NIPES

Advances in Engineering Design Technology

Journal homepage: www.nipesjournals.org.ng

Performance of Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) in Modeling and Prediction of Accident Data

Iyeke, S.D; Ilaboya, I.R and Ehiagbonare, M. O

Department of Civil Engineering, Faculty of Engineering, University of Benin, P.M.B 1154, Benin City, Edo State, Nigeria

Email of Corresponding Author: rudolph.ilaboya@uniben.edu

ARTICLE INFORMATION

Article history:
Received 21 December 2021
Revised 28 February 2022
Accepted 07 March 2022
Available online 27 March 2022

Keywords:
Machine Learning, Wi-Fi Throughput, Livestock Monitoring, Internet of Things, and Data Security and Privacy

https://doi.org/10.37933/nipes.a/4.1.2022.4

https://nipesjournals.org.ng

ABSTRACT

The alarming rate of road traffic accident in the country (Nigeria) is among the most worrisome problems currently facing the nation. Sadly, Nigeria has earned the unenviable distinction of consistently leading all the nations of the world in high road traffic accident and high fatality rate. This study conducts a comprehensive evaluation of selected expert systems such as multiple linear regression and artificial neural network for the modelling and prediction of road accident. The study area is Ugbowo-Lagos Road in Benin City, Edo State Nigeria. A reconnaissance survey was done first to ascertain the geometric characteristic of the road which include; the chainage, the vertical and horizontal curve and the super elevation. Thereafter, primary data which include road accident data was collected from Federal Road Safety Office in Benin City. To investigate the qualities of the primary data, basic preliminary analysis techniques, namely; outlier detection, homogeneity test, test of normality and autocorrelation test were conducted while modelling and prediction of road accident was done with the aid of multiple linear regression and artificial neural network. From the geometric characteristic of the road under study, it was observed that for a chainage of 11.5 to 13.0km, the vertical curve was 12.4% while the super elevation was 4.3%. Calculated Cronbach alpha value of 0.900 as observed in the reliability test revealed that the data are reliable and the computed goodness of fit statistics of reliability gave a maximum Guttman coefficient of 88.10% which further confirm the reliability of the data used. With a computed p-value greater than 0.05 for all the independent variables, the null hypothesis of the Dixon test was accepted and it was concluded that the accident data obtained from FRSC is devoid of outliers. In addition, with a centered VIF < 10, it was concluded that there is the absence of multicollinearity between the dependent and independent variables. With a computed coefficient of determination (R^2) value of 0.9265, artificial neural network (ANN) was acclaimed better. Road accident prediction model compare to multiple linear regression model (MLRM) with a computed R^2 value of 0.0617.

1. Introduction

One of the best ways to understand the occurrence of road accident is to develop accident prediction models which are also standard practices in assessing and improving the safety of our roads [1, 2 and 3]. The main purpose of transportation system is to provide the efficient and safe movement of freight and passenger from one place to another. The economic development is directly and strongly related to the availability of transportation. The soaring number of vehicles on the road had created a major social problem through traffic accidents due to the loss of lives and material [4]. Statistical or crash prediction model have frequently been used in highway safety studies. They can be used to identify major contributing factors or establish relationship between crashes and explanatory variables, such as traffic flows, type of traffic control, and highway geometric variables [5, 6, and 7]. In Nigeria, about 85% of the accounted causes of road accidents are believed to have been constituted by human factors [8]. Many researches carried out in Nigeria revealed that most accidents caused by human factors are the result of driving while drunk, drugs, inexperience or poor driving skills, health problems, psychological problems and temperament are also not left out. These have been shown in different ways by drivers. It is also noted that these human factors are the greatest contribution to the increasing surge of traffic accidents in Nigeria [9]. The attitude towards road traffic accidents includes such behavioral elements of the drivers as: sleeping while driving and tiredness, inadequate preparation for a journey, not been familiar with the highway signs, cutting corners, driving after taking excess alcohol, driving with bad eye sight especially in the night, ignorance of the use of seat belts, the incapability of handling unforeseen circumstances, wrong use of road signs and vehicle signaling, overtaking and incompetent maneuvering [10].

2.0 Methodology

2.1 Description of study area

The study area is Benin City, specifically Benin-Lagos Road. Benin City serves as the principal administrative and socio-economic center for both Oredo Local Government Area and Edo State in Nigeria. Benin City is a humid tropical urban settlement which comprises three Local Government Areas namely Egor, Ikpoba Okha and Oredo. It is located within latitudes $6^{0}20^{'}$N and $6^{0}58^{'}$N and longitudes $5^{0}35^{'}$E and $5^{0}41$’E. It broadly occupies an area of approximately 112.552 sq km. This extensive coverage suggests spatial variability of weather and climatic elements. Benin City lies visibly in the Southern most corner of a dissected margin: a prominent topographical unit which lies north of the Niger Delta, west of the lower Niger Valley, and south of the Western Plains and Ranges [11]. Rainfall, temperature, wind and relative humidity are the most significant climatic elements in Benin City. The rainfall element strongly determines the occurrence of the wet and dry seasons in the study area. As observed during the assessment of the urban troposphere using sensitive rain gauges of the American origin, the rainfall amount, its intensity, duration as well as its distribution throughout the city are determined by the prevailing maritime winds, changing clouds, temperatures and circulating pressures. Two principal air masses prevail in the city. These are the tropical maritime and tropical continental.

The tropical maritime air mass which is essentially humid, warm, moisture-borne, and widely resident in Benin City for almost twelve months, originates from the South Atlantic Zone. It causes rainfall which begins from the late January till its gradual subsidence in mid-November. The arrival of rainfall in the study area brings welcome relief to the urban residents from the prevailing moderately dry and cold wind periods which normally occur between late December and the end of January [11]. Heavy rainfall and the associated floods occur frequently in Benin City and have caused huge economic losses as well as social problems. The base map of the study area is presented in Figure 1.

