JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 36, 1167-1177 (2020)
DOI: 10.6688/JISE.202011_36(6).0002

Forecasting Based on Some Statistical and Machine Learning Methods

AZHARI A. ELHAG[1] AND ABDULLAH M. ALMARASHI[2]
[1]*Mathematics and Statistics Department*
Taif University
Taif, 21974 Saudi Arabia
[2]*Statistics Department*
King Abdulaziz University
Jeddah, 21589 Saudi Arabia
E-mail: azhri_elhag@hotmail.com

Forecasting consists basically of using data to predict the value of the attributes to promote micro- and macro-level decision making. There are many methods to do prediction extending from complexity and data requirement. In this paper, we present the method of an autoregressive integrated moving average (ARIMA), multilayer perceptron artificial neural network (ANN) model and decision tree (DT) method to forecast time-series data, also we use different methods to measure the accuracy of the forecasting of the patient dying after having Ebola virus in the Republic of Liberia over the period of 25 March 2014 to 13 April 2016. The data source is from World Health Organization (WHO).

Keywords: time series, modeling, deep learning, multilayer perceptron, forecasting

1. INTRODUCTION

Predicting a future behavior is an important subject in statistical science because it is important in real life [1, 2], such as predicting weather conditions and temperatures, market and price situation, water flow, energy consumption, *etc.* There has been increasing interest in the subject of prediction [3, 4] in recent years and new methods have emerged especially after the development of computer science, such as deep learning ANNs. These models are able to learn and adapt themselves to any model and they do not need assumptions about the nature of the time series, and have high capabilities in processing large data, high speed and learning efficiency as a result of their ability to accurately answer. Thus, there is a need to study the traditional methods used in time series prediction as well as the method of NNs and compare them to find the most efficient method of prediction [5]. In this paper we propose the multilayer perceptron as part of the deep learning method which has used NN in time-series data to forecast the patient dying after having Ebola virus in the Republic of Liberia.

The term NN is used in a different parameter space family of models, with the flexible structure, resulting from the analysis of brain functioning. Different new models specially under the family grew, is exposed to biological and non-biological applications. The different fields in which NN is applied determine the NN definitions but the single definition which covers variety of type models does not exist [6].

The model assumptions and its structure for NNs need minimal demand to unders-

Received December 18, 2019; revised January 6, 2020; accepted January 14, 2020.
Communicated by Mahmoud Abdel-Aty.

tand the general network structure. The purpose of minimization from prediction error (outputs) multilayer perceptron is employed as target independent variables (inputs). We study machine learning or deep learning in this paper as well as computer experiments from the elements discussed in Hastie [7]. The ANN is exposed as the more important methods for time series prediction, and this more detailed in [8]. The accuracy of the network is dependent on the number of neurons in the network and the testing the distribution of the training data [9].

The problems of forecasting is applied to the ANNs as simple computing frameworks as given in [10]. ANNs are used in the processing of natural language, image recognition and speech recognition [11]. The networks' capabilities are evaluated in a trading simulation, where predictions of exchange rate log-returns are back tested using historical data [12].

In this paper our task is to forecast the patient dying after having Ebola virus. We present the method for prediction of time-series data using MLP. The time-series data (the patient dying after a having Ebola virus) of the Republic of Liberia over the period of the study is described.

The paper is organized as follows, in Section 2, complete description of model formulation, in Section 3, numerical results of real data of Ebola virus in the Republic of Liberia over the period of 25 March 2014 to 13 April 2016. Finally, some comment is discussed in Section 4.

2. MODEL DESCRIPTION

2.1 Multilayer Perceptron (MLP)

Different family of functions are defined with the multi-layer network or perceptron. For the classical case single hidden layer NNs, function of a d-vector to m-vector

$$g(x) = b + W\tanh(c + V_x). \tag{1}$$

Under the points:

The input d-vector $\rightarrow x$
The input-to-hidden weights $k \times d$ matrix $\rightarrow V$
The hidden unit biases k-vector $\rightarrow c$
The output units biases m-vector $\rightarrow b$
The hidden-to-output weights $m \times h$ matrix $\rightarrow W$

and the function $h(x) = \tanh(c + V_x)$ defines the output of the hidden layer. Also, non-linearity case is defined in some network architectures and the hidden layer elements are presented as hidden units.

