https://doi.org/10.48048/tis.2024.7246

Performance Analysis of Autoregressive Integrated Moving Average (ARIMA) and 'earlyR' Statistical Models for Predicting Epidemic Outbreaks: A Case Study on COVID-19 Data in India

Karthick Kanagarathinam[1,*], G. Ponkumar[2] and S. Sendil Kumar[3]

[1]Department of Electrical and Electronics Engineering, GMR Institute of Technology, Andhra Pradesh, India
[2]Department of Electrical and Electronics Engineering, Panimalar Engineering College, Chennai India
[3]Department of Electrical and Electronics Engineering, S.A. Engineering College (Autonomous), Chennai, India

([*]Corresponding author's e-mail: kkarthiks@gmail.com, karthick.k@gmrit.edu.in)

Received: 24 June 2023, Revised: 25 July 2023, Accepted: 1 August 2023, Published: 1 October 2023

Abstract

The objective of this research article is to analyze the performance of the Autoregressive Integrated Moving Average (ARIMA) model and the 'earlyR' statistical model in predicting epidemic outbreaks using COVID-19 data of India. The results of this analysis can help take preventive actions and restrict the spread of the epidemic with the help of projected data. The 'projections' module was utilized to generate the epidemic path by aligning the available COVID-19 data from India, distribution of the time interval between successive cases and reproduction number (R_0) of the corresponding regions with the 'earlyR' epidemic statistical model. The values of (p, d and q) were obtained utilizing the 'auto.arima' function, and an ARIMA time series model was created using the 'forecast' module to forecast future infected occurrences. The 'earlyR' epidemic model yielded a median projected value with an inaccuracy of 35.1 %, while the ARIMA model had a mean error of −1.9 %. A comparison of these methods indicates that the ARIMA model is a superior method compared to the 'earlyR' epidemic model in terms of accuracy.

Keywords: ARIMA model, Prediction model, Forecasting, Statistical model, Time series forecasting

Introduction

COVID-19, caused by the SARS-CoV-2 virus, has rapidly spread worldwide, affecting several countries including China, Spain, Iran and the USA [1]. This virus belongs to the family of coronaviruses, which can cause a range of illnesses from mild respiratory infections to severe diseases like MERS and SARS. Recognizing the severity of the ongoing pandemic, the Expert Committee on Virology Terminology officially named the virus responsible for COVID-19 as SARS-CoV-2 on February 11, 2020.

In the UK, the BA.4 and BA.5 variants have been associated with symptoms such as a runny nose, sore throat, headache, persistent cough and tiredness. The incubation period of the virus varies between 1 and 14 days, with an average of approximately 5 days, while its surface survival time is similar to that of other coronaviruses. Considering the modes of transmission, it is crucial to use masks along with other preventive measures, as breathing can release viral particles, and sneezing and coughing can disperse droplets more forcefully. Close proximity to others in motion, known as slipstreaming, can also increase transmission risk. The R_0 value, representing new incidences produced from each infected individual, can vary based on disease and population characteristics. Effective measures such as contact tracing and quarantine have demonstrated success in controlling COVID-19 within 3 months [2].

Forecasting COVID-19 transmission is of paramount importance for policymakers and public health officials, as it aids in making informed decisions regarding preventive measures and resource allocation in response to the pandemic. By predicting the number of infected cases, hospitalizations and fatalities, authorities can better plan and prepare for the increased demand on medical facilities, equipment and personnel.

This article analyzes 2 models for forecasting COVID-19 transmission: The Autoregressive Integrated Moving Average (ARIMA) model and the 'earlyR' model. The ARIMA model captures underlying patterns and trends in the data to forecast future values, while the 'earlyR' model utilizes assumptions about the virus's behavior and population characteristics to simulate its spread over time. The ‘earlyR’ model leverages the ‘earlyR’ module in R to evaluate R_0, projected R_0, and estimate the daily new infection cases for the subsequent 30-day period. Additionally, this model’s versatility extends beyond COVID-19, making it applicable to analyzing other epidemic diseases such as measles, SARS and pandemic flu. The model’s development relies on a serial interval distribution cited in previous research [3-5], as illustrated in **Figure 1**.

These predictive models provide policymakers with insights into various potential outcomes and scenarios, enabling them to make data-driven decisions when implementing measures to safeguard public health. Such measures may involve social distancing, mask-wearing and travel restrictions. Furthermore, the models help officials determine the timing of imposing or lifting restrictions based on the predicted trajectory of the epidemic, ultimately impacting the course of the epidemic and potentially saving lives.

The emergence of the SARS coronavirus in 2003 from southeast China’s Guangdong region caused a strange pneumonia and confirmed Koch’s idea [6]. To estimate the effects of COVID-19 on mortality rates related to heart and diabetes conditions, controlled trials are essential. These trials should consider pre-existing disparities between individuals with the disease and those without, as well as the evolution of the comparison group over time [7]. Variants of concern (VOCs) known as first-generation neutralization-resistant variants, such as Beta, have been identified with specific mutations present in the receptor binding domain (RBD) at K417N and E484K positions. These mutations have enabled the variants to evade neutralization by both class I and class II antibodies [8]. Host cell penetration, particularly by receptor-mediated endocytosis, is the first step in COVID-19 infection.

