Trends in Health Informatics

www.thi.reapress.com

THI Vol. 1, No. 1 (2024) 9–22.

Paper Type: Original Article

Impact of Statistical Approach on Time-Series Models for Forecasting COVID-19

Shobhana Kashyap[1], Avtar Singh[1,*]

[1] Department of Computer Science and Engineering, Dr. B.R. Ambedkar NIT Jalandhar, Punjab, India, 144008; shobhanak.cs.19@nitj.ac.in; avtars@nitj.ac.in.

Citation:

Received: 5 August 2023 *Revised: 1 October 2023* *Accepted: 16 March 2024*	Kashyap, Sh., & Singh, A. (2024). Impact of statistical approach on time-series models for forecasting COVID-19. *Trends in health informatics, 1* (1), 9-22.

Abstract

For several decades, time-series forecasting has been an engaging research area. It is an essential domain of Machine Learning (ML) that is mainly ignored. It is necessary because prediction problems have a time feature, which makes time series problems more difficult to tackle. Forecasting of many applications such as weather, sales, ECG patterns and even COVID-19 spreads are possible with time series techniques. Inspired by these applications, many scholars have worked on effective forecasting techniques. This paper presents a comparative study of the time series models implemented on India's real-time data of COVID-19. The study aims to estimate the mortality rate of coming 10 days by the interpretation of actual data. Two predictive algorithms, Holt's Linear Exponential Smoothing (HLES) and Autoregressive Integrated Moving Average (ARIMA) have been applied. To accomplish the objective and check the model accuracy, two selection criterion methods, Root Mean Square Error (RMSE) and Akaike Information Criterion (AIC), have been used to calculate the lowest values. The results depict that the HLES model has generally outperformed ARIMA. Adding to this, HLES model has good accuracy in forecasting the mortality rate compared to ARIMA. Moreover, if we face similar circumstances again in the future, then the proposed algorithm can be used to prevent the earlier phase of the outbreak.

Keywords: Time series model, COVID-19, HLES, ARIMA, AIC, RMSE.

1 | Introduction

In 2019 Corona Virus Disease or Severe Acute Respiratory Syndrome (COVID-19 or SARS-CoV-2) was confirmed as a pandemic in China; countless cases were registered globally within a short period [1]. This disease is disrupting the lives of human beings all over the world. The first positive corona virus case was

Corresponding Author:

spot-ted, December 2019 in the Chinese city of Wuhan [2], [3]. On 11 March 2020 World Health Organization (WHO) [4], [5] confirmed SARS-CoV-2 as a pandemic.

The previous studies such as the spread of disease [4], the significant impact of illness [5], digital health solutions [6], mental health [7], solid waste management [8], and many more found that due to COVID-19, the world has been facing challenges. And India has become one of the most impacted countries [9], [10]. The death rate is rising day by day and getting a formidable form [11]. It can cause multiple organ failures and eventually death; if it is not discovered and treated early. Therefore, understanding the evolution of COVID-19 is essential for the prediction of infection and death rates [9], [12]. This prediction is vital for public health officials to allocate financial resources and implement appropriate containment strategies efficiently. Many sorts of research have taken interest in forecasting the deaths due to COVID-19. To make human welfare decisions, researchers have shown interest in this field, helping society prepare itself to fight against pandemic.

Forecasting is a valuable mechanism [13]-[15] that helps to understand the present situation and predict the future better. Much work focused on time series models to forecast the frequency of COVID-19 fatalities on this front. For the prediction of upcoming days, time series models are being used to analyze and estimate the number of deaths. These models predict the general state of the pandemic in the country.

A data-based statistical model such as Autoregressive Integrated Moving Average (ARIMA) [16], [17] has proven effective in predicting short-term forecasts [10], [18], including Tuberculosis [7], comparative case study between India-USA [19], case study of potato price forecasting [20] and consumption of items in Hotel Storage [21] in the past. The research paper [10] predicted the rise in positive cases in India using the ARIMA model; and checked the correctness of the model via evaluation parameters such as R-Squared [9], Akaike Information Criterion (AIC) [22], [23], BIC [24], and MSE [25].

The research paper [16] applied two distinct ARIMA models to get the next ten days data on confirmed cases, recovered cases, and fatalities. For model correctness, statistical metrics such as MAE and Root Mean Square Error (RMSE) have been utilized. The comparison research [17] used sequential data and time series models such as Exponential Smoothing, ARIMA, and Poisson to predict confirmed fatalities in Chile due to COVID-19. In addition, the accuracy of predictions has been evaluated using training and testing methods.

Table 1. Comparison table.

Ref.	Model	Studied Region	Prediction
[26]	ARIMA Model	India	To predict the COVID-19 patients' rise, recovery and death in India based on the daily data
[27]	Logistic linear regression model Susceptible infectious recovered model	Italy and Spain	To investigate the predictive ability of simple mathematical models and provide simple forecasts for the future incidence of COVID-19 in Italy and Spain
[28]	Gaussian distribution model	China	To compute the impact of future prediction and backward on the spread of COVID-19
[29]	Statistical survey study	US	To compute the impact of social distancing on the death count
[30]	Logistic model Weibull model Hills growth model	Global data	To identify individuals who are at the greatest risk

Table 1. Continued.

