TECHNIUM

Technium Vol. 4, Issue 2 pp.64-75 (2022)
ISSN: 2668-778X
www.techniumscience.com

Forecasting Weekly COVID-19 Infection and Death Cases in Iraq Using an ARIMA Model

Mohammed Hussein Jabardi[1], Hasan Thabit Kurmasha[2], and Roaa Razaq Al-Khalidy[1]

[1]Computer Science Department, College of Education, the University of Kufa, Najaf, Iraq

[2]Computer Department, College for Women Education, the University of Kufa, Najaf, Iraq

Corresponding author: **Mohammed Hussein Jabardi,** e-mail: mohammed.naji@uokufa.edu.iq

Abstract. Coronavirus (COVID-19) is a contagious disease by SARS-CoV-2 that causes extreme respiratory disorder. The virus has caused a global crisis that has had repercussions on public health, well-being, and all aspects of public and economic life. Infrastructure, information sources, preventive measures, treatment protocols, and various other resources have been put in place worldwide to combat the growth of this deadly disease. This study used the "AutoRegressive Integrated Moving Average" (ARIMA) forecasting technique to estimate the weekly confirmed cases and deaths from the coronavirus epidemic in Iraq. The data collection period was June 1, 2020, until August 31, 2021. The findings demonstrated the model's high accuracy, with an RMSE of 24.168 for the training data and 32.794 for the testing data.

Keywords. COVID-9, Artificial Intelligent, ARIMA, Coronavirus, time series prediction, forecasting.

Technium Vol. 4, Issue 2 pp.64-75 (2022)
ISSN: 2668-778X
www.techniumscience.com

1. Introduction

A pneumonia outbreak was reported in December 2019 in Wuhan, China [1]. On December 31, 2019, the epidemic was linked to a new strain of coronavirus. Coronavirus is transmitted to oral and respiratory mucosal cells via infected patients' respiratory droplets [2].
Until September 15, 2021, the coronavirus pandemic had resulted in at least 4,663,324 verified mortality, and more than 226,685,202 infected [3]. The Wuhan chain was identified as a novel Beta-coronavirus strain belonging to collection 2B with roughly 70% genetic resemblance to SARS-CoV. Due to the virus's 96% resemblance to a bat coronavirus, it is largely believed to have started in bats as well [4, 5]. The human coronaviruses and their origins and potentially related hosts are shown in Figure 1[6].

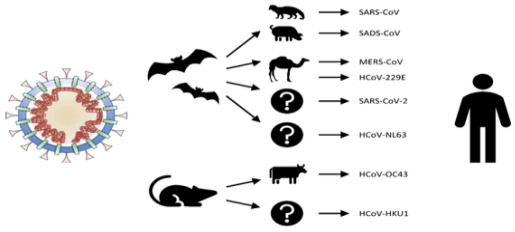

Figure 1. The human coronaviruses and their origins and potentially related hosts.[6]

A time series is a collection of observations made at discrete points in time. They can be captured daily and weekly, monthly, or annual. Time series are used in various situations across a wide variety of disciplines, such as statistics, math, and science. [7].
A time series can be broken down into four parts: Base Level, Trend, Seasonal patterns, and Error. When the slope of the time series increases or decreases, a trend is seen. Whereas seasonality occurs when a distinctive recurring pattern within periodic intervals is noticed due to seasonal variables. However, it is not required that the all-time series exhibit trending, seasonal behavior, or both. A time series may lack a discernible trend but exhibit seasonality; the converse is also true. Thus, a time series can be:
1."Univariate Time Series Forecasting," the past values are used to predict future values.
2."Multivariate Time Series Forecasting" is used When predictors other than the series are included.

One of the major topics to be investigated in time series forecasting is ARIMA models. ARIMA models are the most extensively utilized technique for forecasting time series data because they accurately reflect the relationship between the variables in the data [8].

This research study endeavors to forecast the weekly infected and death tolls of COVID-19 in Iraq. The study's data was collected between May 1, 2020, and August 15, 2021. ARIMA is a robust forecasting model for anticipating the spread and deaths associated with the COVID-19 epidemic. The study offers officials accurate estimations of the epidemic's peak period and intensity based on simple quantitative models. Furthermore, assisting healthcare organizations in logistical planning and making decisions depending on the expected infected numbers.

2. Related Work

Recent studies have adopted various methods to estimate COVID-19 incidence, prevalence, and fatality rate. The majority of prior research has applied ANN, deep learning, and ARIMA techniques for forecasting the infected and death cases of the COVID-19 Pandemic.

