STUDY ON ROAD ACCIDENT PREDICTION USING MULTIPLE LINEAR REGRESSION, AND ARTIFICIAL NEURAL NETWORK

BY

EHIAGBONARE, MICHAEL OMOS.
PG/ENG1715883

DEPARTMENT OF CIVIL ENGINEERING,
FACULTY OF ENGINEERING,
UNIVERSITY OF BENIN,
BENIN CITY, NIGERIA

MARCH, 2023

STUDY ROAD ACCIDENT PREDICTION USING MULTIPLE LINEAR REGRESSION (MLR), AND ARTIFICIAL NEURAL NETWORK (ANN)

BY:

EHIAGBONARE MICHAEL OMOS.
PG/ENG1715883

A REPORT WRITTEN IN THE DEPARTMENT OF CIVIL ENGINEERING IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER IN HIGHWAY AND TRANSPORTATION ENGINEERING OF THE UNIVERSITY OF BENIN, BENIN CITY, NIGERIA

MARCH, 2023

CERTIFICATION

This is to certify that **EHIAGBONARE, Michael Omos** carried out his work in the Department of Civil Engineering, University of Benin, Benin City, Edo State, Nigeria; in partial fulfillment of the requirement for the award of Master's degree in **HIGHWAY AND TRANSPORTATION ENGINEERING.**

____________________ ____________________

ENGR. DR S.D. IYEKE **DATE**
(SUPERVISOR)

____________________ ____________________

ENGR. DR I.R ILABOYA **DATE**
(CO-SUPERVISOR)

____________________ ____________________

ENGR. DR S.D. IYEKE **DATE**
(HEAD OF DEPARTMENT)

DEDICATION

This research work is dedicated to God Almighty

ACKNOWLEDGEMENT

I thank Almighty God for his grace and favor upon my life, and giving me the strength and knowledge for is project.

My profound gratitude goes to my supervisors; Engr. Dr. S.D Iyeke, and co supervisor Engr. Dr. I.R Ilaboya for guiding me through this project.

My sincere appreciation goes to the head of department Engr. Dr. S.D Iyeke, Prof. O.C Izinyon, Prof. O.U Orie, Dr. O. Okiemute, Dr. I.R Ilaboya and other members of staff of the department of Civil engineering.

Also I want to express my profound gratitude to my parents Sir Prof. J.E. EHIAGBONARE and Lady R.E. EHIAGBONARE for their unending support towards my education and to my siblings and friends for their love and encouragement, I am indeed very grateful to you all.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

ANN	-	Artificial neural Network
APMs	-	Accident prediction Models
CV	-	Co-efficient of Variability
FIS	-	Fuzzy Inference System
FL	-	Fuzzy Logic
FRSC	-	Federal Road safety Commissions
JB	-	Jarque-Bera
MSE	-	Mean square error
MLR	-	Multiple Linear Regression
NPIV	-	Number of person injured
NPK	-	Number of person killed
NVI	-	Number of Vehicles involved
SEK	-	Standard error of Kurtosis
UN	-	United Nation
SHSP	-	Strategic highway safety plans

i.

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. 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. The aim of this study is to conduct 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. 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, secondary data which include road accident data was collected from Federal Road Safety Office at lucky way Benin City. To investigate the qualities of the secondary data, basic preliminary analysis techniques, namely; outlier detection, homogeneity test, test of normality and autocorrelation test were done. 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 13km, 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(Vehicle influence factor) < 10, it was concluded that there is the absence of multicollinearity between the dependent (NAC) and independent variables (NPIV, NPIJ, NPK, NVI). 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. The implication of this findings states that if the R^2 value is lesser than 0.0617 it is will not to work .

CHAPTER ONE
INTRODUCTION

1.1 Background of the study

The main purpose of transportation system is to provide the efficient and safe movement of freight and passenger from one place to another. 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 (Helai and Mohamed, 2010). 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 (Gwynn et al 1967; Ghani *et al.,* 2008; Fajaruddin *et al*., 2008).

In Nigeria, about 85% of the accounted causes of road accidents are believed to have been constituted by human factors (Ohakwe *et al*., 2011). 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. 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 (Odumosu, 2005). 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, wrong overtaking and incompetent maneuvering (Osime et al., 2006).

Towards ensuring an accident free highway remains one of the major priorities of a progressive Government and the whole society at large. Presently, the rate of fatality, injury and loss of properties occasioned by road accident is on the increase, a condition that is not only limited to Nigeria alone, but also to most developing countries world-wide (Oladehinde *et al.*, 2007). 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 (Ozgan and Demirci, 2008). Road accident does not just occur but are caused by a number of factors which includes human, mechanical or adverse road conditions. The alarming rate of road traffic accident in the country (Nigeria) is among the most worrisome problems currently facing the nation (Ogwueleka and Ogwueleka, 2010). 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.

An accident prediction model is a mathematical model which describes the relationship between road accident frequencies and various traffic conditions, road geometric features, environmental factor as well as drivers' behaviour (Celikoglua and Cigizoglub, 2007). Considerable research on accident prediction models has been carried out in recent years, and these models can basically be grouped into two main approaches namely: fuzzy logic and artificial neural network (ANN). The fact that road accidents may not be a linear function of various dependent variable for prediction, models have made large room for the use of non-linear approximators such as fuzzy logic and ANN (Esnizah, 2008; Ghani et al., 2008). For example, Xiao *et al*., (1999) developed two fuzzy logic models for predicting the risk of accident that occurred on wet pavement, and the two models were based on Mamdani inference method and Sugeno inference method respectively. The result showed that the fuzzy logic model had superiority over both probabilistic models and non-linear regression models.

Meng et al (2009) employed fuzzy logic to relate urban road accident frequencies with various traffic and road conditions such as; annual average daily traffic (AADT) and traffic limit (TL). These factors were recognized as the prominent influence factor by the models. Haykin, (1999) employed artificial neural network (ANN) to analyze the freeway accident frequencies, and pointed out that ANN method did not require prior-knowledge of any pre-defined underlying relationship between the dependent (accident rate) and independent variables. The study also demonstrated that ANN is a consistent alternative method for analyzing freeway accident frequency. Abdelwahab and Abdel-Aty (2002), Ozgan and Demirci (2008) used series of artificial neural network to model the potential non-linear relationship between injury severity levels and crash related factors. The neural network models were found to have better predictive power compared to traditional method. It is shocking to note that road transportation has the largest number of daily traffic accident in Nigeria (national bureau of statistic year book). In addition, results by Police and road safety has shown that Edo, Lagos, Uyo, and Port Harcourt route account for an estimated one-third of the total accident fatality figures recorded annually in Nigeria. This calls for improvement on the highway safety in Nigeria.

