Hindawi

Research Article

Comparison of Conventional Modeling Techniques with the Neural Network Autoregressive Model (NNAR): Application to COVID-19 Data

Muhammad Daniyal,[1] Kassim Tawiah ⓘ,[2,3] Sara Muhammadullah,[4,5] and Kwaku Opoku-Ameyaw[2]

[1]*Department of Statistics, Islamia University of Bahawalpur, Bahawalpur, Pakistan*
[2]*Department of Mathematics and Statistics, University of Energy and Natural Resources, Sunyani, Ghana*
[3]*Department of Statistics and Actuarial Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana*
[4]*Pakistan Institute of Development Economics, Islamabad, Pakistan*
[5]*National Institute of Health, Islamabad, Pakistan*

Correspondence should be addressed to Kassim Tawiah; kassim.tawiah@uenr.edu.gh

The coronavirus disease 2019 (COVID-19) pandemic continues to destroy human life around the world. Almost every country throughout the globe suffered from this pandemic, forcing various governments to apply different restrictions to reduce its impact. In this study, we compare different time-series models with the neural network autoregressive model (NNAR). The study used COVID-19 data in Pakistan from February 26, 2020, to February 18, 2022, as a training and testing data set for modeling. Different models were applied and estimated on the training data set, and these models were assessed on the testing data set. Based on the mean absolute scaled error (MAE) and root mean square error (RMSE) for the training and testing data sets, the NNAR model outperformed the autoregressive integrated moving average (ARIMA) model and other competing models indicating that the NNAR model is the most appropriate for forecasting. Forecasts from the NNAR model showed that the cumulative confirmed COVID-19 cases will be 1,597,180 and cumulative confirmed COVID-19 deaths will be 32,628 on April 18, 2022. We encourage the Pakistan Government to boost its immunization policy.

1. Introduction

One of the brutal pandemics in human history, coronavirus disease 2019 (COVID-19), has caused millions of human fatalities around the world and continues to rage havoc worldwide since its outbreak in 2019. The pandemic has reshaped scientific thinking and study. Scientists around the world continue to study various variants of this deadly disease to devise strategies to eliminate it from the human race. The different variant of the virus has made it even more hectic for vaccine manufacturers. Vaccinated individuals even get infected with the virus but with a lower risk of dying compared to the unvaccinated [1, 2]. Numerous modeling and forecasting techniques have been proposed for COVID-19 confirmed cases and deaths.

Anwar and Mokhtar [3] utilized an Epidemic Calculator that uses a susceptible, exposed, infected, and recovered (SEIR) compartmental model with information from the Egyptian Ministry of Health and Population. For the most elevated assessed case mortality rate (7.7%), the number of individuals admitted in hospitals was anticipated to top in the middle of June, with a sum of 20,126 in the hospitals and an anticipated death total of 12,303. Statistical modeling and machine learning techniques were applied to foresee and gauge the completion phase of COVID-19 utilizing different time contamination rates and individual numbers of

contacts [4, 5]. Their outcomes indicated that the assessed generation number was 2.2 in Kuwait, with the contact rate among the populace on the high side, denoting an epidemic top value unlikely to be reached and the nation requiring a more severe mediation course of action.

El Desouky [6] forecasted the pinnacle, duration, and reenactment of possible varieties that may be occurring in the social ways and behavior of Egyptians in the sacred season of Ramadan. They recommended three perceived numerical methods (i.e., Euler's method and Runge Kutta method of request two (RK2) and of request four (RK4)) for tackling such conditions of health care globally and subsequently making significant sources of information available. Benkouiten et al. [7] were optimistic that Hajj pilgrims played a key role in the dispersion of the pandemic. Numerical outcomes might be utilized to figure out the number of vulnerable persons to the disease, recuperated, and isolated persons in the long run to help unfamiliar endeavors to develop their mediation benefits and further anticipation. Numerical methodologies and calculated models [8, 9] have been utilized for analyses and understanding of COVID-19.

Pirouz et al. [10] concentrated on the arrangement of confirmed instances of COVID-19 utilizing an Artificial Intelligence (AI) strategy, local area information steering arrangement of the neural network, by adapting a twofold characterization modeling. The proposed model depends on a contextual analysis of China's Hubei territory. A few significant parameters like greatest daily temperature, least daily temperature, normal day-to-to-day temperature, density, relative density, and speed of the wind as well as the quality of the air [11] were parsed as the informational index and picked the number of affirmed cases as the result of information collection for thirty days. They were of the view that the parallel order model gives more prominent ability to exactness in anticipating the announced cases. Besides, they played out the relapse analysis and the example of revealed cases relative to the variety of everyday climatic conditions (speed of the wind, relative density, and normal temperature). Their outcome pointed out that the relative density and the most extreme everyday temperature greatly affected the actual cases. The examination of the observed confirmed COVID-19 cases using machine learning approaches revealed that the variable number of tests in a particular country did not assume any crucial part in the expectation of the aggregate number of confirmed cases [12]. Pham et al. [13] provided a new AI version and a large data application to properly comprehend the situation of COVID-19 and provided alternatives in ceasing COVID-19 outburst to manage the viral mutation spread.

Ranjan [14] compared data on the COVID-19 upsurge in India and multiple countries together with key counties in the United States (US) and noted that India's first number of reproductions, R_0, is anticipated to be around $1.4 \times 10^{3.9}$. At the time, the growth ring of India's infection and that of Washington and California were close. Traditional and integrated models of susceptible-infected-recovered (SIR) model, depending on the data recently organized, were applied to render a recurring short-ring and long-term prognosis. The SIR model estimated India's stability by the end of May, 2020, with a proposed final size of the epidemic around 13,000, although the approximation will be invalid in the instance that India enters the group transmission point. By the application of a similar model, Italy was assumed to reach its pandemic peak on March 21, 2020 [15].

