Characteristics and Prediction of Traffic Accident Casualties In Sudan Using Statistical Modeling and Artificial Neural Networks

Galal A. Ali[1] and Awadalla Tayfour[2]
[1]Professor & Consultant Engineer, School of Civil Engineering, College of Engineering, Sudan University of Science & Technology, PO Box 12881, Khartoum 11111, Sudan
ga03ali@yahoo.com (Corresponding Author)
[2]Department of Electrical Engineering, College of Engineering, Sudan University of Science & Technology, P.O. Box 72, Eastern Daim Khartoum, Sudan 1:

ABSTRACT
Road traffic accident (RTA) casualties in Sudan are among the major causes of death in the age group of 21 to 60.with 61% fatalities. The fatality rate of 35 per 10,000 vehicles is among the highest in the world despite the low car ownership of 1 vehicle to 100 persons. This paper presents accident characteristics and considers road safety management. Crucial issues discussed in the paper include prediction and safety measures. The paper applies Artificial Neural Network (ANN) and regression techniques to comparatively predict traffic accident casualties. Both approaches modeled accident casualties using historical data on population, number of registered cars and other related factors from 1991 to 2009. Comparison of predictions with recorded data was very favorable. Predictions during 2010 – 2014 were determined using projected values for the same predictor variables. ANN forecasts provided the best fit for the data with a maximum difference of 1.84% between predictions and observed data. The study demonstrated that ANNs provide a powerful tool for analysis and prediction of accident casualties. The major causes of accidents were attributed to driver behaviour, vehicle fleet and conditions, road network defects, speed-limit violation, negligence of seat-belt usage and lack of traffic-law enforcement.

1. INTRODUCTION
There is great concern worldwide about the increasimg traffic accidents and casualties (fatalities and injuries) particularly in developing countries. Annually, over three-quarter million people are killed while injured and disabled victims in road traffic accidents

(RTAs) exceed 40 million [1]. Developing countries alone represent 70% of RTA fatalities although they own only 12% of the vehicle fleet. Fatality rates per 10^4 vehicles in some African and Asian countries range between 15 and 65 [2]. In contrast, accident records in many industrial countries indicate that there was decrease in recent annual fatalities. In USA, the fatality rate per 100 million vehicle-miles-traveled (MVMT) decreased by 10 % [3]. However, due to lack of MVMT data in developimg countries, fatality rate per 10^4 vehicles closely estimates the equivalent best measure of fatality rate according to Ali et al. [4]. The importance of continually updating and improving accident and casualty records as well as methods of analyzing such data cannot be overemphasized. Better analysis and evaluation of traffic accident and casualty data will assist policy makers device improved traffic regulations and safety measures to enhance road safety.

An overview of the situation on road traffic accidents in some Arab countries and the Middle East generally attributed the main causes of accidents to speeding, driver negligence and violation of traffic regulations [5,6]. In many cases, the majority of these accidents occur in urban areas, while in some countries about 40% of the casualties involve pedestrians. In Kuwait, a reduction of up to 15% in total fatalities was observed after installation of traffic cameras [7].

2. TRAFFIC ACCIDENT AND CASUALTY CHARACTERISTICS IN SUDAN

In Sudan, casualties are about 10 times more than in many developed countries in spite of the current low car ownership of one vehicle per 100 population. More than 60% of the casualties are in the age group of 21-60 years as shown in Table 1, the major cause of death for about 50% of this age group being RTAs. The high rate of population growth, the large percetage of young drivers, the dramatic and uncontrolled increase in the number of vehicles over recent years compounded with the absence of strict law enforcement and the poor road conditions have all contributed to the high accident rates. In general, drivers have been the main cause in 90 % of the accidents primarily in terms of negligence, high speed and poor driving as illustrated in Table 2, computed from the data provided by Ali [1]. Poor driving reflects carelessly changing lanes or making turns without signaling, their move, etc. Negligence is intended for being indifferent and not concerned with safe driving such as eating, talking to a rider or on phone or playing with radio, etc while driving. This value of 90% is less than the global record of 95 – 98 %. It is also asserted that vehicle age, mechanical defects, tyre failures do have impact on safety. Poor road and vehicle conditions contribute to the remaining 10% (or 2 – 5 % in the latter case) which includes mechanical defects, tyre failure and weather conditions. The corresponding value in Oman, for example, averaged only 3.5 % due to good road and vehicle conditions [8]. However, it cannot be confirmed that analysis of the causes of traffic accidents have been undertaken to investigate the roles of some of these factors on traffic crashes. Such in depth and detailed studies of the causes of accidents could help develop policies to reverse the existing trends and grave situation. Nevertheless, there are indications and various studies in the Arab world and Africa that the main contributors to RTAs are high speeds and the pedestrians.