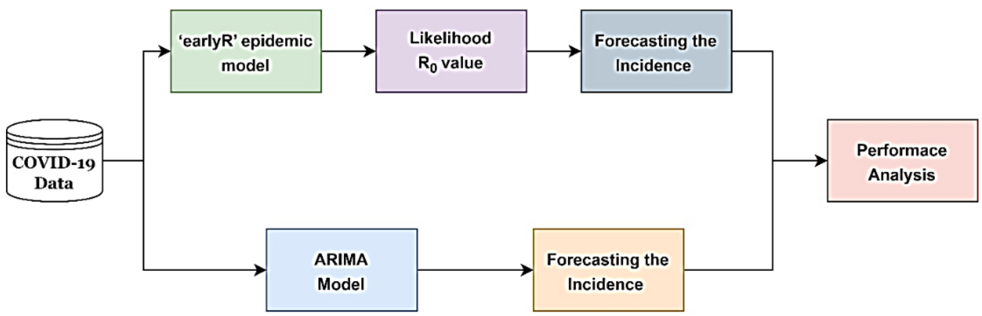


Figure 1 Model analysis flow diagram.

Epidemiologists gather, analyze and evaluate data to determine how to prevent the spread of a new infectious illness or a confirmed infectious disease outbreak. Statistical modeling and prediction are both helpful in gaining knowledge of how diseases are transmitted and how they can be stopped or controlled [9]. Currently, there is no high-quality data for any medication combinations or antivirals, and while a vaccine for COVID-19 is in development, currently no drug blends or antivirals that have been shown to be particularly successful [10]. Early relaxation may cause a further wave of infection [11].

Predicting and projecting the daily count of COVID-19 cases and fatalities through modeling can provide healthcare services with an improved understanding of the magnitude of new infections and deaths, which in turn enables them to prepare for the required equipment and safety protocols ahead of time. Models that forecast the transmission of epidemics are of great interest to the epidemiology and health sciences communities. Various models have been established to forecast the transmission of COVID-19 in different locations, estimate R_0, ascertainment rates, occurrence and incidence over time and locations, and to estimate probable ICU capacity overruns based on future growth [12-16].

The global economy has faced significant adverse impacts due to COVID-19. If positive patients are identified early, the spread of this pandemic disease can be slowed. Predicting COVID-19 disease is useful for identifying patients who are at risk of developing health problems [17]. **Table 1** provides the comparison of the state-of-the-art COVID-19 forecasting methods.

Guo and He [18] employed an artificial neural network (ANN) for the analysis and prediction of confirmed cases and fatalities related to COVID-19. They used the WHO data from January 20 to November 11, 2020, to train and test the ANN. They utilized statistical metrics, such as RMSE and correlation coefficient, to estimate the precision of the ANN model in forecasting the incidences and fatalities related to COVID-19.

Additionally, the study presented ANN-generated projections for incidences and mortality of COVID-19 between June 5, 2020 and November 11, 2020. The results indicate the need for strict and continuous control measures to mitigate the further transmission of the pandemic. He [19] proposed a COVID-19 risk prediction and control system based on inconsistent elements of pandemic estimation. The study investigated the ambiguity attribute of pandemic threat using data on personnel flow and resolved the issue of unpredictability of pandemic risk because of the volatility of personnel flow. The study recommended emergency prevention and surveillance measures for plausible COVID-19 cases. Additionally, the authors developed an approach for evaluating and managing risks associated with public health security through the use of correlation function methodology. This methodology resulted in a unique research approach for preventing and managing epidemic risks. In a similar vein, to enhance the precision of forecasting solar irradiance. Alsharif [20] established a time-series prediction model using ARIMA and compared their findings to those obtained using the Monte Carlo approach.

Table 1 Comparison of state-of-the-art COVID-19 forecasting methods.

Study	Methodology	Data source	Metrics evaluated	Findings
Guo and He [18]	Artificial Neural Network (ANN)	WHO Data (Jan 20 - Nov 11, 2020)	RMSE, correlation coeff.	Forecasted incidence and mortality between June 5 - Nov 11, 2020.
He [19]	Pandemic risk prediction and control system	Personnel flow data	Correlation function methodology	Proposed COVID-19 risk prediction and control system based on inconsistent elements of pandemic estimation
Alsharif [20]	Time-series prediction	Solar irradiance data	Forecast accuracy	ARIMA compared to Monte Carlo approach

Previous studies have contributed significantly to the understanding of COVID-19 and its impact on public health. However, several limitations in these studies need to be acknowledged. Firstly, many early forecasting models relied on limited data, leading to challenges in accurately predicting the dynamic nature of the pandemic. Additionally, some studies used single-source data, which may not capture the full complexity and heterogeneity of COVID-19 transmission patterns across different regions and populations.

Furthermore, the accuracy of predictions in some prior research has been influenced by assumptions made about the virus's behavior and population characteristics. While these models provided valuable insights, they may not have fully accounted for the evolving nature of the pandemic or the emergence of new variants, which can significantly impact transmission dynamics. In contrast, our study contributes to overcoming these limitations by employing a comprehensive and data-driven approach. We utilized multiple sources of data, including WHO data and other reliable databases, to ensure a more robust representation of the COVID-19 outbreak. This allowed us to capture the variability in transmission patterns and make more accurate predictions. Moreover, the 'earlyR' epidemic model and the ARIMA model employed in our study have distinct advantages. The 'earlyR' model leverages information about R_0, projected R_0, and daily new infection cases to forecast transmission, making it a powerful tool in pandemic analysis. On the other hand, the ARIMA model captures underlying patterns and trends in the data, allowing for more nuanced predictions. By comparing the results of both models, our study provides a comprehensive analysis of COVID-19 forecasting.

Another key feature of our research is the incorporation of novel risk prediction and control methods, inspired by the correlation function methodology. This approach addresses the issue of unpredictability of pandemic risk due to factors such as personnel flow, providing a unique perspective on controlling epidemic risks and enhancing preparedness.

Materials and methods

Data

The COVID-19 data for India, specifically from May 25th to June 16th, 2020, along with the data for Tamil Nadu from March 7th to April 16th, 2020, were taken into account for the development and analysis of the model

[21]. To assess methodologies, we use the 'earlyR' pandemic model and the ARIMA model. R has a package called earlyR, designed to help novices get a head start on their projects.