2.2 The Back-Propagation Algorithm

The distribution of multi-layer perceptron under several hidden layer. For given initial input $h_0 = x$, which are denoted by h_i for the ith layer. The output prediction points are denoted by h_K. For the $k = 1, \ldots, k - 1$:

$$h_k = \tanh(b_k = W_k h_{k\text{-}1}) \tag{2}$$

where the biases vector is given by b_k and weights matrix is given by W_k, from $k - 1$ to k. For a single unit i of layer k is

$$h_{k,i} = \tanh(b_{k,i} + \Sigma_j W_{k,i,j} h_{k\text{-}1,j}). \tag{3}$$

And the output layer is given by

$$p = h_k = \text{softmax}(b_k + h_{k\text{-}1}). \tag{4}$$

The loss is also, given by

$$L = -\log p_y, \tag{5}$$

under target class y. $p_y = P(Y = y|x)$, as well as conditional probability estimator of class y give the value of input x.

The chain in this structure is defined by

$$a_k = b_k + W_k h_{k\text{-}1} \tag{6}$$

derivation is also given by

$$\frac{\partial(-\log p_y)}{\partial_{a_{k,i}}} = (p_i - 1_{y=i}), \tag{7}$$

$$\frac{\partial \tanh(u)}{\partial u} = (1 - \tanh(u)^2). \tag{8}$$

The distribution of back-propagation is given

- With the initial output node

$$\frac{\partial L}{\partial L} = 1. \tag{9}$$

- For each $a_{K,i}$ the gradient is computed as

$$\frac{\partial L}{\partial a_{k,i}} = \frac{\partial L}{\partial L}\frac{\partial L}{\partial a_{k,i}} = (p_i - 1_{y=i}). \tag{10}$$

- This is repeated for each layer with $k = K$ down to 1.
- The wrt biases is obtained from gradient

$$\frac{\partial L}{\partial b_{k,i}} = \frac{\partial L}{\partial a_{k,i}}\frac{\partial a_{k,i}}{\partial b_{k,i}} = \frac{\partial L}{\partial a_{k,i}}. \tag{11}$$

- The wrt weights present the gradient:

$$\frac{\partial L}{\partial W_{k,i,j}} = \frac{\partial L}{\partial a_{k,i}}\frac{\partial a_{k,i}}{\partial Wb_{k,i,j}} = \frac{\partial L}{\partial a_{k,i}} h_{k-1,j}. \tag{12}$$

- If $k > 1$, this propagates the gradient back into lower layer:

$$\frac{\partial L}{\partial h_{k-1,j}} = \sum_i \frac{\partial L}{\partial a_{k,i}}\frac{\partial a_{k,i}}{\partial h_{k-1,j}} = \sum_i \frac{\partial L}{\partial a_{k,i}} W_{k,i,j}, \quad \frac{\partial L}{\partial a_{k-1,j}} = \frac{\partial L}{\partial h_{k-1,j}}\frac{\partial h_{k-1,j}}{\partial a_{k-1,j}} = \frac{\partial L}{\partial h_{k-1,j}}(1 - h_{k-1,j}^2). \tag{13}$$

2.3 Time Series

The time-series analysis is a sequence of observation on a variable on time. the most common model for forecasting a time series is Autoregressive Integrated Moving Average (ARIMA) model. This model seems as a best fitting model [13]. The forecasting model is constructed as following.