A hybrid model (The Savitzky Golay Smoothing (SGS) and a Neural Network model with "Long Short-Term Memory" (LSTM-NN) was presented by [9]. The best-fitting model is determined by comparing the outcomes of these two mod-els. The LSTM-NN prediction displays a clear upward trend corresponding to the actual Time Series data. The LSTM-NN curve reveals a rise in mortality and

infected cases, as well as the Time Series. However, the smoothing re-flects a trend to decline. Finally, the LSTM-NN prediction outperforms the SGS Smoothing method.

"Exponential Smoothing" and ARIMA techniques from classical time-series models, as are Feed-Forward Artificial Neural Networks (with a single input unit) method, utilize forecasting models for the COVID-19 spread in Greece. The proposed mixed forecasting model was chosen from (NG combination of TES, ARIMA, and ANN with a single input unit) and the probability distribution (Log-normal). The model generates five situations for the evolution of recorded cases: bad (90 percent quantile), negative (80 percent quantile), slightly negative (70%), expected (forecasts 30%), and sound (25% quantile) [10].

[11] Used the ARIMA model to predicate COVID-19 cases every day for the next 50 days. A nonlinear autoregres-sive (NAR) was utilized to examine the efficiency of predicted models. Based on the "Bayesian Information Criteria" (BIC) results, the optimal ARIMA (1,1,0) model was chosen [11].

The time-series data of confirmed COVID-19 cases in India examined was fed into an ARIMA model. An ACF and a PAF graph were used to determine the ARIMA parameters. Afterward, measure the variation in normality and static properties of these ARIMA models. Next, the models' accuracy is examined using the MAPE, MAD, and MSE values to select the best one. To further narrow down the choices, the best fit ARIMA model is compared to other models such as Single Exponential, S-Curve Trend, Quadratic Trend Linear Trend, Double Exponential. Moving Average using an output accuracy metric such as MAPE, MAD, and MSE [12].

Levenberg-Marquardt and Resilient Propagation were both used in the forecasting process. The Leven-berg-Marquardt approach uses a mix of speed and stability in convergence to minimize a nonlinear function numerically. The model uses two algorithms (steepest descent and Gauss-Newton) to merge training around a complex curvature to make a quadratic approximation and substantially increase speed convergence. The first is responsible for stability, while the second is for convergence speed. Resilient Propagation adapts the weight step directly based on local gradient data [13].

Model and forecast the short-term behavior of COVID-19 using a statistical time series technique. The researchers show how 10-day-ahead forecasts are developed and evaluated throughout four months. They use a quadratic trend for our model, reflecting both the continuity and unpredictability of the two variables we anticipate (global confirmed cases and deaths). The simplified concept provides comparable prediction performance and practical and valuable estimations of ambiguity [14].

ARIMA and FBProphet models were used to develop an approach for forecasting and interpreting COVID-19 cases. ARIMA has performed better than Prophet on a scale of MAE, RMSE, RMSE, and MAPE error matrices [15]. During modelling and forecasting, FBProphet uses a step-by-step process to prevent anomalies. The findings also suggest that FBProphet can fit well with little data, whereas ARIMA needs many data to model and forecast the outcomes.

A simple "Multiple Linear Regression" (MLR) model optimized for forecasting the number of daily cases reported using phone call data presented by [16]. The suggested MLR model uses the association between reported cases and phone call data and other features such as trend, holiday impact, and verified autoregressive lags. The study's draw-back is inaccessibility to more detailed data and supplementary information that could help explain.

The hybrid architecture was established to build a model with LSTM (Long-Short-Term Memory) as an RNN and GRU (Gated Recurrent Unit) type of CNN and a convolutional layer. The first layer of CNN is pattern learning, which detects the essential variables in the input sequence. The RNN's necessary to identify temporal correlations in the input allows for accurate prediction of nonlinear time sequences [17].

[18] proposes a fuzzy logic multiple ensemble neural network model. Ensemble neural networks were used to foresee several possibilities. The outputs of several forecast modules are then dynamically integrated using fuzzy logic to enhance the final prediction.

The Bayesian optimization strategy combined with three deep learning models: LSTM, CNN, and multi-head attention used to anticipate COVID-19. In addition, Bayesian optimization automatically finds the best hyperparameters for each model [19].