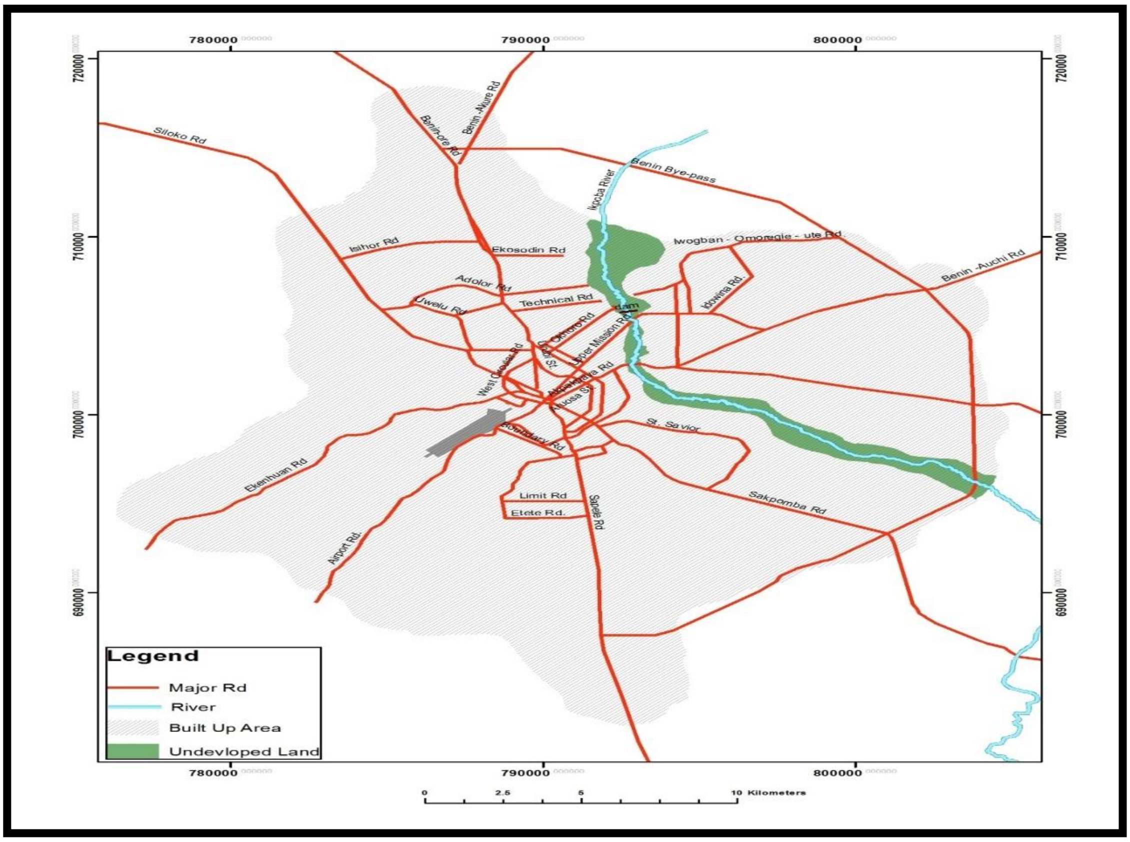

Figure 1: Base map of study area (Adopted from Google Earth)

2.2 Data collection

Two different types of data were employed in this study. The secondary data which include; road accident data was collected from Federal Road Safety Office in Benin City while the primary data which include; traffic volume and the geometric data was collected from the field. For a robust field data, a reconnaissance surveys were carried out at selected points of interest along the study area. For each selected point of interest, detailed information regarding accidents, traffic flow, geometric characteristics, traffic characteristics, road way condition, approach speed, lighting, among others were sourced.

2.3 Preliminary Analysis of Data

2.3.1 Reliability Analysis of the Data

Reliability analysis was done to ascertain the fitness of the data for the selected analysis. The null and alternate hypothesis of reliability was formulated as follows;

H0: Data are reliable

H1: Data are not reliable

Using the Fisher's probability test (F-test), the analysis was conducted at p-value of 0.05. At p-value < 0.05, the null hypothesis was accepted and was concluded that the data are good and can be employed for further analysis.

2.3.2 Assessment of Normality

The Jarque-Bera test for normality is employed to ascertain whether the data follow a normal distribution. Mathematically, the Jarque-Bera test is define as follows"

$$JB = n[(\sqrt{b_1})^2/6 + (b_2 - 3)^2/24] \quad (1)$$

Where:

n is the sample size, $\sqrt{b}$ is the sample skewness and b_2 is the kurtosis coefficient. The null hypothesis for the Jarque-Bera test is that the data are normally distributed while the alternate hypothesis is that the data does not come from a normal distribution. In which case;

H0 = Data follows a normally distributed

H1 = Data do not follow a normal distribution

In general, a large JB value indicates that the residuals are not normally distributed. A value of JB greater than 10 means that the null hypothesis has been rejected at the 5% significance level. In other words, the data do not come from a normal distribution. JB value of between (0-10) indicates that data is normally distributed. To implement the Jarque-Bera test for normality, EVIEWS statistical software was employed.

2.3.3 Rate of Accident Prediction

To predict the rate of accident based on available data, the following prediction models were employed;

i. Multiple linear regression model
ii. Artificial neural network mode

To apply multiple linear regression models, the dependent and independent variables that influences accident rate were first selected and to ascertain the dependence of the selected independent variables on the dependent variable, a regression equation based on equation 2 was developed

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ---------- \beta_5 x_5 + \varepsilon \qquad (2)$$

Where;

X_1, X_2 ---------- X_n are the selected independent variables, Y is dependent variable (Rate of accident), β_0, β_1 are the regression constant and ε is the deviation.

To execute the multiple linear regression modelling and generate the regression equation, statistical software (EViews 9.0) was employed.

2.3.4 Accident Rate Prediction using Artificial Neural Network

To apply neural network, 60% of the data was employed to train a network, 25% of the data was used to validate the network while the remaining 15% was used to test the performance of the network. The neural network modelling and prediction was done with the aid of a neural network modelling tool (MATLAB 10.1). The basic steps involved in the application of the network are as follows:

i. Normalization of the data
ii. Selection of optimum training algorithm or learning rule
iii. Selection of optimum number of hidden neurons
iv. Training of the network
v. Network validation
vi. Network testing and prediction

To avoid the problem of weight variation which can subsequently affect the efficiency of the training process, the input and output data was first normalized to obtain a value of between 0.1 and 1.0 using the normalization equation proposed by [12, 13].

$$x_i = \frac{x - x_{\min}}{x_{\max} - x_{\min}} + 0.1 \qquad (3)$$

Where; x_i is the normalized value of the input and output data, x_{min} and x_{max} are the minimum and maximum value of the input and output data.

Input and output data training resulting in the design of network architecture is of paramount importance in the application of neural network to data modelling and prediction. To obtain the optimal network architecture that possesses the most accurate understanding of the input and output data, two major factors were considered.

i. First is the selection of the most accurate training algorithm and secondly,
ii. The number of hidden neurons.

Based on this consideration, different training algorithm and hidden neurons were tested to determine the best training algorithm and accurate number of hidden neurons that will produce the most accurate network architecture. Selectivity is based on the coefficient of determination (R^2) and the mean square error value (MSE) [13].

2.3.5 Comparism of ANN, MLR .

To compare the performance of artficial neural network (ANN) and multiple linear regression (MLR), the following steps were employed;

i. Prediction of accident rate using selected input variable combinations was done using ANN, MLR.
ii. A regression plot of output between the observed accident rate and the predicted accident rate was generated using ANN, MLR.
iii. Coefficient of determination (r^2) was calculated for ANN predicted values of accident rate, MLR predicted values of accident rate.
iv. The rule of higher the better was employed to select the best model for predicting rate of accident on our highway.