Admittance to real-time information and the powerful use of episode expectation or estimating models are central to getting quick data with respect to the transmission elements of the infection and its ramifications. Besides, every flare-up has novel transmission qualities that are unique in relation to different episodes, which brings up the issue of how standard expectation models would act in delivering precise outcomes. Moreover, different elements including the number of known and obscure factors, contrasts in populace/behavioural intricacies in different geopolitical regions, and variety in control procedures influence the vulnerability of forecast models [16]. Thus, it is challenging for standard epidemiological models like susceptible-infected-recovered (SIR) to give reliable outcomes to long-haul forecasts. Hence, it is vital to not just review the relationship between the parts of the episode data sets but also evaluate the adequacy of the normal sickness expectation models.

As of late, there have been a handful of works that attempt to understand the spread of COVID-19 as well as predict confirmed cases and deaths of COVID-19, especially making use of statistical methodologies. For example, Kucharski et al. [17] investigated a blend of stochastic transmission models on four data sets that caught the everyday number of new cases, the day-to-day number of new internationally sent out cases, the extent of contaminated travelers on departure flight, and the quantity of new confirmed cases to appraise the transmission elements of the illness throughout some time [18]. Machine learning-based model has been applied to analyze and predict the growth of COVID-19 [19]. Guo and He [20] utilized AI to predict cases and deaths attributed to COVID-19 globally. Models of the Markov chain have been availed to predict COVID-19 spread based on secondary data as of March 13, 2020. Xu et al. [21] and Arumugam and Raji [22] utilized Markov models to predict the impact of the coronavirus on the human race using probability matrices and Monte Carlo simulation. Bertozzi et al. [23] opined that the COVID-19 pandemic has put epidemic modeling at the lead of international public policy making.

Al-qaness et al. [24] put forward an updated version of the adaptive neuro-fuzzy inference system (ANFIS) applying an amplified flower pollination algorithm (FPA) after implementing the salp swarm algorithm (SSA). Wu et al. [25] deduced that the COVID-19 epidemic is now filling dramatically in different significant urban areas of China with a fall time behind the Wuhan episode of around one to fourteen days using the susceptible-exposed-infectious-recovered metapopulational model in a Markov Chain Monte Carlo framework. A blended nonlinear assessment approach consolidating the Gaussian process (GP) and unscented Kalman filter (UKF) was suggested to anticipate the dynamic changes in wind speed and further develop the forecasting accuracy [26]. Zhao et al. [27] predicted new COVID-19 cases in a US state using Poisson and gamma distributions.

Hao et al. [28] utilized the advancement pattern investigation of confirmed COVID-19 cumulative cases, cumulative deaths, and cumulative recovered cases in Wuhan from January 23, 2020, to April 6, 2020, by implementing an Elman neural network, long short-term memory (LSTM), and support vector machine (SVM) for future predictions.

Time-series models have been broadly applied to COVID-19 data. Tawiah et al. [29] proposed zero-inflated time-series model for COVID-19 deaths in Ghana. Luo et al. [30] used LSTM and XGBoost algorithms to predict COVID-19 transmission in America using time series. Gecili et al. [31] forecasted COVID-19 confirmed deaths, recovery, and cases in the USA and Italy through the application of novel time-series modeling. Barría-Sandoval et al. [32] predicted COVID-19 cases in Chile by employing time-series techniques. Chyon et al. [33] applied machine learning techniques to autoregressive integrated moving average (ARIMA) models [34] to predict COVID-19 cases. Ali et al. [35] suggested that ARIMA models are suitable for epidemic forecasting. Doornik et al. [36] depicted how to disintegrate the detailed time series of COVID-19 confirmed cases and deaths into a trend, seasonal, and irregular component utilizing machine learning approaches. Nevertheless, forecasting and modeling escalation of COVID-19 persist as a challenge. Therefore, other time-series methods can be explored to forecast confirmed COVID-19 cases and deaths.

In the time-series domain, improving forecasting accuracy is an important and often tricky task confronting data analysts in different areas. Although many time-series models are available in the literature, the study for boosting the ability of prediction models has never stopped. In this paper, we model and forecast the confirmed cumulative COVID-19 cases and deaths in Pakistan based on Box–Jenkins time series, ARIMA model, and neural network autoregressive (NNAR) model vis-a-vis other competing models, thereby comparing them. The proposed model forecast will go a long way to help authorities to develop new strategies to combat the pandemic in Pakistan.

In the subsequent sections of the paper, we present the materials and methods applied, the results, and discussion of the statistical modeling vis-a-vis the conclusions of the study.

2. Materials and Methods

2.1. Data. The data used in this study consist of new confirmed COVID-19 cases and deaths in Pakistan from the first reported case on February 26, 2020, to February 18, 2022, provided by the COVID-19 Health Platform of the Ministry of National Health Services Regulation, Government of Pakistan. We utilized cumulative data on the confirmed cases and deaths. It can be noted from Figure 1 that the cumulative cases and deaths show exponential growth with respect to time, so nonseasonal ARIMA modeling can be used to forecast the trend of current COVID-19 cases and deaths. The summary statistics of the data used in the study are presented in Table 1. It can be observed that the average daily confirmed COVID-19 cases were 2064, and the average

daily deaths attributed to COVID-19 were 41 from February 26, 2020, to February 18, 2022. The minimum daily confirmed cases and daily deaths were 0, respectively, while the maximum confirmed daily cases were 8183 and the maximum daily death was 313.

2.2. Methods. The Box–Jenkins ARIMA (p, d, q) [37] is given by

$$\widehat{X}_t = \mu + \alpha_1 \widehat{X}_{t-1} + \cdots + \alpha_p \widehat{X}_{t-p} + \cdots + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t,$$

(1)

where $\alpha_1 \widehat{X}_{t-1}, \ldots, \alpha_p \widehat{X}_{t-p}$ are the lagged values and $\theta_1 \epsilon_{t-1}, \ldots, \theta_q \epsilon_{t-q}$ are the lagged errors of the series $\widehat{X}_t$. The constants p, d and q represent the order of the autoregressive term, the degree of differencing series, and the order of the moving average term, respectively. ϵ_t is the white noise with mean 0 and variance σ^2. $\widehat{X}_t$ can be differenced once or more.