If $D = 0$: $Z_t = Z_t$
If $D = 1$: $Z_t = Z_t - Z_{t\text{-}1}$
If $D = 2$: $Z_t = Z_t - 2Z_{t\text{-}1} + Z_{t\text{-}2}$

In terms of y, the general forecasting equation is:

$$\hat{y}_t = \mu + \Phi_1 Z_{t\text{-}1} + \ldots + \Phi_p Z_{t\text{-}p} - \theta_q e_{t\text{-}q}. \tag{14}$$

To fit an ARIMA model, the first step is to determine the difference that make the series is stationary (and it is very important). Constructing the linear model of prediction – according to this analysis-passes through four basic stages: Phase I: Identification the model, which means to determine the rank of each of the autoregressive model AR(p) and the model of moving averages MA(q), as the two models comprising a model ARIMA [14, 15]. The second phase: Estimating parameters for the model proposed in the previous step. The third phrase: Testing the of quality the model. The fourth phase: Forecasting, and to choose the best forecasting model we must examine the sample autocorrelation function, partial autocorrelation function plots. The bar in the plot illustrates the value of correlation coefficient in the given lag. The overlay curve represents confidence limits calculated at minus plus standard error. The slow decline of the ACF suggests that first difference may be adjective. When ACF declines slowly at season lags the seasonal difference it becomes adequate [16]. After the models have been estimated, we must select the best one that explains the observed the data. To choose the best ARIMA model amongst many observation which be performed, the next criteria are used, BIC (Bayesian Information criteria), Q statistic, S.E. of regression [17].

2.4 Decision Tree

A decision tree is a decision support tool that uses a model similar to decisions and their likely outcomes, including event outcomes, cost of resources, and tools. This is one way to display an algorithm that contains only conditional controls. Decision trees are commonly used in machine learning. It is a streamlined structure, where each internal node represents a test on an attribute and each branch represents the test result. Each node represents a class label sheet (the decision taken after calculating all attributes). The paths from the root to the sheet represent the classification rules [18, 19].

2.5 The Accuracy Measurement

Have used the symmetric mean absolute percentage error (SMAPE) [20, 21], the mean absolute scaled error (MASE) and the mean absolute percentage error (MAPE) to measure the accuracy of the forecast. SMAPE, MASE and MAPE are given by the following formula [22, 23]:

$$SMAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|e_i|}{|y_i|+|\hat{y}_i|}, \tag{16}$$

$$MASE = \frac{\frac{1}{n}\sum_{i=1}^{n}|e_i|}{\frac{1}{n-1}\sum_{i=2}^{n}|y_i - y_{i-1}|}, \tag{17}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}|\frac{e_i}{y_i}|. \tag{18}$$

3.1 Time Series

Some Descriptive Metrics Data for the patient dying after having Ebola virus in the Republic of Liberia over the period of March 2014 to April 2016, (see Table 1) This table contains the mean and the standard deviation and the range of the data records.

Table 1. The mean and the standard deviation and the range of the data records.

Statistics			
		Total Cases Liberia	Total Deaths Liberia
N	Valid	265	265
	Missing	0	0
Mean		8233.89	3708.95
Std. Deviation		3943.688	1750.824
Range		10706	4811

From Table 1 we can see that from the period of March 2014 to April 2016 the mean of the patient dying after a having Ebola virus is 3709 and the mean of the patient having Ebola virus is 8234. The Measures of association between the total death for the patient dying after a having Ebola virus in the Republic of Liberia and the total case of those who having Ebola virus in the Republic of Liberia is shown in Table 2.

Table 2. The measures of association.

	R	*R* Squared	Eta	Eta Squared
Total Cases, Liberia ∗ WHO report date from 2014 to 2015	0.854	0.729	0.921	0.848
Total Deaths, Liberia ∗ WHO report date from 2014 to 2015	0.855	0.732	0.922	0.850

This model for this data is successful in estimating parameters significance test and it has succeeded in residuals analysis test.

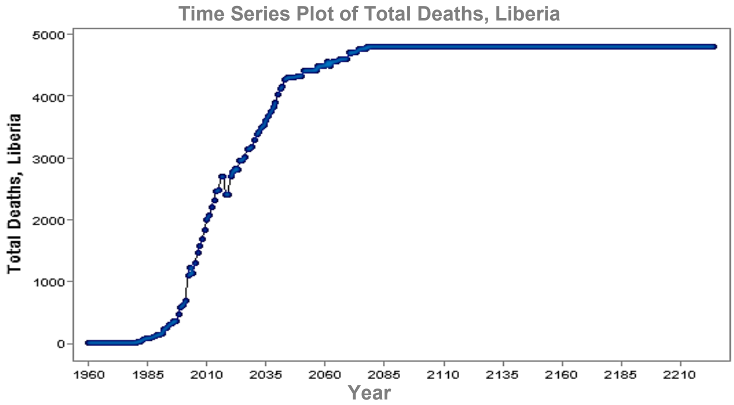

Fig. 1. The time series plot of total death, Liberia.

