Al-Qadisiyah Journal for Administrative and Economic Sciences ISSN$_{Online}$: 2312-9883
QJAE, Volume 27, Issue 2 (2025) ISSN$_{Print}$: 1816-9171
DOI: https:// doi.org/ 10.33916/ qjae.2025 .026068

Comparison between neural networks and time series models to predict the number of traffic accidents in Basrah Governorate For the period (2024-2026)

Muntather J. Mehdi Raissan A . Zalan
montather.jumaa@uobasrah.edu.iq ressan.zalan@uobasrah.edu.iq
University of Basrah

Article history:
Received: 17/3/2025
Accepted: 5/5/2025
Available online: 15 /6 /2025

Corresponding Author : Muntather J. Mehdi Raissan A . Zalan

Abstract: Traffic accidents are considered one of the most dangerous social and security problems facing most countries, including our beloved country, which is a major cause of many deaths and injuries. Therefore, traffic accidents are considered one of the main causes of death in most countries. As a result of the increase in the population in Iraq in general and in Basra in particular, as well as the improvement of living standards, the number of cars is constantly increasing. As a result, the number of accidents is constantly increasing. Therefore, the aim of this study is to attempt to develop a proposed model to predict traffic accidents in Basra Governorate, using several time series models and comparing them with neural networks in order to reach the best model to predict the number of traffic accidents in the near future. The importance of this study comes from following an advanced scientific method to determine the best and most efficient statistical model in order to use it in predicting the number of traffic accidents in later periods so that the authorities responsible for roads and planning authorities can know what this phenomenon will lead to in the future in order to reduce it. The study concluded that the neural network model was better than other models and was adopted in predicting later periods.

Keywords: traffic accidents, time series, neural networks, prediction.

INTRODUCTION: Traffic accidents are a major source of concern in most countries of the world, including Iraqi society and Basra society, as they are one of the reasons for the depletion of human and material resources in society, and they also cause psychological and social problems among members of society. Therefore, it is necessary to predict their numbers in a scientific manner in order to confront this problem, reduce its effects, and develop solutions and proposals to treat, mitigate, or prevent them. The number of these accidents increases from one year to another for many reasons, the most important of which will be mentioned in this research. Therefore, statistics on the number of accidents were obtained from the Traffic Directorate in Basra Governorate in order to predict their numbers for the coming period using the best prediction methods, which are the Box-Jenkins method and neural networks, in order to reduce their effects.

Some previous Arab studies:

- A study titled: Using Time Series Analysis to Predict the Number of Traffic Accidents in the City of Misrata from 2012 to 2021 by researcher (Amhamed Milad Baouh). The study aimed to find the best and most efficient statistical models for predicting the number of traffic accidents in the city of Misrata, relying on monthly data. The study concluded that the best prediction model was ARIMA (0,1,1), as it was the most appropriate and accurate in prediction.
- A study titled: Predicting Traffic Accidents in the City of Lattakia Using Artificial Neural Networks by researchers (Dr. Shaza Alwan and Noura Hamdan). The study aimed to build a model for predicting traffic accidents using neural networks for the city of Lattakia. The study results demonstrated the high reliability of artificial neural networks in predicting the number of traffic accidents in the city of Lattakia. - A research entitled "Developing a Model for Predicting Traffic Accidents Using the Fuzzy Logic Approach" "Case Study: Lattakia City" by the researchers (Dr. Shaza Alwan and Nagham Saleh). The research aims to develop a model for predicting traffic accidents using the

fuzzy logic method on urban roads in order to reduce the possibility of many future accidents occurring. The study concluded that the road's engineering specifications did not have a clear impact on traffic accidents.

Study problem:

As a result of the continuous increase in traffic accidents in Basra Governorate and the accompanying increase in the number of deaths and injuries, which may make them compete with other major factors in the causes of death such as heart disease, cancer, and others, it is necessary to be aware of what the future numbers of traffic accidents will lead to in order to control them by the competent authorities.

Study objective:

Propose a model to predict the number of traffic accidents in Basra Governorate by comparing the ARIMA model and neural networks.

Concept of traffic accident[(3)] :

A traffic accident is defined as an interceptive event that occurs without prior planning by one or more cars and results in damages, injuries and deaths.