1.2 Statement of the problem

Every year the lives of approximately 1.35 million people are cut short as a result of a road traffic crash. Between 20 and 50 million more people suffer non-fatal injuries, with many incurring a disability as a result of their injury (WHO report). Road traffic injuries cause considerable economic losses to individuals, their families, and to nations as a whole. These losses arise from the cost of treatment as well as loss of productivity for those killed or disabled by their injuries, and for family members who need to take time off work or school to care for the injured. Road traffic crashes cost most countries 3% of their gross domestic product (Ozgan and Demirci, 2008).

1.3 Aim and Objectives of the study

The aim of the study is to analyze and predict accident rate using the multiple linear regression and artificial neural network.

The specific objectives are to;

ii. Compare the road geometry data with results of similar work on the study area

iii. Develop an accident prediction model using multiple linear regression and artificial neural network (ANN).

iv. Validation of the models was done before comparing them.

v. Compare the performance of multiple linear regression and artificial neural network (ANN) as accident prediction model.

1.4 Scope of Work

The overall scope of the study includes;

Assessment of geometrical features influencing accident rate on our highways.

Collection of road accident data from Federal Road Safety Office in Benin City

Preliminary analysis of road accident data using selected statistical techniques

Modelling of road accident data using supervised learning algorithm

Comparison and selection of best fit model for road accident prediction.

1.5 Justification for the study

The study is relevant for the following reasons

The rate of accident on the highways is becoming very alarming, accident has claimed lives of prominent people and future leaders and destroyed valuable properties. This study is expected to develop accident prediction model that will facilitate the reduction of accident occurrence rate on our highways.

In addition, development of accident prediction models will help practitioners to realistically improve intersection design and ensure that more judicious use is made of the usually limited budgeting allocation to road safety activities.

Finally, the study will also help to monitor and reduce accident rate thereby improving the social and economic wellbeing of the people.

CHAPTER TWO
LITERATURE REVIEW

2.1 Road Accidents

The 19th Century industrial revolution resulted in some fundamental changes in the transport sector and provided more flexibility of movement, speed and timing. Since then, there has been an upsurge in both human and vehicular motor movement, a situation that has also resulted in more fatal road accident (Meng *et al.*, 2006). Different circumstances precipitate fatal vehicle accidents. For so many years in the past, road accident was a major source of concern to the major causes of death and severe injuries. It causes the mortality rate to be on the high side. In several parts of the world, measures have been taken and implemented with all seriousness and today, statistical report shows that accidents occurrence globally is on the decline and the number being killed or injured are dwindling (Lugarde, 2007). This is not the case of Nigeria. Political and socio economic and some other reasons are good indicators to why road accidents has remained a leading cause of death in Nigeria.

A study by Chen (2010) showed that the fatality rate in African Countries ranges from 10 to 100 folds more than that in the United States. Also, Lugarde (2007) reported that Africa has an average rate of 28.3 per 100,000 population road traffic mortality compared with Europe.

Concerns about the rising incidence of fatal road accidents compelled stakeholders, including the United Nations (UN) Assembly, into seeking means to curb road facilities. In 2011, the UN adopted the period 2011- 2020 as the UN Decade of Action for Road Safety, with which all efforts will concentrate on stabilizing and then reducing global road traffic fatalities by 2020. According to the former UN Secretary General, Banki-moon, lives will be saved through this decade of Action (A speech at the event, 2011). Following the declaration by the UN in 2011, the Federal Road Safety Commission (FRSC) in Nigeria set out to adopt and domesticate the UN action plan by developing a number of programs suitable for every road user in the country.

Poor road structure and population growth have greatly led to an increase in accident rate. The establishment of the Federal Road Safety Corps (FRSC) by the government of the Federal Republic of Nigeria in 1988 (Udes Decree 45 of 1988 as amended by Decree 35 of 1992, with effect from 18th February, 1988) was to reduce the rate of accidents on our highways. The commission was given the following responsibilities, policy making, organization and administration of road safety in Nigeria.

Mr Osita Chidoha, the FRSC Corps Marshal and Chief Executive, estimated that Nigeria currently loses three billion naira every year to read crashes. Road crashes cost Nigeria 13% of her gross National products (GNP) which inhibits economic and social development. Nigeria loses about 3% of GDP from Road traffic cases that is about 17% of the current National reserves. Nigeria is ranked second – highest in the rate of road accidents among 193 countries of the world (Aghonkhese *et, al*, 2013). (Delen *et al*., (2006)) had argued that three – quarters of all accidents on Nigerian roads involve fatalities. Aside from the Boko haram crisis, accidents are currently by far the main causes of violent death in Nigeria (Nigeria Watch fourth Report on violence in Nigeria 2006 – 2014). The WHO adjudged Nigeria the most dangerous country in Africa with 33.7 death per 100,00 population every year (WHO 2013 report on Accident in Africa). According to their report, one in every four road accident deaths in Africa occurs in Nigeria. The WHO survey and the FRSC report of 5,693 fatal accidents in 2009 (FRSC 2009) leaves no doubt about the dangerous situation on Nigeria roads.

The measures to prevent accident maybe from speed reduction, widening of the road, speed enforcement. These different types of factors can be identified to prevent accident on the highway for future purposes; it can also help our road agencies like the ministry of road and transport and the road safety of Nigeria in investigation process and planning process. In essence, causes of road accident can be regrouped into three broad categories, which are road users' behavior, road defects and vehicular defects (Anyaoku, 2009).

2.2 Impact of Road Accident

The repercussions of accidents have been colossal. Despite the happiness and change of quality of family lives associated with owing a vehicle. Its possessions have left many families bereaved on their breadwinners or loved ones (Anyaoku, 2009). The socio – economic costs of Road Traffic Accidents (RTA) in Nigeria are immense and the direct cost of traffic causalities can perhaps be understood best in terms of the labour cost to the nation's economy. (Anyaoku, 2009); argued that persons injured in accidents on Nigeria highways and streets no longer participate in the economic mainstream, and this amounts to a loss of labour of millions of persons years to the nation.