3.0 Discussion of Results

Result of the geometry features of the road under study is presented in Table 1. The components of the result which include; chainage (km), vertical curve (%), horizontal curve (m) and super elevation were selected in reference to [2, 3 and 4].

Table 1: Geometric Characteristics of Ugbowo Benin-Ore Road

Chainage KM	Vertical curve %	Horizontal curve (M)	Super elevation %
11.5-13.0	12.4	2440.56	4.3
13.7-14.6	0	425.67	5.1
24.7-39.3	8.7	0	0
59.5-62.3	2.6	3642.47	1.5
72.8-73.8	0	2088.15	2.6
74.0-76.6	0	3290.26	1.2
78.0-81.0	1.3	5726.30	2.1
84.0-85.0	0	1022.96	0.4
86.0-87.0	0	5087.89	1.3
90.0-90.5	0	904.28	0.5

The geometry features of Ugbowo Benin-Ore road such as design speed, road width, and median Shoulders were 100km/hr, 10.5m, and 1.5m respectively. The length of the road is 94km and the AADT is 1850. In order to attain the primary goal of road transportation, road designers and their disciplines need to use different emerging technologies and techniques. Analysis of road geometric design consistency has been used widely to improve the safety of the roads. Geometric design consistency can be demarcated as how a driver expectation and the road performance match up (i.e. when the road with good constituency level matches a driver expectation, the road user is not amazed while driving along it). Design constituency corresponds to reliving the design speed with actual driving behaviour, which is expressed by the 85th percentile speed of passenger cars under free-flow conditions are prerequisites to accident free highways. Descriptive statistics of the data employed for the analysis is presented in Table 2.

Table 2: Descriptive statistics of accident data

Variable	Observations	Obs. with missing data	Obs. without missing data	Min.	Max.	Mean	Std. deviation
NAC	60	0	60	8.000	36.000	19.550	6.863
NPIV	60	0	60	55.000	365.000	175.917	74.651
NPIJ	60	0	60	23.000	180.000	67.150	31.338
NPK	60	0	60	2.000	33.000	13.083	7.158
NVI	60	0	60	11.000	76.000	37.050	14.611

NAC: Number of accident cases, NPIV: Number of persons involved, NPIJ: Number of persons injured, NPK: Number of persons killed and NVI: Number of vehicles involved.

Based on the results of Table 2, it was observed that the minimum case of accident from 2014 to 2018 was 8 while the maximum was 36 cases of accident. On the number of vehicles involved, the minimum number was 11 while the maximum number of vehicles was 76. On the number of persons involved, the minimum number was 55 while the maximum was 365. On the number of persons injured, the minimum was 23 while the maximum was 180. On the number of persons killed, the minimum was 2 while the maximum was 33.

The correlation matrix of regression which shows how the individual variables relates to the others is presented in Table 3

Table 3 Correlation Matrix

Variables	NAC	NPIV	NPIJ	NPK	NVI
NAC	**1**	0.782	0.777	0.429	0.920
NPIV	0.782	**1**	0.741	0.376	0.862
NPIJ	0.777	0.741	**1**	0.394	0.773
NPK	0.429	0.376	0.394	**1**	0.367
NVI	0.920	0.862	0.773	0.367	**1**

Result of Table 3 revealed that the individual variables are strongly positively correlated with one another. For example, with a correlation coefficient of 0.777 it was concluded that the number of persons involved in an accident (NPIV) is strongly correlated with the number of persons injured (NPIJ). With a correlation coefficient of 0.429, the number of persons involved in an accident is poorly correlated with the number of persons killed (NPK). With a correlation coefficient of 0.920, it was concluded that the number of persons involved in an accident (NPIV) is most strongly correlated with the number of vehicles involved (NVI). For reliability analysis, it is important that analysis of variance is significant at the 5% confident limit. The computed analysis of variance is presented in Table 4.

Table 4: Analysis of variance

Source	DF	Sum of squares	Mean squares	F	Pr > F
Between subjects	59	176265.850	2987.557	3.081	**< 0.0001**
Within subjects	240	1298026.400	5408.443		
Between measures	4	1069161.733	267290.433	275.624	**< 0.0001**
Residual	236	228864.667	969.766		
Total	299	1474292.250	4930.743		

Computed against model Y=Mean(Y)

With probability p-value <0.0001 as observed in Table 4, it was concluded that the model is significant, hence the Cronbach alpha value for assessing reliability was calculated and presented in Table 5

Table 5: Cronbach's alpha statistics

Cronbach's alpha	Standardized Cronbach's Alpha
0.675	0.900

For reliability, the Cronbach alpha value must be greater than 0.65. For standardized Cronbach alpha values of 0.900 as observed in Table 5, it was concluded that the accident data are reliable. Finally, the goodness of fit statistic of reliability were calculated and presented in Table 6.

Table 6: Goodness of fit statistics of reliability

Variable	<Scale/deleted item> Mean	<Scale/deleted item> Variance	<Scale/deleted item> Correlation	<Scale/deleted item> R^2	<Scale/deleted item> Cronbach's alpha	<Scale/deleted item> Guttman L6
NAC	293.200	13528.468	0.853	0.866	0.661	0.873
NPIV	136.833	2816.311	0.827	0.764	0.721	0.855
NPIJ	245.600	9271.702	0.776	0.753	0.787	0.851
NPK	299.667	14189.277	0.609	0.814	0.693	0.881
NVI	275.700	11874.214	0.895	0.901	0.786	0.857

Results of Table 6 further confirmed that the accident data are reliable with Guttman L6 coefficient of 0.851 to 0.881, coefficient of determination of 0.753 to 0.901, Cronbach alpha value of 0.661 to 0.787 and correlation coefficient of 0.609 to 0.895.

For data analysis using the method of linear regression, it is expected that the observed data be statistically normally distributed. To test the normality assumption of the accident data, Jarque-Bera approach was employed and result is presented in Figures 2, 3, 4, 5 and 6 respectively.