This study forecasted the amount of COVID-19 infected and fatality cases expected in Iraq over the following few days. We conceived an Auto-Regressive Integrated Moving Average (ARIMA) model using data obtained from June 1, 2020, to April 30, 2021, and validated it using data acquired from May 1, 2021, to August 15, 2021.

3. Methodology

The proposed model is divided into five tasks: 1-Data collection and preprocessing, 2-Transform data to be stationary, 3- Identify the best ARIMA(p,d,q), Fitting an ARIMA Time Series Model, 4- Validating Forecasts (model evaluation), 5- Producing and Visualizing Forecasts). The flowchart of the proposed model is shown in Figure 2.

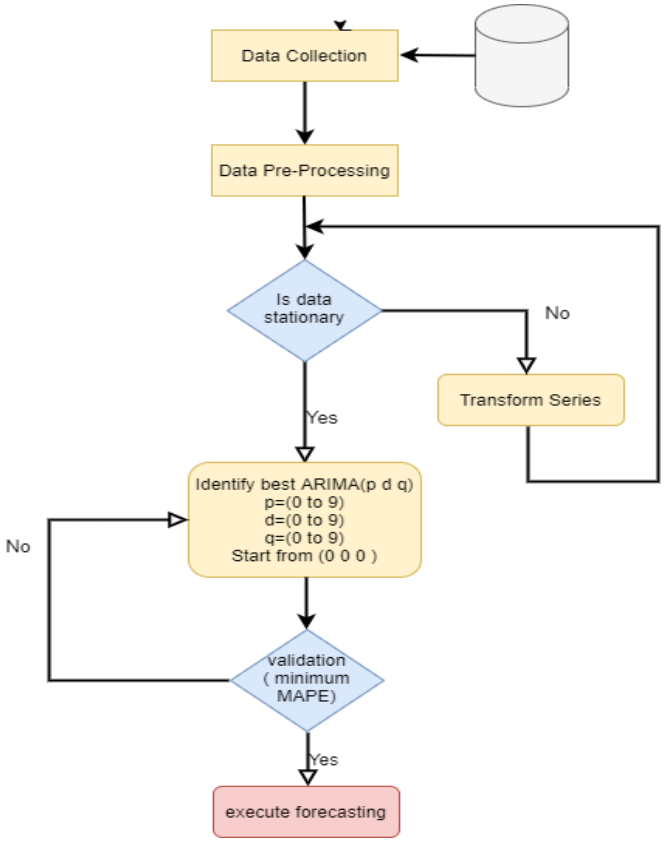

Figure 2. Flowchart of the proposed model

3.1 Data

The confirmed infected cases and deaths of COVID-19 data are derived from the central database at "Johns Hopkins University" (JHU) [20] as shown in Table 1. Data is divided into 67% for training data and 33% for testing data. It is typical to split the data into training and test sets with models. Training sets are used to estimate the parameters of a forecasting method, while test sets are used to evaluate the accuracy.

Table 1. A dataset from April 1, 2020, to October 24, 2021, according to COVID-19 Data Repository maintained by the "Johns Hopkins University" (JHU)

iso_code	continent	date	new cases	new deaths
IRQ	Asia	4/1/2020	34	2
IRQ	Asia	4/2/2020	44	2
IRQ	Asia	4/3/2020	48	0
IRQ	Asia	….	…..	…..
IRQ	Asia	….	…..	…..
IRQ	Asia	1/1/2021	902	11
IRQ	Asia	1/2/2021	840	5
IRQ	Asia	1/3/2021	741	5
IRQ	Asia	….	…..	…..
IRQ	Asia	….	…..	…..
IRQ	Asia	10/20/2021	1388	26
IRQ	Asia	10/21/2021	1882	39
IRQ	Asia	10/22/2021	1846	26
IRQ	Asia	10/23/2021	1064	36
IRQ	Asia	10/24/2021	1247	24

3.2 Methods

Time series forecasting is the AI technique for forecasting the time to come using past information. Numerous epidemics and infectious diseases have been denoted using time series models, notably SARS, Ebola, influenza, and dengue. The ARIMA is a fundamental time series forecasting method that has remained popular. ARIMA was first proposed in 1970 by George Box and Gillim Jenkins in their book "Time Series Analysis, Forecasting, and Control."[21]

ARIMA predicts future values based on the information stored within it. ARIMA is a good choice for forecasting time series that have enough information in the past to predict future values, and it is not recommended to use a series with repetition patterns.