Causes of traffic accidents[(6)]:

There are several causes of traffic accidents:

If we want to know the causes of traffic accidents, we must know that all causes are either due to a mechanical error in the car or due to human errors due to the driver or due to the road and the surrounding conditions. These causes can be limited to two main reasons:

1- Direct causes:

These are the causes that actually lead to accidents, so they are called direct, such as:

1. The driver's failure to estimate the correct traffic situation.
2. The driver's lack of sufficient background about the road.
3. The driver's lack of sufficient skill and experience that the driver should have.
4. Excessive speed.

2- Indirect causes:

These are the causes that contribute to the occurrence of accidents and are therefore called indirect, such as:

1. The presence of a poor road.
2. Lack of traffic awareness among drivers.
3. Lack of actual application of traffic regulations.

So we conclude from the above that there are direct and indirect causes that lead to traffic accidents.

Traffic accidents in Basra for the period 20020-2023:

The following is a table showing the number and type of traffic accidents for the period 2020-2023:

Table (1) Numbers and types of traffic accidents for the period 2020-2023

The year / The accident	2020	2021	2022	2023
Collision	44	86	416	452
Run-over	6	12	80	83
Rollover	14	11	58	53
Complex	-	-	65	35
Total	64	109	619	623

source: Statistics of Basra Governorate Traffic Directorate

We note from the table that in terms of number, accidents are taking a general increasing trend, but in terms of the type of accident, we note that the type of collision has the largest number throughout these years among other types.

Traffic accident losses in Basra for the period from 2020-2023:

Table (2) Number of traffic accident losses in Basra for the period from 2020-2023

The year / Type of loss	2020	2021	2022	2023
Deceased	60	89	248	242
Injured	84	159	893	901
Total	144	248	1141	1143

Source: Basra Traffic Directorate Statistics

We note from the table that the number of deaths due to traffic accidents in 2022 was more in terms of numbers, but in terms of the number of injured, we note that 2023 was more than the rest of the other years.

The concept of time series [2],[5]:

They are the values of a phenomenon that are arranged according to time or are a set of values that the phenomenon takes in time periods that are consecutive and equal. It is also known as a sequence of observed values of a random phenomenon arranged over time.

Exponential Smoothing Models[11]:

The basis of the exponential smoothing method in forecasting is to weight or smooth past observations of the time series to obtain a prediction of the future. Exponential smoothing depends on all previous values based on what is called the weighted moving average, which uses a specific coefficient called the exponential smoothing constant, and its value is limited between zero and one. The smoothing method is distinguished from the moving average in that it involves all the values of the series in the prediction, and the name exponential means that the coefficients decrease under the influence of the exponential power.

There are several methods for forecasting using the exponential smoothing method, and the use of these methods depends on the nature of the data, whether it has a trend or not, is it seasonal or non-seasonal. In the case of research data, the data has a trend, so the Holt exponential method was used, which is suitable for such data.

Exponential Holt Method[7] :

This method is used to predict the values of the time series in the event of a trend in the data, and it consists of three equations with two exponential smoothing constants, which are (α, β), as their values are between zero and one, as the first constant is used to smooth the time series, while the second constant is used to get rid of the trend in the data. The following are the equations of this method:

$F_t = \alpha Y_t + (1-\alpha)(F_{t-1} + b_{t-1})$, t=1,2,…,n

$b_t = \beta(F_t - F_{t-1}) + (1-\beta)b_{t-1}$

$\hat{Y}_t = F_n + b_n * m$,$m \geq 1$

Where: m is the time period to be predicted

F_t : the smoothed value of the time series

b_t: the magnitude of the linear trend of the data

$\hat{Y}_t$: the predictive value of the series

he exponential smoothing values (α,β) are extracted according to trial and error, as different values are given until we obtain the lowest error in estimating the data.

Box-Jenkins models[8] :

The Box-Jenkins package is usually called ARIMA models, which are models of moving averages and autoregression after taking the differences for the time series when they are not stationary. Since the series are divided into stable, unstable, seasonal and non-seasonal, we will only discuss the non-seasonal Box-Jenkins models.