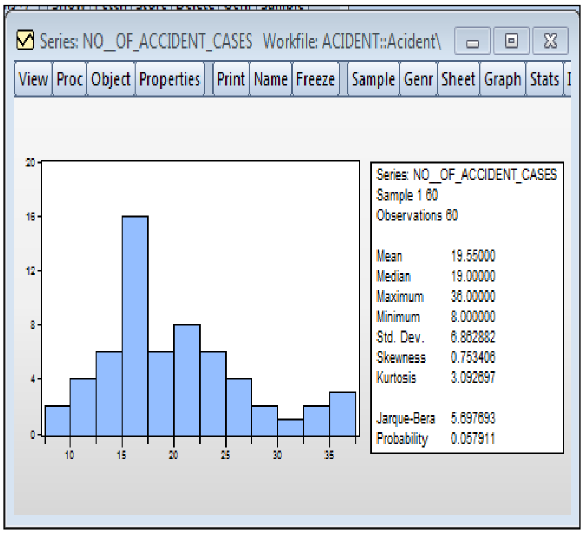

Figure 2: J-B normality test for NAC

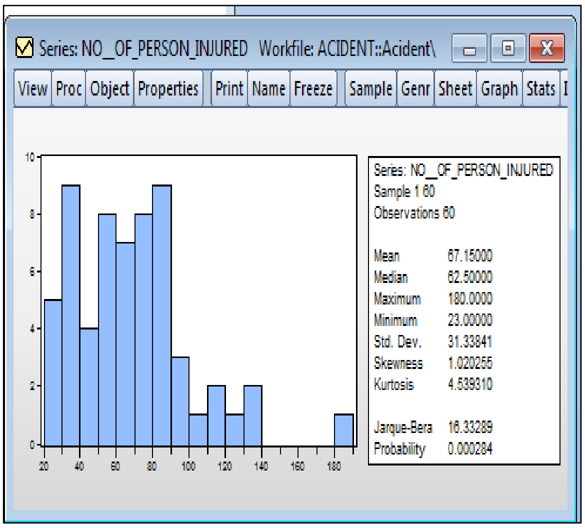

Figure 3: J-B normality test for NPIJ

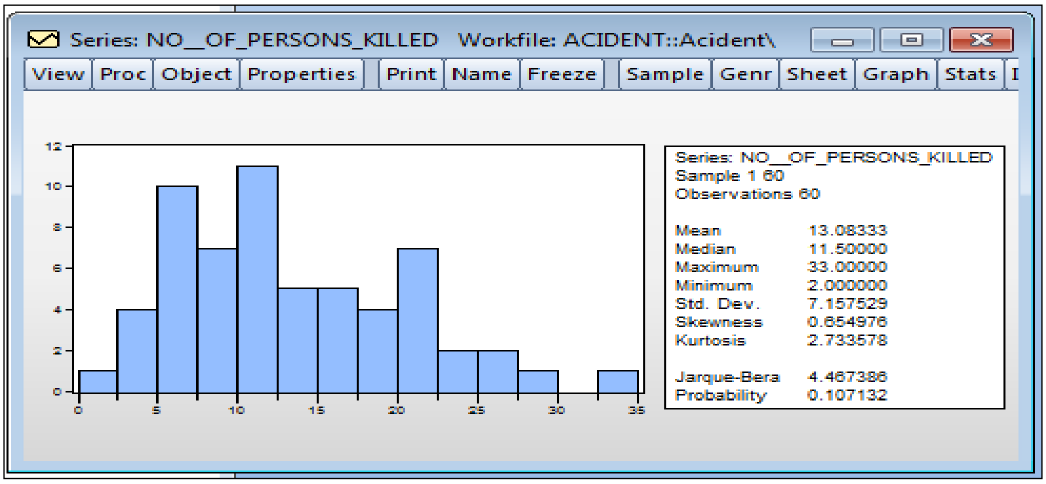

Figure 4 J-B normality test for NPIV

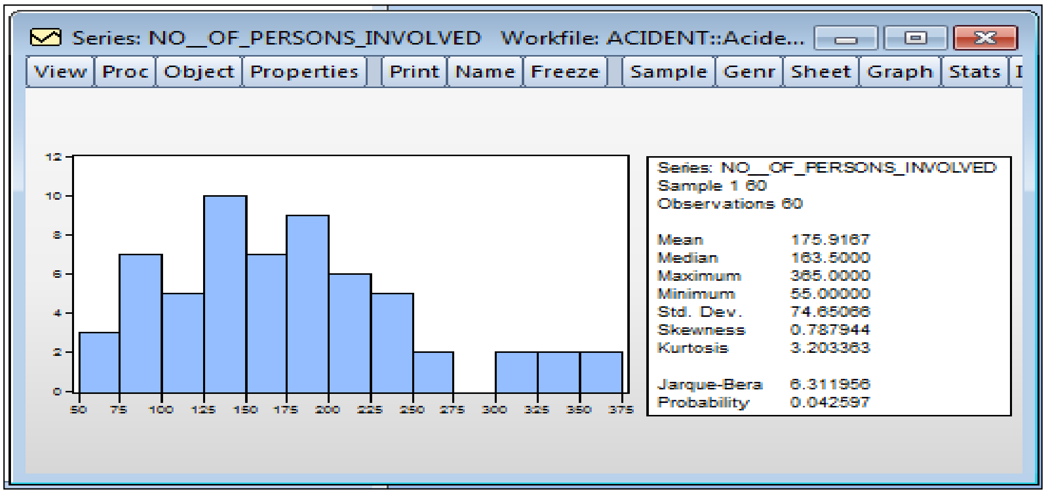

Figure 5: J-B normality test for NPK

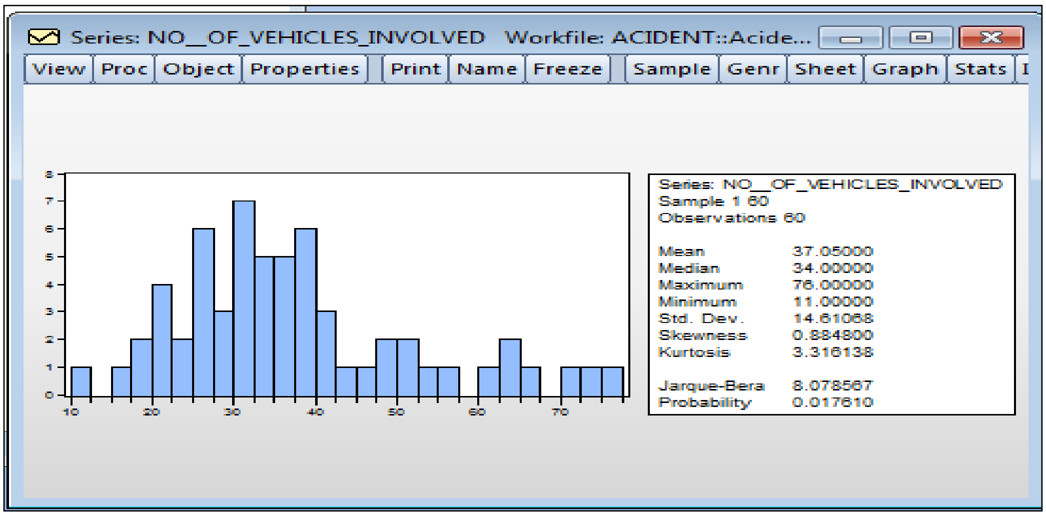

Figure 6: J-B normality test for NVIV

In Figure 2, the computed Jarque-Bera test statistics was 5.697693 with a probability p-value of 0.057911. Although, the Jarque-Bera tests value is less than 10, the computed p-value was greater than 0.05. Hence, it was concluded that the data on number of accident cases is not normally distributed. For number of persons injured as observed in Figure 3, the calculated Jarque-Bera value was observed to be 16.33299 with a probability p-value of 0.002284. Although, the Jarque-Bera value is greater than 10, the calculated p-value is less than 0.05. Hence, it was concluded that the data on number of persons injured follows a normal distribution. For number of persons involved as observed in Figure 4, the calculated Jarque-Bera value was observed to be 6.311956 with a probability p-value of 0.042597. Since the Jarque-Bera value is less than 10, and the calculated p-value is less than 0.05, it was concluded that the data on number of persons involved follows a normal distribution. For number of persons killed as observed in Figure 5, the calculated Jarque-Bera value was observed to be 4.467386 with a probability p-value of 0.107132. Although, the Jarque-Bera value is less than 10, the calculated p-value is greater than 0.05. Hence, it was concluded that the data on number of persons killed did not obey normality. For number of vehicles involved as observed in Figure 6, the calculated Jarque-Bera value was observed to be 8.078567 with a probability p-value of 0.017610. Since the Jarque-Bera value is less than 10, and the calculated p-value is less than 0.05, it was concluded that the data on number of vehicles involved follows a normal distribution.

The dependence of the dependent variable on the selected independent variables was evaluated using the coded least square regression equation presented as follows;

NAC C NVI NPI NPIJ NPK (4)

Where;
NAC is the number of accident cases;
C is the constant of regression;
NVI is the number of vehicles in involved;
NPIJ is the number of persons injured; and
NPK is the number of persons killed.

The coded regression equation was implemented using statistical software and results obtained is presented in Table 7

Table 7: Output of Regression Analysis

Equation: EQ1 Workfile: ACCIDENT DATA SPSS::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: NO__OF_ACCIDENT_CASES
Method: Least Squares
Date: 07/06/21 Time: 10:59
Sample: 1 60
Included observations: 60

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.895824	0.979829	2.955437	0.0046
NO__OF_VEHICLES_INVOLVED	0.403358	0.049764	8.105442	0.0000
NO__OF_PERSON_INJURED	0.034617	0.017725	1.952974	0.0559
NO__OF_PERSONS_INVOLVED	-0.010164	0.009223	-1.101991	0.2753
NO__OF_PERSONS_KILLED	0.089668	0.051892	1.727978	0.0896

R-squared	0.866475	Mean dependent var	19.55000
Adjusted R-squared	0.856764	S.D. dependent var	6.862882
S.E. of regression	2.597364	Akaike info criterion	4.826527
Sum squared resid	371.0466	Schwarz criterion	5.001056
Log likelihood	-139.7958	Hannan-Quinn criter.	4.894795
F-statistic	89.22679	Durbin-Watson stat	2.032629
Prob(F-statistic)	0.000000		

From the result of Table 7, the following observations were made

i. With a regression (p-value) of 0.0046, it was concluded that the regression analysis was significant at 0.05 degree of freedom

ii. Independent variables, such as number of vehicles involved was observed to have a very strong influence on the dependent variable compared to other variables