'earlyR' epidemic statistical model

Cori *et al.* [22] established a complex epidemic spread model, but a new model called "earlyR" has been introduced. **Figure 2** displays the process involved in the "earlyR" model. Using "earlyR" and incidence, we can estimate the R_0 of COVID-19. An accurate estimate of the collective figure of short-term new incidences can improve public health crisis decisions, and the serial interval distribution is a vital component.

Li *et al.* [23] assessed the COVID-19 incubation time at 5.2 days and found evidence of close contact transmission, while Nishiura *et al.* [24] determined a 95 % credible interval the mean and standard deviation as 4.7 and 2.9 days, respectively and it is used to determine the Maximum-Likelihood R_0 value to develop our model. We resample the data 1,000 times using the bootstrap method to obtain plausible R_0 values. This R code uses the 'projections' package to anticipate the collective figure of short-term new incidences over the following 10 days [25]. By adjusting the current COVID-19 data of Tamil Nadu, India, serial interval distribution and the resulting R_0 value, we can predict the future course of the pandemic. The daily occurrence follows a Poisson distribution with a parameter dictated by the infectiousness of the population on any given day (Eq. (1)).

$$\lambda\,(\mathrm{t}) = \sum_{n=1}^{t-1} \mathrm{X_n}\, \mathrm{P}\,(\mathrm{t} - \mathrm{n}) \tag{1}$$

This occurs when $\mathrm{P}\,(\mathrm{t} - \mathrm{n})$ represents the vector of probability distribution and X_n denotes the actual rate of occurrence at time n. Values can be analyzed to determine their mean, median, minimum and maximum.

ARIMA model

The ARIMA model is defined by a set of 3 parameters, namely p, d and q, which are abbreviated as ARIMA (p, d and q). In particular, the 'p' value refers to autoregressive constraints count, the 'd' value refers to non-seasonal differences count, and 'q' value corresponds to the count of parameters of the moving average component.

The random variable X_t denotes the unidentified incidences in the time series under consideration at instant t. The term AR(p) refers to a p^{th} order autoregressive process and can be defined as follows:

$$\mathrm{X_t} - \sum_{k=1}^{p} \Phi_k X_{t-k} = \in_t \tag{2}$$

where, $\in_t$ represents the uncertainty that impacts our observations, $\Phi_{1,} \Phi_2, \dots \dots \Phi_p$ are the AR coefficients. The description of the parameters for the q^{th} order moving average component, denoted as MA(q), is as follows:

$$\mathrm{X_t} = \in_t + \sum_{k=1}^{q} \theta_k \in_{t-k} \tag{3}$$

here, θ_1, θ_2, …, θ_q are the parameters of the moving average component.

The level of differentiation, denoted by d, indicates the extent to which the time series cases deviate from the stationary state, and is described as follows:

$$\Phi(B)\nabla^d \mathrm{X_t} = \theta(B) \in_t \tag{4}$$

The term $\nabla \mathrm{X_t}$ represents the first-order differencing operator $(1 - \mathrm{B})$ applied to $\mathrm{X_t}$, where B is the backward shift operator defined as $B^n X_t := X_{t-n}$.

The building process of the ARIMA model is illustrated in **Figure 3**. The missing data is filled with zeros, and the values are estimated using the "AUTOARIMA" function (p, d and q). The Partial Autocorrelation function (PACF) and Autocorrelation function (ACF) are utilized, along with the MAPE parameter, for identifying the best model.

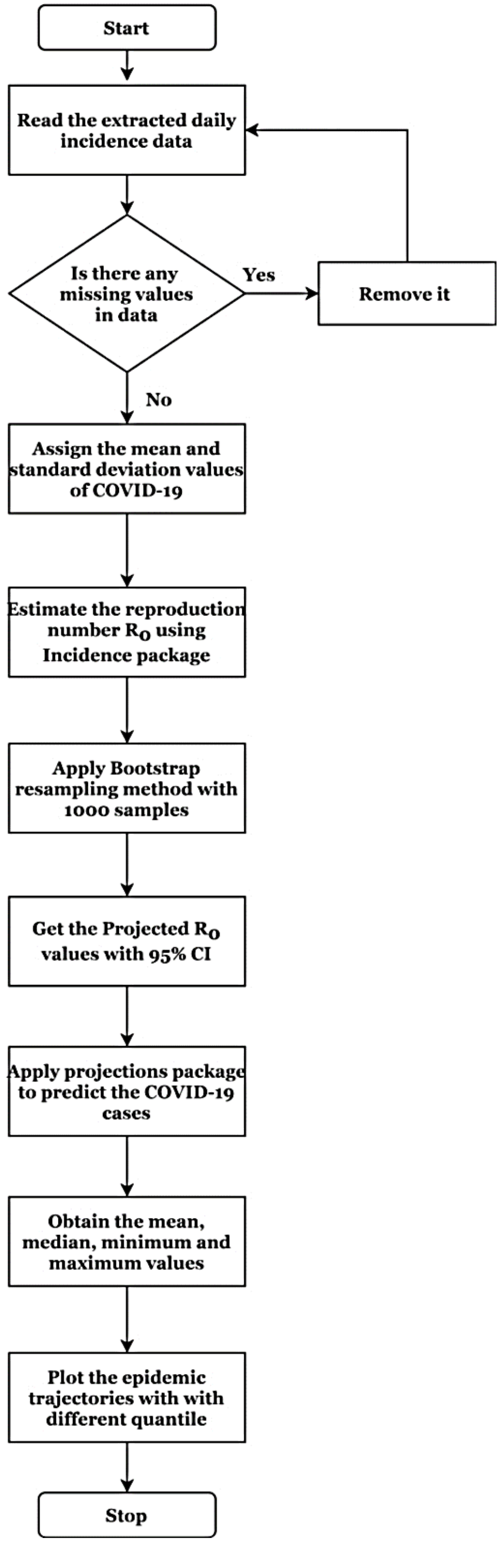


Figure 2 Flowchart of 'earlyR' COVID-19 prediction model.

