Journal of Economics and Administrative Sciences (JEAS)

Available online at http://jeasiq.uobaghdad.edu.iq

Using a hybrid SARIMA-NARNN Model to Forecast the Numbers of Infected with (COVID-19) in Iraq

Ayat Ahmed Hamel
University of Baghdad, Baghdad, Iraq
ayat.ahmed1201b@coadec.uobaghdad.edu.iq

Baydaa Ismael Abdulwahhab
University of Baghdad, Baghdad, Iraq
baidaa_29@coadec.uobaghdad.edu.iq

Received:5/1/2022 **Accepted: 23/1/2022** **Published: June / 2022**

Abstract:

Coronavirus disease (COVID-19) is an acute disease that affects the respiratory system which initially appeared in Wuhan, China. In Feb 2019 the sickness began to spread swiftly throughout the entire planet, causing significant health, social, and economic problems. Time series is an important statistical method used to study and analyze a particular phenomenon, identify its pattern and factors, and use it to predict future values. The main focus of the research is to shed light on the study of SARIMA, NARNN, and hybrid models, expecting that the series comprises both linear and non-linear compounds, and that the ARIMA model can deal with the linear component and the NARNN model can deal with the non-linear component. The models were applied in the health sector to predict the numbers of people infected with the Covid-19 virus in Iraq where the data were collected via the website of the Iraqi Ministry of Health through the daily epidemiological situation of all Iraqi provinces for the period (2021\3\28 to 2021\8\15). When analyzing, studying, and comparing these models, the researcher noted that the hybrid model outperformed other models because it had the lowest value for the MSE, RMSE, MAE, and MAPE so it was used to predict future values.

Keywords: Time series, SARIMA, NARNN, Hybrid SARIMA-NARNN.

1. Introduction

The analysis of time series is one of the mathematical and statistical methods used to explain natural phenomena and their behavior over a specific period. The neural network is one of the vital aspects of artificial intelligence which indicates an essential improvement effectively in a human attitude and revolves around the concept of neural networks that mock human thinking by illustrating the computers. Predictable progress in this topic could be attributed to various studies in the field of neural processing simulation that have solved difficulties and followed self-learning processes that rely on the experience stashed in the network to produce improved results. It is intended that seasonal time series models are a set of viewing values associated with each generated sequentially over time and refer to replicate the pattern of movement of the time series in opposing months over successive years.[1][4]

COVID-19 has had an impact on Iraq, resulting in death. Coronavirus is an irresistible infection brought about by the SARS-CoV-2 infection. Most of those contaminated with the infection will encounter gentle to direct respiratory manifestations and will recuperate without clinical assistance. Then again, certain individuals will turn out to be hazardously sick and require clinical consideration. The older and those with hidden ailments like cardiovascular infection, diabetes, ongoing lung sickness, or malignant growth are bound to foster major ailments. Coronavirus can make anybody debilitated, make them exceptionally wiped out, or make them perish at whatever stage in life. The best procedure for keeping away from and easing back transmission is to be very much informed with regards to the infection and how it spreads. To ensure yourself as well as other people against contamination, keep no less than 1 meter away from others, wear a well-fitting veil, and clean up or utilize an alcohol put-together rub concerning an ordinary premise. Get vaccinated and heed neighborhood guidance when it's your move. The infection spreads in minute fluid particles from a contaminated individual's lips or nose when they hack, sniffle, talk, sing, or relax. The particles range in size from bigger respiratory drops to small vapor sprayers. Assuming you're unwell, it's indispensable to rehearse great respiratory cleanliness, like hacking into a flexed elbow and to remain at home and isolate yourself until you feel good. One of the most critical challenges to solve when an epidemic strikes is establishing the pandemic's course at an inflection point. As a result, the purpose of this research is to provide a short- to a medium-term projection of COVID-19 disease transmission and inflection points.[5][7][8][9]

2. Methodology

a.<u>Seasonal Autoregressive Integrated Moving Average model.</u>

The SARIMA (P,D,Q)s model in general is as follows:

$$\phi_p(B)\phi_p(B^s)z_t = \delta + \theta_q(B)\theta_Q(B^s)a_t$$

Where:

- $\phi_p(B) = \left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right)$

Is the p-order non-seasonal autoregressive operator

- $\phi_p(B^L) = \left(1 - \phi_{1,L} B^L - \phi_{2,L} B^{2L} - \cdots - \phi_{P,L} B^{PL}\right)$

Is the order P non-seasonal autoregressive operator

- $\theta_q(B) = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right)$

Is the non-seasonal moving average operator of order q

- $\theta_Q(B^L) = \left(1 - \theta_{1,L}B^L - \theta_{2,L}B^{2L} - \cdots - \theta_{Q,L}B^{QL}\right)$

Is the order Q seasonal moving average operator

- $\delta = \mu\phi_p(B)\phi_P(B^L)$

Is a constant term where μ is the mean of stationary time series

The tentative model identification is based on ACF and PACF patterns that are recognized. A member of the SARIMA family provides the best approximation to the data. At the outset, more than one potential model is usually recognized.

- **Maximum likelihood is used to estimate SARIMA models.**

Diagnostics can sometimes produce inconsistent results, and no single model is always superior to all others. In such cases, we can explore mixing models. [1][3][6]

b. Artificial Neural Network

Among soft computing system models, Artificial Neural Networks (ANNs) function as human brain system operations. Neurons are commonly acknowledged as one of the human body's most fundamental components. Neural networks take in data from various sources, mix and analyze it in unique ways, and then apply nonlinear operations to the output. According to the structure and function of ANN, it is a system that accepts inputs, processes data, and outputs. It has three layers: an input layer, a hidden layer, and an output layer, with the following function:

$$Y_j = f\left(\sum_i W_{ij}X_{ij}\right)$$

In engineering and other industries, nonlinear autoregressive ANN can be utilized to solve a range of challenges. [2][11]

c.The NARNN model

The Nonlinear Autoregressive Neural Network model, yt =f(yt1,yt2,,ytd), can predict a simple time series given past values of the same time series. NARNN utilizes a two-layer FFBP with a sigmoid exchange work in the secret layer and direct exchange work in the result layer as its default. The NARNN organization's results, y(t), are taken care of once again into the organization's contribution (through delays). The arrangement is displayed in Figure 1.[12][10]