Table 3 summarizes the historical data for Sudan, while Table 4 compares the various 2005 accident and casualty rates with those of 2009 according to Directorate General of Traffic [9]. Table 4 indicates that most of the parameter rates of accidents and casualties (fatalities and injuries) have increased. The only rates that dropped were fatality and injury rates per 10^3 accidents. This was largely because the increase in fatality from 2005 to 2009 was low compared to that of the accidents (Table 1). This may be attributed to more accidents occurring in urban areas with less severity combined with more improvements of roads. Key ingredients for successful traffic improvement programs and prerequisites for traffic safety management are the availability of sufficient and reliable data and thereon the capability to predict traffic accident casualties and safety situations. Accident and casualty rates per 10^4 vehicles were computed as in Table 5. Improvement and effective safety management programs can then be developed for implementation and assessment. A crucial requirement, therefore, for the development of successful traffic improvement programs is the availability of reliable predictive models that incorporate the essential factors related to accidents and casualties.

Table 1. causalities in Sudan by age group - 2009

Age Group	Casualties	Casualties,%
≤10	79	8
11 to 20	116	12
21 to 30	199	21
31 to 40	156	16
41 to 50	116	12
51 to 60	109	11
61 to 70	76	8
71 to 80	56	6
≥81	47	5
Total	954	100

Source: The Directorate General of Traffic. annual 2009 statistical book [9]

Table 2. driver-attributed causes of accidents in Sudan – 2008

Driver-related cause of accidents	Percentage
Negligence	6
Speeding	48
Poor driving	34
Alcohol	2
Vehicles, road and environment	10

Source: Computed from Ali (2010)

Table 3. Sudan historical data required for model training and determination of casualty rates

Year	Predictor variables				Dependent variables			
	Pop*103 (SDG)*	GDP*103 (Vehicles)	NOCR (km)	NKPR	Accidents	Injuries	Fatalities	Casualties
1991	24366	18.0	124720	950	4385	2912	395	3307
1992	25040	18.7	151053	1010	5455	4444	413	4857
1993	25734	20.4	172933	1057.2	6080	3543	404	3947
1994	26445	22.1	163053	1126.8	6241	3641	395	4036
1995	27177	22.4	123389	1196.4	6510	4532	580	5112
1996	27489	26.6	164929	1243.3	9661	5245	589	5834
1997	28701	28.5	175722	1298.2	7187	5452	608	6060
1998	29495	29.1	182587	1339.5	7893	5725	715	6440
1999	30281	30.4	190000	1687.9	10730	7113	864	7977
2000	31088	32.6	198000	2036.3	13707	8115	905	9020
2001	31913	35.7	220000	2384.7	15827	8820	952	9772
2002	32592	49.3	259852	2733.1	16038	6207	490	6697
2003	33393	52.9	325897	3081.5	18481	6222	784	7006
2004	34203	71.0	330804	3429.9	21958	5858	651	6509
2005	35009	76.2	364107	3778.3	27712	7198	751	7949
2006	35814	85.9	275350	4126.7	34029	8998	810	9808
2007	36620	97.5	368954	4475.1	37402	10193	870	11063
2008	38242	107.8	395720	4562.3	39176	9307	810	10117
2009	39800	111.0	397318	4740.4	44668	9411	899	10206

*US$1 = 3 Sudanese pounds (SDG)

Source: Directorate General of Traffic. annual statistical books 1991-2009, Khartoum, Sudan - 2010