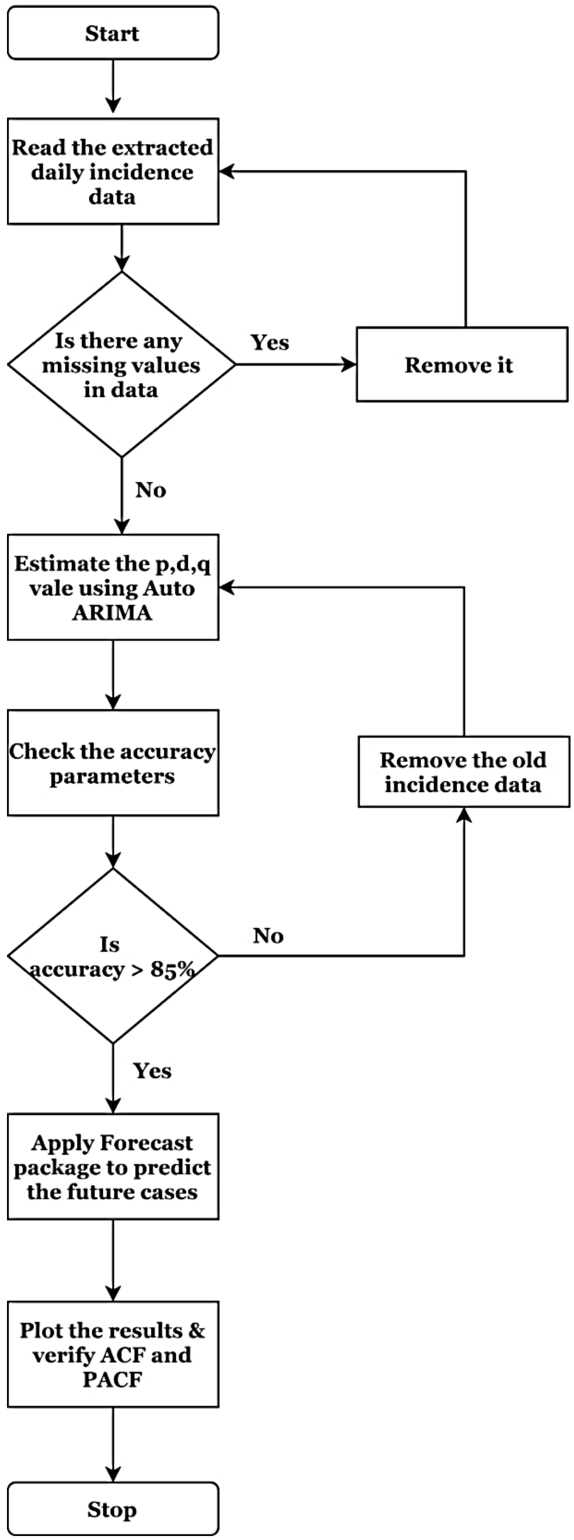


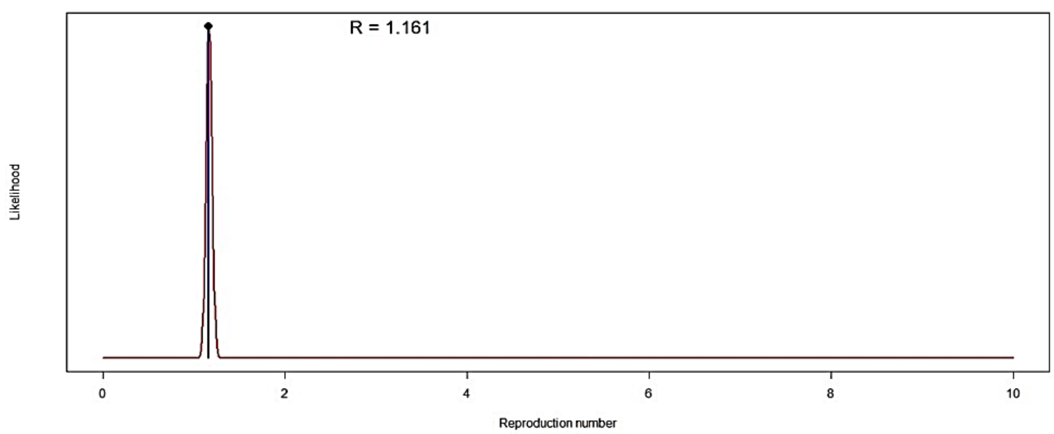
Figure 3 Flowchart of ARIMA model.