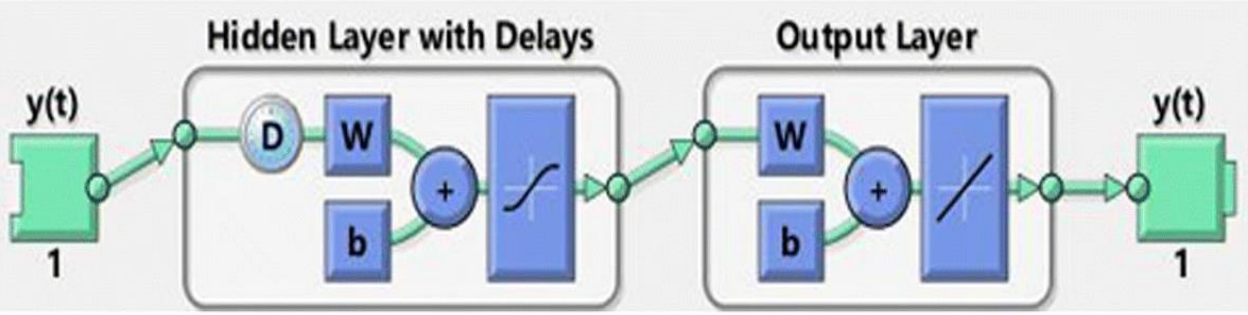

Figure: 1 The NARNN in its current configuration. The NARNN is made up of a 1 unit output layer and a unit hidden layer with D delays.

d. The hybrid SARIMA-NARNN model.

SARIMA-NARNN, a hybrid model, was created in two stages. The main goal of the SARIMA model stage was to find linear correlations between the actual data. The residuals were then calculated using the SARIMA model. The main goal of the NARNN model stage was to recreate the nonlinear relationships that were observed in the residuals. The sum of SARIMA model forecasts and corrects residuals from the NARNN model yielded. The final combined forecasting values for the time series:

$$\hat{y}_t = \hat{L}_t + \hat{N}_T$$

e. Index of performance statistics

The wellness and forecast execution of the SARIMA, NARNN and SARIMA-NARNN models were assessed utilizing, displaying and testing mistakes. For mistake assessment, three lists were utilized: RMSE, MAE, and MAPE. The recipes for ascertaining are as per the following:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t|$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|y_t - y_t^{\wedge}|}{y_t}$$

3. Application

○ **SARIMA model analysis**

The data were collected from the official website of the Iraqi Ministry of Health through the daily epidemiological situation of all Iraqi provinces for the period (28/3/2021 to 15/8/2021). The time series was split into two sets before the model process: estimating data set and testing data set; time-series plots are presented in figure (2). When analyzing and forecasting a time series, it's a good idea to plot the data and pay attention to the time series' distinctive characteristics. It directs the researcher in selecting a suitable model strategy that captures the identified traits directly. The time series must first be made stationary before the procedure can begin. The ADF test is utilized to examine the stationarity of time series, and the findings are displayed in Table (1). The first difference in a series, d=1, is ideal for making a series reasonably stationary. Without making a difference, reject the hypothesis of time series stationarity. The ACF and PACF of series are also shown in figure (3) and figure (4) to estimate another two parameters of the candidate model. Table (2) lists all models and their AIC and BIC values. Finally, for predicting, we choose SARIMA $(2,1,2)_{12}$. Figure (5) depicts the model's residual. The residual is subjected to a white noise test after fitting the SARIMA (2, 1, 2)12 model. Figures (6) and (7) show the autocorrelation and partial autocorrelation function graphs for the residual series. White noise can be seen in the residual, indicating that the model is valid. Table (3) and figure (8) residuals show the projected value for 28 periods.

○ **NARNN model analysis**

By applying the NARNN model to the residuals of the SARIMA model, you can set the hidden units (1-10) and proportions for preparing (80%) and testing (20%). The NAR structure is portrayed in Figure 9. Figure 10 shows neural organization preparing execution, while Figure 11 shows neural organization preparing state. Figure 12 shows neural organization preparing mistake histogram, Figure 13 shows relapse for preparing, approval, and testing, just as every one of the three relapses consolidated. Figure 14 shows simulation results, and Figure 15 shows the error autocorrelation function plot. The predicted value for 28 periods is shown in table (4) and graphic (16).

○ **Hybrid SARIMA-NARNN model analysis**

The hybrid model's predictions are derived by combining the SARIMA model's predictions with the residuals provided by the NARNN model. Table (5) and figure (17) show the projected value for 28 periods. Table 6 shows the results of the three models' prediction performance.