Table 4. accident and casualty rates in Sudan – 2005 and 2009

Accident/Casualty Rates	2005	2009
Accidents per 10^4 vehicles	761.09	1124.24
Accidents per 10^5 population	79.16	112.23
Accidents per day	75.92	122.38
Fatalities per 10^3 accidents	27.10	20.13
Fatalities per 10^5 population	2.15	2.26
Fatalities per month	62.58	74.92
Injuries per 10^3 accidents	259.74	210.69
Injuries per 10^5 population	20.56	23.65
Injuries per day	19.72	25.78

Source: Table 1

Table 5. accident, casualty, injury and fatality rates per 104 vehicles

Year	Accidents per 10^4 vehicles	Casualties per 10^4 vehicles	Injuries per 10^4 vehicles	Fatalities per 10^4 vehicles
1991	351.59	265.15	233.48	31.67
1992	361.13	321.54	294.20	27.34
1993	351.58	228.24	204.88	23.36
1994	382.76	247.53	223.30	24.23
1995	527.60	414.30	367.29	47.01
1996	585.77	353.73	318.02	35.71
1997	409.00	344.86	310.26	34.60
1998	432.29	352.71	313.55	39.16
1999	564.74	419.84	374.37	45.47
2000	692.27	455.56	409.85	45.71
2001	719.41	444.18	400.91	43.27
2002	617.20	257.72	238.87	18.86
2003	567.08	214.98	190.92	24.06
2004	663.78	196.76	177.08	19.68
2005	761.09	218.31	197.69	20.63
2006	931.41	268.45	246.28	22.17
2007	1013.73	299.85	276.27	23.58
2008	989.99	255.66	235.19	20.47
2009	1124.24	256.87	236.86	22.63

Source: Computed using data of Table 3, Directorate General of Traffic.
annual statistical books 1991-2009 [9]

Figure 1 depicits the unique characteristics of the distinct annualy increasing accidents while the casualties are almost without change. These trends indicate that effective safety measures are not taken. In contrast, comparisons made in Figures 2 and 3, respectively, for fatality and injury rates with three other countries show that in the latter cases there are clear reductions in casualty rates. It is noticed that these three countries are an induatrial country (Newzealand), a rich Gulf developimg country in Asia (Oman) and a developing country in Africa (Malawi).

Prediction of accidents and casualties today relies on various models and relationships, based largely on expert systems and statistical techniques. They predict accidents and casualties either as a function of registered vehicles per population as reported by Al-Suleiman and Al-Masaeid [10] or traffic flow and road type proposed by Jadaan and Nicholson [11]. Ali et al.[4] used time while Pattnaik and Sreedar [12] considered several exogenous variables. Causes of traffic accidents are numerous, interelated and complex, and safety may be related to several factors such road, vehicle and traffic characteristics, user behaviour and traffic law enforcement [13]. Sometimes the developed models may not fit the data as claimed by Wong-Toi [14], or only address linear relationships between the dependent and independent variables. More comprehensive approaches are required to account for the important variables and their relationships in analyzing and predicting traffic accident casualties. The main objective of this investigation was to predict casualties for Sudan up to the year 2014 applying ANNs and regression models.

3. ARTIFICIAL NEURAL NETWORKS

ANNs are computer models designed for knowledge processing and prediction. The ability of ANNs to spontaneously learn from examples, to reason over historical fuzzy data, and to provide adequate responses to new information not previously stored in memory has gained growing interest in this technology [15,16]. As a result of numerous applications in various engineering fields, this new technique has attracted increasing acceptance and demonstrated remarkable success.