Methodology evaluation

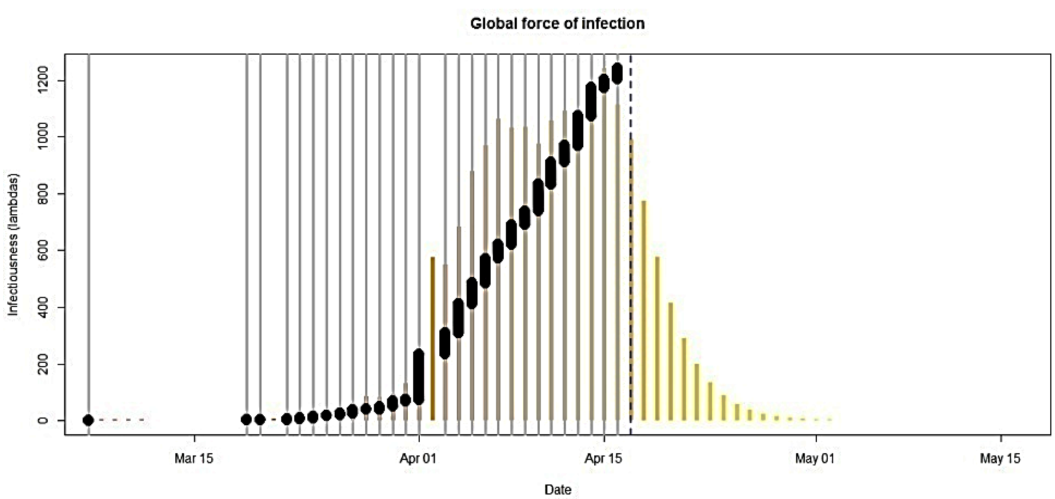
Real-time data assessment of the 'earlyR' pandemic model

The data of Tamil Nadu state from March 7th to April 16th, 2020, were taken into account for the development. We adopted the COVID-19 and utilized the average and variability of COVID-19 durations, specifically 4.7 days for the mean and 2.9 days for the standard deviation, as reported by Nishiura *et al.* [24]. **Figure 4(a)** demonstrates that for the aforementioned time period has a probability R_0 value of 1.161. **Figure 4(b)** shows the global force of infection, and **Figure 4(c)** displays the expected R_0 values of the likelihood ratio test based on a bootstrap resampling procedure with 1,000 samples. The minimum, median, mean and maximum values with a 95 % CI are 1.061, 1.161, 1.162 and 1.281, respectively. **Figure 4(d)** depicts the epidemic trajectory over the following 10 days in different quantiles. Within a 95 % confidence interval (CI), the incidence rates for the following 10 days are as follows: Lowest = 607; median = 806; mean = 804.9; maximum = 1,028. Reports of actual cases throughout the expected time frame totaled 1,242. With a median projected value, the 'earlyR' epidemic model yields an inaccuracy of 35.1 %. The percentage error is calculated using the Eq. (5) given below.

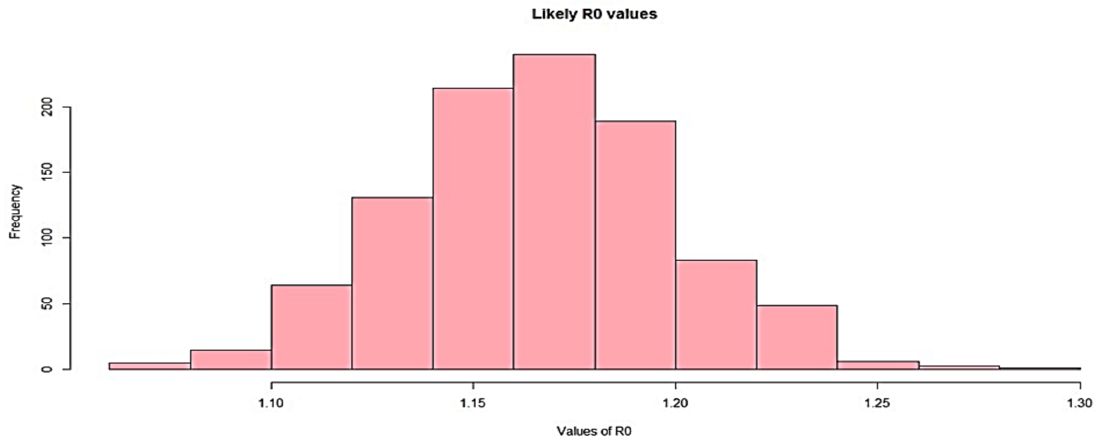
$$\%\ \text{error} = \frac{Actual\ Cummulative\ Incidence - Predicted\ Cummulative\ Incidence}{Actual\ Cummulative\ Incidence} \times 100 \quad (5)$$



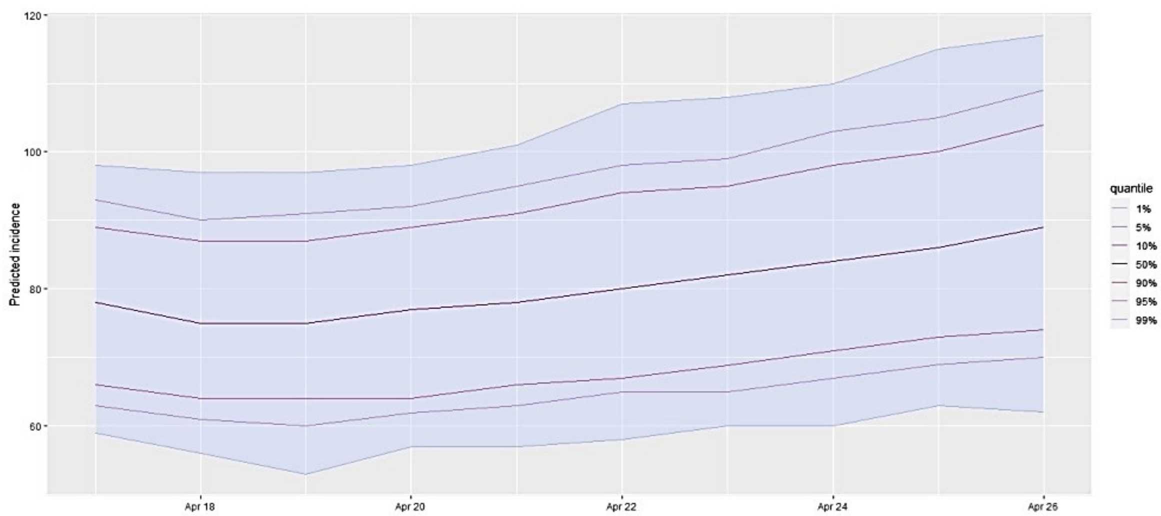
(a)



(b)



(c)



(d)

Figure 4 (a) The likelihood R_0 value of Tamil Nadu data, (b) global force of infection, (c) forecasted likely R_0 values for the next 10 days and (d) incidence estimation for various quantiles for the next 10 days.