a.Results and Discussion

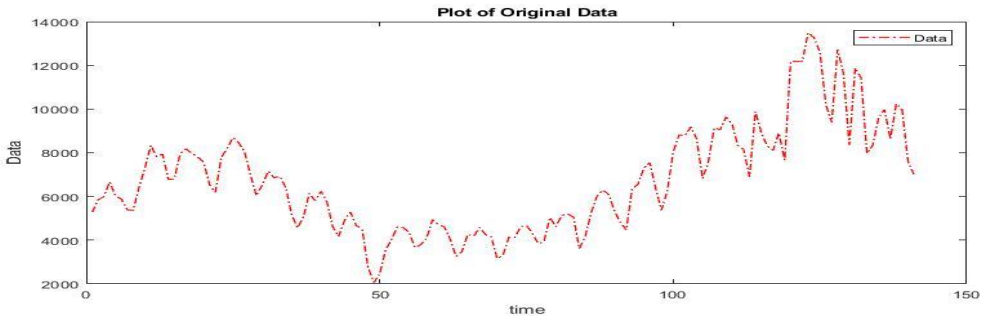

Figure 2: Time series plots

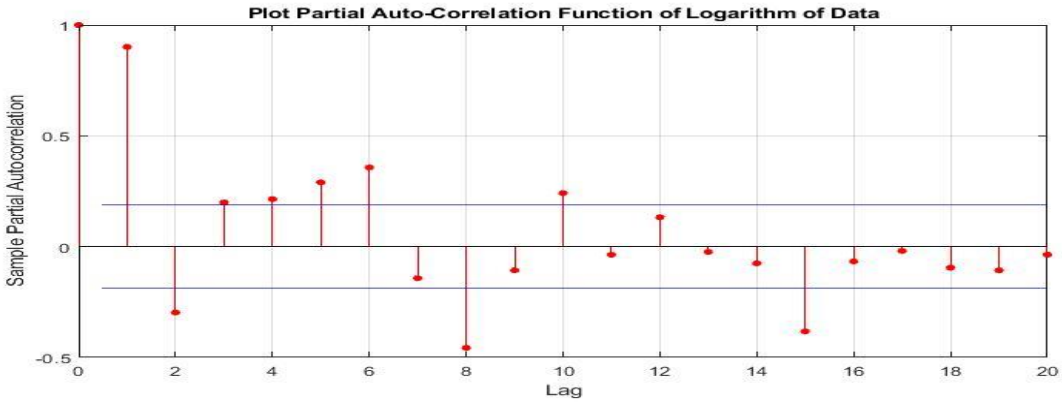

Figure 3: ACF plot of the time series

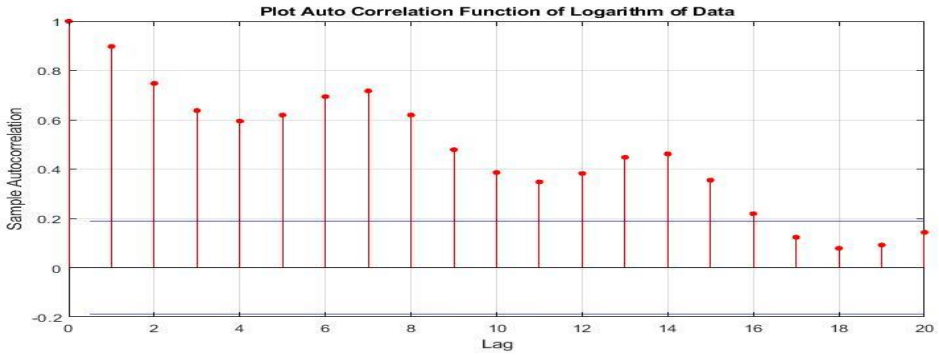

Figure 4: PACF plot of the time series

Table 1: Table of p-values and t-test results from the ADF test.

Number of difference	ADF (t-test)	ADF (p-value)
d=0	0.0962	0.6849
d=1	-8.2781	1e-3

Table 2: Table of models and AIC, BIC values

model	p	d	q	s_AR	S_MA	LogLikelihood	aic	bic
1	0	1	1		12	62.3922	-118.784	-110.602
2	0	2	1		12	36.1072	-66.2144	-58.0322
3	0	3	1		12	16.5497	-27.0995	-18.9173
4	1	1	1		12	62.3941	-116.788	-105.879
5	1	2	1		12	39.3881	-70.7762	-59.8667
6	1	3	1		12	17.0979	-26.1958	-15.2862
7	0	1	2		12	62.4257	-116.851	-105.942
8	0	2	2		12	47.2249	-86.4498	-75.5402
9	0	3	2		12	16.6029	-25.2058	-14.2962
10	1	1	2		12	63.4539	-116.908	-103.271
11	2	1	2		12	81.9684	-151.937	-135.572
12	2	2	2		12	48.9632	-85.9264	-69.5621
13	0	1	2	12	12	61.1576	-114.315	-103.406
14	0	1	1	12	12	67.4528	-126.906	-115.996
15	0	2	1	12	12	43.2662	-78.5324	-67.6228
16	1	1	1	12	12	67.5094	-125.019	-111.382
17	2	1	1	12	12	73.4504	-134.901	-118.537
18	2	1	2	12	12	90.4727	-166.945	-147.854
19	2	2	2	12	12	58.6177	-103.235	-84.1438
20	2	1	0	12	12	69.4145	-128.829	-115.192