iii. The strong regression statistics such as coefficient of determination (R^2 = 0.866) and adjusted coefficient of determination (Adj. R^2 = 0.857) supports the application of linear regression as a model for accident data analysis and prediction.

Using the result of Table 7, the overall linear regression equation was thereafter generated and presented as follows;

$$NAC = 2.8958 + 0.4034(NVI) + 0.0346(NPIJ) - 0.0102(NPI) + 0.0897(NPK) \quad (5)$$

Where;
NAC is the number of accident cases;
NVI is the number of vehicles in involved;
NPIJ is the number of persons injured; and
NPK is the number of persons killed.

The classification of the data into input and target variables for ANN modelling is presented in Table 8.

Table 8: Classification of data for ANN modelling

Input variables for ANN modeling	Number of accident cases (NAC)
	Number of persons involved (NPI)
	Number of persons injured (NPIJ)
	Number of vehicles involved
Target variable for ANN modeling	No of persons killed

Input and output data training resulting in the design of network architecture is of paramount importance in the application of neural network to data modelling and prediction. To obtain the optimal network architecture that possesses the most accurate understanding of the input and output data, two factors were considered. First was the selection of the most accurate training algorithm and secondly, the number of hidden neurons. Based on this consideration, different training algorithm and hidden neurons were selected and tested to determine the best training algorithm and accurate number of hidden neurons that will produce the most accurate network architecture. Selectivity was based on coefficient of determination (r^2) and mean square error (MSE). Table 9 shows the performance of the different training algorithm tested.

Table 9: Selection of optimum training algorithm for ANN

S/No	Training Algorithm (Learning Rule)	Training MSE	Cross Validation MSE	R-Square (r^2)
1	Hopfield	0.005672	0.00278	0.75
2	Generalized Regression	0.007677	0.00249	0.88
3	Gradient and rate of change of gradient (Quick prop)	0.003843	0.002711	0.78
4	Adaptive step sizes for gradient plus momentum (Delta Bar Delta)	0.004487	0.00534	0.87
5	Second order method for gradient (Conjugate gradient)	0.06322	0.00507	0.81
6	Improved second order method for gradient (Levenberg Marquardt)	0.0000333*	0.0000451*	0.95*

Result of Table 9 revealed that improved second order method of gradient also known as Levenberg Marquardt Back Propagation training algorithm was the best learning rule and was therefore adopted in designing the network architecture. To determine the exact numbers of hidden neuron, different numbers of hidden neurons were selected to create a trained network using Levenberg Marquardt Back Propagation training algorithm. Performance of the trained network was assessed using mean square error (MSE) and coefficient of determination r^2. The number of hidden neurons corresponding to the lowest MSE and the highest r^2 as presented in Table 10 was selected to design the network architecture.

Table 10 Selection of optimum number of hidden neurons for ANN

S/No	Number of Hidden Neurons	Training MSE	Cross Validation MSE	R-Square (r^2)
1	2	0.00442	0.00788	0.81
2	4	0.00411	0.00912	0.76
3	6	0.00700	0.00133	0.82
4	8	0.00355	0.00966	0.80
5	10	0.000103*	0.0000224*	0.92

Based on the results of Table 9 and 10, Levenberg Marquardt Back Propagation training algorithm having 10 hidden neurons in the input layer and output layer was used to train a network of 4 input processing elements (PEs) and 1 output processing elements. The input layer of the network uses the hyparbolic targent (tan-sigmoid) transfer function to calculate the layer output from the network input while the output layer uses the linear (purelin) transfer function. The number of hidden neuron was set at 10 neurons per layer and the network performance was monitored using the mean square error of regression (MSEREG).
A performance evaluation plot which shows the progress of training, validation and testing is presented in Figure 7.