ARIMA combines two models: "Autoregressive" (AR) and "Moving Average" (MA). The AR model takes advantage of the mutualistic relationship between an event and a certain number of delayed events. While the MA.

The model exploits the relationship between an event and the remaining error of a daily average model when used to delay events. Transforming nonstationary time series into stationary series is accomplished using the integrated(I) method. ARIMA model is typically represented as ARIMA(p,d,q), where the parameters are replaced with real numbers to identify the specific ARIMA model utilized[22]. The parameters of the ARIMA model are interpreted as follows:

- p: The auto-regressive part of the reference to the lags of the stationary series., also called the lag order.
- d: It refers to the lags of the prediction error, also known as the degree of the difference.
- q: When we look at this model, the moving average part shows how big the moving average window is or how many times it has been taken.

To estimate such parameters(p,d,q) for any given data, we ran numerous models with varying parameters and chose the one with the lowest "Mean Absolute Percentage Error" (MAPE).

A linear regression model uses differencing to stabilize the model by removing trends and seasonal structures. The seasonality may negatively impact the prediction model [23]. ARIMA (0,0,0) indicates model has no parameters. As a result, the ARIMA model with zero parameters acts an ARMA model or even a simple AR or MA model. The "Augmented Dickey-Fuller" (ADF) and unit-root tests can be used to determine whether or not a time series is stationary.

An AR model is one in which Its lags solely determine yt. in other words, The "lags of Yt determine y_t". The AR equation is formulated as in equation 1[24][25].

$$Yt = c + \beta 1 Yt{-}1 + \beta\, 2Yt{-}2 + \cdots + \beta\, pYt{-}p + \varepsilon t \qquad (1)$$

Where, Y_{t-1} is the lag_1 of the series, $beta_1$ is the coefficient of lag_1 that the model estimates and alpha is the intercept term, also estimated by the model.

Similarly, a pure Moving Average (MA alone) model in which Y_t is solely determined by lagged forecast errors[26].

$$Y_t = c + \Phi_1 \varepsilon_{t-1} + \Phi_2 \varepsilon_{t-2} + \cdots\cdots + \Phi_q \varepsilon_{t-q} + \varepsilon_t \qquad (2)$$

To use an ARIMA model, there must have been at least one differences in the data to render the time series stable. As a result, the equation 3 is

$$Yt = c + \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p}\, \varepsilon t + \Phi_1 \varepsilon_{t-1} + \Phi_q \varepsilon_{t-q} \qquad (3)$$

Where β_p represents the AutoRegressive parameter and Φ_q refers to Moving Average parameters.

If the autocorrelations are positive over many lags, the series requires additional differencing to be effective. If, on the other hand, it is determined that the lag1 autocorrelation is excessively negative, then the series is most likely over-differentiated.

In summary, ARIMA combines two models: Autoregressive (A.R.) and Moving Average (M.A.). ARIMA (p, q, d) is a linear regression model based on prior p values and q errors following d times of difference. The process of determining the values of p, d, and q that are best for a given time series can be achieved through three different approaches [27]:

- "Plots of autocorrelations" (ACF) and "partial autocorrelations plot" (PACF)
- "Akaike's Information Criterion" (AIC) and "Bayesian Information Criterion" (BIC) have the lowest AIC values and BIC values.
- "AutoArima" function from library(forecast).
- Loop ARIMA(p,d,q) technique

The selection of the best ARIMA(p d q) parameters is made using the Loop ARIMA(p,d,q) technique. The number of iteration is p×d×q, and The first iteration parameters are ARIMA(0,0,0), The second iteration parameters are ARIMA(0,0,1), while the last iteration parameters are ARIMA(9,9,9).

4-Evaluation

An accurate forecasting model relies heavily on assessment procedures. We used the "Mean Absolute Error" (MAE) and "Root Mean Squared Error" (RMSE) to evaluate the performance of the various techniques (RMSE). MAE and RMSE are often used to assess forecasting techniques' ability to predict time series [28]. If a model's forecast error is measured using these two metrics, it's possible to describe that mistake in the same units as the forecasting variable. The lower the value, the better. MAE is easy to understand because it expresses absolute inaccuracy. Since the prediction errors are squared, the RMSE penalizes big mistakes. To avoid huge predicting mistakes, it follows that the RMSE is useful.