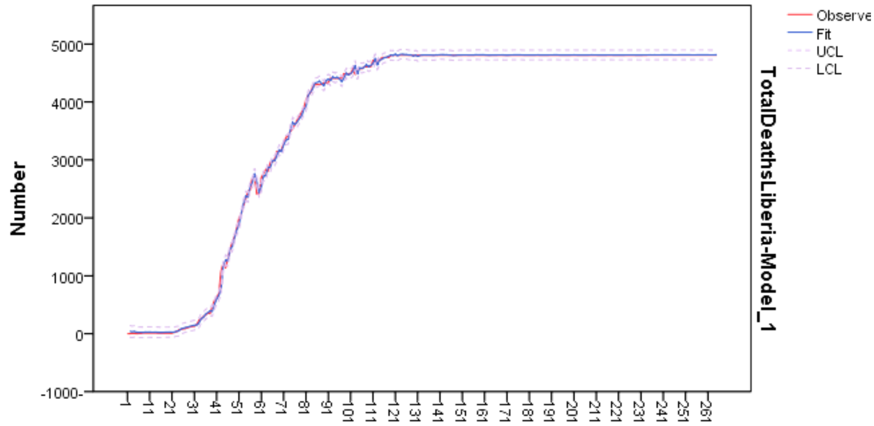

Fig. 2. A relation for the model predicted for the patient dying after a having Ebola virus in the Republic of Liberia.

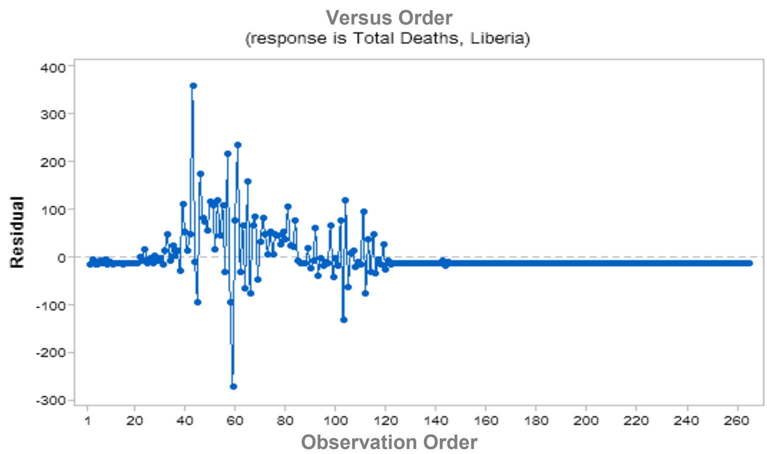

Fig. 3. Residual of response.

Table 3. Final estimates of parameters.

Model Description			
			Model Type
Model ID	Total Deaths, Liberia	Model_1	ARIMA (1, 1, 16)

Table 4. Modified Box-Pierce (Ljung-Box) Chi-Square statistic.

Model Statistics													
Model	Number of Pre-dictors	Model Fit Statistics									Ljung-Box Q(18)		
		Stationary R-squred	R-squred	RMSE	MAPE	MAE	MaxAPE	MaxAE	Normal-ized BIC	Statistics	DF	Sig.
Total Deaths, Liberia-Model_1	1	0.411	0.999	42.242	21.315	20.822	1357.581	316.231	7.888	11.561	1	0.001

We use the *p*-value (*P*) to determine whether the model meets the assumption that the residuals are independent, that means the model fits the data. The above table shows that the *p*-value is less than 0.05 indicating significance.