Ref.	Model	Studied Region	Prediction
[31]	Statistical model	UK	Estimation of excess 1- year mortality from COVID-19 with underlying conditions
[32]	SIRD model	China, Italy	Forecasting numbers of COVID-19 patients
[15]	SEIR model	Italy	Forecasting numbers of COVID-19 patients
Current Study	ARIMA model HLES model	India	Estimate the morality rate of coming 10 days

Many recent studies have observed that when dealing with infectious diseases, it's important to know how many people are infected and how many people will die as a result. For a nation like India, predicting the number of COVID-19 confirmed cases using mathematical and statistical models are essential. As the world's second most populated nation, there is a large country wide spread risk in India. The long-lasting and deadly COVID-19 pandemic has caused havoc throughout India, impacting almost every industry. Because of this, it is imperative that we take a look at this issue, review the previous method of solving the problem, and devise a way to deal with it. *Table 1* represents the comparison between different studies to the current study. In this study, two-time series algorithms have been evaluated to forecast deaths on ten forthcoming days.

Time series analysis demonstrates how data changes and correct forecasting reveals the data's developing tendency. The research's primary goal is to establish 10-day estimates of the total number of deaths caused by COVID-19 from 03 August to 12 August 2021. Next is to discover which time series model gives the best fit with the respective selection method criteria.

The remaining overall structure is written as follows: the materials and methods part explains the data and methodology used to predict the total number of deaths from COVID-19 in India. The next section covers the experimental results. The conclusions section summarizes the entire work and possible future scope.

2 | Materials and Methods

This section explains the dataset description from where we collect the data. Data pre-processing technique for the selection of univariate attribute need of time series analysis, time series models and performance measure parameters.

2.1 | Data Description

We have used the dataset from Kaggle[1] which includes the daily level of cases in the covid-19-india.csv file. The dataset is taken from 01 Jan 2020 to 02 Aug 2021 and feature description given in *Table 2*. After examining the dataset, from *Figs. 1(a)-1(c)*, we discovered that the number of cured cases, death rate and confirmed cases are relatively high from April 2021 to July 2021.

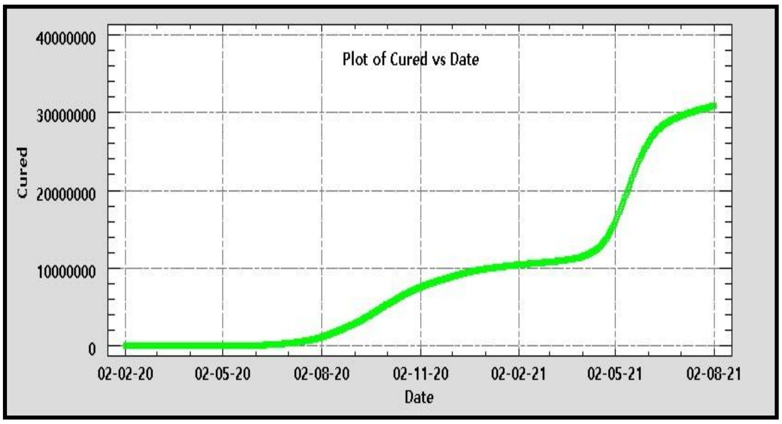
2.2 | Dataset Preprocessing

We performed dataset cleaning process [33], [34] to extract the correct information for prediction. This study removed some columns such as S.No., Confirmed India National and Confirmed Foreign Nationals from the dataset because their rankings were too low and included too many "-" symbols. We generate the new dataset from original COVID-19 dataset. The new data COVID-19 contains four columns and 551 occurrences, with

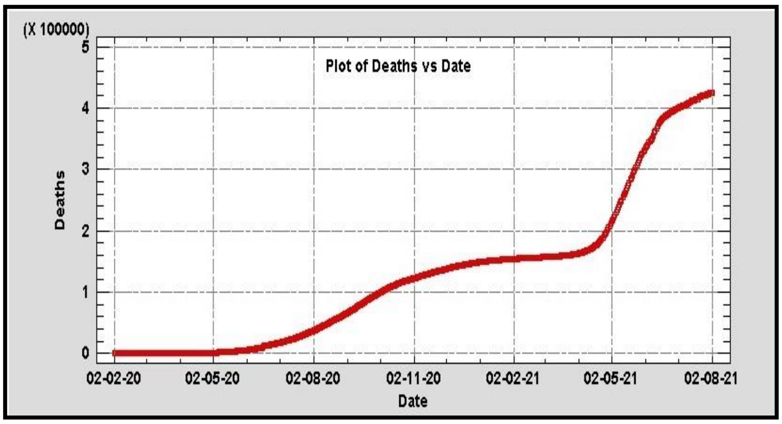
[1] https://www.kaggle.com/sudalairajkumar/covid19-in-indiaselectcovid_19_india.csv

each row reflecting data collected over a single day. Although ARIMA only takes univariate data [24], the number of recovered, fatalities, and confirmed cases must be separated. This research focuses on predicting the number of deaths during the next 10 days.

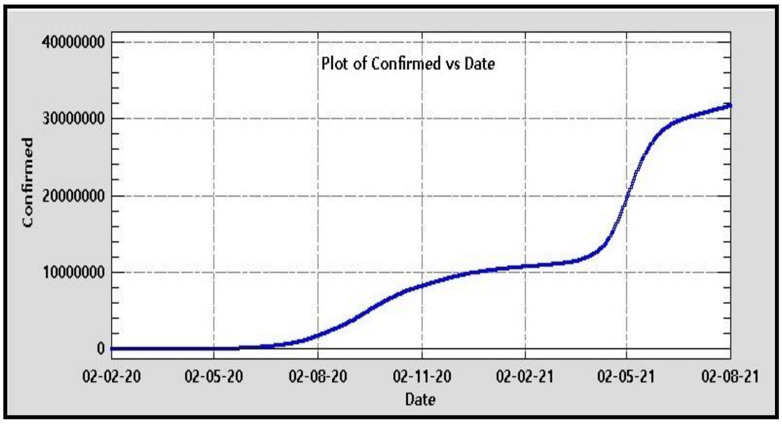
Therefore, this work select the one univariate attribute i.e., deaths for the prediction. The plot *Fig. 2*, here X-axis depicts the date variable, and Y-axis displays the deaths variable, which has 519 occurrences. The simple regression component is used in this scatter plot to fit a curve to the given data.



a. Cured cases.



b. Death cases.



c. Confirmed cases.