4.1 Root Mean Square Error (RMSE)

The " Root Mean Square Errors " (RMSE) measures the variations between expected and observed values. The RMSE is one of the most often used criteria for assessing prediction quality [29]. The steps of calculating the RMSE is 1- Take the difference between each pair of the observed and predicted value and figure out how much it is 2- Take the difference value and square it. 3-It's now time to add up every single difference in squared value. In step 4, divide the sum of all observations together by the total number of observations. 5- Finally, take the root square of the average value. The RMSE is calculated using equation 4:

$$\boldsymbol{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(\boldsymbol{X_i} - \widehat{\boldsymbol{X}}_{\boldsymbol{i}})^2}{\mathbf{N}}} \qquad (4)$$

Where N is the data points number, X_i is the actual observations, and $\widehat{\boldsymbol{X}}_{\boldsymbol{i}}$ is a predicate value. RMSE is one of the most commonly used measures for evaluating the quality of the predictions model. An estimate of the data's concentration around the line of best fit is provided by RMSE. RMSE is frequently used to validate forecasting, climatology, and statistical analysis testing results.

Technium Vol. 4, Issue 2 pp.64-75 (2022)
ISSN: 2668-778X
www.techniumscience.com

5-Results

The ARIMA approach was established in order to forecast the COVID-19 pandemic in terms of infections and deaths of people. Figure 4 visualize the predicted and actual cases, from July 1, 2020, to August 31, 2021. From the forecasting visualization graph, we can prove that the ARIMA model has the best performance according to prediction on training data (blue line) and testing data (yellow line) compared to original data (red line).

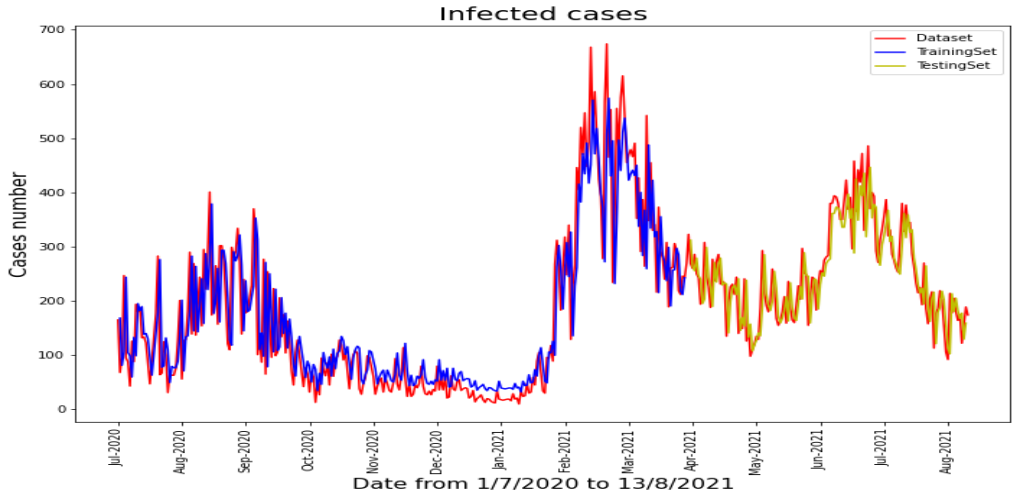

Fig. 4. The actual and predicted infected cases from July 1, 2020, to October 24, 2021.

The residuals (forecasting errors) indicate how far the data points are from the regression line. The evaluation metrics RMSE measure how dispersed these residuals are. Put another way, showing how closely the data is packed around the best-fit line. In order to get the RMSE, the squared errors are added together. This means that the RMSE gives more important mistakes more weight. In this metric, the lower the value, the better the model did at what it did. The forecasting of newly infected cases (unseen cases) from 25-31 Oct 2021 compared to actual values is shown in Figure 5. The evaluation metrics Root Mean Square Errors (RMSE) results are 215.653.

Date	Actual value	Predicate value
25-Aug	1537	1630
26-Aug	1429	1548
27-Aug	1412	1528
28-Aug	1471	1752
29-Aug	1397	1269
30-Aug	1206	806
31-Aug	1046	1229
RMSE	**215.653**	

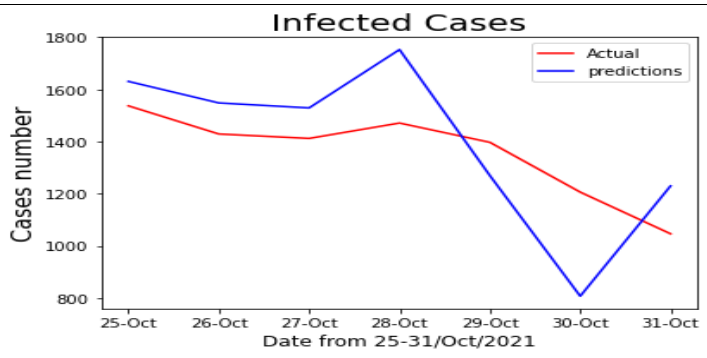

Fig 5. The actual and predicted number of infected cases for seven days (25-31 Oct 2021).