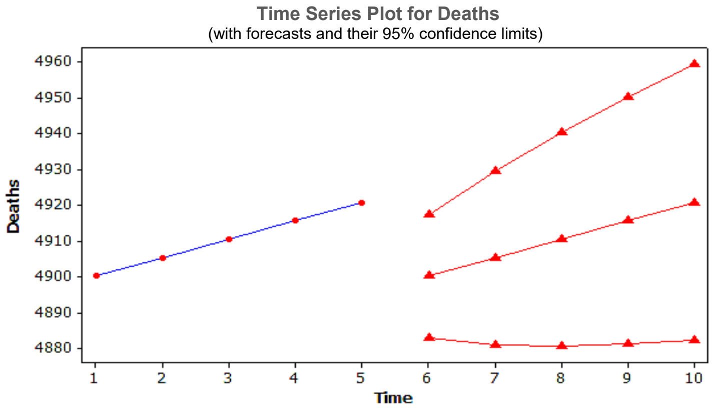

Fig. 4. Time series plot for the patient dying after a having Ebola virus.

Table 5. Forecasts from period 265.

		95% Limits	
Period	Forecast	Lower	Upper
266	4823.86	4725.38	4922.34
267	4842.21	4675.99	5008.42
268	4860.36	4647.86	5072.86
269	4878.55	4628.04	5129.06
270	4896.73	4613.29	5180.18

3.2 Decision Tree

To predict the patient dying after having Ebola virus in the Republic of Liberia based on total cases.

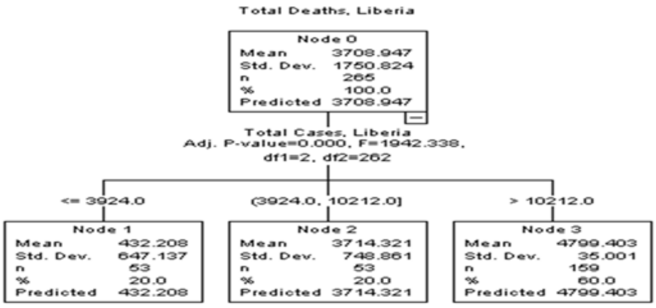

Fig. 5. The predict values of the patient dying after a having Ebola virus in the Republic of Liberia.

3.3 Neural Network

After loading the data set we use Multilayer Perceptron (MLP) taking 0:00:00.02 second to build the model. Sum of squares error for the training is 0.064 and the relative error 0.001 and the sum of squares error for the testing is 0.026 and the relative error is 0.002.

Table 6. Sum of squares, relative errors and squares error for the testing.

Model Summary		
Training	Sum of Squares Error	0.064
	Relative Error	0.001
	Stopping Rule Used	Training error ratio criterion (.001) achieved
	Training Time	0:00:00.02
Testing	Sum of Squares Error	0.026
	Relative Error	0.002
Dependent Variable: Total Deaths, Liberia		

Table 7. The network information.

Input Layer	Factors 1	Total Cases, Liberia
	Number of Units*	70
Hidden Layer(s)	Number of Hidden Layers	1
	Number of Units in Hidden Layer 1*	2
	Activation Function	Hyperbolic tangent
Output Layer	Dependent Variables 1	Total Deaths, Liberia
	Number of Units	1
	Rescaling Method for Scale Dependents	Standardized
	Activation Function	Identity
	Error Function	Sum of Squares

* Excluding the bias unit

Units in hidden Layers for the Hidden Layer the activation function is a Hyperbolic tangent, for the output Layer the dependent variable (Total Deaths, Liberia) used the Sigmoid function and the Sum of Squares are used as the error Function.

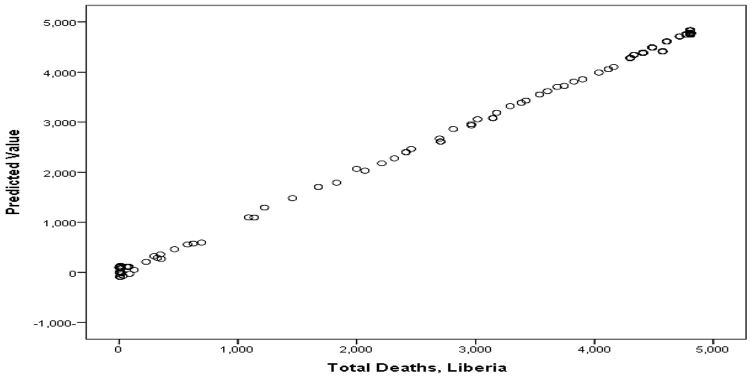

Fig. 6. The prediction results using the proposed Neural Network.