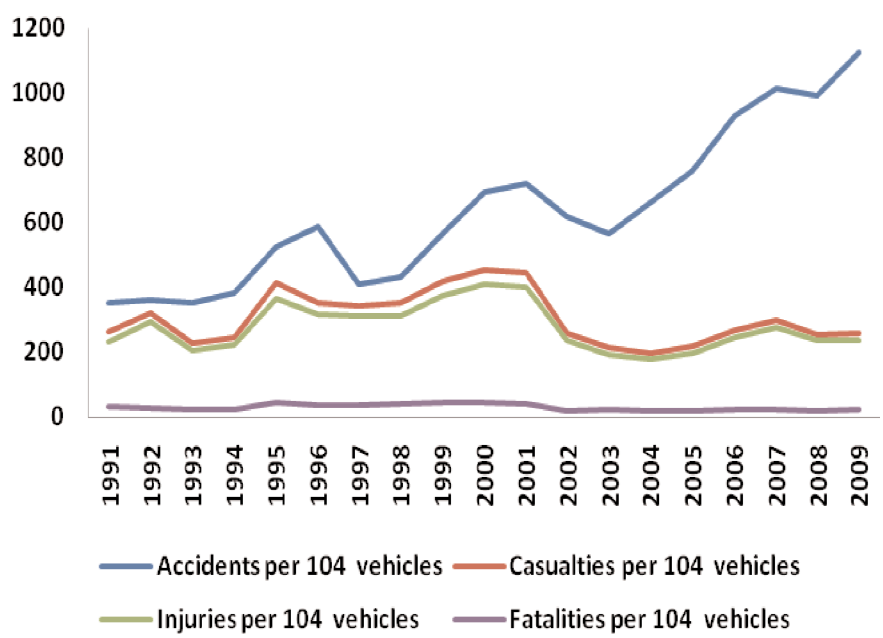


Figure 1. Accidents and casualties in Sudan per vehicle rate

Source: Tables 3 and 5

ANNs are relatively new in the field of traffic engineering and accident analysis. The new approach has been sparsely demonstrated in traffic congestion forecasting by Taber et al [17] and a few other applications. Al-Alawi et al. [8] applied computer-based techniques and artificial neural networks to the analysis and prediction of road traffic accidents. Artificial Neural Networks have also been used in problems that were traditionally solved by statistical methods. A number of researchers have conducted