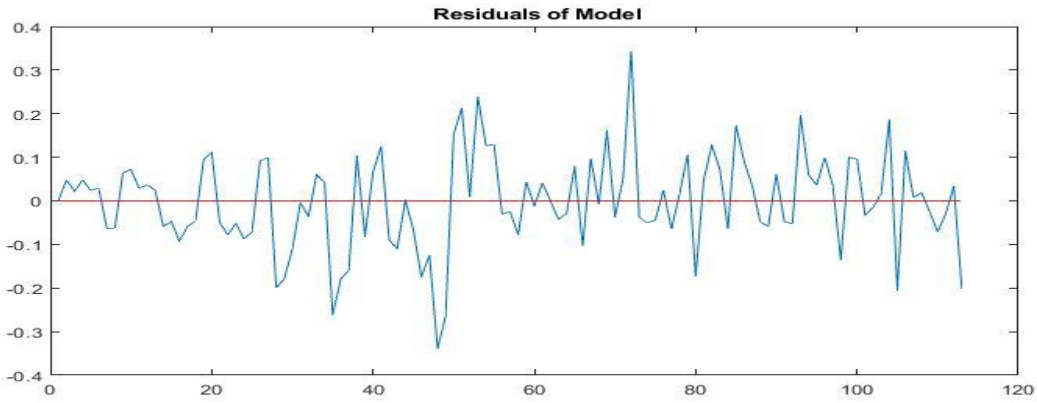

Figure 5: Residual of the model

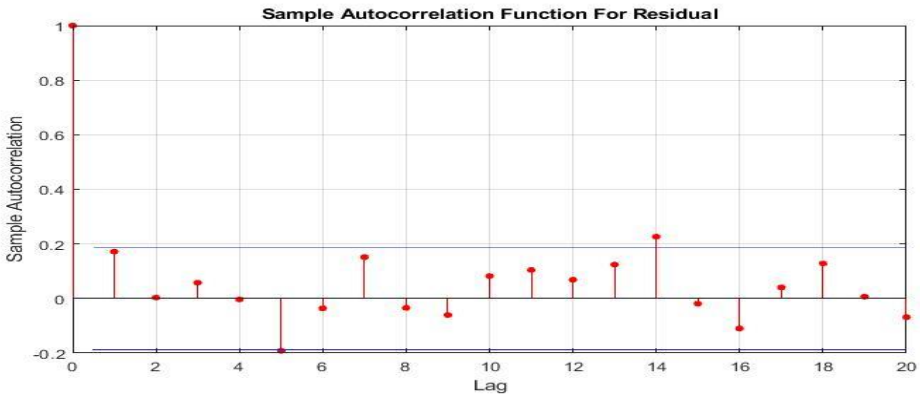

Figure 6: ACF of the residual

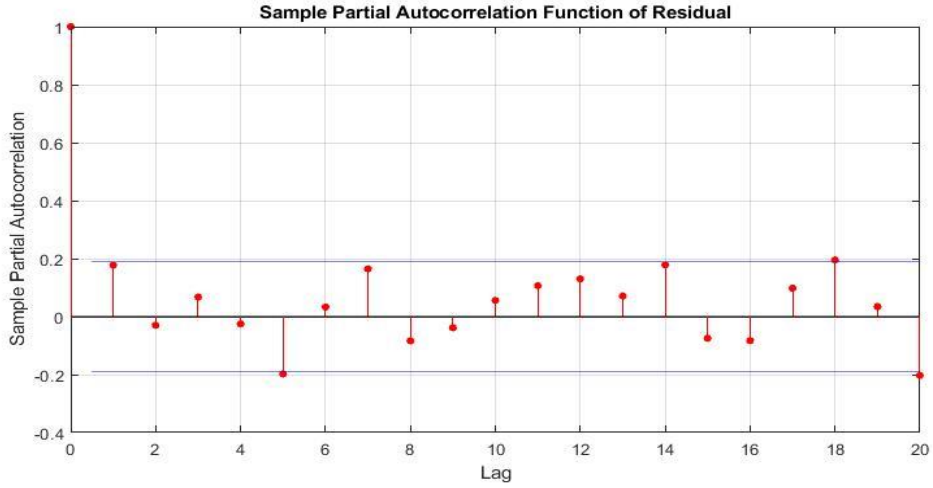

Figure 7: PACF of the residual

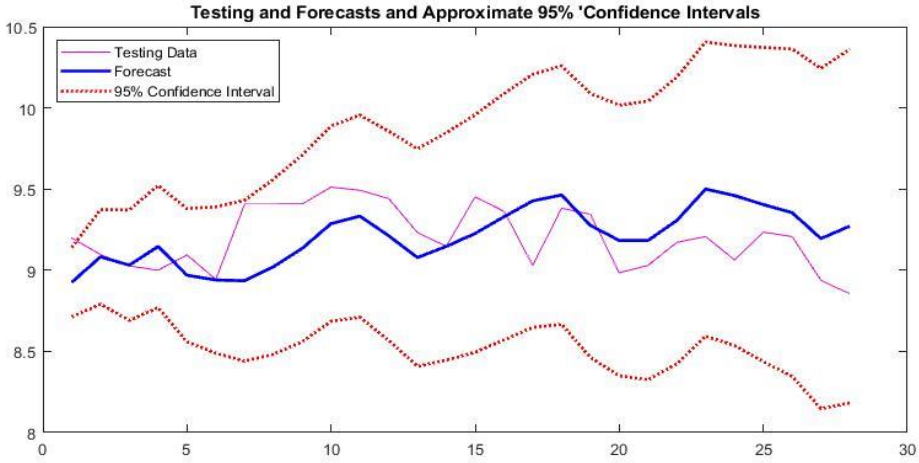

Figure 8: The forecast value for 28 period

Table 3: Table of forecast value for 28 period

Period	Forecast	Mean forecast error	Period	Forecast	Mean forecast error
1	8.925583664	0.011805809	16	9.329053371	0.149724441
2	9.082201402	0.022107551	17	9.427015193	0.157997641
3	9.030997374	0.030145167	18	9.463150399	0.164925318
4	9.146161962	0.036860716	19	9.275608981	0.172053266
5	8.969262118	0.043925443	20	9.18323179	0.180807198
6	8.93953214	0.05281831	21	9.182981096	0.191578474
7	8.935292468	0.063846773	22	9.306607052	0.203136014
8	9.021147105	0.075564615	23	9.499812234	0.213547288
9	9.134520899	0.085903727	24	9.459739126	0.221923933
10	9.286762591	0.094054875	25	9.404326673	0.243745703
11	9.333137017	0.100870056	26	9.354558798	0.264565456
12	9.212111651	0.107956942	27	9.195220965	0.285859322
13	9.077738036	0.116786171	28	9.2712945	0.30876897
14	9.146645134	0.127692573			
15	9.226411438	0.13933904			