2.3 Causes of Road Accidents

The term accident is commonly accepted as an occurrence involving one or more transportation vehicles in a collision that results in property damage, injury or death. The term accident also implies a random event that occurs for no apparent reason other than "It Just Happened" (Lester *et al.,* 2010). Accident can be caused by a number of factors among these factors are;

2.3.1 Driver Factors

The major contributing factor in most accident situations is considered to be the performance of the driver of one or both (in multiple vehicles accident) of the vehicles involved. Driver error can occur in many ways, including inattention to the roadway and surrounding traffic, failure to yield the rights of way and disobedience of traffic rules. These "failures" can occur due to unfamiliarity with roadway conditions, travelling at high speeds, drowsiness, drinking, using a cell phone or dealing with other distractions within the vehicles (Lester *et al*., 2010).

The driver behaviour and attitude is very important in judging the driver's actions. Human factors without doubt are the most complex and difficult to isolate as they are almost all very important in nature. Consider sensory capabilities, knowledge, attitude, alertness, health, driving skills, age, customs, habits, weight, strength and freedom of movement of these emotional factors are the greatest valuable attributes and the most difficult to identify.

2.3.2 Age

Old driver which is in ages of 60 and above, have a bad vision which were not clear and they tend to drive slowly. While younger driver which is in ages of 16 -25 years tend to drive fast and have lack of experience in driving, with lack of skills in handling motor vehicle especially when they are facing an accident (Esnizah, 2008).

2.3.3 Alcohol and Drugs

When alcohol or drugs are involved in the crash, it is more likely to end as a high severity crash in both types of highways as the relevant variables has positive parameters in both of the case. The alcohol involvement has been recorded as whether alcohol presented or alcohol contributed towards the crash based on the judgment made by the police officer. Drivers and motorcyclists with blood alcohol content greater than zero are at higher risk of an accident than those whose blood alcohol content increases from zero. Many types of drugs detected in accident victims are liable to impair driving skills, there is still uncertainty as to whether this translates to an increased accident risk.

2.3.4 Driver's Behaviour (Speed)

The speed of motor is at the core of the road traffic injury problem. Speed influences both crash risk and crash consequence. It is becoming more difficult at shorter time for a driver to stop and avoid an accident when the vehicle is at a higher speed. Accident risk increases as speed increases, especially at road junctions and while overtaking as road users underestimate the speed and overestimate the distance of an approaching vehicle.

2.3.5 Vehicle Factors

A small percentage of accidents are caused by mechanical failure of a vehicle, such as some forms of tyre failure or steering failure (Aworemi *et al.*, 2010). Faulty brakes can cause accident between vehicles or vehicles with other things. Worn out tyres also can cause the vehicle to get involved in an accident.

2.3.6 Road Factors

The condition and quality of the road, which include the pavement, shoulders, intersection and the traffic control system, can be a factor, in accident. The road must be designed to provide adequate sight distance at the design speed else motorist will be unable to take remedial actions to avoid an accident. The road side equipment such as street light, markings or signs and all equipment for road must be provided to ensure safety for the road users. Traffic signals must provide adequate decision sight distance when the signal goes from green to red. The super elevation of highway must carefully lay out the comet radius and the appropriate transition sections to ensure that vehicles can negotiate curves safely (Lester et al., 2010). Irrespective of the crash occurrence area, the variables related with the roadway geometry results in a positive parameter. This implies the fact that when the roadway is not leveled and straight it is more likely to result in a high severity crash. When a crash occurs on an urban or rural interstate or local road, the probability of having a more severe injury is less compared to artenals and collectors (Esnizah, 2008). This may be due to the fact that when people drive in local roads, they might be more careful and also there might be lesser vehicular interactions due to the low traffic volumes on those highways. On interstates, the decreasing trend in having more severe injuries may be due to the high safety attributes available on those highways coupled with the uniform travel speed conditions.

2.3.7 Environmental Factors

The climatic and environmental conditions can also be a factor in road transportation crashes. The most is weather transportations function. Weather on roads can contribute to crashes; for example, wet pavement reduces friction and flowing or standing water can cause the vehicle to hydroplane. Many severe crashes have occurred during conditions of smoke or fog which can greatly reduce visibility (Aworemi *et al.*, 2010). When the crash occurs on a wet road surface which indeed has less skid resistance, it seems to be ended with a lesser severe crash in both urban and rural roadways as the variable related to the road surface condition gives a negative parameter. This may be due to the fact that drivers are more cautious under severe weather conditions and try to maintain lower driving speeds under these conditions. On the other hand, when the crash occurs under dark or unlit conditions in urban areas, the severity of the crash is going to be higher. However, this variable is non – significant in rural areas.

2.4 Hazard of Road Accident

The effect/hazard of road accident to a Nation affects people, property and environment. Road accident conditions are caused by poor road infrastructure, overloading of vehicles, law impunity. Road accidents increase dependency, destruction of properties and loss of lives, injuries and permanent disabilities. In general, societies are faced with a great loss, in terms of human, material and financial resources. As the number of accidents increases, so is the number of destroyed facilities, injuries increasing steadily causing malfunctioning of families and community at large.

2.5 Control/Prevention of Road Accidents

Road safety reduce the number of road traffic crashes or injury on the roads and this can be achieved through multi-disciplinary approaches which involve inviting road and Traffic Engineer, Education and training of road users, and vehicle design. Extensive remedial measures aimed at improving road safety have been developed in the field of engineering, education, and enforcement "Three Es" (Gbadmosi, 1994).

2.5.1 Engineering Measures to Road Accident Prevention

Many countries now regularly implement low-cost measure to "Black spot", (places where accidents cluster). The subsequent savings in accident are substantial (not uncommonly up to three quarter), with Economic benefits of several times more than the cost of the measures in the first year. Successful treatments have included: changes in layout at junctions to define priorities, more wide spread use of road makings to delineate traffic laws and waiting areas for turning vehicles, improvements in skidding resistance of wet roads, more uniform street lighting, and more highly flammable and legible direction, uniform, and warnings signs.