Fig. 1. Cured, death, and confirmed cases from 01 January 2020 to 02 August 2021.

Table 2. Feature description.

S. No.	Feature Name	Description	Missing Values	Type
1	S. no.	Serial number	0	Numeric
2	Date	Date of observation	0	Date-time
3	Time	Time of observation	0	Date-time
4	State/Union territory	Name of the State/Union territory	0	Numeric
5	Confirmed Indian national	Cumulative number of confirmed Indian nationals	0	Numeric
6	Confirmed foreign national	Cumulative number of confirmed foreign nationals	0	Numeric
7	Cured	Cumulative number of cured people	0	Numeric
8	Deaths	Cumulative number of death cases	0	Numeric
9	Confirmed	Cumulative number of confirmed cases	0	Numeric

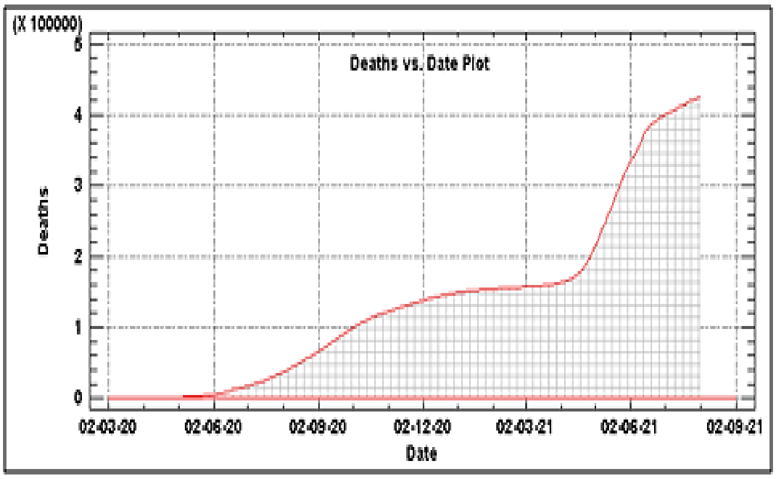


Fig. 2. Plot of deaths vs. date.

2.3 | Time Series

Time is the only variable that is used in time series analysis [35]. Now, a question arises: why is it necessary to have a time series algorithms when there are already many methods available? The answer of this question can be understood by this example. Let's suppose the supervised learning technique, which includes linear regression or logistic techniques with an independent and dependent variable. To derive a function or mapping function that describes how one variable is connected to another. Then we may go on to the analysis. However, in time series analysis, we only have one variable, which is time. A time series collects data points or observations recorded for a specific time [16]. Time series analysis is a Machine Learning (ML) application that utilizes data and time sequences to estimate future occurrences. This series is a set of data gathered at regular intervals. Based on previous time-series data, this approach gives relatively close projections of future patterns.

2.4 | Time Series Model

This section explains the different time series model which includes AR, MA, ARIMA and HELS methods.

2.4.1 | Autoregressive model

Autoregressive (AR) model is a kind of model that calculates the regression of previous time series and estimates the current or future values in the series. We can forecast the values of interest in this model by utilizing a linear mixture of previous values of the given attribute. This denotes a regression of the parameter inverse to it.

Observations at multiple time slots time-1, t-2, t-3,…, t-k affect the time period at t. It is determined by the coefficient factor at a certain point in time. In the long run, the share price of any specific business X may be influenced by the price history of all of its past shares. *Eq. (1)* describes an AR model of finite order (ar).

$$O_{time} = \beta 1 * o_{time-1} + \beta 2 * o_{time-2} + \ldots + \beta k * o_{time-k}. \quad (1)$$

2.4.2 | Moving average model

In the Moving Average (MA) model, the historical values of the forecast variable are used in conjunction with its prior prediction error values. External influences at distinct slots t-1, t-2, t-3,....., t-k affect the time period at t. Errors or residuals are the terms used to describe these unanticipated consequences. At every given moment, the coefficient factor determines the influence of earlier points in time on the current situation. It's possible that the price of a share in a business X may be affected by an overnight merger or a firm's bankruptcy. The MA model uses the residuals or mistakes from previous time series to predict the current and future values of the series and can be formulated using equation as shown in *Eq. (2)*.

$$O_{time} = \alpha 1 * e_{time-1} + \alpha 2 * e_{time-2} + \ldots + \alpha k * e_{time-k}. \quad (2)$$

2.4.3 | Auto regressive moving average model

This model suggests the time series forecasting framework. Its primary role is short-term forecasting, which necessitates more than 40 previous data points. ARIMA forecasting technology forecasts future values completely based on its previous values [10].

It performs effectively when the data has a robust or predictable trend over time with few outliers. When the information is relatively lengthy and the correlation between prior observations is consistent, ARIMA is typically preferable to exponential smoothing methods. The basic stage while using the ARIMA approach is to look for stationary points. The term "stationary" [36] denotes that the statistic maintains a generally consistent level across time.

Data is non-stationary if there is a trend, like in most relates to the financial applications. This is a forecasting method that assumes prior values contain intrinsic information and can be used to estimate values. When we combine the difference technique with the AR and MA models, we get an ARIMA model. ARIMA (ar, diff, ma) is a general signs, for easily specify the integer values that are replaced by variables.