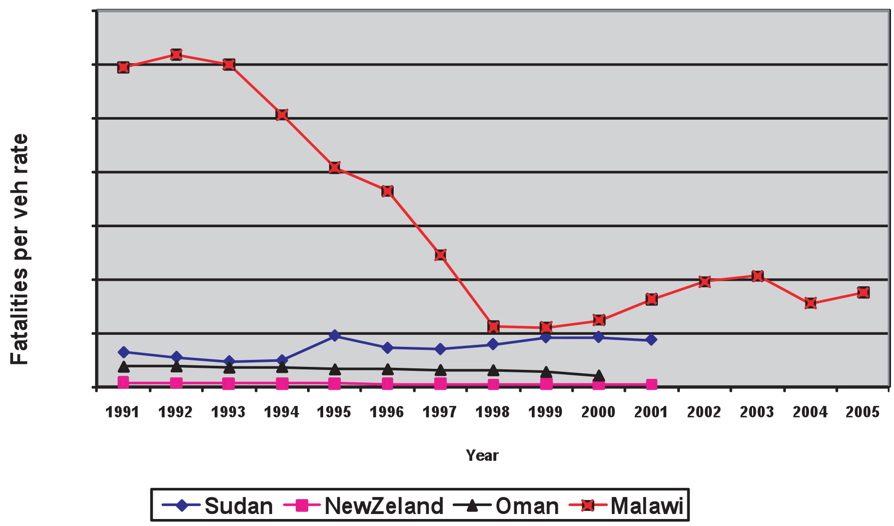


Figure 2. Fatality rates per 104 vehicles – four countries

Source: Tables 3 and 5

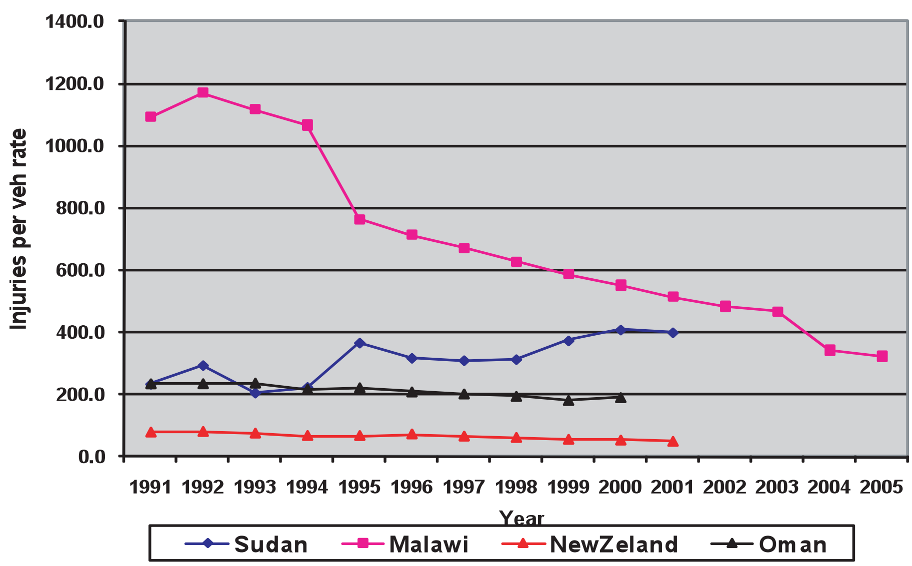


Figure 3. Injuries per vehicle rate - four countries

Source: Tables 3 and 5

comparative studies of statistical methods with ANNs [18,19]. Their studies have shown that, if trained on medium to large data sets, ANNs can be quite useful in prediction. Assumptions are not required regarding the functional form of the relationship between predictor and response variables as the case is with the statistical methods.

3.1 Data Preparation and Rehabilitation for Model Building

The main factors considered to contribute to annual accident casualty figures included annual population growth (POPG), the gross domestic product (GDP) of Sudan, the number of registered cars (NOCR) and the total length of paved roads (NKPR). The number of driver licenses issued (NOLI) was originaly considered to be included but the recorded data were poorly documented. However, elemination of this factor from the model was considered not affect the output as it is related to POPG and NOCR with possible confounding. Other factors could have been incorporated, such as road and vehicle conditions and driver behaviour. However, some of these are difficult to quantify while others have incomplete historical data over the period of study. Data on traffic and accident casualty statistics were readily obtained from Sudan Directorate General of Traffic (DGT) [9]. Data on population and GDP, and the length of paved roads were obtained, respectively, from the Ministry of Finance and Economic Planning and the Ministry of Roads and Bridges of the Government of Sudan.

After carefully scrutinizing the values of the candidate variables, correlation analysis was performed to assess the linear association between the variables. Results of the correlation analysis are shown in Table 6. The strength of the correlation coefficients between the predictor variables and the casualties are as given in the last column of the Table. The results show that all have high positive correlation to casualties.

3.2 Development of the ANN Model

The ANN model shown in Figure 3 was developed for modeling and prediction of the number of casualties in Sudan. The variables selected for developing the ANN model were as mentioned above. All four variables were assumed to be functions of the year (Y), and were chosen as input parameters to the proposed ANN architecture. The dependent variable, the number of car accident casualties (NOAC), was chosen as the model output. Since the relationship of the above variables to the number of casualties may not be linear, the following nonlinear model was proposed:

Table 6. correlation matrix

Variable	POPG	GDP	NOCR	NKPR	No. of casualties, NOAC
POPG	1.0	0.94	0.94	0.57	0.91
GDP		1.0	0.94	0.98	0.88
NOCR			1.0	0.96	0.90
NKPR				1.0	0.96
No. of Casualties, NOAC					1.0

$$NOAC_Y = f\,(POPG_Y, GDP_Y, NOCR_Y, NKPR_Y) \quad (1)$$

where Y = 1,2,3,......n; n being the number of years which the ANN model was to be trained, and $NOAC_Y$ the predicted number of accident casualties for year Y.