The Box–Jenkins multiplicative seasonal ARIMA model [37–40] represented by ARIMA $(p, d, q) \times (P, D, Q)$ is given by

$$\beta_p(K)\gamma_P(K^f)(1-K)^d(1-K^f)^D \widehat{X}_t = \omega + \alpha_q(K)\varphi_Q(K^f)\epsilon_t, \quad (2)$$

with $\beta_p(K) = 1 - \beta_1 K - \cdots \beta_p K^P$; $\gamma_P(K^f) = 1 - \gamma_1 K^f - \cdots -\gamma_P K^{Pf}$; $\alpha_q(K) = 1 - \alpha_1 K - \cdots \alpha_q K^q$; and $\varphi_Q(K^f) = 1 - \varphi_1 K^f - \cdots - \varphi_Q K^{Qf}$, where K is the operator balanced shift and f is the frequency of seasonality. D and d are the seasonal difference and ordinary differencing degrees, respectively. $\beta_p(K)$ and $\gamma_P(K^f)$ are the regular autoregressive polynomial of order p and seasonal autoregressive polynomial of order P, respectively. Also, $\alpha_q(K)$ and $\varphi_Q(K^f)$ are the polynomials of regular moving average of order q and seasonal moving average of order Q, respectively. Similarly, $\omega = \rho(1 - \beta_1 - \cdots \beta_p)(1 - \gamma_1 - \cdots - \gamma_P)$, where the mean of the process $(1 - K)^d(1 - K^f)^D \widehat{X}_t$ is ρ. ϵ_t is the white noise with mean 0 and variance σ^2.

Shunway and Stoffer [38] proposed that to maintain casualty and investibility, the solution set of all polynomials in the multiplicative model must be outside the unit circle. For simplicity, we assumed $\omega = 0$. Thus, we selected the most apt values of p, d, q, P, D, and Q by calculating and examining the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of our data by graphing the time series and identifying any unusual data points as well as selecting the appropriate transformation of the variance stabilization. We determined the order of p, q, P, and Q by the examination of the ACF and PACF [41]. We employed the portmanteau test for the residual analysis to check for autocorrelation. For an adequate model, the errors are expected to be uncorrelated or white noise [42]. The portmanteau test confirms the ACF residual plots, PACF residual plot, and the normal probability plot.

The model with the least root mean square error (RMSE) and mean absolute error (MAE) is selected as the most appropriate for our data. The expressions RMSE and MAE are
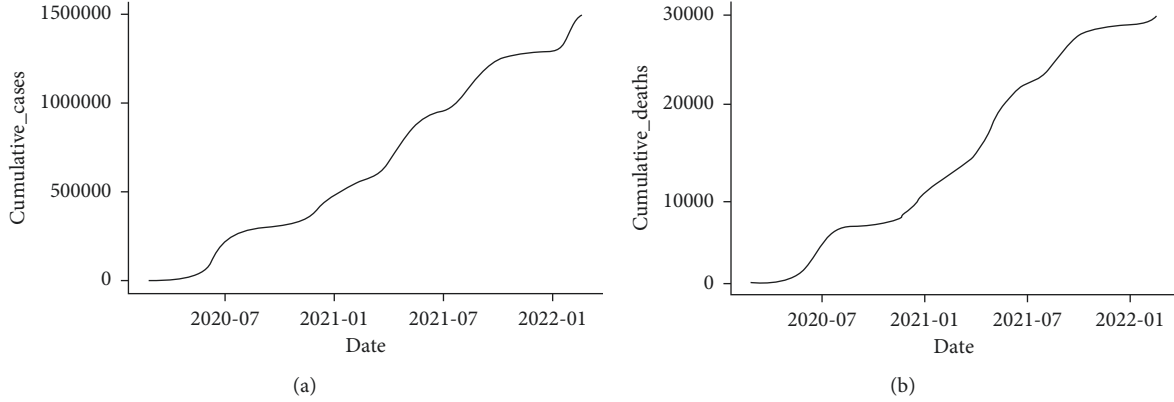
FIGURE 1: Confirmed cumulative daily cases (a) and cumulative daily deaths (b) of COVID-19 in Pakistan from February 26, 2020 to February 18, 2022 (source: https://covid.gov.pk).

TABLE 1: Summary statistics of daily confirmed cases and deaths of COVID-19 from February 26, 2020, to February 18, 2022.

	Daily cases	Daily deaths
Minimum	0	0
Maximum	8183	313
Mean	2064	41
Median	1589	31
Lower quartile	626	10
Upper quartile	3101	62
Standard deviation	1712	37

$$
\begin{aligned}
\text{RMSE} &= \sqrt{\frac{1}{T-N} \sum_{t=N+1}^{T} \left(X_t - \widehat{X}_t\right)^2}, \\
\text{MAE} &= \frac{1}{T-N} \sum_{t=N+1}^{T} \left|X_t - \widehat{X}_t\right|,
\end{aligned}
\tag{3}
$$

where $X_1, \ldots, X_N$ and $X_{N+1}, \ldots, X_T$ are the partitions of the data. These metrics summarize as well as assess the quality of the model. The smaller the value, the better the model with a superior quality for forecasting.

We used the Dickey–Fuller (DF) test, the Phillips–Perron (PP) test, and Augmented Dickey–Fuller (ADF) test, which are unit root tests, to check whether our data are stationary or not. Violations were corrected to meet all necessary assumptions of the model.