Real-time data assessment of the ARIMA model

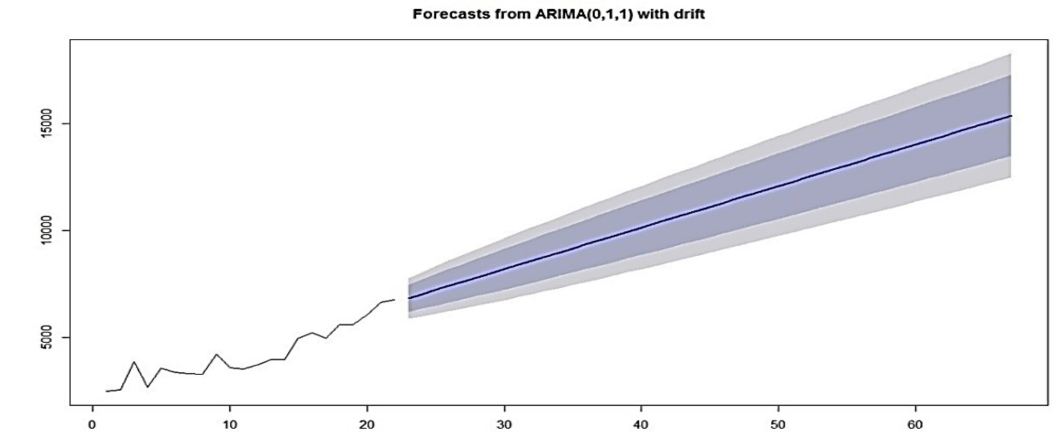
We utilized the 'auto.arima' function in R to develop the forecasting model, the COVID-19 data of India from May 25th, 2020 to June 16th, 2020 has been used to develop ARIMA model. The actual and predicted new cases, along with 80 and 95 % CIs, are illustrated in **Figure 5(a)**, with the black line representing actual occurrences and the blue line depicting the predicted ones. The blue and grey shaded areas correspond to 80 and 95 % confidence intervals, respectively. The ACF and PACF plots of the ARIMA (0, 1 and 1) model are displayed in **Figures 5(b)** and **5(c)**, respectively. **Table 2** summarizes the performance metrics of the ARIMA (0, 1 and 1) model, with abbreviations such as MAE, MASE, ME, MPE and RMSE. **Table 3** provides actual reported cases and predicted values, with a deviation range of −11.27 to 11.61 %.

Table 2 Performance parameters of training set.

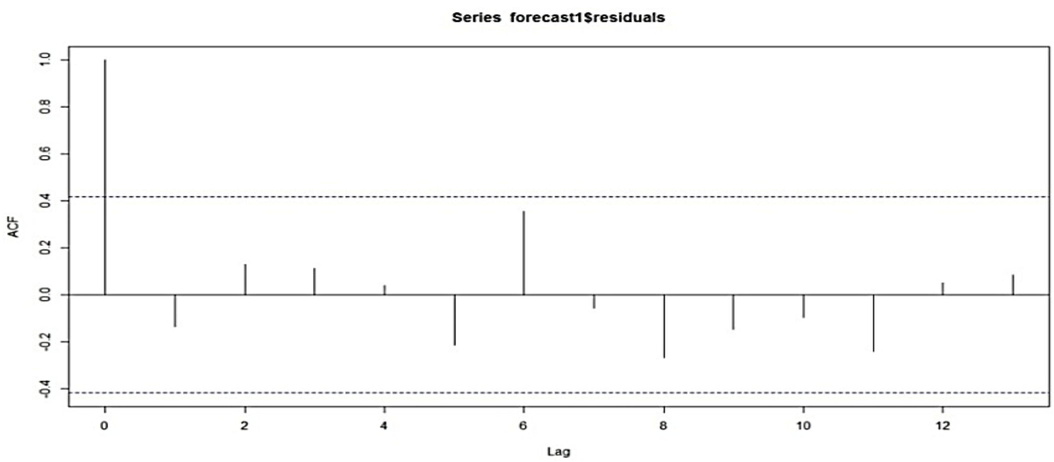
Model	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
ARIMA (0, 1 and 1) training set	5.89	449.67	375.49	−1.54	9.51	0.86	−0.135

Table 3 Comparison of reported and anticipated incidents.