The following are the ARIMA model's parameters. "ar" signifies the degree of lag samples in the analysis. "diff" denotes the degree of differencing. "ma" is frame's magnitude. *Eq. (3)* is general term for the complete model.

$$O_{time} = \beta 1 * o_{time-1} + \beta 2 * o_{time-2} + \alpha 1 * e_{time-1} + \alpha 2 * e_{time-2} + \ldots + \beta k * o_{time-k} + \alpha k * e_{time-k}. \quad (3)$$

2.4.4 | Holt linear exponential smoothing method

Holt expanded the simple exponential smoothing approach to anticipate data with a trend in 1957. A prominent smoothing technique for predicting data with trend is HLES two-parameter approach [2], [17]. It consists of three distinct formulas that interact to provide a complete forecast. The initial approach is simple smoothing formula that immediately modifies the very last smoothed number for the trend of the preceding time. The trend is modified during time using the second equation, which expresses the trend as the variance between the latest two smoothed readings.

Therefore, the finalized forecast is generated using the third formula. This method includes one prediction equation, as shown in *Eq. (4)*, as well as two smoothing equations, as shown in *Eqs. (5)* and *(6)*.

$$\hat{O}_{time+h|time} = l_{time} + hb_{time}. \quad (4)$$

$$l_{time} = \alpha O_{time} + (1-\alpha)(ł_{time-1} + \beta_{time-1}). \quad (5)$$

$$b_{time} = \beta * (ł_{time} - ł_{time-1}) + (1-\beta^{*})\, b_{time-1}. \quad (6)$$

2.5 | Comparison Criteria

This section discusses the performance measures used to evaluate the effectiveness and to check the comparability of the supplied models.

2.5.1 | Akaike information criterion

This approach assisted in the selection of parameters for regression and is also useful in determining the functional sequence of an ARIMA model. *Eq. (7)* illustrates the AIC formula.

$$AIC = -2\log(L) + 2(P + Q + K + 1). \quad (7)$$

Eq. (7) shows the revised AIC for the ARIMA model.

$$AICc = AIC + \frac{2(P+Q+K+1)(P+Q+K+2)}{(T-P-Q-K-2)}. \quad (8)$$

2.5.2 | Root-mean-square error

In various natural sciences, RMSE has been employed as a normal statistical notation to assess model effectiveness. It is used to calculate the difference between predicted and actual numbers from the context. These discrepancies are referred to as error terms or RMSE. The metrics represent the standard deviation of the error terms or how distant the endpoints are from the regression or modelled line. It computes the amount of inconsistency between two datasets. *Eq. (9)* is used to get the RMSE value.

$$RMSE = \sqrt{\frac{1}{n}\sum_{n=1}^{n}(fi - oi)\hat{}2}, \quad (9)$$

where n is the no. of elements, f and o represents the forecasted and observed data.

3 | Experimental Results

The experiment has been conducted on the statgraphics tool. This paper uses HLES and the ARIMA model to forecast new deaths per day using two different selection criteria methods.

3.1 | Forecast Data with HLES Method

In this method, both models are executed with AIC criteria. We found that HLES method (with alpha value 0.9083 and beta value 0.5038) has been selected to forecast future deaths values. This framework suggests that the linear trend with all previous data values predicted best forecast for the future.

In *Fig. 3*, we plot different graphs of the selected model. *Fig. 3(a)* represents the time sequence plot of deaths between actual and forecast values. *Fig. 3(b)* plots the forecast values for deaths. *Fig. 3(c)* draw the points for residual autocorrelations for deaths, and *Fig. 3(d)* plots the residual periodogram for deaths.

In *Table 3*, we summarize the selected model's performance parameters, which fit the historical data. One ahead forecast error represents the dissimilarity between the input value at time t and the forecast produced at time t-1.

The size of the error is measured by the first two statistics. A better model yields a lower value. The final statistics assess bias, and for better out- comes, the model returns a number near to zero. Different values of the ARIMA model giving us major value for comparison. Holt's Method gives the lowest value of AIC, which has been used to generate the predictions. In all cases of the ARIMA model second method having the lowest RMSE and AIC value. So, this is the best fit for our model to forecast.

The actual comparison of data is shown in *Table 4* between original values and forecasted values. This model also calculates the values and the Lower Confidence Interval (LCI) factor with 95% of the limit and the Upper Confidence Interval (UCI) factor with 95% of the limit. We observed that the forecasted values are very close to the original values.

3.2 | Forecast Data with ARIMA Method

Both models are run using RMSE criterion in this approach. According to the experimental findings, ARIMA (2, 0, 2) has been used to anticipate future deaths values. This contribute to the advancement that a modeling approach connecting the most current data value to past data values and historical noise delivers the best estimate for future data. In *Fig. 4*, we plot different graphs of the selected model. *Fig. 4(a)* represents the time sequence plots of deaths between actual and forecast values. *Fig. 4(b)* plots the forecast values for deaths. *Fig. 4(c)* draw the points for residual autocorrelations for deaths, and *Fig. 4(d)* plots the residual periodogram for deaths. *Table 5* highlights the performance of the chosen model, displaying RMSE, MAE, and ME. Here, ARIMA model, calculates standard error, t-test, and p-value. We have selected individual values of ARIMA, where p=2 (AR value), d=2 (d value), and q=2(MA value). The given of different parameters, compute AR(1), AR(2), MA(1), and MA(2).

The AR(2) term has a p-value less than 0.05, indicating that it is substantially different from 0. The MA(2) term has a p-value less than 0.05, indicating that it is substantially different from 0.309604 is the calculated standard deviation of the input white noise.