2.2.1. Neural Network Autoregressive Modeling.

We focused on the NNAR model with a hidden layer selected automatically throughout the modeling process. Lagged values of the time series can really be employed as input data to a neural network with time-series data, exactly as it is done with lagged values in a linear autoregressive model. When this is done, the model is referred to as an NNAR model. An NNAR (p, kp, k) denotes the hidden layer has pp delayed inputs and kk nodes. Moreover, NNAR $(p, 0p, 0)$ model is the same as an ARIMA $(p, 0p, 0)$ but without parameter

limitations that assure stationarity. The NNAR (p, kp, k) [43, 44] is represented by

$$
f(X) = \beta_0 + \sum_{k=1}^{K} \beta_k g\left(w_{k0} + \sum_{j=1}^{p} w_{kj} X_j \right).
\tag{4}
$$

The expression is constructed in two stages. The K activations come first. In the activation, $A(k)$, $k = 1, \ldots, K$, the hidden layer is calculated as a function of the input characteristics $X_j = X_{t-1}, \ldots, X_{t-p}$, with

$$
A(k) = h(k) = g\left(w_{k0} + \sum_{j=1}^{p} w_{kj} X_j \right),
\tag{5}
$$

where g is a previously defined nonlinear activation function. Each $A(k)$ may be seen as a separate $h_k(X)$ transformation of the unique characteristics. The output layer receives these K instigations from the hidden layer.

$$
f(X) = \beta_0 + \sum_{k=1}^{K} \beta_k A(k).
\tag{6}
$$

Our survival dependent variable contains the output in the form of 0 (fatal) and 1 (alive). In NNAR modeling, the sigmoid activation function (identical to logistic regression) is used to translate a linear function that converts the probability from 0 to 1 [45]. This sigmoid activation function is of the form

$$
\begin{aligned}
g(z) &= \frac{\exp(z)}{1 + \exp(z)}, \\
&= \frac{1}{1 + \exp(-z)}.
\end{aligned}
\tag{7}
$$

All modeling and forecasting were done in R [46].

3. Results and Discussion

As illustrated in Figure 2, the cumulative series of the confirmed COVID-19 cases retain a trend after detrending the data set. The new daily confirmed cases of COVID-19 series (Figure 2) can also reflect a unit root problem. In other
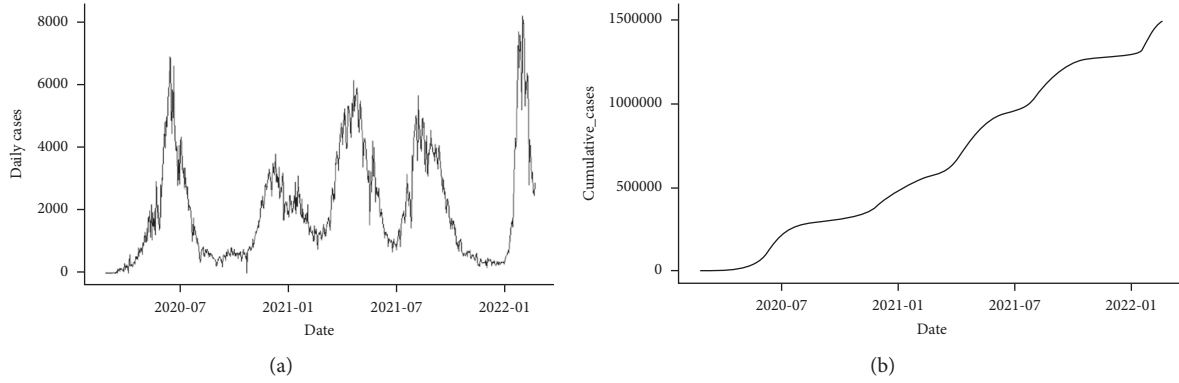
(a)

(b)

Figure 2: Trend of COVID-19 daily cases (a) and cumulative daily cases (b) in Pakistan from February 26, 2020, to February 18, 2022.
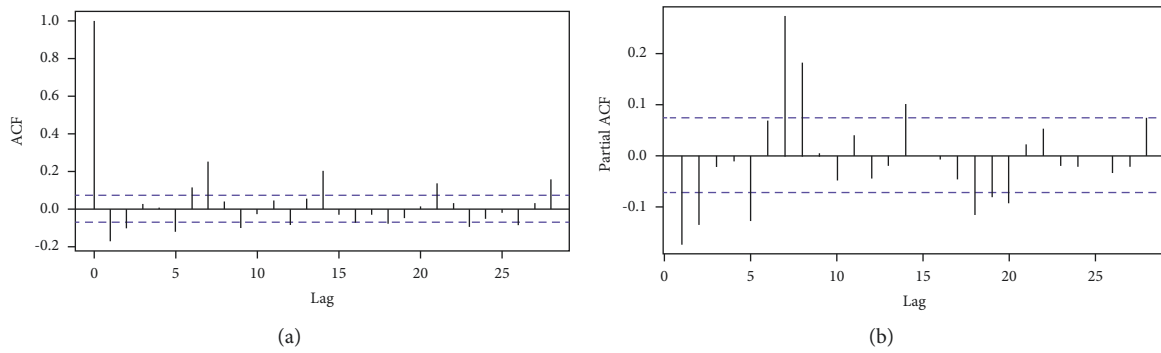

(a)

(b)

Figure 3: COVID-19 daily cases with ACF (a) and PACF (b).

words, the statistical properties such as mean, variance, and covariance of the original series are not constant over time. To remove this pattern from the data, we take the difference of the new daily cases. In Figure 3, the new case ACF plot illustrates a moving average (MA) pattern and the PACF plot indicates an autoregressive (AR) pattern. This, therefore, calls for the application of stationary series in further modeling and forecasting. A specific pattern in the ACF and PACF plots corresponds to a particular order of p and q. We decoupled our data set into two parts, namely, training and testing, with 90% of the data for training and 10% for testing to access the model accuracy [47].