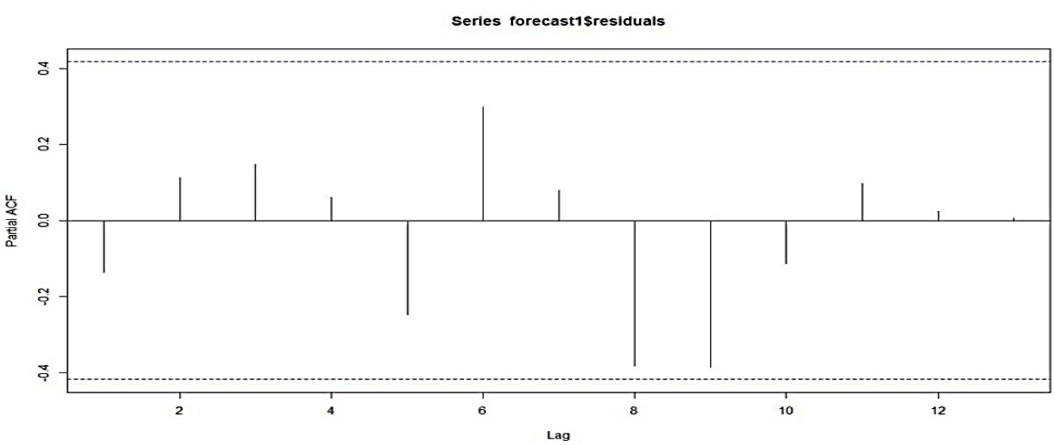
Day	Predicted cases	Actual cases	% error
25-05-2020	6,838	6,977	−2.03
26-05-2020	7,032	6,535	7.06
27-05-2020	7,226	6,387	11.61
28-05-2020	7,420	6,566	11.50
29-05-2020	7,614	7,466	1.94
30-05-2020	7,809	7,964	−1.98
31-05-2020	8,003	8,380	−4.71
01-06-2020	8,197	8,392	−2.37
02-06-2020	8,391	8,171	2.62
03-06-2020	8,585	8,909	−3.77
04-06-2020	8,779	9,304	−5.98
05-06-2020	8,973	9,851	−9.78
06-06-2020	9,167	9,887	−7.85
07-06-2020	9,362	9,971	−6.50
08-06-2020	9,556	9,983	−4.46
09-06-2020	9,750	9,987	−2.43
10-06-2020	9,944	9,985	−0.41
11-06-2020	10,138	9,996	1.40
12-06-2020	10,332	10,956	−6.03
13-06-2020	10,526	11,458	−8.85
14-06-2020	10,720	11,929	−11.27
15-06-2020	10,914	11,502	−5.38
16-06-2020	11,109	10,667	3.97
	Mean error		−1.9 %



(a)



(b)



(c)

Figure 5 (a) Model evaluation results based on historical data and projections, (b) ACF plot of ARIMA model and (c) PACF plot.

Discussion of findings

Our findings suggest that the ARIMA model outperforms the 'earlyR' epidemic model in terms of accuracy. The reproduction rate is the primary factor determining the expected daily incidence in the future. Unless strict measures are taken to monitor and control infections, the number of future cases may reach its maximum value. Our research can aid government officials in taking necessary preventive measures, such as implementing quarantines, travel restrictions, social distancing and restricting mass gatherings or employing no-contact greetings to prevent the spread of the epidemic. Quarantine is an effective measure for protecting the public, involving the separation and restriction of contagious disease-exposed individuals. Organizations, universities, societies and governments implement public health and social measures (PHSMs) to control or mitigate the spread of contagious illnesses such as COVID-19. Social stigma may arise from fear, worry and lack of understanding of the disease [26]. Rapid diagnosis and patient isolation may help slow the transmission of COVID-19 and reduce the reproduction number.

Table 4 compares the accuracy of existing models for predicting the spread of COVID-19, including a neural network-powered model by Wieczorek *et al.* [27] and ensemble models by Kallel *et al.* [28], such as AdaBoost, Random Forest and Gaussian Naive Bayes. Our proposed approach, an ARIMA model, achieved a mean error of −1.9 % and a mean absolute error of 483.56. The limitations of our study are as follows: Early forecasting models might be influenced by limited data, which can affect the accuracy of predictions, especially considering the dynamic nature of the COVID-19 pandemic. The 'earlyR' epidemic model relies on assumptions about the virus's behavior and population characteristics, which might not fully capture the evolving nature of the pandemic or the impact of new variants.

Table 4 Comparison of COVID-19 spread forecasting models.

Author(s)	Model	Accuracy
Wieczorek *et al.* [27]	Neural network powered COVID-19 spread forecasting model	87.70 %
Kallel *et al.* [28]	AdaBoost (n = 100)	89.35 %
	Random Forest (n = 50)	90.04 %
	Gaussian Naive Bayes	89.81 %
Proposed approach	ARIMA Model	Mean Error is −1.9 %, and MAE is 483.56

Conclusions

This study developed and compared 2 epidemic models, the ARIMA and the 'earlyR' epidemic model, to predict the occurrence of COVID-19 in India. The ability to evaluate R_0 and predict the probable cases in the near future is critical to restricting the spread of the disease. It is important to note that R_0 may differ for different infectious diseases and populations. Using the 'earlyR' model, we determined the minimum, median, mean and maximum projected values of R_0 with a 95 % CI, as well as the expected incidences for different quantiles. However, our comparison showed that the ARIMA model provides more accurate predictions. Our proposed ARIMA model was compared with existing forecasting models for predicting the spread of COVID-19, including a neural network-powered model and ensemble models. Our model achieved a mean error of −1.9 % and a mean absolute error of 483.56, outperforming the other models.

Our study suggests that health policymakers can use the ARIMA model to forecast the incidence of COVID-19 in India and take required actions to prevent the transmission of the virus. Furthermore, our findings highlight the importance of accurately estimating R_0 and predicting probable cases for controlling the spread of epidemics. Future research could focus on improving the accuracy of epidemic models and exploring the use of deep learning and other advanced techniques in predicting the incidence of infectious diseases.