3.4 Forecasting Performance

Fig. 7 shows the results of predicted values using the above methods.

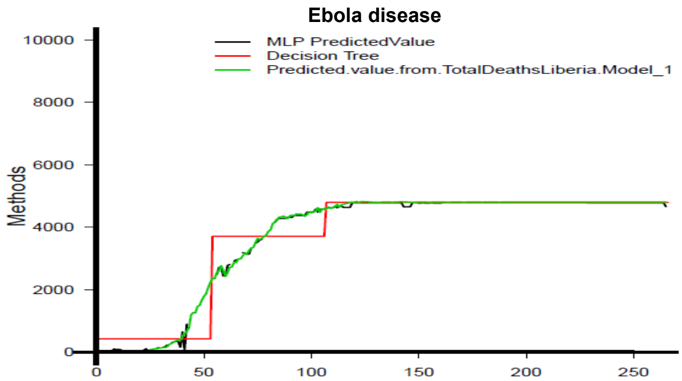

Fig. 7. The results of predicted values using the above methods.

4. CONCLUSION

This paper uses Box-Jenkins and decision tree models and investigates the application of neural network problem of prediction the patient dying after having Ebola virus in the Republic of Liberia where is represents a huge problem. It is clear to show that the both ARIMA and MLP are best models to use, according to the value of SMAPE, MASE and MAPE (see Table 8). It must be pointed out that the implementation of such a mechanism

Table 8. The performance of the forecasting.

	MAPE	MASE	sMAPE
ARIMA	0.2131	0.9312	0.0063
CART	4.3833	10.9341	0.1529
MLP	1.1128	1.6912	0.0286

to predict the patient dying after having Ebola virus in the Republic of Liberia is extremely useful. Finally, there must be other studies on other countries that have the same conditions as Ghana, Sierra Leone.

ACKNOWLEDGEMENT

This project was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, Saudi Arab under Grant No. KEP-35-130-39. The authors, therefore acknowledge with thanks DSR technical and financial support.

REFERENCES

1. M. M. Mohie El-Din, M. M. Amein, A. M. Abd El-Raheem, H. E. El-Attar, and E. H. Hafez, "Estimation of the coefficient of variation for Lindley distribution based on progressive first failure censored data," *Journal of Statistics Applications & Probability*, Vol. 8, 2019, pp. 83-90.
2. M. A. Sabry, H. Z. Muhammed, A. Nabih, and M. Shaaban, "Parameter estimation for the power generalized Weibull distribution based on one- and two-stage ranked set sampling designs," *Journal of Statistics Applications & Probability*, Vol. 8, 2019, pp. 113-128.
3. A. Yusuf and S. Qureshi, "A five parameter statistical distribution with application to real data," *Journal of Statistics Applications & Probability*, Vol. 8, 2019, pp. 11-26.
4. D. Kumar, U. Singh, S. K. Singh, and P. K. Chaurasia, "A new lifetime distribution: Some of its statistical properties and application," *Journal of Statistics Applications & Probability*, Vol. 7, 2018, pp. 413-422.
5. A. Tavanaei, M. Ghodrati, S. R. Kheradpisheh, T. Masquelier, and A. Maida, "Deep learning in spiking neural networks," *Neural Networks*, Vol. 111, 2019, pp. 47-63.
6. I. Argatov, "Artificial Neural Networks (ANNs) as a novel modeling technique in tribology," *Frontiers of Mechanical Engineering*, Vol. 29, 2019, pp. 1-9.
7. T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning Data Mining*, *Inference and Prediction*, 2nd ed., Springer Series in Statistics, Berlin, 2008.
8. D. Samek and P. Varacha, "Time series prediction using artificial neural networks: single and multi-dimensional data," *International Journal of Mathematical Models and Methods in Applied Sciences*, Vol. 1, 2013, pp. 38-46.
9. Y. Li and W. Yufa, "Machine learning methods of bankruptcy prediction using accounting ratios," *Open Journal of Business and Management*, Vol. 6, 2018, pp. 1-20.
10. H. C. Koh and L. N. Killough, "The use of multiple discriminant analysis in the assessment of the going-concern status of an audit client," *Journal of Business Finance & Accounting*, Vol. 17, 2010, pp. 179-192.
11. S. K. Gouda, S. Kanetkar, D. Harrison, and M. K. Warmuth, "Speech recognition: Key word spotting through image recognition," arXiv:1803.03759v1[stat.ML], 2018, 1-9.
12. M. Khashei and M. Bijari, "A new hybrid methodology for nonlinear time series forecasting," *Modelling and Simulation in Engineering*, Vol. 4, 2011, pp. 1-5.
13. H. Hedman, "Performance evaluation of artificial neural networks in the foreign exchange market," Master's Thesis, Department of Mathematics, Royal Institute of Technology, 2012.