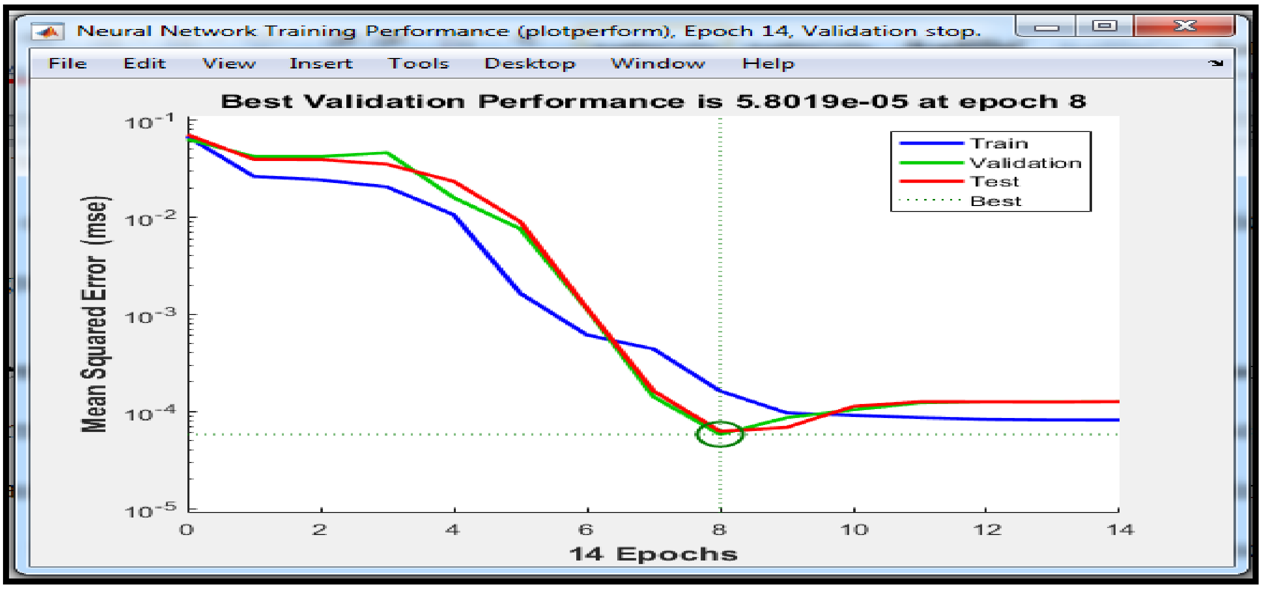

Figure 7: Performance curve of trained network for predicting number of persons killed

From the performance plot of Figure 7, no evidence of over fitting was observed. In addition, similar trend was observed in the behaviour of the training, validation and testing curve which is expected since the raw data were normalized before use. Lower mean square error is a fundamental criterion used to determine the training accuracy of a network. An error value of 5.8019e-05 at epoch 8 is an evidence of a network with strong capacity to predict number of persons killed. The training state, which shows the gradient function, the training gain (Mu) and the validation check, is presented in Figure 8.

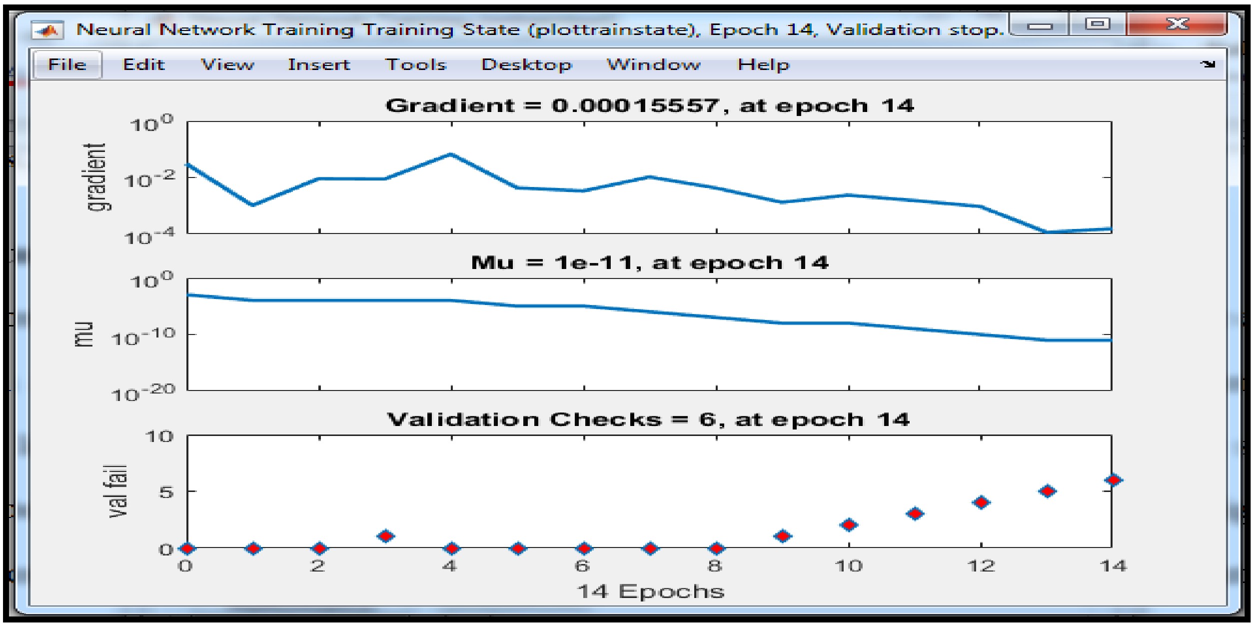

Figure 8: Neural network training state for predicting number of persons killed

Back propagation is a method used in artificial neural networks to calculate the error contribution of each neuron after a batch of data training. Technically, the neural network calculates the gradient of the loss function to explain the error contributions of each of the selected neurons. Lower error is better. Computed gradient value of 0.00015557 as observed in Figure 3.2 indicates that the error contribution of each selected neuron is very minimal. Momentum gain (Mu) is the control parameter for the algorithm used to train the neural network. It is the training gains and its value must be less than one. Momentum gain of 1.0e-11 shows a network with high capacity to predict the number of persons killed.

The regression plot which shows the correlation between the input variables; number of accident cases (NAC), number of persons involved (NPI), number of vehicles involved (NVI) and number of persons injured (NPIJ) and the target variable number of persons killed (NPK) coupled with the progress of training, validation and testing is presented in Figure 9.