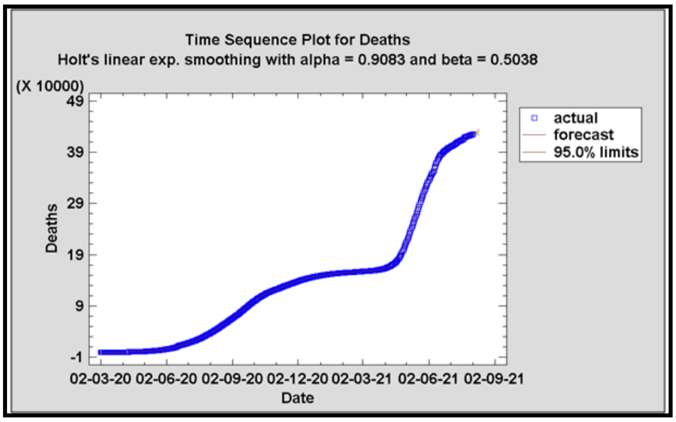
The fair values of the chosen model are shown in the comparative findings in *Table 6*. The ARIMA(2,0,2) model, which was utilized to create the predictions, has the lowest RMSE value. As a result, this is the best fit for our forecasting model.

The actual data comparison in *Table 7* illustrates the differences in values between the original and predicted data. This model also computes the values as well as the LCI factor with a 95% CI and the UCI factor with a 95% CI. We discovered that the predicted values are pretty similar to the actual ones.

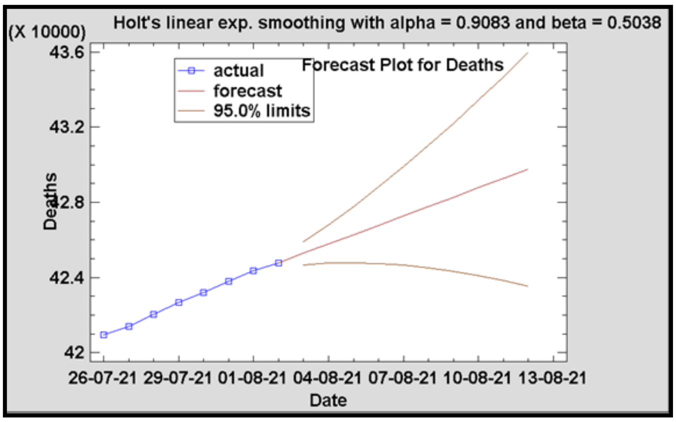
The results of HLES Model with the AIC selection method give more accurate values than the ARIMA(2,0,2) model with the RMSE selection method. *Fig. 5* shows how well the values of Holt's model correspond to the real data.

Table 3. Different models comparisons.

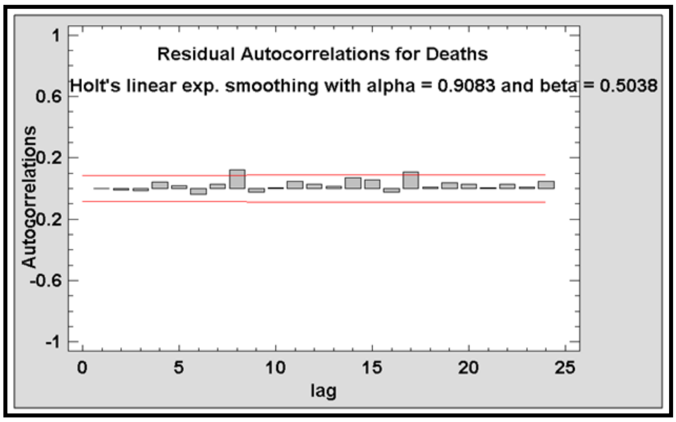
Model	RMSE	MAE	ME	AIC
HLES	309.494	116.574	2.09632	11.4776
ARIMA(0,2,2)	310.097	116.995	2.10177	11.4815
ARIMA(1,2,1)	310.18	117.111	2.14009	11.482
ARIMA(1,1,2)	309.668	116.332	12.3354	11.4826
ARIMA(0,2,1)	311.151	118.57	2.29498	11.4844



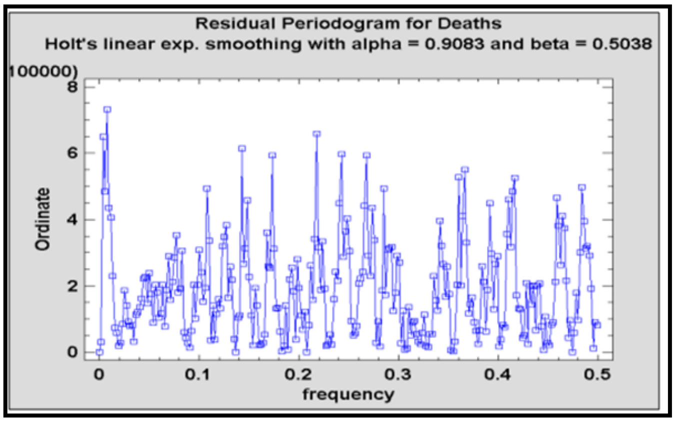
a. Plot of deaths forecast.



b. Residual autocorrelations for deaths.



c. Residual periodogram.



d. Plot for deaths.

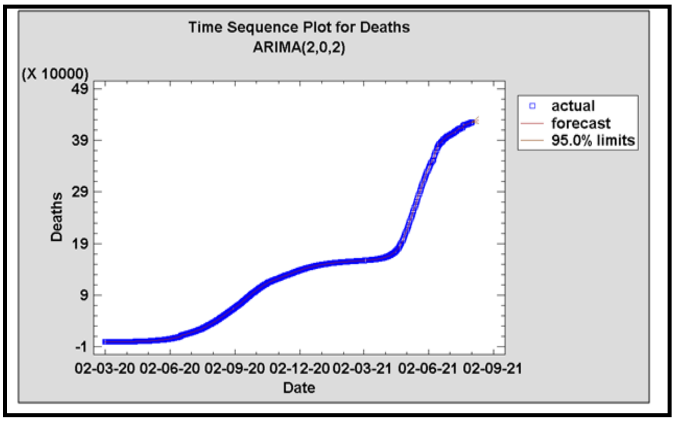
Fig. 3. Plot between actual and forecast data using HLES method with alpha=0.9083 and beta=0.5038.