The forecasting of new death cases (unseen cases) from 25-31 Oct 2021 compared to actual values is shown in Figure 6. The evaluation metrics RMSE results are 5.164.

Technium Vol. 4, Issue 2 pp.64-75 (2022)
ISSN: 2668-778X
www.techniumscience.com

Date	Actual value	Predicate value
25- Oct	32	29
26- Oct	31	30
27- Oct	37	31
28- Oct	22	33
29- Oct	28	28
30- Oct	27	27
31- Oct	32	27
RMSE	5.164	

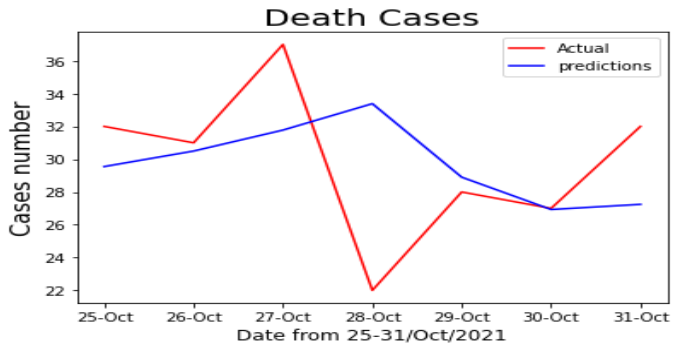

Fig.6. The actual and predicted number of death cases for seven days (25-31 Oct 2021).

The proposed system can be applied for any specific period of time; for example, the proposed approach is used to forecast the time series for 14 days from August 16, 2021, to August 31, 2021. As shown in Figures 7 and 8, the forecasting model curve improves the quality of the proposed system. The curve of predicate results is close to the original dataset.

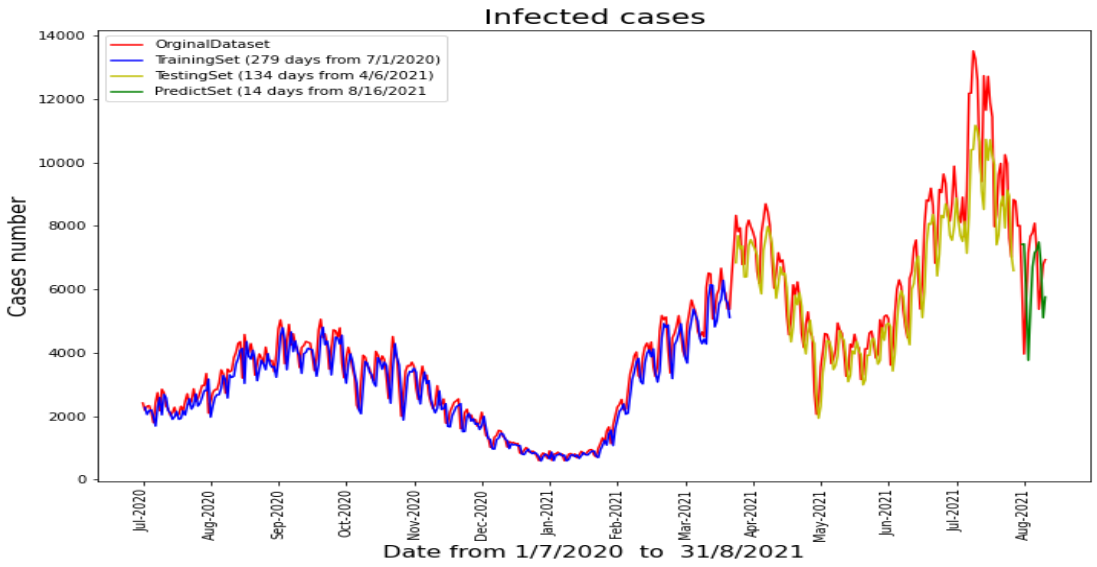

Figure 7. Forecasting of infected cases for 14 days.