2.5.2 Protective Measures to Road Accident Prevention

Use of seat belt reduces the risk of death or services injury by about 45per cent (Jacob,1995) publicity has also played a major part in increasing wearing rules, but for full effect it needs to be bunched by legislation. Legislation for compulsory wearing was first introduced in the state of Victoria, Australia in 1971. Today all the major developed and developing countries of the world including Nigeria have being enforcing the warning of seat belts, with reported levels of compliance of over 90 percent in some countries, notably for front seat passengers in the United Kingdom. Other protective measure that are gaining support are the wearing of helmet by cyclists and the used of crash protective burrier on the central reserve of high-speed motorways, and to guard rigid object on the road side (utility poles, sign supports, Bondge Abutment, and Ties). Many advances have been made in vehicle design to protect occupant. There is also potential for greater protection for the Vulnerable pedestrian and cyclists who come into contact with motor vehicles, but promising developments have not yet been fully exploited. Other protective measures developed include:

i. restrain on mobile phone usage when driving;
ii. alcohol intake
iii. enforcement of speed unit and others

2.6 Strategic Highway Safety Plans

A Strategic Highway Safety Plan (SHSP) developed by the State Department of Transportation (DOT) is a new Federal requirement of SAFETEA-LU, 23, and is a major part of the core Highway Safety Improvement Program (HSIP). The purpose of an SHSP is to identify the State's key safety needs and guide investment decisions to achieve significant reductions in highway fatalities and serious injuries on all public roads. It is a statewide-coordinated safety plan that provides a comprehensive framework, and specific goals and objectives, for reducing highway fatalities and serious injuries on all public roads. This statewide document, developed by the State DOT in a collaborative process, includes input from public and private safety stakeholders (US Department of Transportation, 2006).

2.6.1 Benefits of Strategic Highway Safety Plans (SHSP)

The benefits of SHSP include:

i. Better coordination of state-wide goals and safety programs that most effectively reduce highway fatalities and serious injuries on all public roads through a comprehensive approach.
ii. Scheduling and implementation of safety improvement programs, comprehensive initiatives, and projects to be coordinated throughout the State.
iii. Establishing common state-wide safety goals and priorities.
iv. Strengthening existing partnerships.
v. Building new safety coalitions.
vi. Sharing data, knowledge, and resources.

vii. Quantifying the existing and needed resources and activities to meet the State's safety goal.

viii. Avoiding redundant activities and leveraging limited existing resources such as funds, people, and leadership attention, toward common objectives.

ix. Communicating the impact of investing additional resources for highway safety countermeasures, and

x. Incorporating both behavioural and infrastructure strategies and countermeasures to have a greater impact on reducing highway fatalities and serious injuries on all.

2.6.2 Developing the SHSP

The following suggested activities will help to create a process and identify milestones for the development of the SHSP;

i. Gain Leadership Support and Initiative

ii. Identify a Champion

iii. Initiate the Development Process

iv. Gather Data

v. Analyse Data

vi. Establish a Working Group

vii. Bring Safety Partners Together

viii. Adopt a Strategic Goal

ix. Identify Key Emphasis Areas

x. Form Task Groups

xi. Identify Key Emphasis Area Performance Based Goals

xii. Identify Strategies and Countermeasures

xiii. Determine Priorities for Implementation

xiv. Write the SHSP

2.7 Theory of Intersections

Intersection is an area shared by two or more roads. This area is designated for the vehicles to turn to different directions to reach their desired destinations. Its main function is to guide vehicles to their respective directions. Traffic intersections are complex locations on any highway. This is because vehicles moving in different direction want to occupy same space at the same time. In addition, the pedestrians also seek same space for crossing. Drivers have to make split second decision at an intersection by considering his route, intersection geometry, speed and direction of other vehicles etc. A small error in judgment can cause severe accidents. It also causes delay and it depends on type, geometry, and type of control. Overall traffic flow depends on the performance of the intersections. It also affects the capacity of the road. Therefore, both from the accident perspective and the capacity perspective, the study of intersections is very important for the traffic engineers especially in the case of urban scenario (Tom Matthew 2009).

2.7.1 Conflicts at an intersection

Conflicts at an intersection are different for various types of intersection. The essence of the intersection control is to resolve these conflicts at the intersection for the safe and efficient movement of both vehicular traffic and pedestrians. Two methods of intersection controls are there: time sharing and space sharing. The type of intersection control that has to be adopted depends on the traffic volume, road geometry, cost involved, importance of the road etc.

2.7.2 Levels of intersection control

The control of an intersection can be exercised at different levels. They can be either passive control, semi control, or active control. In passive control, there is no explicit control on the driver. In semi control, some amount of control on the driver is there from the traffic agency. Active control means the movement of the traffic is fully controlled by the traffic agency and the drivers cannot simply maneuver the intersection according to his choice.

2.7.2.1 Passive control

When the volume of traffic is less, no explicit control is required. Here the road users are required to obey the basic rules of the road. Passive control like traffic signs, road markings etc. are used to complement the intersection control

2.7.2.2 Semi control

In semi control or partial control, the drivers are gently guided to avoid conflicts. Channelization and traffic rotaries are two examples of this.

1. Channelization: The traffic is separated to flow through definite paths by raising a portion of the road in the middle usually called as islands distinguished by road markings. The conflicts in traffic movements are reduced to a great extent in such a case. In channelized intersections, as the name suggests, the traffic is directed to flow through different channels and this physical separation is made possible with the help of some barriers in the road like traffic islands, road markings etc.
2. Traffic rotaries: It is a form of intersection control in which the traffic is made to flow along one direction around a traffic island. The essential principle of this control is to convert all the severe conflicts like through and right turn conflicts into milder conflicts like merging, weaving and diverging. It is a form of `at-grade' intersection laid out for the movement of traffic such that no through conflicts are there. Free-left turn is permitted whereas through traffic and right-turn traffic is forced to move around the central island in a clock-wise direction in an orderly manner. Merging, weaving and diverging operations reduces the conflicting movements at the rotary.

2.7.2.3 Active Control

Active control implies that the road user will be forced to follow the path suggested by the traffic control agencies. He cannot maneuver according to his wish. Traffic signals and grade separated intersections come under this classification.

2.7.3 Types of Highway Intersection

Highway intersections are classified into three categories, viz;

i. Grade-separated without ramps
ii. grade-separated with ramps (commonly known as interchanges and
iii. At-grade

2.7.4 Types of At-Grade Intersections

The basic types of at-grade intersections are T or three-leg intersections which consist of three approaches; four-leg or cross intersections, which consist of four approaches; and multi-leg intersections, which consist of five or more approaches.