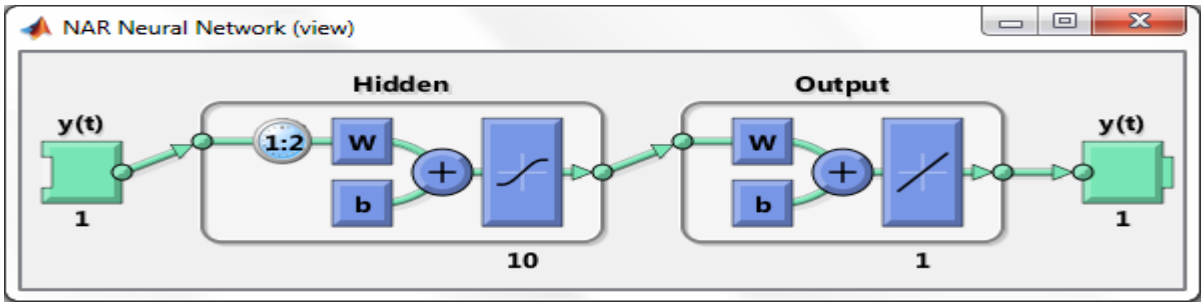

Figure 9: The NAR structure

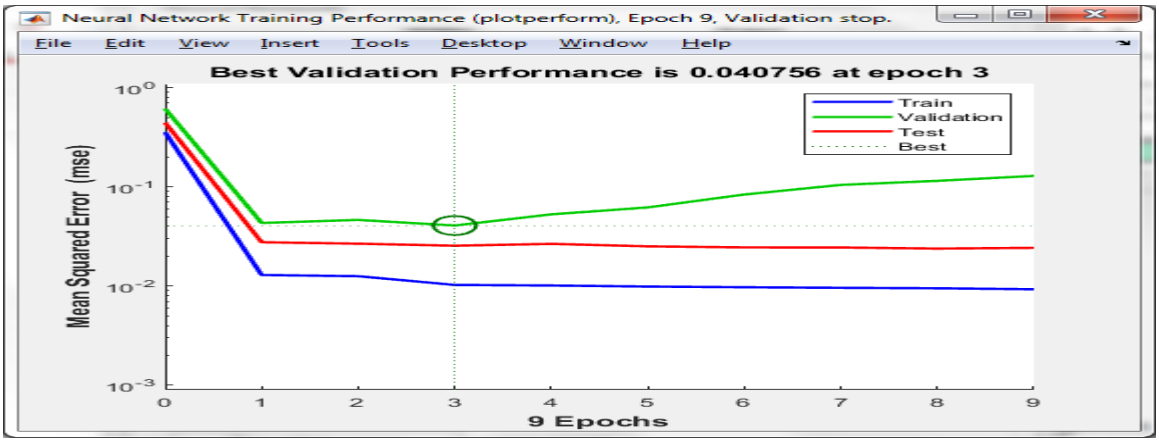

Figure 10: Performance of Neural Network Training

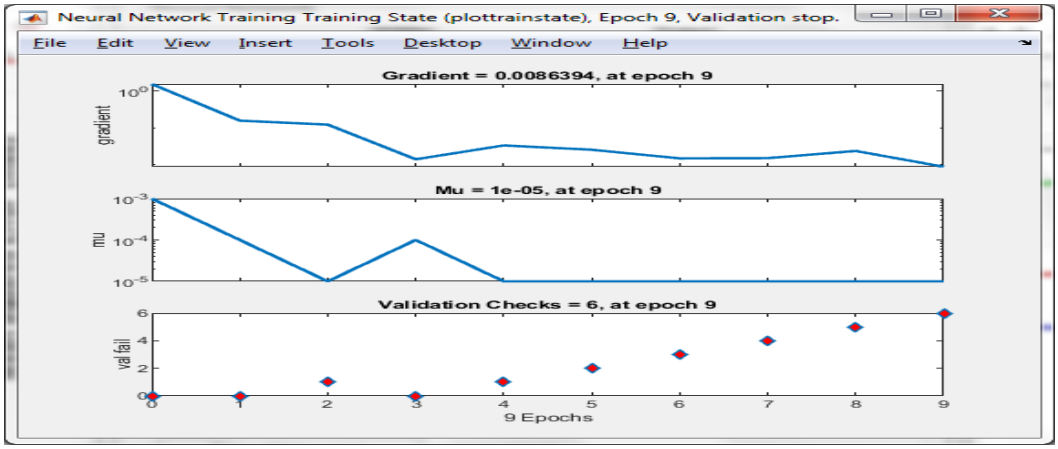

Figure 11: The Training State of a Neural Network