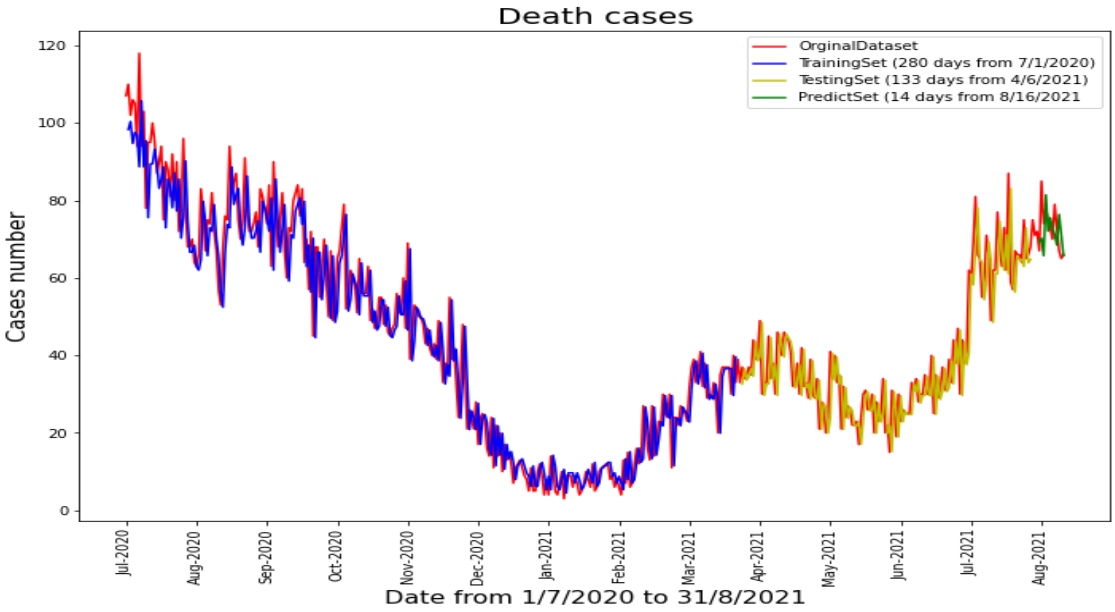

Figure 8. Forecasting of death cases for 14 days.

The three different forecasting techniques (LSTM-RNN, Exponential Smoothing and Naïve) are used to compare ARIMA technique performance as shown in the table.

Table 6. ARIMA vs LSTM-RNN, ES and Naïve forecasting techniques

Forecasting Techniques	Infected cases	Death cases
	RMSE	RMSE
RNN	569.71	8.43
Exponential Smoothing	1608.5	43.361
Naïve	913.725	8.680
ARIMA	215.653	5.164

6-Conclusions

We proposed a univariate time series model in this work to forecast the number of COVID-19 infection and mortality cases expected in the following days in Iraq. We developed an Auto-Regressive Integrated Moving Average (ARIMA) model using data from April 1, 2020, to October 31, 2021, and validated it using data from May 1, 2021 to October 24, 2021. Experiments are conducted to validate the approach's performance and system evaluation using the root-mean-square error (RMSE) measure described in Section 2.1 and draw the significant findings. The experiment's objective was dual. First, we sought to determine which technique is better appropriate for predicting disease-infected individuals, such as KOVID-19. Second, we sought to determine the best ARIMA parameter value that should be employed to obtain accurate forecasts.