Crashes often occur at intersections because these are the locations where two or more roads cross each other and activities such as turning left, crossing over, and turning right have the potential for conflicts resulting in crashes. The National Motor Vehicle Crash Causation Survey data collected at crash scenes between 2005 and 2007 is used in statistical analyses such as descriptive analysis, generalized logit model, and configural frequency analysis. Descriptive statistics are first used to highlight characteristics of the intersection-related crashes. The results from this analysis provide guidelines for in-depth analysis. Close associations of crash factors with critical reasons of an event that made the crash imminent are revealed through the analysis of generalized logit model (US Department of Transportation, 2010)

2.7.5 Classification of Intersections

i. Road segment

ii. Traffic control

iii. Lane design

2.7.5.1 Road Segement

One way to classify intersections is by the number of road segments (arms) that are involved.

i. A three-way intersection is a junction between three road segments (arms): T junction when two arms form one road, or a Y junction. The latter also known as a fork if approached from the stem of the Y.

ii. A four-way intersection, or crossroads, usually involves a crossing over of two streets or roads. In areas where there are blocks and in some other cases, the crossing streets or roads are perpendicular to each other. However, two roads may cross at a different angle. In a few cases, the junction of two road segments may be offset from each when reaching an intersection, even though both ends may be considered the same street.

iii. Five-way intersections are less common but still exist, especially in urban areas with non-rectangular blocks. An example of this is the intersection for which the Five Points district in Atlanta is named.

iv. Six-way intersections usually involve a crossing of three streets at one junction; for example, a crossing of two perpendicular streets and a diagonal street is a rather common type of 6-way intersection.

v. Seven or more approaches to a single intersection, such as at Seven Dials, London, are rare.

2.7.5.2 Traffic control

Another way of classifying intersections is by traffic control technology. This includes;

i. Uncontrolled intersections without signs or signals (or sometimes with a warning sign). Priority (right-of-way) rules may vary by country: on a 4-way intersection traffic from the right often has priority; on a 3-way intersection either traffic from the right has priority again, or traffic on the continuing road.

ii. Yield-controlled intersections which may or may not have specific "Yield" signs known as "Give Way" signs in some countries)

iii. Stop-controlled intersections have one or more "Stop" signs.

iv. Signal-controlled intersections depend on traffic signals, usually electric, which indicate which traffic is allowed to proceed at any particular time.

2.8 Laws and Regulation on Road Accident in Nigeria

Law enforcement has an important part to play in improving road safety. Polite activity in this respect is most effective where technological aids are available, and when the laws are acceptable to the majority of road users. The most dramatic reductions in accidents due to law enforcement lie in the area of drinking and driving. A successful law has been based on prescribing a limit to the amount of alcohol in the blood in the range from 50mg/100ml. the application of the law, new technology to aid enforcement, enhanced publicity, education of teenagers. In schools, and the developments of rehabilitation course are very good measures in accident reduction and prevention scheme.

2.9 Education and Training

Educating the young ones about the menace of road traffic crashes is a very good step in preventing crashes. Starting from schools, road traffic education should be part of civic education or better still a subject on its own. This will entrench a good culture of traffic into the children right from their primary school days. Training and retraining of drivers, pedestrians and other people connected with road usage is also another strategy that is very important. There is supposed to be an established training institute for drivers where they would be a special training for intending drivers and also provide retraining for others so as to refresh their knowledge and also keep them abreast of new development in road safety all over the world. This will go a long way in tackling the menace of road traffic crashes. In Nigeria for example, FRSC laid down some rules which are as follows: (FRSC)

i. Must know road safety rules
ii. Keep loyal to your side /law
iii. Keep calm and handle all turns carefully
iv. Place attention to traffic rule
v. Never exceed the speed limit
vi. Keep a sole distance
vii. Memorize and adhere to road signs

2.10.1 Linear Regression Model

Regression problems are of many dimensions. One aspect of regression which is mostly employed is the one that deals with the prediction of the dependent variable (x). This is a case of simple regression analysis. In the case were more than two independent variables are involved, we speak of multiple Regression. If two variables x and y are supposed to be related (linearly), the relationship may be tested as follows.

Let x_1 x_2 ………… x_n be the observed values of (x)

y_1 y_2 ……. y_n be the corresponding values of (y).

For instance, let (x_1, x_2, x_3 and x_4) the sets of independent variables and y is the corresponding rate of accident. The generalized multiple linear regression equation that relates a single dependent variable to numerous independent variables is presented as shown in Equation 2.1:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ---------- \beta_5 x_5 + \varepsilon \quad (2.1)$$

where

X_1, X_2, X_3… X_5 are input variables

Y, β_0, β_i, β_{ii}, and β_{ij} are the known parameters

2.10.2 Method of Least Square Regression

This is based on a linear prediction equation in which we must consider the problem of deriving computational formula for determining the point estimate (a) and (b) as contained in two linear simultaneous equations as shown in Equation 2.2 & 2.3:

$$an + \sum x = y \quad (2.2)$$

$$a\sum x + b\sum x^2 = \sum xy \quad (2.3)$$

2.10.3 Estimation of Regression Parameters

The least-square estimates of regression parameters (a) and (b) can be obtained from the generalized linear equation are shown in Equation 2.4

$$y = a + bx \quad (2.4)$$

where;

a is the intercept and

b is the slope

From:

$$a = \bar{y} - \mathrm{b}\,\bar{\mathrm{x}} \quad (2.5)$$

$$b = \sum y - nxy \quad (2.6)$$

$$n\sum x^2 - \left(\sum x\right)^2 = \sum x^2 - nx^2 \quad (2.7)$$

Again;

$$S^2 = \frac{1}{n-2}\left(xy^2 - ny^2\right) - b\left(\sum xy - nxy\right) \tag{2.8}$$

$$Sa^2 = S^2\left(\frac{1}{n} + \frac{x^2}{\sum x - n\bar{x}^2}\right) \tag{2.9}$$

$$Sb^2 = \frac{S^2}{\sum x^2 - nx^2} \tag{2.10}$$

$$Var(y) = \frac{S^2}{n}\left[\frac{(x_0 - x)S^2}{\sum x^2 - nx^2}\right] \tag{2.11}$$