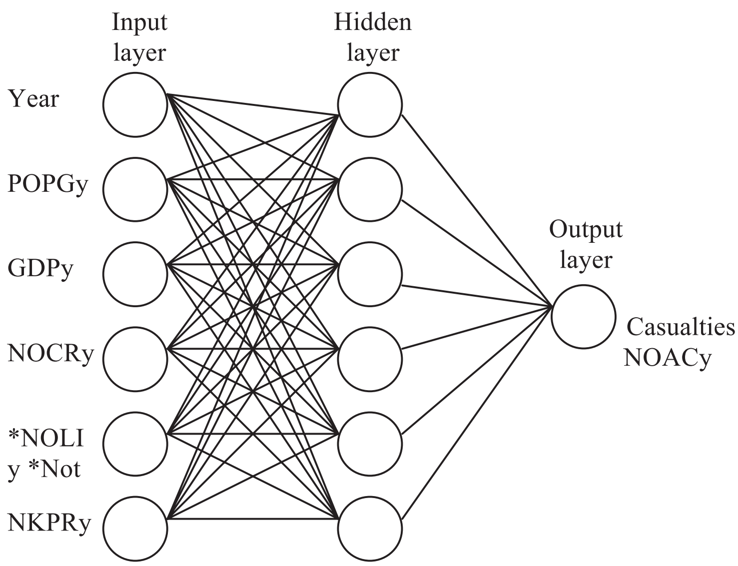
As illustrated by the ANN architecture in Figure 3, each network comprises many simple processing elements that are organized into a sequence of layers. These are the input, the hidden, and the output layers. The neurons in the input layer receive 5 input signals representing the above input variables. Hence, five neurons were used for the input layer in the ANN architecture. The output layer consists of one output neuron representing the casualties. Between the input and output layers, Hecht-Nielsen [20] showed that one or two hidden layers with an adequate number of neurons is sufficient to model any solution surface of practical interest. One hidden layer containing five neurons was found adequate to develop the model.

Historical data covering the period from 1991 to 2009 were used in training the proposed model. The training data set are shown in columns 2 – 5 of Table 4. These data were presented repeatedly to the ANN model, and weights were adjusted by small amounts that were dictated by the General Delta Rule. The training process of the ANN models was performed using the NeuroShell® simulator. A description of how the learning process is performed by the Backpropogation (BP) algorithm and the detailed training process, respectively, may be found elsewhere [21]. As shown in Figure 4 for the plot of casualties against year, the R^2 of the trained model was 0.9335. This indicates that 93.35 % of the variability in the number of casualties could be explauined by the selectedc independent variables and the historical data used. Having trained the network successfully, the next step was to test the network in order to assess its performance and to examine its capabilities. Once the model was developed, the contributions of the different independent variables to the variation in the values of the dependent variable (casualties) were obtained using the NeuroShell® utility. Examination of these contributions in Table 7 revealed that the change in population ($POPG_Y$). and the number of cars ($NOCR_Y$) had substantial influence on the increasing number of car accident casualties (43% and 32% respectively). The increase in $NKPR_Y$ accounted for 18%, while the contribution of GDP_Y.to the variation in the number of accident casualties was marginal (7%). These contributions were computed by the NeuroShell® program as measures of each input strength in relation to those of the other input parameters of the model being developed.

Table 7. the contribution of the input parameters to the output results

Parameter	POPGY*103	GDPY	NORCY	NKPRY
Contribution, %	43	7	32	18

In order to use the developed model to predict the number of car accident casualties for the years to come, it was necessary first, to forecast the model input variables. A regression analysis, with year as the independent variable, was applied to project these variables by curve-fitting the historical data using eqns (3 - 5). Needless to mention, these projections are one of many plausible scenarios that could evolve in the future. With the many developments in road construction and upgrading, in addition to the continuous improvement in traffic management in Sudan, it would be difficult to make more accurate projections of these variables.



* Not included (see text)