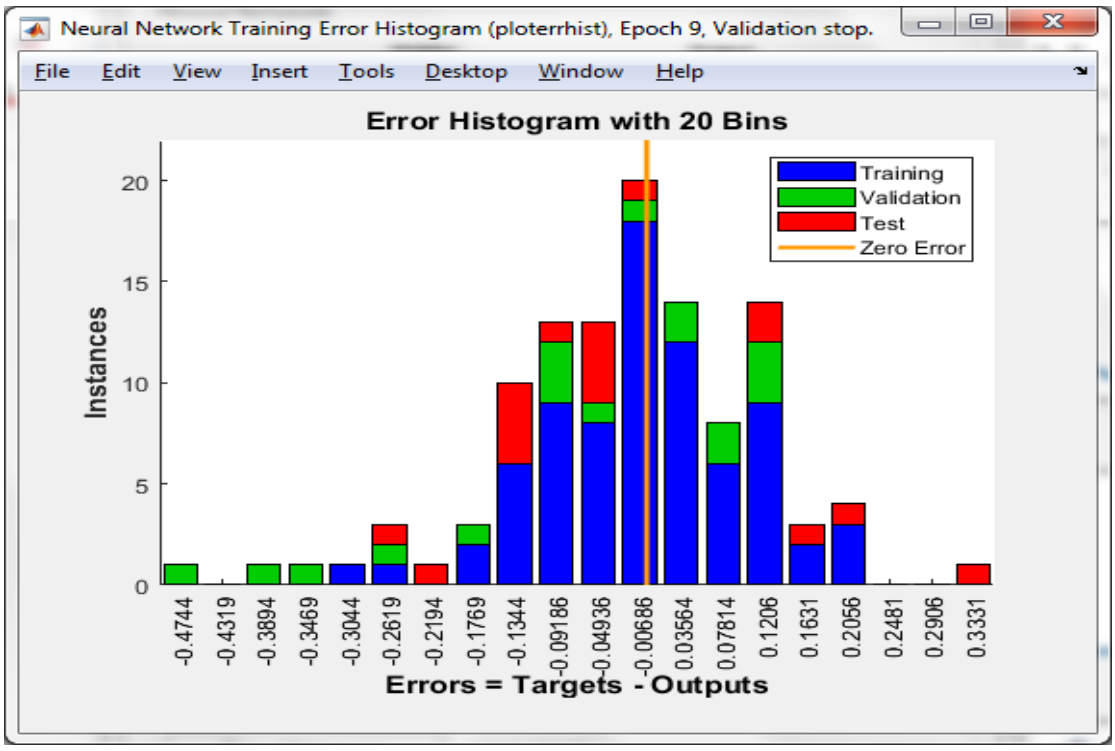

Figure 12: Histogram of Neural Network Training Errors

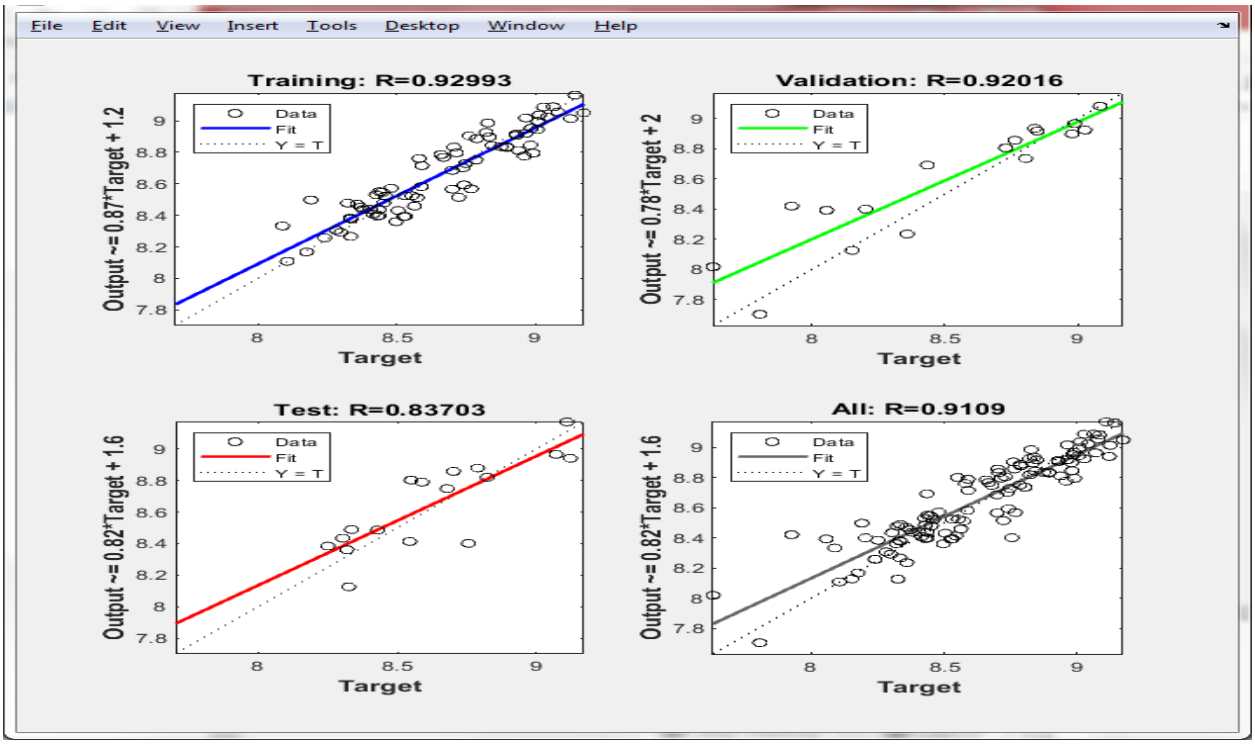

Figure 13: Training, validation, and testing regressions, as well as all three regressions