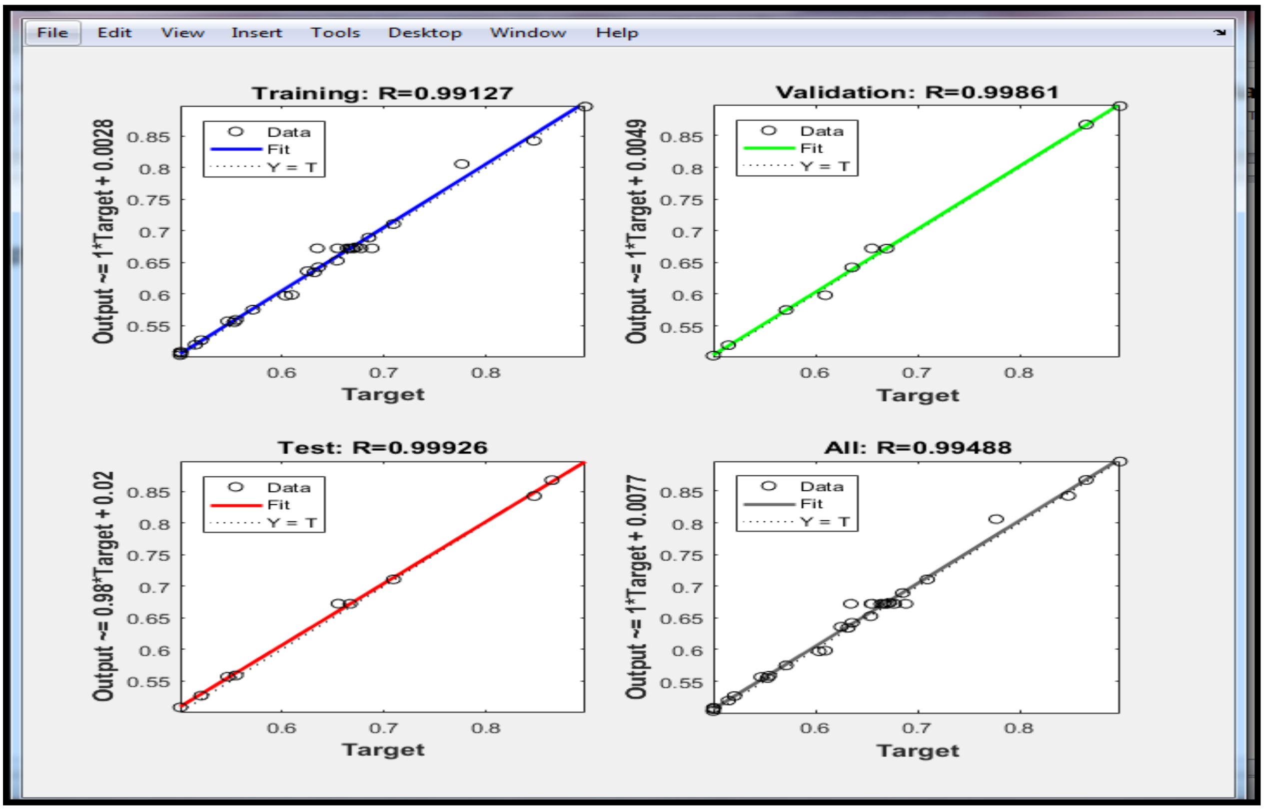

Figure 9: Regression plot showing the progress of training, validation and testing

Based on the computed values of the correlation coefficient (R) as observed in Figure 9, it was concluded thet the network has been accurately trained and can be employed to analyzed accident data and subcequently predict the number of persons killed. To test the reliability of the trained network, the network was thereafter employed to predict its own values of number of persons killed using the same sets of input parameters viz; number of accident cases (NAC), number of persons involved (NPI), number of vehicles involved (NVI) and number of persons injured (NPIJ). Based on the observed and the ANN predicted values, a regression plot of outputs was thereafter generated as presented in Figure 10.

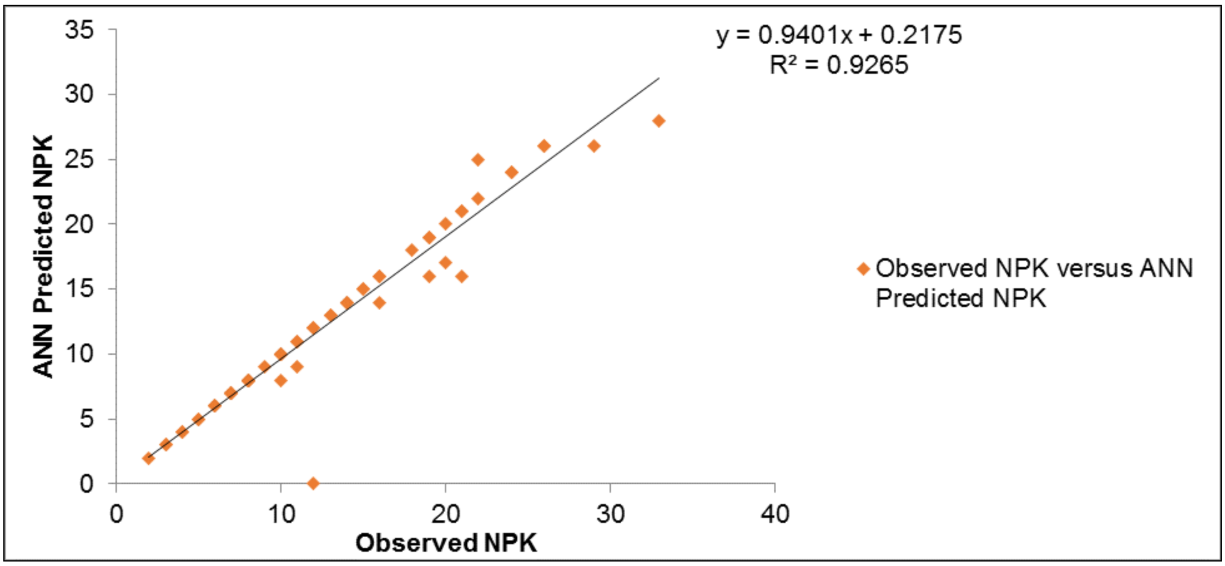

Figure 10: Regression plot of observed versus ANN predicted NPK

Based on the observed and the linear regression predicted values, a regression plot of outputs was also generated as presented in Figure 11.