Just as in the daily confirmed cases and cumulative confirmed cases, a similar pattern was observed in the daily confirmed deaths and cumulative confirmed deaths due to COVID-19 (Figures 4 and 5). As a result, the same modeling and forecasting procedure was applied here just as in the confirmed cases above.

The estimated ARIMA model for daily confirmed cases, possessing two autoregressive (AR) and two moving average (MA) terms as illustrated in Figure 3 and integrated of order 1, is given by

$$\widehat{x} = -0.202x_{t-1} - 0.0587x_{t-2} - 0.094\,e_{t-1} - 0.1014\,e_{t-2}. \quad (8)$$

Moreover, the estimated ARIMA model for daily confirmed deaths, possessing two autoregressive (AR) and two moving average (MA) terms as illustrated in Figure 3 and integrated of order 1, is given by

$$\widehat{x} = -0.869x_{t-1} - 0.0919x_{t-2} - 0.089\,e_{t-1} - 0.579\,e_{t-2}. \quad (9)$$

3.1. Forecast of Cumulative Confirmed Daily Cases of COVID-19 from February 19, 2022, to April 18, 2022.

From Table 2, the NNAR model had RMSE and MAE values of 195.3010 and 143.5501, respectively, for the training data set. For the testing data set, the NNAR model had RMSE and MAE values of 2136.0690 and 1589.5690, respectively. The NNAR model's RSME and MAE values for both the training and testing data sets were the least among the ARIMA models and the other competing models. This shows that the NNAR model has higher forecast quality and assesses the data better than the rest of the models [48], making it the most appropriate candidate model for predicting the cumulative daily confirmed COVID-19 cases. We, therefore, used the NNAR model to predict the cumulative daily confirmed cases of COVID-19 from February 19, 2022, to April 18, 2022. The forecast value as illustrated in Figure 6 shows that the cumulative daily cases of COVID-19 in Pakistan will be 1,597,810 on April 18, 2022.

3.2. Forecast of Cumulative Confirmed Daily Deaths of COVID-19 from February 19, 2022, to April 18, 2022.

The process for predicting cumulative deaths is identical to that for cumulative confirmed cases. From Table 3, the NNAR model had RMSE and MAE values of 10.36647 and 5.065002, respectively, for the training data set. For the
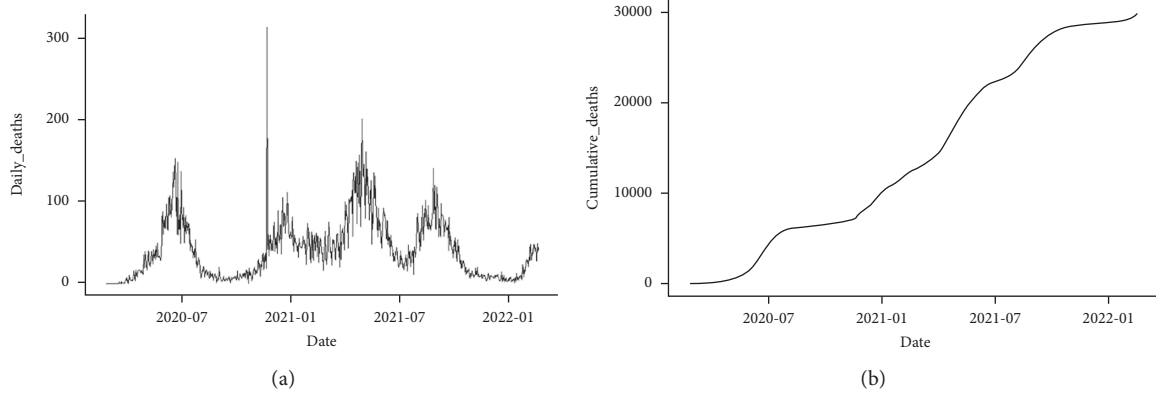
(a)

(b)

Figure 4: Trend of COVID-19 daily deaths (a) and cumulative daily deaths (b) in Pakistan from February 26, 2020, to February 18, 2022.
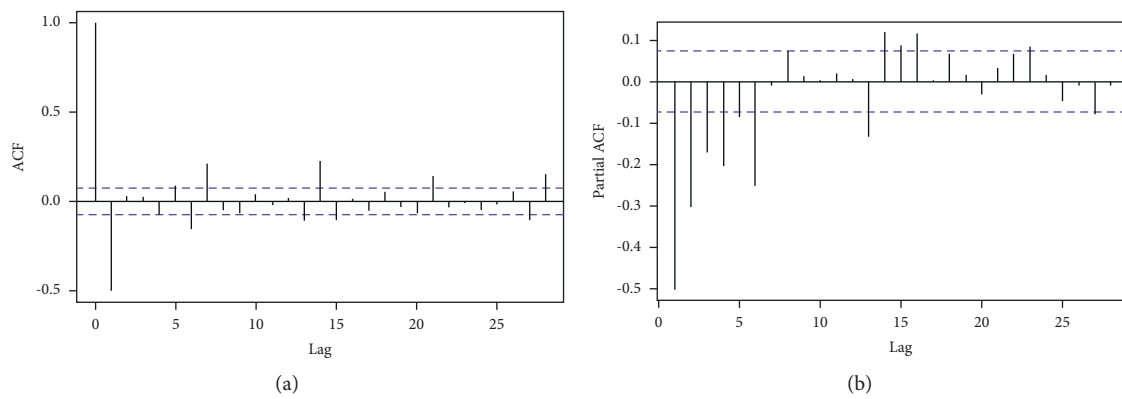

(a)

(b)

Figure 5: COVID-19 daily deaths with ACF (a) and PACF (b).

Table 2: Accuracy of different time-series models for predicting COVID-19 cumulative daily cases.