The non-seasonal Box-Jenkins models are divided into two types[9] :

Stable time series: These are models that are valid in representing time series without taking any number of differences for them, i.e. they are originally stable.

A- Unstable time series.: These are models that are not valid in representing time series except after taking a number of differences for them, i.e. they are originally unstable.

The Box-Jenkins methodology is based on three parts in its formulation:

- The auto-regressive models:

It is defined as the current value in the time series being in terms of the weighted sum of the previous values plus the random error term, and the general formula for this model is explained as follows:

$$Y_t = \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \cdots + \theta_p Y_{t-p} + \epsilon_t \ , \qquad \epsilon_t \sim N(0, \sigma_\epsilon^2)$$

Where P symbolizes the degree of autoregression, which is symbolized by the symbol: AR(P).

- The moving average models:

This model is defined as the value of the current series being in terms of the values of the current and previous random variables plus the random error term, and the general formula for this model is explained as follows:

$Y_t = \epsilon_t + \emptyset\ \epsilon_{t-1} + \emptyset_2\ \epsilon_{t-2} + \cdots + \emptyset_p\ \epsilon_{t-q}$

Where q symbolizes the degree of moving averages, which is symbolized by the symbol: MA(q)

- Mixed auto-regressive moving average models:

This model is a composite of the AR model and the MA model and consists of (P+q) of parameters and takes the following general formula:

$Y_t = \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \cdots + \theta_p Y_{t-p} + +\emptyset \in_{t-1} + \emptyset_2 \in_{t-2} + \cdots + \emptyset_p \in_{t-q}$

It is symbolized by the symbol ARMA(P,q).

The Box-Jenkins method depends on four stages in order to reach the prediction, which are[7] :

1- Identification stage
2- Estimation stage
3- Checking Diagnostic stage
4- Forecasting stage

Artificial neural networks[1],[10] :

It is a system used to process and deal with data in a way that resembles and simulates the way the human biological nervous system works, as these networks are characterized by the same characteristics of the nervous system that depends on the idea of storing information and learning it and then benefiting from it later. The neural network is one of the most important statistical methods used in the prediction process. Neural networks generally consist of four layers:

A- Input Layer
2- Hidden Layer
3- Summation Layer
4- Output Layer

First: Input Layer:

It is the first layer that the neural network consists of, and this layer may consist of one layer or several layers. For example, in the linear model, this layer consists of one layer) an internal variable + an external variable (There is also a unit that represents the bias value (Bias Unit) which is equal to (1). As shown in the following figure:

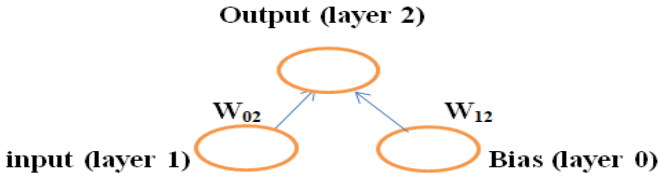


Figure (1) Neural network for linear model (input 1 + output 1)

Second: The hidden layer:

This layer consists of one node for each variable or input vector, and therefore if there is more than one input variable, it will consist of several nodes according to the number of these variables in the training set, as each node of these nodes in this layer is trained only once. The idea of this layer is based on multiplying each input value by its weight, and then the result of the multiplication is passed to the activation function, as different functions are used, such as the (Hyperbolic Tangent) function, as this function enables us to spread the output values between [-1, 1] in a more streamlined manner, or use the sigmoid function, which is widely used because it is more streamlined in spreading the output values between [0, 1], which leads to giving the non-linear model in the neural network greater flexibility to match the most appropriate shape for the curve among the inputs and outputs, as in the following figure:

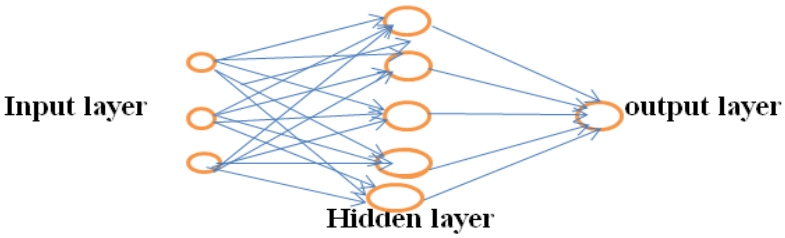


Figure (2) The neural network of the non-linear model (several inputs + one output + several hidden units)

Third: The sum layer:

In this third layer, which is the sum layer, each summation node receives the outputs from the layer before it, which is the hidden layer, for each of the existing variables.