References

[1] World Health Organization. *Coronavirus disease 2019 (COVID-19) situation report - 90.* World Health Organization, Geneva, Switzerland, 2020.
[2] J Hellewell, S Abbott, A Gimma, NI Bosse, CI Jarvis, TW Russell, JD Munday, AJ Kucharski and WJ Edmunds. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Global Health* 2020; **8**, E488-E496.
[3] FM Guerra, S Bolotin, G Lim, J Heffernan, SL Deeks, Y Li and NS Crowcroft. The basic reproduction number (R_0) of measles: A systematic review. *Lancet Infect. Dis.* 2017; **17**, E420-E428.
[4] M D'Arienzo and A Coniglio. Assessment of the SARS-CoV-2 basic reproduction number, R_0, based on the early phase of COVID-19 outbreak in Italy. *Biosafety Health* 2020; **2**, 57-9.
[5] G Chowell, H Nishiura and LMA Bettencourt. Comparative estimation of the reproduction number for pandemic influenza from daily case notification data. *J. R. Soc. Interface* 2007; **4**, 154-66.
[6] YC Wu, CS Chen and YJ Chan. The outbreak of COVID-19: An overview. *J. Chin. Med. Assoc.* 2020; **83**, 217-20.
[7] E Rezel-Potts, A Douiri, X Sun, PJ Chowienczyk, AM Shah and MC Gulliford. Cardiometabolic outcomes up to 12 months after COVID-19 infection. A matched cohort study in the UK. *PloS Med.* 2022; **19**, e1004052.
[8] CK Wibmer, F Ayres, T Hermanus, M Madzivhandila, P Kgagudi, B Oosthuysen, BE Lambson, TD Oliveira, M Vermeulen, KVD Berg, T Rossouw, M Boswell, V Ueckermann, S Meiring, AV Gottberg, C Cohen, L Morris, JN Bhiman and PL Moore. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *Nat. Med.* 2021; **27**, 622-5.
[9] SK Yadav and Y Akhter. Statistical modeling for the prediction of infectious disease dissemination with special reference to COVID-19 spread. *Front. Publ. Health* 2021; **9**, 645405.
[10] J Pang, MX Wang, IYH Ang, SHX Tan, RF Lewis, JIP Chen, RA Gutierrez, SXW Gwee, PEY Chua, Q Yang, XY Ng, RKS Yap, HY Tan, YY Teo, CC Tan, AR Cook, JCH Yap and LY Hsu. Potential rapid diagnostics, vaccine and therapeutics for 2019 novel coronavirus (2019-nCoV): A systematic review. *J. Clin. Med.* 2020; **9**, 623.
[11] S Xu and Y Li. Beware of the second wave of COVID-19. *Lancet* 2020; **395**, 1321-2.
[12] L Zhong, L Mu, J Li, J Wang, Z Yin and D Liu. Early prediction of the 2019 novel coronavirus outbreak in the mainland China based on simple mathematical model. *IEEE Access* 2020; **8**, 51761-9.
[13] K Mizumoto and G Chowel. Transmission potential of the novel coronavirus (COVID-19) onboard the diamond Princess Cruises Ship, 2020. *Infect. Dis. Model.* 2020; **5**, 264-70.
[14] C Quick, R Dey and X Lin. Regression models for understanding COVID-19 epidemic dynamics with incomplete data. *J. Am. Stat. Assoc.* 2021; **116**, 1561-77.
[15] M Ritter, DVM Ott, F Paul, JD Haynes and K Ritter. COVID-19: A simple statistical model for predicting intensive care unit load in exponential phases of the disease. *Sci. Rep.* 2021; **11**, 5018.
[16] GR Shinde, AB Kalamkar, PN Mahalle, N Dey, J Chaki and AE Hassanien. Forecasting models for coronavirus disease (COVID-19): A survey of the state-of-the-art. *SN Comput. Sci.* 2020; **1**, 197.
[17] D Haritha, N Swaroop and M Mounika. Prediction of COVID-19 cases using CNN with x-rays. *In*: Proceedings of the 5th International Conference on Computing, Communication and Security, Patna, India. 2020.
[18] Q Guo and Z He. Prediction of the confirmed cases and deaths of global COVID-19 using artificial intelligence. *Environ. Sci. Pollut. Res.* 2021; **28**, 11672-82.
[19] P He. Study on epidemic prevention and control strategy of COVID-19 based on personnel flow prediction. *In*: Proceedings of the International Conference on Urban Engineering and Management Science, Zhuhai, China. 2020, p. 688-91.
[20] MH Alsharif, MK Younes and J Kim. Time series ARIMA model for prediction of daily and monthly average global solar radiation: The case study of Seoul, South Korea. *Symmetry* 2019; **11**, 240.
[21] Ministry of Health and Family Welfare, Available at: https://www.mohfw.gov.in, accessed January 2023.
[22] A Cori, NM Ferguson, C Fraser and S Cauchemez. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am. J. Epidemiol.* 2013; **178**, 1505-12.

[23] Q Li, X Guan, P Wu, X Wang, L Zhou, Y Tong, R Ren, KSM Leung, EHY Lau, JY Wong, X Xing, N Xiang, Y Wu, C Li, Q Chen, D Li, T Liu, J Zhao, M Liu, ..., Z Feng. Early transmission dynamics in Wuhan, China, of novel coronavirus - infected pneumonia. *New Engl. J. Med.* 2020; **382**, 1199-207.
[24] H Nishiura, NM Linton and AR Akhmetzhanov. Serial interval of novel coronavirus (COVID-19) infections. *Int. J. Infect. Dis.* 2020; **93**, 284-6.
[25] T Jombart, P Nouvellet, S Bhatia and ZN Kamvarprojections. Project future case incidence, Available at: https://CRAN.R-project.org/package=projections, accessed January 2023.
[26] World Health Organization. Advice for the public: Coronavirus disease (COVID-19), Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public, accessed January 2023.
[27] M Wieczorek, J Siłka and M Woźniak. Neural network powered COVID-19 spread forecasting model. *Chaos Solitons Fractals* 2020; **140**, 110203.
[28] A Kallel, M Rekik and M Khemakhem. Hybrid-based framework for COVID-19 prediction via federated machine learning models. *J. Supercomput*. 2022; **78**, 7078-10.