Method	Source	RMSE	MAE
Mean	Training	1566.384	1309.385
	Testing	2207.310	1862.713
Drift	Training	412.2189	284.6854
	Testing	2114.5660	1713.2369
Naïve	Training	412.2303	284.5769
	Testing	2177.6843	1688.9792
Holt	Training	394.6241	277.7294
	Testing	2191.3800	1740.4823
SES	Training	394.5822	277.6888
	Testing	2180.9038	1744.2653
ARIMA order SES	Training	392.7844	275.1783
	Testing	2182.1964	1754.7370
ARIMA (2, 2, 2)	Training	392.7844	275.1783
	Testing	2182.1964	1754.7370
ARIMA auto	Training	393.0668	275.3658
	Testing	2182.9865	1760.4807
NNAR	Training	195.301	143.5501
	Testing	2136.069	1589.569

Figure 6: Forecast of cumulative COVID-19 confirmed cases from February 19, 2022, to April 18, 2022.

testing data set, the NNAR model had RMSE and MAE values of 12.89895 and 8.009270, respectively. The NNAR model's RSME and MAE values for both the training and testing data sets w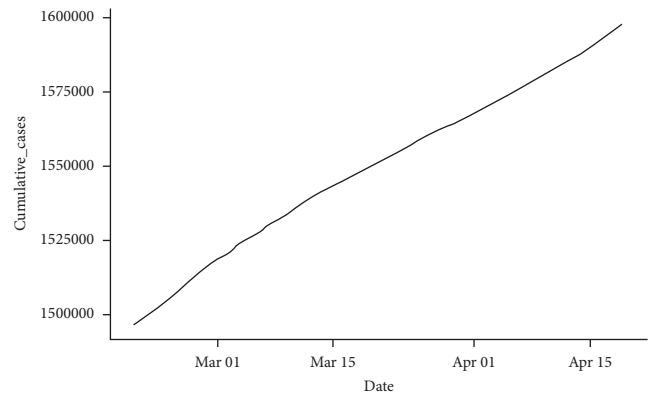ere the least among the ARIMA models and the other competing models. This shows that the NNAR model has higher forecast quality and assesses the data better than the rest of the models [48], making it the most appropriate candidate model for predicting the cumulative daily deaths. We, therefore, used the NNAR model to predict the cumulative daily cases of COVID-19 from February 19, 2022, to April 18, 2022. The forecast value as illustrated in Figure 7 shows that the cumulative daily deaths of COVID-19 in Pakistan will be 32,628 on April 18, 2022.

TABLE 3: Accuracy of different time-series models for predicting COVID-19 cumulative daily deaths.

| Method	Source	RMSE	MAE
Mean	Training	38.17407	30.06568
	Testing	34.05141	31.50445
Drift	Training	25.76414	14.11472
	Testing	33.95330	31.50954
Naïve	Training	25.76424	14.10708
	Testing	29.07091	26.64583
Holt	Training	20.28812	11.70446
	Testing	123.38359	98.10785
SES	Training	20.27235	11.62327
	Testing	37.35009	34.89225
ARIMA order SES	Training	20.20818	11.58092
	Testing	36.33789	33.86672
ARIMA order (2, 2, 2)	Training	20.20818	11.58092
	Testing	36.33789	33.86672
ARIMA auto	Training	20.06109	11.54640
	Testing	22.33402	20.61141
NNAR	Training	10.36647	5.065002
	Testing	12.89895	8.009270
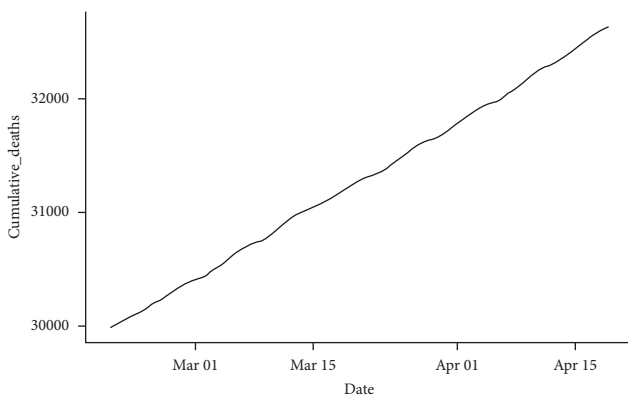
FIGURE 7: Forecast of cumulative COVID-19 daily deaths from February 19, 2022, to April 18, 2022.

4. Conclusion

The COVID-19 pandemic continues to destroy human life around the world. Almost every country throughout the globe suffered from this pandemic, forcing various governments to apply different restrictions to reduce its impact. This study used COVID-19 data in Pakistan from February 26, 2020, to February 18, 2022, as a training and testing data set to compare different time-series models. We estimated and assessed models on the training set and assessed them on the testing set. We computed the RSME and MAE for the ARIMA model, the NNAR model, and other competing models. The out-of-sample RMSE and MAE of the NNAR model were the least among all other models, indicating that the NNAR model outperforms the ARIMA model and the other competing models in terms of forecasting. That is, the NNAR model has better forecast, assessment, and quality compared to the rest of the models. Based on the NNAR forecasted values, the cumulative number of confirmed COVID-19 cases will be 1,597,810 and the cumulative deaths attributed to COVID-19 will be 32,628 on April 18, 2022. We, therefore, suggest that the NNAR model can be adopted to model and forecast COVID-19 cases and deaths as well as other time-series data just like the multigene genetic programming by Niazkar and Niazkar [49]. It is worth noting that other machine learning techniques for time-series data can be considered and used in a similar manner. As COVID-19 has prolonged for more than two years and with the prevailing virus mutation, lockdown is not a feasible solution in current circumstances. Although more than half of Pakistan's population is immunized, if the government's current immunization policy continued, the cumulative cases and cumulative deaths would decrease in the coming months. It is paramount for the Government of Pakistan to boost the immunization policy and ease restrictions to flatten the curve.