Table 4. Results of 10-days ahead forecasts (03 August to 12 August 2021) using HLES for the number of deaths.

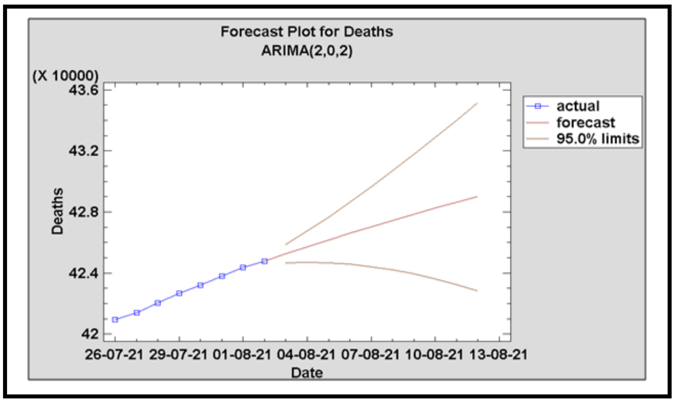
Dates	Actual Data	HLES Method		
	Deaths	Forecast_Data	Lower 95% Limit	Upper 95% Limit
03-08-2021	425215	425284	424697	425872
04-08-2021	425777	425782	424788	426777
05-08-2021	426310	426280	424818	427742
06-08-2021	426774	426778	424795	428761
07-08-2021	427391	427276	424722	429830

Table 4. Continued.

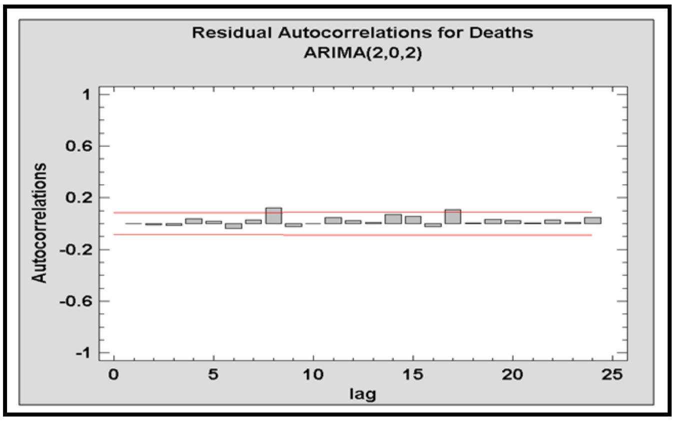
Dates	Actual Data	HLES Method		
	Deaths	Forecast_Data	Deaths	Forecast_Data
08-08-2021	427882	427774	424603	430944
09-08-2021	428329	428272	424443	432100
10-08-2021	428702	428769	424242	433297
11-08-2021	429199	429267	424004	434531
12-08-2021	429689	429765	423730	435801



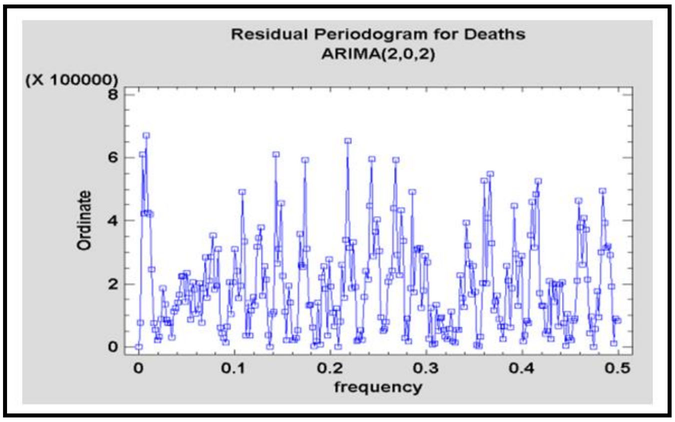
a. Plot of deaths forecast.



b. Residual autocorrelations for deaths.



c. Residual periodogram.



d. **Plot for deaths.**

Fig. 4. Plot between actual and forecast data for deaths using (ARIMA(2,0,2)).

Table 5. ARIMA model summary.

Parameter	Estimate	Standard Error	t	P-value
AR(1)	1.9968	0.00739474	270.03	0.000000
MA(1)	0.634751	0.0443708	14.3056	0.000000
AR(2)	-0.996825	0.00743468	-134.078	0.000000
MA(2)	-0.0913635	0.0445278	-2.05183	0.040691

Table 6. Different models comparisons.

Model	RMSE	MAE	ME	AIC
HLES	309.604	116.574	2.09632	11.4776
ARIMA(2,0,2)	309.494	115.828	2.00177	11.4764
ARIMA(1,1,2)	309.668	116.332	12.3345	11.4826
ARIMA(2,1,1)	309.777	116.343	12.1595	11.4833
ARIMA(0,2,2)	310.097	116.995	2.10177	11.4815

Table 7. Actual and predicted data using ARIMA(2,0,2) model for the number of deaths.