Fourth: The output layer:

This layer is considered the last layer in the neural network, which is specific to the outputs, as the output nodes are represented by dual neurons, and they are responsible for the forward propagation of signals entering the network, so if the neural network is feed forward (NN), then the spread of signals entering the network is always forward, and thus

any signal coming out of any cell will depend on the incoming signal only. However, if the neural network is feed backward (NN), then the network will have one or more feed back loops, and thus the output of the signal from any neuron is due to the input of all the remaining neurons, which are rarely used.

Training:

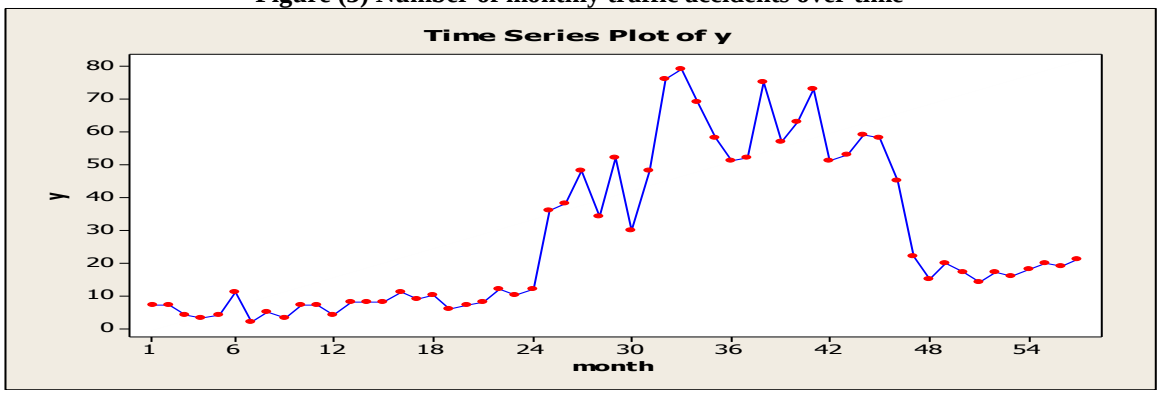
The process of training the neural network is done by determining initial values for the weights and biases so that the network becomes ready for training. Through the training process, these weights and biases are changed repeatedly until the smallest value for the performance function is reached, as the mean square error (MSE) is considered the default performance function for the feed-forward network.

The practical aspect:

First: Forecasting using the Box-Jenkins method:

Before starting to use the Box-Jenkins method for forecasting, it is necessary to verify the stability of the time series in the mean and variance as follows:

Figure (3) Number of monthly traffic accidents over time



Source: Prepared by the researcher according to the outputs of the Minitab program

It is initially clear from the shape of the series that it is not stable in the mean and variance due to the presence of a trend in the series.

To confirm this, the autocorrelation function and the partial autocorrelation function were used as follows:

Figure (4) Autocorrelation function for the time series (y)