2.10.4 Regression Test of Significance

There are three major tests of significance for multiple linear regressions which are normally applied to assess the suitability of regression method to data analysis. They include

i. R^2-test; which is the coefficient of determination test

ii. F-test; which is the variance ratio test.

iii. T-test; which is the test for the significant of parameter based on ANOVA

2.10.4.1 Coefficient of Determination (R^2)

This test indicates the percentage contribution of the model to changes in independent variables. For example, if $R^2 = 0.85$, it indicates that the changes in x_1, x_2 x_3 and x_4 (the independent variables) account for about 85% of the changes in y (the dependent variables). The closer the value is to one, the better and more reliable is the model. Mathematically, r^2 can be obtained from Equation 2.12:

$$R^2 = \frac{SST}{SSR} = \frac{1 - SSE}{SST} \tag{2.12}$$

where;

SST is total sum of square,

SSR is regression sum of square and

SSE is sum of square error, is shown in equation 2.13a and 2.13b

$$R^2 = 1 - \frac{6\sum d^2}{n(n^2 - 1)} \qquad (2.13a)$$

Where n is the number of variables

2.10.4.2 Fishers F- Test

Fishers F test is used to test the overall significance of the model. The F statistic seeks to find out if the explanatory variables have significant influence on the dependent variables. Mathematically, the fishers F test is defined as follows

$$F - Ratio = \frac{(SSR)/(K-1)}{(SSE)/(n-1)} = \frac{(SSR)/2}{(SSE)/(n-1)} \qquad (2.13b)$$

Where; k is the number of parameters. As explained above, the F-test is used for the overall significance of the model. For example, using 5% level of significance, this means that any value for an F-test obtained which lead to a level of significance higher than 5% will make the model insignificant, but F-test value which leads to a level of significance less than 5% make a model adequate or significant. We can therefore proceed to checking or testing for the significance of each parameter which is the t-test. This may not be necessary unless the F-test fails.

2.10.5 Artificial Neural Network

Artificial neural network (ANN) is a computational model based on the structure and functions of biological neural networks. Information that flows through the network can affect the output of the network since a neural network changes - or learn in a manner that is controlled by input and output. ANNs are considered nonlinear statistical data modelling tools which are capable of modelling the complex relationships between inputs and outputs. ANN is also known as a neural network (Rajurkar *et al*., 2002).

Artificial Neural Networks (ANNs) have been motivated right from their inception by the recognition that the human brain computes in an entirely different way from the conventional digital computer. The brain is a highly complex, non-linear, and parallel computer (information-processing system). It has the capability to organize its structural constituents, known as neurons in a massively distributed and parallel network, so as to perform certain computations (e.g., pattem recognition, perception, prediction and simulation) many times quicker than the fastest digital computer in existence today (Persson and Berndtsson, 2001).

ANNs consist of a large number of simple processing elements called neurons or nodes. Each node is then connected to other nodes by means of direct links. Each link is associated with a weight that represents the strength of outgoing signal. The processing of each node is carried out in two steps, that is, the weighted sum of the inputs is taken, and is followed by the application of the activation function as presented in Figure 2.1.

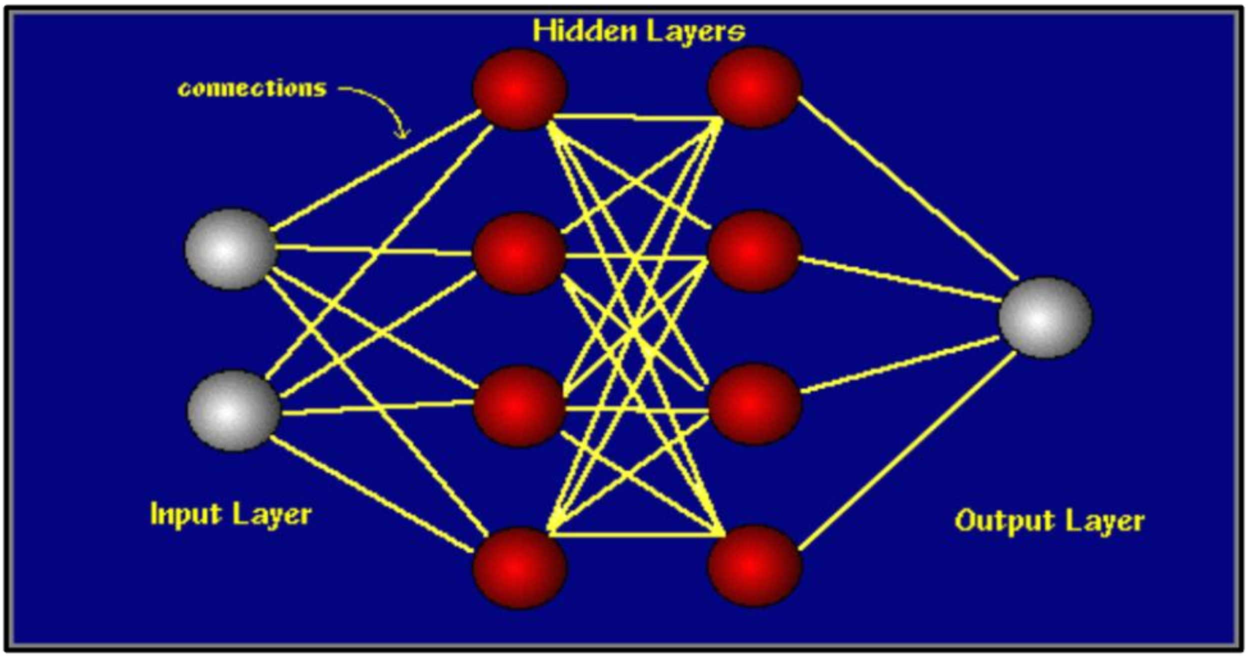


Figure 2.1: Schematic Representation of a Neural Network Unit (Persson and Berndtsson, 2001)

There is various activation function employed in ANNs, such as; the unipolar binary function or tan- sigmoid function (S), the bipolar binary function (B), the hyperbolic tangent function (T) and the linear function (L). Tan sigmoid activation or transfer function being the most commonly used. The weights of connections encode the knowledge embedded in the network.

The “intelligence” of a neural network emerges from the collective behaviour of neurons, each of which performs only very limited operation. Each individual neuron finds a solution by working in parallel. The following list describes the overall tasks involved in constructing an ANN;

i. determine the network properties or architecture: This includes the network connectivity, the types of connections, the order of the connections (if any), and the weight range values.

ii. determine the system dynamics: this entails the weight initialization method, the activation- calculating formula, and the learning rule.