14. S. P. Shakti, *et al.*, "Annual automobile sales prediction using ARIMA model," *International Journal of Hybrid Information Technology*, Vol. 10, 2017, pp. 13-22.
15. A. P. Ansuj, "Sales forecasting using time series and neural network," *Computer & Industrial Engineering*, Vol. 31, 1996, pp. 421-424.
16. J. Wang, J. Liang, J. Che, and D. Sun, "ARIMA model identification using particle swarm optimization algorithm," in *Proceedings of International Conference on Computer Science and Information Technology*, 2008, pp. 223-227.
17. A. Wadia and M. T. S. Ismail, "Selection wavelet transforms model," *Applied Mathematical Science*, Vol. 5, 2011, pp. 315-326.
18. B. Kamiński, M. Jakubczyk, and P. Szufel, "A framework for sensitivity analysis of decision trees," *Central European Journal of Operations Research*, Vol. 26, 2017, pp. 135-159.
19. K. Karimi and H. J. Hamilton, "Generation and interpretation of temporal decision rules," *International Journal of Computer Information Systems and Industrial Management Applications*, Vol. 3, 2011, pp. 314-323.
20. M. R. Abonazel, "Different estimators for Stochastic parameter panel data models with serially correlated errors," *Journal of Statistics Applications & Probability*, Vol. 7, 2018, pp. 423-434.
21. M. G. M. Ghazal, "Prediction of exponentiated family distributions observables under type-II hybrid censored data," *Journal of Statistics Applications & Probability*, Vol. 7, 2018, pp. 307-319.
22. E. Torsen, P. N. Mwita, and J. K. Mungatu, "Nonparametric estimation of the error functional of a location-scale model," *Journal of Statistical and Econometric Methods*, Vol. 7, 2018, pp. 1-19.
23. S. Spiliotis and E. Assimakopoulos, "Statistical and machine learning forecasting methods: Concerns and ways forward," *PLoS ONE*, Vol. 13, 2018, pp. 1-26, e019.

Azhari A. Elhag received Ph.D. degree in Statistics in 1993 from Saint Petersburg State University of Economics and Finance in Department of Statistics. His research interests include factor analysis. He is the author of several articles published in different international scientific journals and is a member of different working groups. He is an Assistant Professor, of Applied Statistics, International University of Africa, Khartoum, Sudan, previously. Currently, is an Assistant Professor, of Applied Mathematics, Mathematics Department, Faculty of Science, Taif University, Taif, Saudi Arabia.

Abdullah M. Almarashi received the Ph.D. degree in Statistical Sciences (Time Series Analysis) from the School of Science, University of Strathclyde Glasgow, UK in 2014. He is currently an Associate Professor with the Faculty of Science, Statistics Department, King Abdul Aziz University (KAU), and Saudi Arabia since 2019. Moreover, and since 2019, he has been a Consultant of the Vice Presidency of Educational Affairs for Strategic Planning, King Abdulaziz University. His research interests include time series analysis, statistical inference, medical statistics, and regression models.