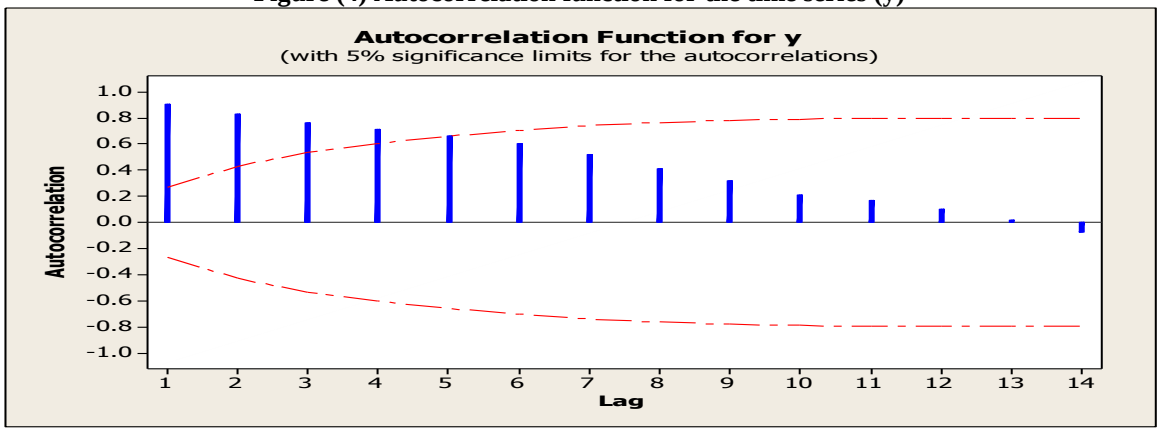
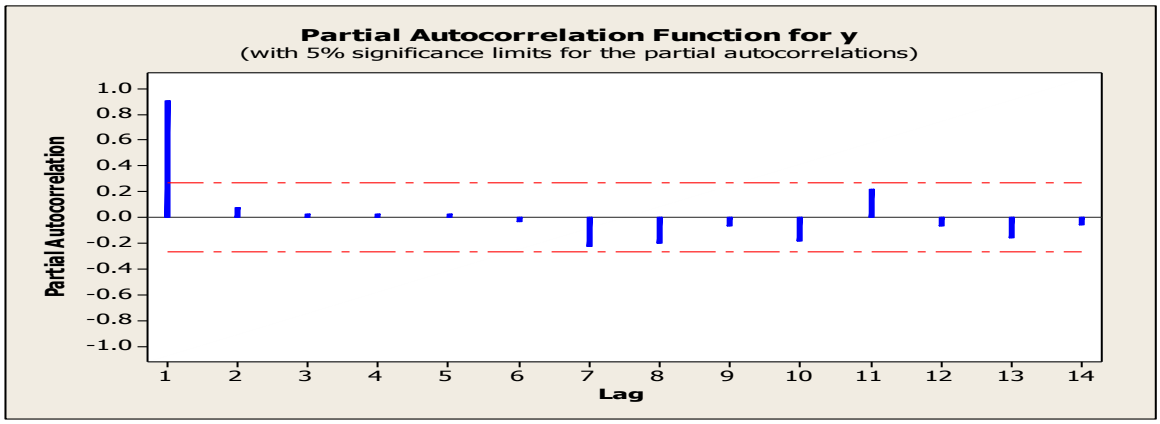


Figure (5) Partial autocorrelation function of the time series (y)



We notice from the two figures that the time series is not stable.

The Dickey Fuller unit root test was also used to verify whether the time series is stable or not, as follows:

H_0: P=0 There is a unit

H_1: P>0 There is no unit root

Table (3) Dickey Fuller unit root test

The series	with constant and direction		without constant and direction		with constant	
y	t	Prob.	t	Prob.	t	Prob.
	-0.957083	0.2983	-1.586607	0.7858	-1.708939	0.4213

We note from the table that the value of Prob. for the t-test is greater than 0.05, which means accepting the null hypothesis, i.e. the instability of the time series.

The natural logarithm of the time series was taken, then the differences were taken so that the time series would be stable with the mean and variance, and the autocorrelation function of the series was found after taking the logarithm and the first difference, as follows:

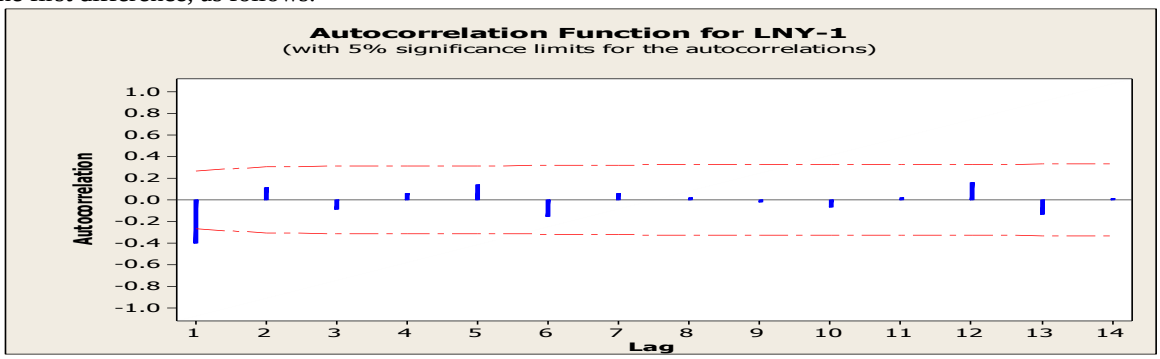


Figure (6) Autocorrelation function of the time series y after taking the logarithm and taking the first difference

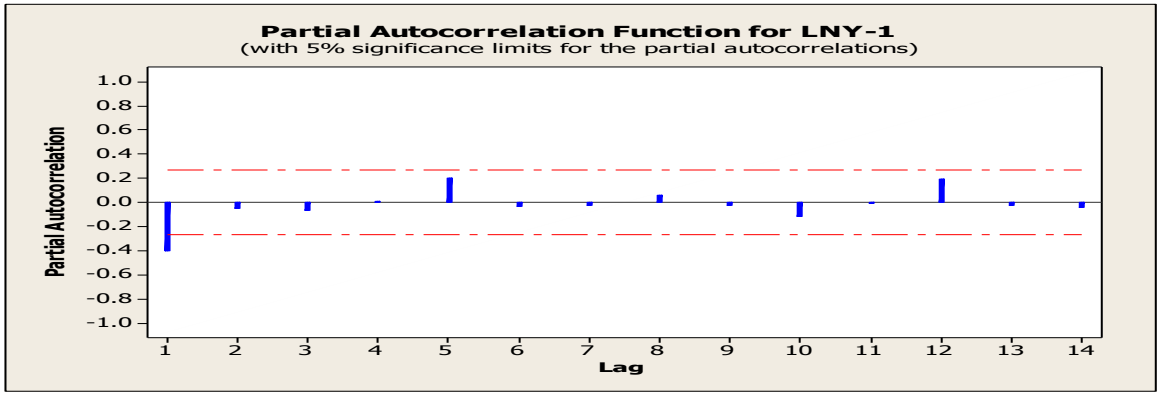


Figure (7) Partial autocorrelation function of the time series y after taking the logarithm and taking the first difference

We notice from Figures (6) and (7) that the series is stable. To confirm, the unit root of the time series was found after taking the natural logarithm, as follows:

Table (4) The expanded Dickey-Fuller test for the time series y after taking the logarithm and taking the first difference

The series	with constant and direction		without constant and direction		with constant	
y	t	Prob.	t	Prob.	t	Prob.
	-8.379398	0.0000	-8.305288	0.0000	-8.307241	0.0000

We note from Table ( ) that the value of Prob. for the t-test is less than 0.05, which means rejecting the null hypothesis, i.e. the stability of the time series.

After the time series was stabilized after taking the first difference, several models were nominated for the Box-Jenkins method and then compared using the MSE criterion to choose the best model for prediction.

Table (5) The nominated models for the ARIMA model and the mean square error for each of them

The model	mse
ARIMA(1,1,0)	0.03342
ARIMA(0,1,1)	0.03360
ARIMA(1,1,1)	0.03391
ARIMA(2,1,0)	0.03396
ARIMA(2,1,1)	0.03455
ARIMA(0,1,2)	0.03393
ARIMA(1,1,2)	0.03452
ARIMA(2,1,2)	0.03519

From Table (5) we notice that the best model for prediction is ARIMA(1,1,0) because it has the lowest mean square error which is (MSE=0.03342).

Second: Prediction using the exponential smoothing method:

Before starting to forecast using the exponential smoothing method, it is necessary to note the shape of the time series plot in Figure (4) in order to determine the appropriate method for forecasting using this method. We notice from the figure that there is a trend in the series, and therefore the best method for forecasting is the Holt exponential method, as the simulation method was used in the QSB program to predict the constants (α, β) if they were: (α=0.62, β=0)

The value of the mean square error was determined, which was equal to(MSE=0.03267702)

Third: Prediction using neural networks:

the first stage is Building the network:

The network goes through several stages: The first is determining the inputs, as this is considered the first step in order to build the network for prediction purposes, and the number of appropriate inputs for the prediction network will be determined according to the data of the number of traffic accidents entered.