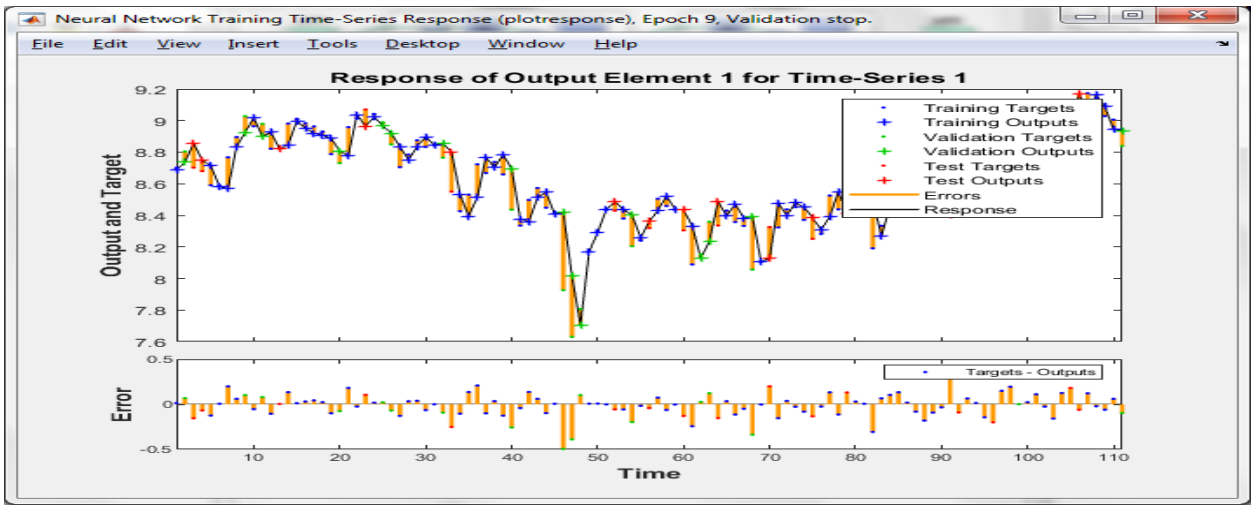

Figure 14: Simulation results

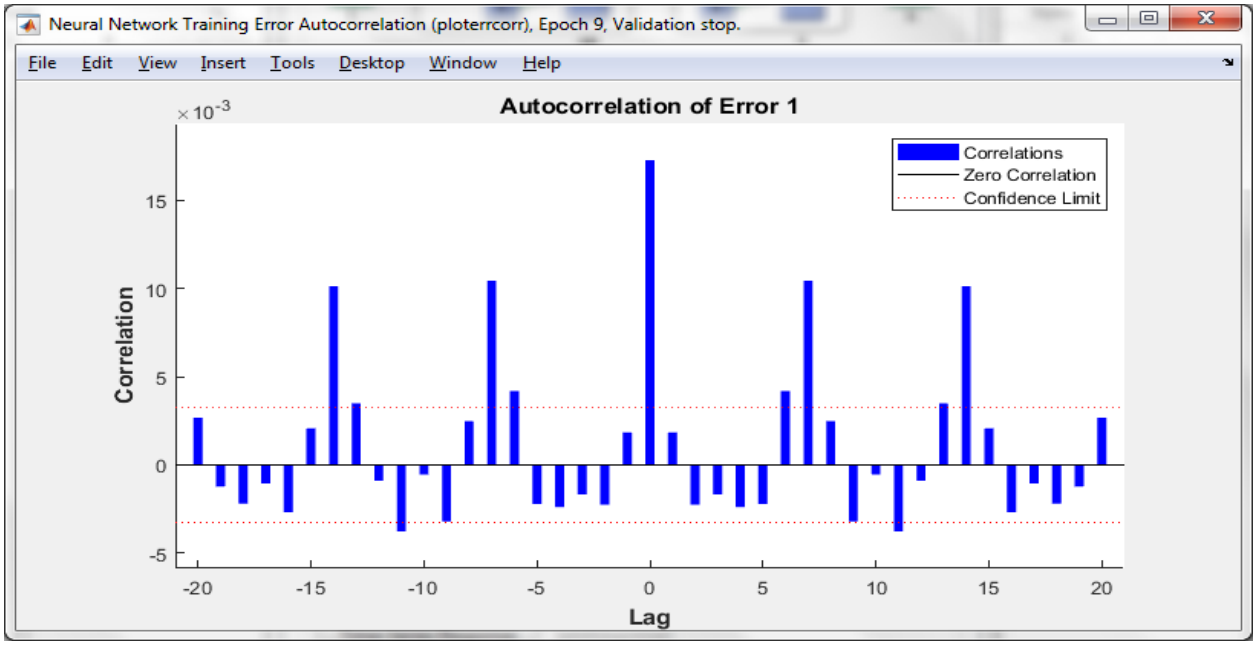

Figure 15: The error autocorrelation function

Table 4: Table of forecast value for 28 period

t	NARNN_Forecast	t	NARNN_Forecast
1	8.804656107	16	8.619118564
2	8.748637471	17	8.619108892
3	8.68656546	18	8.61910048
4	8.643547286	19	8.619093164
5	8.625918769	20	8.619086801
6	8.620904417	21	8.619081264
7	8.619664438	22	8.619076447
8	8.619353256	23	8.619072256
9	8.619260406	24	8.61906861
10	8.619220248	25	8.619065436
11	8.619194566	26	8.619062675
12	8.619174343	27	8.619060272
13	8.619157243	28	8.619058181
14	8.619142489		
15	8.619129688		

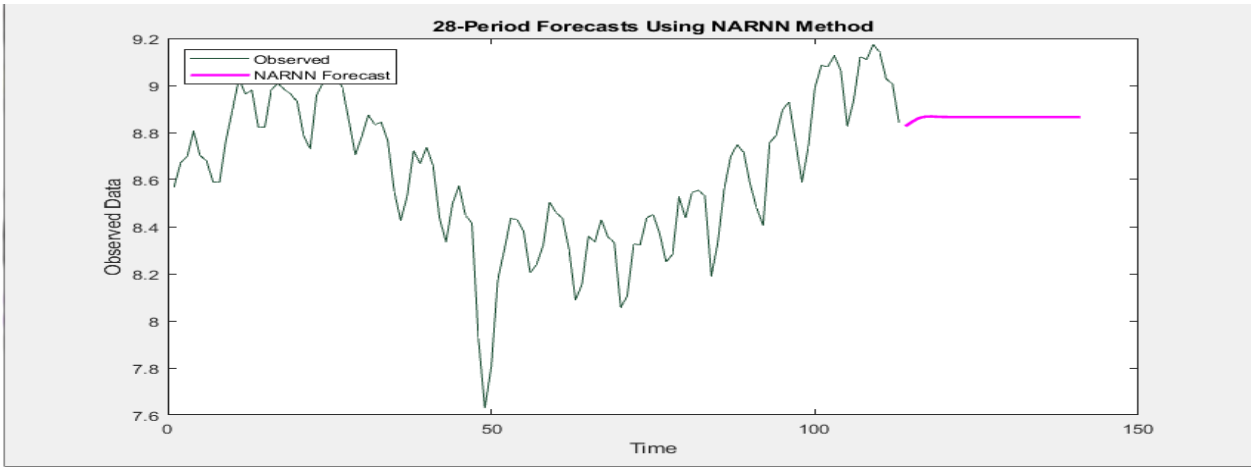

Figure 16: Forecast value for 28 period

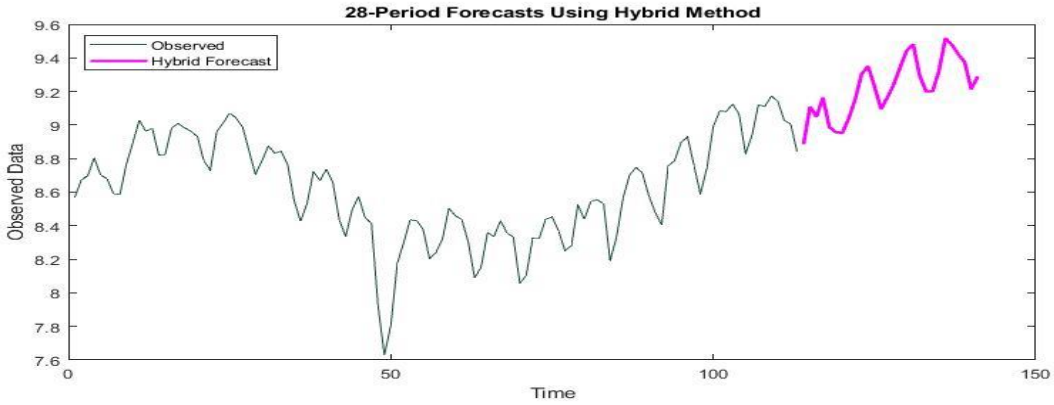

Figure 17: Forecast value for 28 period

Table 5: Table of forecast value for 28 period

T	Hybrid Forecast	T	Hybrid Forecast
1	8.885611365	16	9.346708299
2	9.109598786	17	9.44467012
3	9.049011156	18	9.480805326
4	9.163632747	19	9.293263908
5	8.986931684	20	9.200886718
6	8.9571889	21	9.200636023
7	8.952946897	22	9.32426198
8	9.038802056	23	9.517467162
9	9.152175833	24	9.477394054
10	9.304417517	25	9.4219816
11	9.350791944	26	9.372213726
12	9.229766578	27	9.212875893
13	9.095392963	28	9.288949427
14	9.164300061		
15	9.244066366		