The topology of a neural network is specified by the number of layers, the number of units per layer and the weighted connections among all the units. These types of layers are the Input layer, the Hidden layer and the Output layer as presented in Figure 2.2

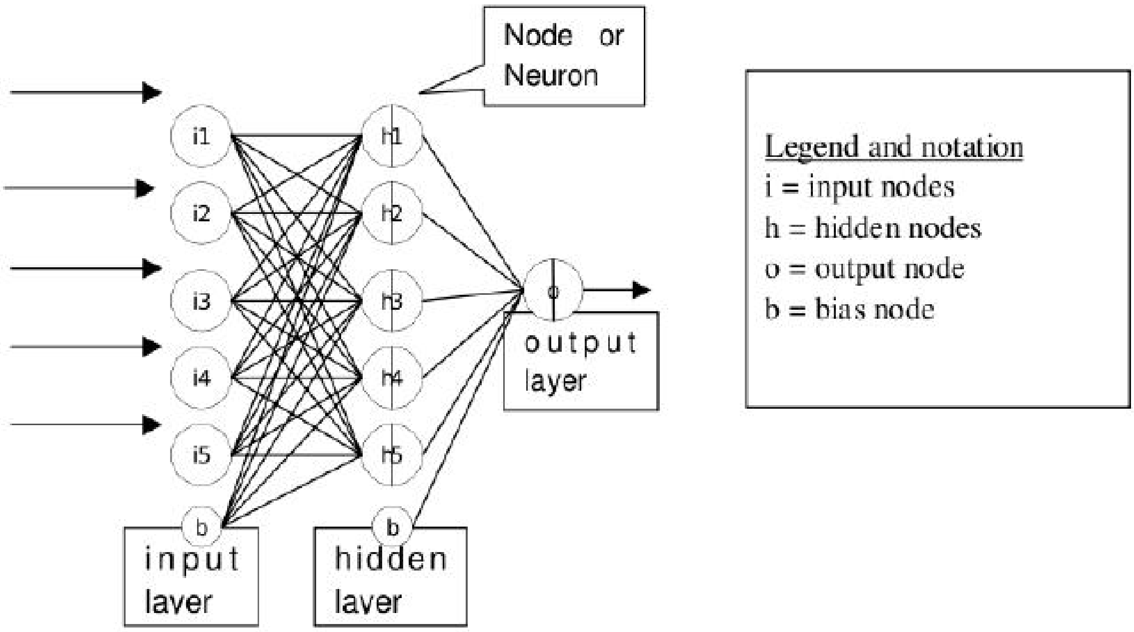


Figure 2.2: Neural Network Layers (Persson and Berndtsson, 2001)

Most ANNs contain some form of 'learning rule' which modifies the weights of the connections according to the input patterns that it is presented with. In a sense, ANNs learn by example as do their biological counterparts. In reality, a child learns to recognize dogs from examples of dogs. Although there are many different kinds of learning rules used by neural networks. Improved second order method of gradient (Lavenberg Marquardt) is often utilized by the most common class of ANNs called 'back propagation neural networks' (BPNNs). Back propagation is an abbreviation for the backwards propagation of error. Lavenberg Marquardt as with other types of back propagation, learning is a supervised process that occurs with each cycle or 'epoch' (i.e. each time the network is presented with a new input pattern) through a forward activation flow of outputs, and the backwards error propagation of weight adjustments. More simply, when a neural network is initially presented with a pattern it makes a random 'guess' as to what it might be. It then sees how far its answer was from the actual one and makes an appropriate adjustment to its connection weights. More graphically, the process looks something like this:

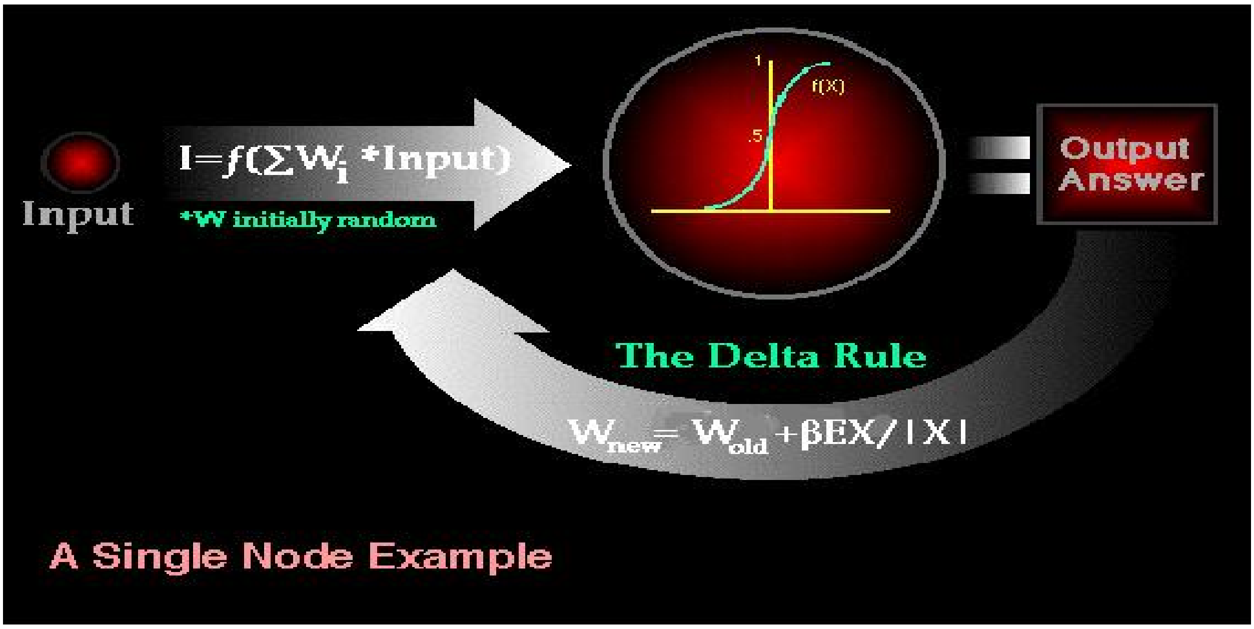


Figure 2.3: Back propagation neural network (Rajurkar et al., 2002).