Data Availability

The data used in this study are made up of confirmed daily cases and confirmed daily deaths of COVID-19 in Pakistan from February 26, 2020, to February 18, 2022, provided by the COVID-19 Health Platform of the Ministry of National Health Services Regulation, Government of Pakistan (https://covid.gov.pk).

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Authors' Contributions

MD, KT, SM, and KOA conceived the idea. MD, KT, and SM suggested the statistical methodology, performed the statistical analysis, and drafted the manuscript. KOA reviewed the manuscript. All authors read and approved the final manuscript.

References

[1] World Health Organization, "Episode #49," 2019, https://www.who.int/emergencies/diseases/novel-coronavirus-2019/mediaresources/science-in-5/episode-49-can-i-get-infected-after-vaccination.

[2] World Health Organization, "Coronavirus disease 2019 (COVID-19):Situation report, 61," 2020, https://apps.who.int/iris/handle/10665/331605?show=full.

[3] W. A. Anwar and A. Mokhtar, "Forecasting the peak of novel coronavirus disease in Egypt using current confirmed cases and deaths," *medRxiv*, 2020.

[4] L. A. Amar, A. A. Taha, and M. Y. Mohamed, "Prediction of the final size for COVID-19 epidemic using machine learning: A case study of Egypt," *Infectious Disease Modelling*, vol. 5, pp. 622–634, 2020.

[5] A. M. Almeshal, A. I. Almazrouee, M. R. Alenizi, and S. N. Alhajeri, "Forecasting the spread of COVID-19 in Kuwait using compartmental and logistic regression models," *Applied Sciences*, vol. 10, no. 10, p. 3402, 2020.

[6] E. D. El Desouky, "Prediction of the epidemic peak of Covid19 in Egypt, 2020," *medRxiv*, vol. 395, Article ID 931e934, 2020.

[7] S. Benkouiten, J. A. Al-Tawfiq, Z. A. Memish, A. Albarrak, and P. Gautret, "Clinical respiratory infections and pneumonia during the Hajj pilgrimage: A systematic review," *Travel Medicine and Infectious Disease*, vol. 28, pp. 15–26, 2019.

[8] A. Ahmed, B. Salam, M. Mohammad, A. Akgül, and S. H. A. Khoshnaw, "Analysis coronavirus disease (COVID-19) model using numerical approaches and logistic model," *AIMS Bioengineering*, vol. 7, no. 3, pp. 130–146, 2020.

[9] M. Arif, D. M. Khan, M. Aamir, U. Khalil, R. A. R. Bantan, and M. Elgarhy, "Modeling COVID-19 data with a novel extended exponentiated class of distributions," *Journal of Mathematics*, vol. 2022, Article ID 1908161, 14 pages, 2022.

[10] B. Pirouz, S. Shaffiee Haghshenas, S. Shaffiee Haghshenas, and P. Piro, "Investigating a serious challenge in the sustainable development process: Analysis of confirmed cases of COVID-19 (new type of coronavirus) through a binary classification using artificial intelligence and regression analysis," *Sustainability*, vol. 12, no. 6, p. 2427, 2020.

[11] Q. Guo, Z. Wang, Z. He et al., "Changes in air quality from the COVID to the post-COVID era in the beijing-tianjin-tangshan region in China," *Aerosol and Air Quality Research*, vol. 21, no. 12, Article ID 210270, 2021.

[12] M. H. R. Khan and A. Hossain, "Machine learning approaches reveal that the number of tests do not matter to the prediction of global confirmed COVID-19 cases," *Frontiers in Artificial Intelligence*, vol. 3, Article ID 561801, 2020.

[13] Q.-V. Pham, D. C. Nguyen, T. Huynh-The, W.-J. Hwang, and P. N. Pathirana, "Artificial intelligence (AI) and big data for coronavirus (COVID-19) pandemic: A survey on the state-of-the-arts," *IEEE Access*, vol. 8, Article ID 130839, 2020.

[14] R. Ranjan, "Predictions for COVID-19 outbreak in India using epidemiological models," *medRxiv*, 2020.

[15] D. Fanelli and F. Piazza, "Analysis and forecast of COVID-19 spreading in China, Italy and France," *Chaos, Solitons & Fractals*, vol. 134, Article ID 109761, 2020.

[16] S. F. Ardabili, A. Mosavi, P. Ghamisi et al., "COVID-19 outbreak prediction with machine learning," *Algorithms*, vol. 13, no. 10, p. 249, 2020.

[17] A. J. Kucharski, T. W. Russell, C. Diamond et al., "Early dynamics of transmission and control of COVID-19: A mathematical modelling study," *The Lancet Infectious Diseases*, vol. 20, no. 5, pp. 553–558, 2020.

[18] K. Ganasegeran, M. F. A. Jamil, M. R. Appannan, A. S. H. Ch'ng, I. Looi, and K. M. Peariasamy, "Spatial dynamics and multiscale regression modelling of population level indicators for COVID-19 spread in Malaysia," *International Journal of Environmental Research and Public Health*, vol. 19, no. 4, p. 2082, 2022.

[19] S. Tuli, S. Tuli, R. Tuli, and S. S. Gill, "Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing," *Internet of Things*, vol. 11, Article ID 100222, 2020.

[20] Q. Guo and Z. He, "Prediction of the confirmed cases and deaths of global COVID-19 using artificial intelligence," *Environmental Science and Pollution Research*, vol. 28, no. 9, Article ID 11682, 2021.

[21] Z. Xu, H. Zhang, and Z. Huang, "A continuous markov-chain model for the simulation of COVID-19 epidemic dynamics," *Biology*, vol. 11, no. 2, p. 190, 2022.

[22] R. Arumugam and R. Raji, "A Markov model for prediction of corona virus COVID-19 in India-A statistical study," *Xi'an Dianzi Keji Daxue Xuebao/Journal of Xidian University*, vol. 14, pp. 1422–1426, 2020.