Table 6: Table of Prediction performance results of three models

MAPE	MAE	RMSE	Method
0.021109	0.194325	0.235021138	SARIMA
0.042413	0.393954	0.43597	NARNN
0.018927	0.174539	0.219252	Hybrid

4. Conclusion

From the preceding results, it's clear that the SARIMA and SARIMA-NARNN models were predicted to be superior. While the produced discoveries for anticipating execution were analyzed utilizing RMSE, MAE, and MAPE, the hybrid model was found to have the best prediction when compared to other models employed separately. When compared to SARIMA and SARIMA-NARNN models, it was sometimes clear that the NARNN model did not produce improved estimates.

References

[1] A.E. McLeod, K.W. Hipel, and W. C. Lennox, , 1977 "Advances in Box- Jenkins modeling, applications," Water Resources Research, vol.13, pp.577-586.

[2] Bimenyimana, S., Norense, G., Asemota, O., Lingling, L., & Li, L. (2017). Output power prediction of photovoltaic module using nonlinear autoregressive neural network. *Researchgate.Net*, *2*(4), 32–40.

[3] G.E.P. Box, G. Jenkins, Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco,CA, 1970

[4] I. A. Iwok. (2017). Handling Seasonal Autoregressive Integrated Moving Average Model with Correlated Residuals. *American Journal of Mathematics and Statistics*, *7*(1), 1–6.

[5] Melin, P., Monica, J. C., Sanchez, D., & Castillo, O. (2020). A new prediction approach of the COVID-19 virus pandemic behavior with a hybrid ensemble modular nonlinear autoregressive neural network. *Soft Computing*, *0123456789*.

[6] Permanasari, A. E., Awang Rambli, D. R., & Dominic, P. D. D. (2009). Prediction of Zoonosis Incidence in Human using Seasonal Auto Regressive Integrated Moving Average (SARIMA). In *International Journal of Computer Science and Information Security* (Vol. 5, Issue 1, pp. 103–110).

[7] Perone, G. (n.d.). *Title : ARIMA forecasting of COVID-19 incidence in Italy, Russia, and the USA Gaetano Perone 1*. *December 2019*.

[8] Poonia, N., & Azad, S. (2020). *Short-term forecasts of COVID-19 spread across Indian states until 1 May 2020*. *May*.

[9] Ribeiro, M. H. D. M., da Silva, R. G., Mariani, V. C., & Coelho, L. dos S. (2020). Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. *Chaos, Solitons and Fractals*,

[10] Wang, K. W., Deng, C., Li, J. P., Zhang, Y. Y., Li, X. Y., & Wu, M. C. (2017). Hybrid methodology for tuberculosis incidence time-series forecasting based on ARIMA and a NAR neural network. *Epidemiology and Infection*, *145*(6), 1118–1129.

[11] Zhou, L., Xia, J., Yu, L., Wang, Y., Shi, Y., Cai, S., & Nie, S. (2016). Using a hybrid model to forecast the prevalence of schistosomiasis in humans. *International Journal of Environmental Research and Public Health*, *13*(4).

[12] Zhou, L., Zhao, P., Wu, D., Cheng, C., & Huang, H. (2018). Time series model for forecasting the number of new admission inpatients. In *BMC Medical Informatics and Decision Making* (Vol. 18, Issue 1).

استعمال الانموذج الهجين (hybrid SARIMA-NARNN) للتنبؤ بأعداد المصابين بفايروس (كوفيد-19) في العراق

الباحث/ ايات احمد هامل
جامعة بغداد ،كلية الادارة والاقتصاد
ayat.ahmed1201b@coadec.uobaghdad.edu.iq

أ. م. بيداء اسماعيل عبد الوهاب
جامعة بغداد ،كلية الادارة والاقتصاد
baidaa_29@coadec.uobaghdad.edu.iq

Received:5/1/2022 **Accepted: 23/1/2022** **Published: June / 2022**

مستخلص البحث:

فايروس كورونا هو احد الامراض التنفسية الحادة ظهرت رسميا في مدينة ووهان الصينية في كانون الثاني 2019 وفي اواخر شباط بدأ المرض ينتشر بسرعة في جميع انحاء العالم مما تسبب في حالات طوارئ صحية واجتماعية واقتصادية. تعد السلاسل الزمنية من الاساليب الاحصائية المهمة المتبعة لدراسة وتحليل ظاهرة معينة والتعرف على نمطها والعوامل المؤثرة فيها واستعمالها للتنبؤ بالقيم المستقبلية. ان المحور الاساسي للبحث هو تسليط الضوء على دراسة النماذج المنفردة SARIMA و NARNN والنماذج الهجينة hybrid SARIMA-NARNN وعلى فرض ان السلسلة تضم المركبتين الخطية وغير الخطية بأعتبار ان انموذج SARIMA له القدرة على التعامل مع المركبة الخطية و انموذج NARNN له القدرة على التعامل مع المركبة الغير خطية. تم تطبيق النماذج في القطاع الصحي للتنبؤ بأعداد المصابين بفايروس (كوفيد-19) في العراق وتم الحصول على البيانات من الموقع الرسمي لوزارة الصحة العراقية عن طريق الموقف الوبائي اليومي لكل محافظات العراق. وعند تحليل ودراسة هذه النماذج والمقارنة بينهم لاحظ الباحث تفوق الانموذج الهجين على النماذج الاخرى لأمتلاكه ادنى قيمة للمعايير (MSE,MAPE,MAE,RMSE), لذا تم استخدامه للتنبؤ بالقيم المستقبلية.

المصطلحات الرئيسة للبحث: فايروس كورونا, السلاسل الزمنية, الانموذج الهجين, التنبؤ

*البحث مستل من رسالة ماجستير