Figure 4. The architecture, input and output of the proposed ANN model

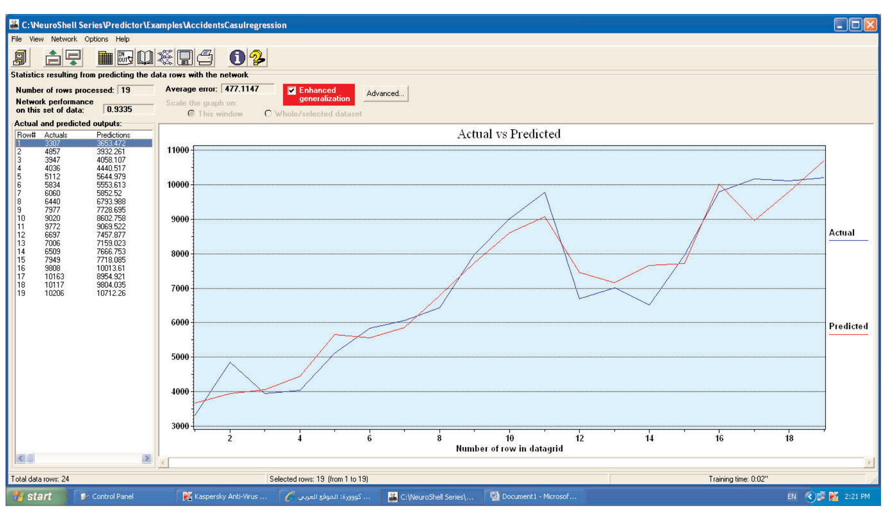


Figure 5. Results of training the ANN model (1991 – 2009)

		R²	
$POPG_Y$	$= 11.294y^2 + 589.66y + 23867$	0.9967	(2)
GDP_Y	$= 0.3882y^2 - 2.4352\ y + 23.152$	0.9862	(3)
$NOCR_Y$	$= 68.7y^3 + 2716.7\ y^2 - 14311\ y + 161751$	0.9186	(4)
$NKPR_Y$	$= - 10.98y^2 + 17.156y + 835.7$	0.9826	(5)

Using these equations, forecast data for the years 2009 to 2014 were generated for each variable. The projected values of the predictor variables were then fed as input to the ANN model in order to predict the number of car accident casualties for each year until the year 2014. The actual vlaues of the casualties and the predicted ones are presented in Figure 4 as a plot of casualties against the year. As demonstrated by the figure and the associated table for the observed and predicted values, the comparison was found very favoudable. Thae comparison between the observed and predicted casualties is also summarized in Table 8. Also shown in the table, the difference between them ranged between - 0.74 % and a maximum of 1.84 % reflecting very close agreement.

Table 8. Model predictions of the number of casualties for the years 2010-2014

Year	2010	2011	2012	2013	2014
Actual observed	10, 432	10, 789	11, 283	11, 912	12, 676
ANN forecast	10, 353	10, 805	11, 403	11, 946	12, 909
% difference	- 0.74	+ 0.15	+ 1.06	+ 0.29	+ 1.84

4. SUMMARY AND CONCLUSIONS

Within the scope of this investigation and accounting for data accuracy and completeness, the following conclusions and recommendations may be drawn from this research study:

In this paper, the authors have attempted to investigate and compare the predictive capabilities of ANN with with observed values for annual car accident casualties in Sudan where traffic accidents are among the major causes of death in spite of its low motorization level. Thus, there is need for such investigations to contribute to the understanding of the underlying features of the problem and to the development of better methods of analysis and assessment of new safety measures. The accident characteristics of Sudan indicated that effective safety measures need to be introduced to reduce the accident and casualty rates by time. Drivers' negligence, poor driving and speeding are the main causes of accidents and the resulting casualties, the road users being responsible for about 90 %.

The models used the number of car accident casualties as the dependent variable and the annual population of the country, GDP, the number of registered cars and the length of paved roads as the independent variables. The response variable (NOAC) was fitted using ANNs. Preliminary examination of the data indicated that the predictor variables were collinear. However, each of the four variable had its own contribution to the number of casualties although with marginal value of 7 for GDP. It was found that the

Figure 6. ANN casualties development and prediction models (1991 – 2014)

forecasts obtained from the ANN model compared very favorably with observed data with a maximum fdifference of 1.84%.

More safety measures need to be adopted to attain comparable rates. Actions to be taken to bring accidents and casualties down would include speed control, mandatory use of safety belts, training of drivers and rigorous licensing, control of driver behavior and pedestrian crossing, special attention to adherence of buses and trucks to safety regulations, and improved and signed and marked road and intersection design. Furthermore, for road safety enhancement, it is crucial to establish effective National Road Safety Council.

REFERENCES

[1] G. A. Ali, Traffic accidents and road safety management: a comparative analysis and evaluation in industrial, developing and rich-developing countries, *29th Southern African Transport Conference*, Pretoria, South Africa, Proc: www.satc2010.za.co, 2010, 530-540.