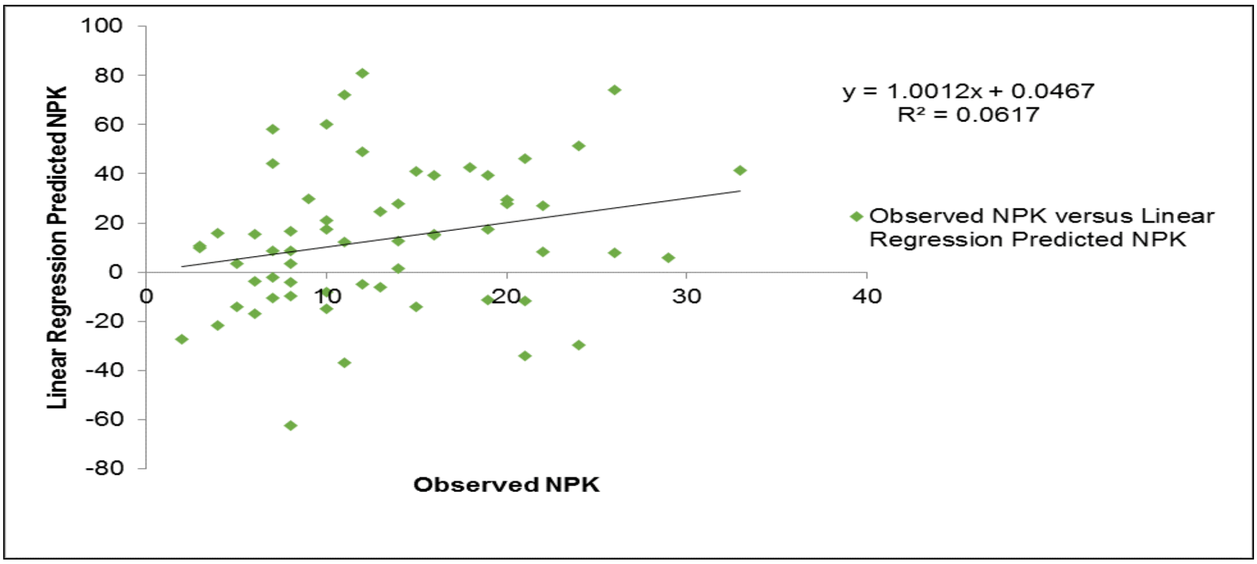

Figure 11: Regression plot of observed versus LRM predicted NPK

Coefficient of determination (r^2) values of 0.9265 as observed in Figures 11 was employed to draw a conclusion that ANN is a better model compared to linear regression for the analysis and prediction of accident data.

4.0 Conclusion

In this study, a comprehensive analysis of accident data was done using linear regression and artificial neural network. Five years' monthly accident data were employed for the study and the statistical properties of the data was assessed by means of selected preliminary statistical techniques such as reliability test, test of normality and descriptive statistics. Based on the outcome of the analysis, it was concluded that the data are not only devoid of outliers, they are also reliable and normally distributed. Coefficient of determination (r^2) values of 0.9265 as observed in Figures 4.10

was employed to draw a conclusion that ANN is a better model compared to linear regression for the analysis and prediction of accident data.

References

[1] Ozgan, E. and Demirci, R. (2008); Neural Networks-based Modeling of Traffic Accidents in Interurban Rural Highways, Duzce Sampling; Journal of Applied Science, vol. 8, pp. 146-151

[2] Oyedepo, O. J and Makinde, O. O, (2010), Accident Prediction Models for Akure – Ondo Carriageway, Ondo State Southwest Nigeria; Using Multiple Linear Regressions, An International Multi-Disciplinary Journal, Ethiopia, Vol. 4, No. 2, Pp. 30-49

[3] Popoola, O. M. Abiola, O. S. Odunfa, S. O and Ismaila, S. O (2017); Comparison of Road Traffic Accident Prediction Models for Two-Lane Highway Integrating Traffic and Pavement Condition Parameters, Journal of Natural Science, Engineering and Technology, Vol.16 No.2, Pp. 1-10.

[4] Helai, H. and Mohamed, A. A. (2010); Multilevel data and Bayesian analysis in traffic safety. Accident Analysis and Prevention, Retrieved on October 18, 2010, from www.elsevier.com/locate/aap, vol. 42, pp: 1556–1565

[5] Gwynn, D.W. (1967); Relationship between Road Accident and Hourly Volumes, Traffic Quarterly, Vol. 14(1), pp: 407-418

[6] Ghani, A.R.A., Daniel, B.D., Yusof, I., Zamhari, K.A., Abdullah, M.E., Omar, N. and Termida, N.A. (2008); Traffic Engineering and Safety Module. 2nd Ed. Johor: Penerbit UTHM, pp" 67-78

[7] Fajaruddin, M., Ismail, Y., Ismail, R., Aziz, A. S., and Esnizah, M. S. (2008); Black spot Study and Accident Prediction Model Using Multiple Liner Regression. Proceedings of the First International Conference on Construction in Developing Countries; Karachi, Pakistan: Universiti Tun Hussein Onn Malaysia, vol. 1, pp: 345-366

[8] Ohakwe, J., Iwueze, I. S., Chikezie, D. C. (2011); Analysis of Road Traffic Accidents in Nigeria: A Case study of Obinze/Nekede/Iheagwa Road in Imo State, southeastern, Nigeria. Asian Journal of applied Sciences, 4(2): pp. 166-175

[9] Odumosu, A. O. (2005) Demeanour Transposition as Strategy for Traffic Accident Reduction in Nigeria: Case Study of Niger State, Nigeria, Proc. 13th International Conference on Road Safety on Four Continents, Warsaw, Poland, p. 18

[10] Osime, O. C., Ehikhamenor, E., Oludiran, O. O., Iribhogbe, P. E., Ighedosa, S. E. and Elusoji, S. O. (2006) Road Traffic Accident Trends in Nigeria. International Journal of Injury Control and Safety Promotion, 13(4), pp. 251-253

[11] Okhakhu, P.A. (2010). The Significance of Climatic Elements in Planning the Urban Environment of Benin City, Nigeria, Unpublished PhD Thesis, Ekpoma-Nigeria: Department of Geography and Regional Planning, Ambrose Alli University.

[12] Sinan, M.T; Beytullah, E. and Asude, A (2011), Prediction of adsorption efficiency for the removal of Ni (II) ions by zeolite using artificial neural network (ANN) approach, Fresenius Environmental Bulletin, vol. 20, No. 12, pp, 3158-3165

[13] Rezaie, F., Moghaddam, S., Afandizadeh, H. and Ziyadi, M (2011), Prediction of accident severity using artificial neural networks, International Journal of Civil Engineering, Vol. 9, No. 1, pp:23.35

[14] Kin, C; Luk, J; Ball, E. and Sharma, A. (2001): An Application of Artificial Neural Networks for Rainfall Forecasting; Meteorological and Atmospheric Physics Journal, vol. 33, pp: 883-699.