortal Wkly Rep*, *69*(43), 1569–1570. https://doi.org/10.15585/mmwr.mm6943e1

Rahmatizadeh, S., Valizadeh-Haghi, S., & Dabbagh, A. (2020). The role of artificial intelligence in management of critical COVID-19 patients. *J Cell Mol Anesth*, *5*(1), 16–22.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386–408. https://doi.org/10.1037/h0042519

Santosh, K. C. (2020). AI-driven tools for coronavirus outbreak: Need of active learning and cross-population train/test models on multitudinal/multimodal data. *J Med Syst*, *44*(5), 1–5. https://doi.org/10.1007/s10916-020-01562-1

Shinde, G. R., Kalamkar, A. B., Mahalle, P. N., Dey, N., Chaki, J., & Hassanien, A. E. (2020). Forecasting models for coronavirus disease (COVID-19): A survey of the state-of-the-art. *SN Comput Sci*, *1*(4), 197. https://doi.org/10.1007/s42979-020-00209-9

Sultana, J., Mazzaglia, G., Luxi, N., Cancellieri, A., Capuano, A., Ferrajolo, C., de Waure, C., Ferlazzo, G., & Trifirò, G. (2020). Potential effects of vaccinations on the prevention of COVID-19: Rationale, clinical evidence, risks, and public health considerations. *Expert Rev Vaccines*, *19*(10), 919–936. https://doi.org/10.1080/14760584.2020.1825951

Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., Wang, B., Xiang, H., Cheng, Z., Xiong, Y., Zhao, Y., Li, Y., Wang, X., & Peng, Z. (2020). Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan China. *J Am Med Assoc*, *323*(11), 1061–1069. https://doi.org/10.1001/jama.2020.1585

Zarrin, P., Maleki, M., Khodadai, Z., & Arellano-Valle, R. B. (2019). Time series models based on the unrestricted skew-normal process. *J Stat Comput Simul*, *89*(1), 38–51. https://doi.org/10.1080/00949655.2018.1533962