[2] G. A. Ali and A. Shigidi, A comparative analysis of traffic accidents and road safety management, *soric' 02: Safety on Roads: International Conference*, Manama, Bahrain, 2002.

[3] NHTSA, 2008 Accidents data summary, US Department of Transportation, *cited in TRB E-Newsletter,* TRB@newsletters.nas.edu, 2010.

[4] G. A. Ali, C. K. Bakheit and N. Sivakugan, Traffic accidents in Oman: characteristics and comparative analysis of fatality rates**.** *World Safety Journal,*. 1994, VIII(1-94), 15-19.

[5] M. Abdel Aty and H. Abdel-Wahab, Analysis and prediction of traffic fatalities resulting from angle collisions including the effect of vehicles' configuration and compatibility. *Accidents Analysis and Prevention, 2003.*

[6] G. A. Ali, S. Al-Alawi and C. K. Bakheit, A Comparative nalysis of raffic accident prediction using statistical methods and artificial neural network, *SQU Scientific Journal of Science and Technology*, 1998, Vol. 3, 13-22.

[7] A. H. Aljassar, Ali, M. A. and M. S.Al-Anzi, An investigation on the effect of traffic cameras on road safety in Kuwait, *2nd Gulf Road Conf Paper No. E5* Abu Dhabi, UAE, 2004.

[8] S. Al-Alawi, Ali, G. A. and C. S. Bakheit, A novel approach for traffic accident analysis and prediction using Artificial Neural Network, *Journal of Road & Transport Research, Australian Research Board,* 1996, 5(2), 118 – 128.

[9] Directorate General of Traffic, (2010). Annual statistics books 1991-2009. Ministry of Interior, Khartoum, Sudan.

[10] T. I. Al-Suleiman and H. R. Al-Masaeid, Descriptive model for fatality rates of traffic accidents in Jordan. *ITE Journal*, 1992, 62(4), 37-39.

[11] K. S, Jadaan and A. J. Nicholson, Relationships between road accidents and traffic flows in an urban network. *Traffic Eng and Control,* 1992, 33(9), 507-511.

[12] S. B. Pattnaik and M. A. Sreedar, 1993. Accident analysis with an expert system for remedial measures. *IE(I) Journal*, 1993, 74, 122-127.

[13] E. Petredou and M. Moustaki, 2001. Human factors in the causation of road traffic crashes. *European J of Epidemio*, Kluwer Acad Pub, the Netherlands, 2001, 16: 819-826.

[14] D. Wong-Toi, Macroscopic models relating traffic volumes and fatalities - A case study of NZ & Australia, . *Road & Transport Research, ARRB*, 1994, 3(4), 28-40.

[15] S. Lee, T. Lee, H. J. Kim and Y. K. Lee, Development of accident prediction model with intel system theory. Gervasi et al., eds, *ICCSA, LNCS 3481*, 2005, 880-888.

[16] L. Mussone, A. Ferrar and M. Oneta, An analysis of urban collisions using an artificial intelligence model. *Accident Analysis and Prevention*, Vol. 31, 1999, 705-718.

[17] J. T. Taber, W. J. Grenney and T. E. Senti, A knowledge-base and simulation algorithm hybrid model for traffic congestion at intersections, *Developments in Artificial Intelligence for Civil & Struc Eng,* Topping,ed., Civil-Comp Press, Edin, UK, 1995.

[18] B. D. Ripley and N. Hjort, Pattern recognition and neural networks: A statistical approach, *Cambridge University Press,* Cambridge, 1994.

[19] H. S. Stern, Neural networks in applied statistics (with discussion), *Technometrics,* 1996, 38 (3), 205 - 220.

[20] R. Hecht-Nielsen, Theory of back-propagation neural network. *Proc of Int Joint Conf on Neural Networks,* IEEE. Washington, D.C., 1989, Vol. 1, 593-605.

[21] P. K. Simpson, Artificial neural systems. *Foundations Paradigms, Applications, and Implementations*. 1st ed., *Pergamon Press Inc*., Elmsford, NY, 1990.