References

[1] Wu, Y.-C., C.-S. Chen, and Y.-J. Chan, *The outbreak of COVID-19: An overview*. Journal of the Chinese medical association, 2020. **83**(3): p. 217.
[2] Al-Rohaimi, A.H. and F. Al Otaibi, *Novel SARS-CoV-2 outbreak and COVID19 disease; a systemic review on the global pandemic*. Genes & Diseases, 2020. **7**(4): p. 491-501.
[3] worldometers. COVID-19 CORONAVIRUS PANDEMIC. 2021 October 14, 2021 [cited 2021 August 15, 2021]; Available from: https://www.worldometers.info/coronavirus/.
[4] Lelli, D., et al., *Detection of coronaviruses in bats of various species in Italy*. Viruses, 2013. **5**(11): p. 2679-2689.
[5] Wong, M.C., et al., *Evidence of recombination in coronaviruses implicating pangolin origins of nCoV-2019*. BioRxiv, 2020.
[6] Rabi, F.A., et al., *SARS-CoV-2 and Coronavirus Disease 2019: What We Know So Far*. Pathogens, 2020. **9**(3): p. 231.
[7] Siebert, J., J. Groß, and C. Schroth. *A Systematic Review of Packages for Time Series Analysis*. in *Engineering Proceedings*. 2021. Multidisciplinary Digital Publishing Institute.
[8] Kumar, N. and S. Susan. *Covid-19 pandemic prediction using time series forecasting models*. in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. 2020. IEEE.
[9] Rasjid, Z.E., R. Setiawan, and A. Effendi, *A Comparison: Prediction of Death and Infected COVID-19 Cases in Indonesia Using Time Series Smoothing and LSTM Neural Network*. Procedia computer science, 2021. **179**: p. 982-988.
[10] Katris, C., *A time series-based statistical approach for outbreak spread forecasting: Application of COVID-19 in Greece*. Expert Systems with Applications, 2021. **166**: p. 114077.
[11] Khan, F.M. and R. Gupta, *ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India*. Journal of Safety Science and Resilience, 2020. **1**(1): p. 12-18.
[12] Tandon, H., et al., *Coronavirus (COVID-19): ARIMA based time-series analysis to forecast near future*. arXiv preprint arXiv:2004.07859, 2020.
[13] de Oliveira, L.S., S.B. Gruetzmacher, and J.P. Teixeira, *COVID-19 Time Series Prediction*. Procedia Computer Science, 2021. **181**: p. 973-980.
[14] Petropoulos, F., S. Makridakis, and N. Stylianou, *COVID-19: Forecasting confirmed cases and deaths with a simple time series model*. International journal of forecasting, 2020.
[15] Rostami-Tabar, B. and J.F. Rendon-Sanchez, *Forecasting COVID-19 daily cases using phone call data*. Applied soft computing, 2021. **100**: p. 106932.
[16] Castillo, O. and P. Melin, *Forecasting of COVID-19 time series for countries in the world based on a hybrid approach combining the fractal dimension and fuzzy logic*. Chaos, Solitons & Fractals, 2020. **140**: p. 110242.
[17] Hawas, M., *Generated time-series prediction data of COVID-19' s daily infections in Brazil by using recurrent neural networks*. Data in brief, 2020. **32**: p. 106175.
[18] Melin, P., et al. *Multiple ensemble neural network models with fuzzy response aggregation for predicting COVID-19 time series: the case of Mexico*. in *Healthcare*. 2020. Multidisciplinary Digital Publishing Institute.
[19] Abbasimehr, H. and R. Paki, *Prediction of COVID-19 confirmed cases combining deep learning methods and Bayesian optimization*. Chaos, Solitons & Fractals, 2021. **142**: p. 110511.
[20] University, J.H., *COVID-19 Data Repository*. 2020, Center for Systems Science and Engineering (CSSE) at Johns Hopkins University: github.com.
[21] Box, G.E., et al., *Time series analysis: forecasting and control*. 2015: John Wiley & Sons.
[22] Lai, Y. and D.A. Dzombak, *Use of the autoregressive integrated moving average (ARIMA) model to forecast near-term regional temperature and precipitation*. Weather and Forecasting, 2020. **35**(3): p. 959-976.

[23] Nau, R. *Statistical forecasting: notes on regression and time series analysis. [cited 2021; linear regression and time series forecasting models]. Available from: https://people.duke.edu/~rnau/411arim.htm.*

[24] Nau, R., *The mathematical structure of Arima models.* Duke University Online Article, 2014.

[25] Angco, Robert Jay N., et al. "Time series approach on Philippines' three economic participation using ARIMA Model." *Technium Soc. Sci. J.* 25 (2021): 304.

[26] Mahmoudi, M.R. and S. Baroumand, *Modeling the stochastic mechanism of sensor using a hybrid method based on seasonal autoregressive integrated moving average time series and generalized estimating equations.* ISA transactions, 2021.

[27] Angco, Robert Jay N., et al. "Time series approach on Philippines' three economic participation using ARIMA Model." *Technium Soc. Sci. J.* 25 (2021): 304.

[28] Sitohang, Sunarsan, and Very Karnadi. "Forecasting Method Using the Minitab Program." *Technium Soc. Sci. J.* 16 (2021): 618.

[29] Chicco, D., M.J. Warrens, and G. Jurman, *The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation.* PeerJ Computer Science, 2021. **7**: p. e623.