Dates	Actual Data	ARIMA Method		
	Deaths	Forecast_Data	Lower 95% Limit	Upper 95% Limit
03-08-2021	425215	425259	424668	425849
04-08-2021	425777	425721	424724	426718
05-08-2021	426310	426172	424709	427635
06-08-2021	426774	426612	424630	428593
07-08-2021	427391	427039	424491	429587
08-08-2021	427882	427455	424297	430613
09-08-2021	428329	427859	424050	431668
10-08-2021	428702	428252	423754	432749
11-08-2021	429199	428633	423411	433855
12-08-2021	429689	429002	423022	434982

4 | Conclusion

At present, COVID spread is a big issue; it can lead to death if not recognized and treated timely. Therefore, it is essential to understand the spread behaviour of this disease and the projection of infections and deaths. The research aim is to use time series models to predict the death rate in India due to COVID-19. The evaluation was done using AIC and RMSE statistics. Comparative studies of the significance among these models, it was discovered that HLES model predicts more accurately than ARIMA model. This time series model can serve as a forecasting framework for assessing the present status of illness and seriousness, allowing administration and health care personnel to make more informed decisions in order to lower India's mortality rate.

The first gap in the study is that the presented dataset lacked several features, resulting in unimproved outcomes. The findings may have been better if we had previously known the corona symptoms, which would have benefited in greater forecasting. This data collection was compiled from all of India's states. De-spite this, due to a research deficit, it did not have comprehensive knowledge of every state of the cases that came in every day. The second gap was that we couldn't link the data to any environmental variables since each state's environment was distinct, making it impossible to connect the dataset. We could choose a specific location to estimate future scenarios including environmental factors.

The deep learning Recurrent Neural Network time series fore- casting model can improve COVID-19 mortality rate generalization ability in the future. Furthermore, for modeling the highly accurate prediction model for the COVID-19 pandemic, this study may be improved by incorporating many aspects or variables such as geographical factors, demographical factors, and meteorological factors (humidity, temperature, wind, rainfall, and speed).

Reference

[1] Singh, S., Parmar, K. S., Kumar, J., & Makkhan, S. J. S. (2020). Development of new hybrid model of discrete wavelet decomposition and autoregressive integrated moving average (ARIMA) models in application to one month forecast the casualties cases of covid-19. *Chaos, solitons & fractals*, *135*, 109866. DOI:10.1016/J.CHAOS.2020.109866

[2] Poonia, N., & Azad, S. (2020). *Short-term forecasts of covid-19 spread across indian states until 1 may 2020*. DOI:10.20944/preprints202004.0491.v1

[3] Rahimi, I., Chen, F., & Gandomi, A. H. (2023). A review on COVID-19 forecasting models. *Neural computing and applications*, *35*(33), 23671-23681. DOI:10.1007/s00521-020-05626-8

[4] Chakraborty, I., & Maity, P. (2020). COVID-19 outbreak: Migration, effects on society, global environment and prevention. *Science of the total environment*, *728*, 138882. DOI:10.1016/J.SCITOTENV.2020.138882

[5] Kitajima, M., Ahmed, W., Bibby, K., Carducci, A., Gerba, C. P., Hamilton, K. A., ... & Rose, J. B. (2020). SARS-CoV-2 in wastewater: State of the knowledge and research needs. *Science of the total environment*, *739*, 139076. DOI:10.1016/J.SCITOTENV.2020.139076

[6] Fagherazzi, G., Goetzinger, C., Rashid, M. A., Aguayo, G. A., & Huiart, L. (2020). Digital health strategies to fight COVID-19 worldwide: challenges, recommendations, and a call for papers. *Journal of medical internet research*, *22*(6), e19284. *https://www.jmir.org/2020/6/e19284*

[7] Choudhari, R. (2020). COVID 19 pandemic: mental health challenges of internal migrant workers of india. *Asian journal of psychiatry*, *54*, 102254. DOI:10.1016/J.AJP.2020.102254

[8] Sharma, H. B., Vanapalli, K. R., Cheela, V. S., Ranjan, V. P., Jaglan, A. K., Dubey, B., … & Bhattacharya, J. (2020). Challenges, opportunities, and innovations for effective solid waste management during and post covid-19 pandemic. *Resources, conservation and recycling*, *162*, 105052. DOI:10.1016/J.RESCONREC.2020.105052

[9] Dhamodharavadhani, S., Rathipriya, R., & Chatterjee, J. M. (2020). COVID-19 mortality rate prediction for india using statistical neural network models. *Frontiers in public health*, *8*(8), 1–12. DOI:10.3389/fpubh.2020.00441

[10] Kumar, S. L., Sarobin, M, V. R., & Anbarasi, L, J. (2021). Predictive analytics of covid-19 pandemic: statistical modelling perspective. *Walailak journal of science and technology (WJST)*, *18*(16), 15583. DOI:10.48048/wjst.2021.15583

[11] Kumar, S., Viral, R., Deep, V., Sharma, P., Kumar, M., Mahmud, M., & Stephan, T. (2021). Forecasting major impacts of covid-19 pandemic on country-driven sectors: challenges, lessons, and future roadmap. *Personal and ubiquitous computing*, *27*, 807-830. DOI:10.1007/s00779-021-01530-7

[12] Rustam, F., Reshi, A. A., Mehmood, A., Ullah, S., On, B. W., Aslam, W., & Choi, G. S. (2020). COVID-19 future forecasting using supervised machine learning models. *IEEE access*, *8*, 101489–101499. DOI:10.1109/ACCESS.2020.2997311

[13] Bećirović, E., & Ćosović, M. (2016, September). Machine learning techniques for short-term load forecasting. *2016 4th international symposium on environmental friendly energies and applications (EFEA)* (pp. 1-4). IEEE. DOI:10.1109/EFEA.2016.7748789

[14] Perc, M., Gorišek Miksić, N., Slavinec, M., & Stožer, A. (2020). Forecasting covid-19. *Frontiers in physics*, *8*, 127. DOI:10.3389/fphy.2020.00127