[23] A. L. Bertozzi, E. Franco, G. Mohler, M. B. Short, and D. Sledge, "The challenges of modeling and forecasting the spread of COVID-19," *Proceedings of the National Academy of Sciences*, vol. 117, no. 29, Article ID 16738, 2020.

[24] M. A. A. Al-qaness, A. A. Ewees, H. Fan, and M. A. E. Aziz, "Optimization method for forecasting confirmed cases of COVID-19 in China," *Journal of Clinical Medicine*, vol. 9, no. 3, p. 674, 2020.

[25] J. T. Wu, K. Leung, and G. M. Leung, "Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study," *The Lancet*, vol. 395, no. 10225, pp. 689–697, 2020.

[26] X. Zhao, H. Wei, C. Li, and K. Zhang, "A hybrid nonlinear forecasting strategy for short-term wind speed," *Energies*, vol. 13, no. 7, p. 1596, 2020.

[27] H. Zhao, N. N. Merchant, A. McNulty et al., "COVID-19: Short term prediction model using daily incidence data," *PLoS One*, vol. 16, no. 4, Article ID e0250110, 2021.

[28] Y. Hao, T. Xu, H. Hu, P. Wang, and Y. Bai, "Prediction and analysis of corona virus disease 2019," *PLoS One*, vol. 15, no. 10, Article ID e0239960, 2020.

[29] K. Tawiah, W. A. Iddrisu, and K. A. Asosega, "Zero-inflated time series modelling of COVID-19 deaths in Ghana," *Journal of Environmental and Public health*, vol. 2021, Article ID 5543977, 9 pages, 2021.

[30] J. Luo, Z. Zhang, Y. Fu, and F. Rao, "Time series prediction of COVID-19 transmission in America using LSTM and XGBoost algorithms," *Results in Physics*, vol. 27, Article ID 104462, 2021.

[31] E. Gecili, A. Ziady, and R. D. Szczesniak, "Forecasting COVID-19 confirmed cases, deaths and recoveries: Revisiting established time series modeling through novel applications for the USA and Italy," *PLoS One*, vol. 16, no. 1, Article ID e0244173, 2021.

[32] C. Barría-Sandoval, G. Ferreira, K. Benz-Parra, and P. López-Flores, "Prediction of confirmed cases of and deaths caused by COVID-19 in Chile through time series techniques: A comparative study," *PLoS One*, vol. 16, no. 4, Article ID e0245414, 2021.

[33] F. A. Chyon, M. N. H. Suman, M. R. I. Fahim, and M. S. Ahmmed, "Time series analysis and predicting COVID-19 affected patients by ARIMA model using machine learning," *Journal of Virological Methods*, vol. 301, Article ID 114433, 2022.

[34] S. Roy and G. S. Bhunia, "Spatial prediction of COVID-19 epidemic using ARIMA techniques in India," *Modeling Earth Systems and Environment*, vol. 7, no. 2, pp. 1385–1391, 2021.

[35] M. Ali, D. M. Khan, M. Aamir, U. Khalil, and Z. Khan, "Forecasting COVID-19 in Pakistan," *PLoS One*, vol. 15, no. 11, Article ID e0242762, 2020.

[36] J. A. Doornik, J. L. Castle, and D. F. Hendry, "Modeling and forecasting the COVID-19 pandemic time-series data," *Social Science Quarterly*, vol. 102, no. 5, pp. 2070–2087, 2021.

[37] G. Box, G. Jenkins, and G. Reinsel, *Time Series Analysis*, Wiley, Hoboken, NJ, USA, 4th edition, 2008.

[38] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and its Applications: With R Examples*, Springer, New York, NY, USA, 3rd edition, 2011.

[39] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, Springer, New York, NY, USA, 2nd edition, 2002.

[40] P. Ramos and J. Oliveira, "A procedure for identification of appropriate state space and ARIMA models based on time-series cross-validation," *Algorithms*, vol. 9, no. 4, p. 76, 2016.

[41] W. S. Wei, *Time Series Analysis: Univariate and Multivariate Methods*, Addison-Wesley, Boston, MA, USA, 2nd edition, 2005.

[42] W. R. W. Mahiyuddin, N. I. Jamil, Z. Seman et al., "Forecasting ozone concentrations using box-jenkins ARIMA modeling in Malaysia," *American Journal of Environmental Sciences*, vol. 14, no. 3, pp. 118–128, 2018.

[43] A. Tealab, H. Hefny, and A. Badr, "Forecasting of nonlinear time series using ANN," *Future Computing and Informatics Journal*, vol. 2, no. 1, pp. 39–47, 2017.

[44] G. Benrhmach, K. Namir, A. Namir, and J. Bouyaghroumni, Nonlinear autoregressive neural network and extended kalman filters for prediction of financial time series," *Journal of Applied Mathematics*, vol. 2020, Article ID 5057801, 6 pages, 2020.

[45] J. J. Hopfield, "Artificial neural networks," *IEEE Circuits and Devices Magazine*, vol. 4, no. 5, pp. 3–10, 1988.

[46] R. Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2019.

[47] S. Muhammadullah, A. Urooj, F. Khan, M. N. Alshahrani, M. Alqawba, and S. Al-Marzouki, "Comparison of weighted lag adaptive LASSO with autometrics for covariate selection and forecasting using time-series data," *Complexity*, vol. 2022, Article ID 2649205, 10 pages, 2022.

[48] N. D. Bokde, Z. M. Yaseen, and G. B. Andersen, "ForecastTB-an R package as a test-bench for time series forecasting-application of wind speed and solar radiation modeling," *Energies*, vol. 13, no. 10, p. 2578, 2020.

[49] H. R. Niazkar and M. Niazkar, "COVID-19 outbreak: Application of multi-gene genetic programming to country-based prediction models," *Electronic Journal of General Medicine*, vol. 17, no. 5, Article ID em247, 2020.