As for the second stage, it is the analysis stage, as the network analysis process was carried out, and the following was shown:

Groups	Training group	Test group	Total observations
Percentage	%85	%15	%100
Number	48	9	57

Source: Prepared by the researchers based on the outputs of the R program

As for the next stage, it is the processing stage, relying on the R program, the data under study was represented.

The data	The input	The output
The representation	(1-‹1)	(1‹0)

In the training phase, the Support vector machine algorithm was used, and the learning rate was fixed at : Learning Rate=0.1 and the momentum constant at Momentum Constant = 0.2. The number of iterations = 500. Then, the prediction was made and the mean square error was extracted to predict the neural networks, which was: MSE = 0.03097629

Comparison between models:

In order to predict the number of traffic accidents, the three models will be compared using the mean square error and using the model that has the lowest average error in prediction, as follows:

Table (6) Comparison between prediction models

The model	MSE
ARIMA(1,1,0) Model	0.03342
Holt Exponential Model	0.03268
Neural Network Model	0.03098

As we note from Table (6) that the best model for prediction is the neural network model because it has the lowest mean square error, so traffic accidents were predicted according to the months, as the actual data were as follows:

Table (7) Number of traffic accidents in Basra Governorate for the years 2020-2024 according to the months

The Month	The years				
	2020	2021	2022	2023	2024
1	7	8	36	52	20
2	7	8	38	75	17
3	4	8	48	57	14
4	3	11	34	63	17
5	4	9	52	73	16
6	11	10	30	51	18
7	2	6	48	53	20
8	5	7	76	59	19
9	3	8	79	58	21
10	7	12	69	45	
11	7	10	58	22	
12	4	12	51	15	

It was converted to the logarithm as follows:

Table (8) Number of traffic accidents in Basra Governorate for the years 2020-2024 according to the months in logarithm

The Month	The years				
	2020	2021	2022	2023	2024
1	0.8451	0.9031	1.5563	1.716	1.301
2	0.8451	0.9031	1.5798	1.8751	1.2305
3	0.6021	0.9031	1.6812	1.7559	1.1461
4	0.4771	1.0414	1.5315	1.7993	1.2305
5	0.6021	0.9542	1.716	1.8633	1.2041
6	1.0414	1	1.4771	1.7076	1.2553
7	0.301	0.7782	1.6812	1.7243	1.301
8	0.699	0.8451	1.8808	1.7709	1.2788
9	0.4771	0.9031	1.8976	1.7634	1.3222
10	0.8451	1.0792	1.8389	1.6532	
11	0.8451	1	1.7634	1.3424	
12	0.6021	1.0792	1.7076	1.1761	

The prediction was made using the neural network model and the prediction results were as follows: Table (9) Predictive values for the months for the coming years according to the months

The prediction was made using the neural network model and the

The year	The month	Logarithm predictive value	Integer predictive value
2024	10	1.28745600	19
	11	1.29778500	20
	12	1.31018100	20
2025	1	1.32523000	21
	2	1.34367900	22
	3	1.36641000	23
	4	1.39431200	25
	5	1.42793800	27
	6	1.46681400	29
	7	1.50861400	32
	8	1.54946400	35
	9	1.58577000	39
	10	1.61558300	41
	11	1.63861600	44
	12	1.65562300	45
2026	1	1.66777600	47
	2	1.67626000	47
	3	1.68208900	48
	4	1.68604900	49
	5	1.68871900	49
	6	1.69051100	49
	7	1.69170900	49
	8	1.69250900	49
	9	1.69304100	49
	10	1.69339600	49
	11	1.69363100	49

	12	1.69378800	**49**

We notice from the table an increase in traffic accidents in the coming years. The following figure shows this:

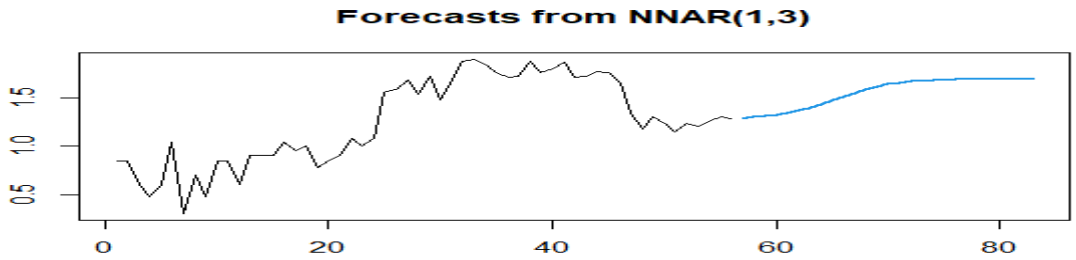


Figure (8) Predictive values of traffic accidents

Conclusions :

1- The number of accidents is increasing during the following years.
2- The time series of traffic accidents is not stable in the mean and variance.
3- The series of traffic accidents stabilized at the first difference after taking the logarithmic transformation.
4- The best model for predicting the number of traffic accidents is the neural network model, then the Holt exponential model, then the Box-Jenkins model.
5- The number of traffic accidents is increasing for the years 2025-2026

Recommendations:

1- We recommend that the administration responsible for traffic management use the results of the model that was reached to control accidents and know their direction in the future.
2- We recommend paying attention to accurately recording all details of the number of traffic accidents, such as the age of the driver and the type of vehicle, etc.
3- Improving the roads of the province and maintaining the damaged ones and establishing modern roads in order to reduce the number of accidents in the future.
4- Using hybrid models to predict the number of traffic accidents and comparing them with the research models above.

Sources:

1- Asaad, Shaza, and Hamdan. Nora, 2021, "Predicting traffic accidents in the city of Lattakia using artificial neural networks", Tishreen University Journal for Scientific Research and Studies - Engineering Sciences Series, Volume (43), Issue (1).
2- Bari, Adnan, Majed Abdul Rahman, 2002, "Statistical prediction methods - Part One" King Saud University, available at http://www.abarry.ws/books.pdf.
3- Hassan, Zuhair Abdul Wahab, 2009, Study of motorcycle accidents in the city of Karbala, Karbala University Scientific Journal, Volume (7), Issue (3).
4- Al-Sarhan, Hussein Ali Hashem, 2018, "Using some statistical methods to predict lost electrical energy - an applied study". Master's thesis in statistical sciences, College of Administration and Economics, University of Karbala.
5- Shoman, Abdul Latif Hassan, Al-Sarraf, Nizar Mustafa, 2013, "Time Series and Index Numbers", Dar Al-Doctor for Administrative and Economic Sciences, First Edition, Baghdad.
6- Dahd, Sabiha Nima, 2015, "A Survey Study on the Phenomenon of Traffic Accidents in Dhi Qar Governorate - Causes and Solutions", Journal of the College of Basic Education for Educational and Human Sciences / University of Babylon, Issue (20).
7- Abdul Ahad, Manahil Daniel, and Younis, Nadwa Salem, (2012), "Predicting the Sales Quantity of the Medical Product Using the Triple Exponential Smoothing Method", Journal of Education and Science, Volume (25), Issue (4).
8- Imran, Kholoud Musa, Zalan, Risan Abdul Imam, 2012, "Using Some Statistical Methods to Predict Electricity Consumption in the Kingdom of Saudi Arabia", Journal of Economic Sciences. Volume (8), Issue (29).
9- Vandal, Walner. 1992, "Time Series from the Application Interface and Box-Jenkins Models", Mars Publishing House.
10- Mohamed, Monem Aziz, et al., 2016, "Predicting Car Accident Deaths Using Generalized Autoregressive Conditional (GARCH) and Neural Networks (ANN) Techniques", The Fifth International Scientific Conference of Arab Statisticians (Cairo - Arab Republic of Egypt).
11- Al-Wardi, Adnan Hashem, 1990, "Statistical Forecasting Methods - Methods and Applications", Dar Al-Hikma Press. First Edition. Iraq.