[15] Russo, L., Anastassopoulou, C., Tsakris, A., Bifulco, G. N., Campana, E. F., Toraldo, G., & Siettos, C. (2020). Tracing day-zero and forecasting the COVID-19 outbreak in Lombardy, Italy: A compartmental modelling and numerical optimization approach. *Plos one*, *15*(10), e0240649. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0240649

[16] Ali, M., Khan, D. M., Aamir, M., Khalil, U., & Khan, Z. (2020). Forecasting COVID-19 in Pakistan. *Plos one*, *15*(11), e0242762. DOI:10.1371/journal.pone.0242762

[17] Barría-Sandoval, C., Ferreira, G., Benz-Parra, K., & López-Flores, P. (2021). Prediction of confirmed cases of and deaths caused by COVID-19 in Chile through time series techniques: A comparative study. *Plos one*, *16*(4), e0245414. DOI:10.1371/journal.pone.0245414

[18] Lee, D. H., Kim, Y. S., Koh, Y. Y., Song, K. Y., & Chang, I. H. (2021). Forecasting covid-19 confirmed cases using empirical data analysis in korea. *Healthcare (Switzerland)*, *9*(3), 1–17. DOI:10.3390/healthcare9030254

[19] Shastri, S., Singh, K., Kumar, S., Kour, P., & Mansotra, V. (2020). Time series forecasting of covid-19 using deep learning models: india-usa comparative case study. *Chaos, solitons and fractals*, *140*, 110227. DOI:10.1016/j.chaos.2020.110227

[20] Şahinli, M. A. (2020). Potato price forecasting with holt-winters and arima methods: a case study. *American journal of potato research*, *97*(4), 336–346. DOI:10.1007/s12230-020-09788-y

[21] Chandra, C., Fajrin, A. A., & Eko Suharyanto, C. (2021). Forecasting the items consumption in the hotel storage with the autoregressive integrated moving average method. *Engineering, mathematics and computer science (EMACS) journal*, *3*(1), 13–19. DOI:10.21512/emacsjournal.v3i1.6979

[22] Sharma, V. K., & Nigam, U. (2020). Modeling and forecasting of covid-19 growth curve in india. *Transactions of the indian national academy of engineering*, *5*(4), 697–710. DOI:10.1007/s41403-020-00165-z

[23] Harous, S., Al Harmoodi, M., & Biri, H. (2019). A comparative study of clustering algorithms for mixed datasets. *2019 amity international conference on artificial intelligence (AICAI)* (pp. 484-488). IEEE.

[24] Khan, F. M., & Gupta, R. (2020). ARIMA and nar based prediction model for time series analysis of covid-19 cases in india. *Journal of safety science and resilience*, *1*(1), 12–18. DOI:10.1016/j.jnlssr.2020.06.007

[25] Lydia, M., Kumar, S. S., Selvakumar, A. I., & Kumar, G. E. P. (2016). Linear and non-linear autoregressive models for short-term wind speed forecasting. *Energy conversion and management*, *112*, 115-124. DOI:10.1016/j.enconman.2016.01.007

[26] Katoch, R., & Sidhu, A. (2021). An application of ARIMA model to forecast the dynamics of COVID-19 epidemic in India. *Global business review*, 0972150920988653. https://journals.sagepub.com/doi/abs/10.1177/0972150920988653

[27] Chu, J. (2021). A statistical analysis of the novel coronavirus (COVID-19) in Italy and Spain. *PloS one*, *16*(3), e0249037. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0249037

[28] Li, L., Yang, Z., Dang, Z., Meng, C., Huang, J., Meng, H., … & Shao, Y. (2020). Propagation analysis and prediction of the covid-19. *Infectious disease modelling*, *5*, 282. DOI:10.1016/J.IDM.2020.03.002

[29] Bayham, J., & Fenichel, E. P. (2020). Impact of school closures for COVID-19 on the US health-care workforce and net mortality: a modelling study. *The lancet public health*, *5*(5), e271-e278. DOI:10.1101/2020.03.09.20033415

[30] Hembram, K. P. S. S., & Kumar, J. (2021). Epidemiological study of novel coronavirus (COVID-19): macroscopic and microscopic analysis. *International journal of community medicine and public health*, *8*, 1364-1369.

[31] Banerjee, A., Pasea, L., Harris, S., Gonzalez-Izquierdo, A., Torralbo, A., Shallcross, L., … & Hemingway, H. (2020). Estimating excess 1-year mortality associated with the covid-19 pandemic according to

underlying conditions and age: a population-based cohort study. *The lancet*, *395*(10238), 1715–1725. DOI:10.1016/S0140-6736(20)30854-0

[32] Caccavo, D. (2020). *Chinese and Italian COVID-19 outbreaks can be correctly described by a modified SIRD model.* https://doi.org/10.1101/2020.03.19.20039388

[33] Theerthagiri, P., Jacob, I. J., Ruby, A. U., & Vamsidhar, Y. (2020). *Prediction of COVID-19 possibilities using KNN classification algorithm.* https://doi.org/10.21203/rs.3.rs-70985/v2

[34] Arpaci, I., Huang, S., Al-Emran, M., Al-Kabi, M. N., & Peng, M. (2021). Predicting the covid-19 infection with fourteen clinical features using machine learning classification algorithms. *Multimedia tools and applications*, *80*(8), 11943–11957. DOI:10.1007/s11042-020-10340-7

[35] Chatfield, C. (2000). Multivariate forecasting methods. *Time-series forecasting*, *2*, 28–39. DOI:10.1201/9781420036206.ch5

[36] Iwok, I. A., & Okpe, A. S. (2016). A comparative study between univariate and multivariate linear stationary time series models. *American journal of mathematics and statistics*, *6*(5), 203–212. DOI:10.5923/j.ajms.20160605.02