2.10.5.1 Relationship between ANN and the Human Brain

A neural network resembles the human brain in two aspects:

i. Knowledge is acquired by the network through a learning process.
ii. Inter-neuron connection strengths known as synaptic weights are used to store the knowledge.

The most interesting fact about neural networks is the possibility of learning. The learning ability and performance of an ANN depends on the suitability of its architecture which is needed to be pre-specified, more specifically, the number and configuration of its hidden nodes.

A neural network is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. It is used to learn patterns and relationships in data. The learning rule enables a network to gain adequate knowledge from available data and apply that knowledge to train the network. There are three major learning paradigms, each corresponding to a particular abstract learning task. These are:

i. Supervised learning
ii. Unsupervised learning
iii. Reinforcement learning

In supervised learning, the understanding rate and sensitivity of the network is usually monitored by checking the error difference between the predicted data (network output) and the input-target data. A commonly used function for monitoring is the mean-squared error, which tries to minimize the average squared error between the network's output, f(x), and the target value y over all the pairs. This is not the case for unsupervised learning where no emphasis is placed on the difference between the network output and the target variable. In which case, no special attention is placed on the sensitivity of the network to perturbation in the input-target variable.

In supervised learning, the output datasets are provided which are used to train the network and get the desired outputs whereas in unsupervised learning no datasets are provided, instead the data is clustered into different classes as in the case of face recognition problem. In addition, for unsupervised learning, since there is no desired output provided, categorization is done so that the algorithm differentiates correctly between the face of a horse, cat or human (clustering of data).

When problems have a relatively small number of state (controlling variables) and the underlying random structure is relatively simple, one can use dynamic programming. But when the problem has a very large number of states (>> 1000) and has a complex stochastic structure, then reinforced learning is needed. Reinforced learning is a technique useful in solving control optimization. In control optimization you are interested in optimizing some objective function by recognizing the best action in every state visited by the system.

2.10.5.2 Types of Neural Network

Currently, many types of neural networks are known. These networks are mutually different in architecture and in the way the weights are adjusted. The most popular type of neural network is the multiple-layer perceptron (MLP). This network has a simple interpretation as a form of input-output model, with the weights and thresholds being the main parameters of the model. Such networks can model functions of almost arbitrary complexity provided the number of layers, and the number of units in each layer is known. Important issues in MLP design include specification of the number of hidden layers and the number of units in these layers.

A Radial Basis Function network (RBF) has a hidden layer of radial units, each modeling a Gaussian response surface. Since these functions are non-linear, it is not actually necessary to have more than one hidden layer to model any shape of function: sufficient radial units will always be enough to model any function.

Generalized Regression Neural Network (GRNN) and Probabilistic Neural Network (PNN) are variants of the radial basis function (RBF) network. Unlike the standard RBF, the weights of these networks can be calculated analytically. PNNs are designed for classification tasks and GRNNs for regression.

In the PNN, there are at least three layers: input, radial, and output layers. The radial units are copied directly from the training data, one per case. Each model a Gaussian function centered at the training case. There is one output unit per class. Each is connected to all the radial units belonging to its class, with zero connections from all other radial units.

Generalized Regression Neural Networks (GRNNs) work in a similar fashion to PNNs. As with the PNN, Gaussian kernel functions are located at each training case. Each case can be regarded as evidence that the response surface is a given height at that point in input space, with progressively decaying evidence in the immediate vicinity. The GRNN copies the training cases into the network to be used to estimate the response on new points. The output is estimated using a weighted average of the outputs of the training cases, where the weighting is related to the distance of the point from the point being estimated. The first hidden layer in the GRNN contains the radial units. A second hidden layer contains units which help to estimate the weighted average. Each output has a special unit assigned in this layer which forms the weighted sum for the corresponding output. To get the weighted average from the weighted sum, the weighted sum must be divided through by the sum of the weighting factors. A single special unit in the second layer calculates the latter value and then the output layer performs the actual divisions (using special “division” units). Hence, the second hidden layer always has exactly one more unit than the output layer. In regression problems only a single output is estimated, and so the second hidden layer usually has two units (Hung et al., 2008).

The most widely used kind of neural network is the linear neural network. In neural network terms, a linear model is represented by a network having no hidden layers, but an output layer with fully linear units (that is, linear units with linear activation function). The weights correspond to the matrix, and the thresholds to the bias vector. When the network is executed, it effectively multiplies the input by the weights matrix and then adds the bias vector (Rabunal, 2005).

2.10.5.3 ANN implementation

ANNs are usually implemented by using electronic components or are simulated in software on a digital computer. They are characterized by;

1. Their pattern of connections between the neurons (called its architecture)
2. Their methods of determining the weights on the connections (called their training, or learning, algorithm),
3. Their activation function, and
4. Their number of layers: single (Hopfield nets); bilayer (Carpenter/Grossberg adaptive resonance networks); and multilayer (most back-propagation networks).

If all the signals flow only in one direction, an ANN is called a Feed-Forward Network (FFN). Otherwise, they are called Recurrent Networks.

2.10.5.4 Advantages of ANN

i. Artificial neural networks are algorithms that can be used to perform non-linear statistical modeling and provide a new alternative to logistic regression.
ii. Neural networks offer a number of advantages, including requiring less formal statistical training, ability to implicitly detect complex non-linear relationships between dependent and independent variables, ability to detect all possible interactions between predictor variables, and the availability of multiple training algorithms.

iii. In higher dimensional settings ANNs are able to represent complex non-linear behaviour. So when there is no information on the form of the function to be predicted and the task of specifying the functional form from the data is computationally complex, then ANNs are superior over polynomial expansions used in linear regressions, as the errors for the later are more dispersed than ANN.

iv. In addition, ANN is a non-parametric tool, thus eliminates the error in parameter estimation, while most statistical tool such as regression analysis are parametric models that need higher background of statistic.

2.10.5.5 Limitations of ANN

i. Disadvantages include its “black box” nature, greater computational burden, proneness to over fitting, and the empirical nature of model development

ii. Another prominent problem that is associated with ANN is the problem of curse of dimensionality resulting from the large volume of input data and the associated weight variation thus making the network want to memorize instead of learning

2.10.6 Theory of Fuzzy Logic

Fuzzy logic has become an important tool for number of different applications ranging from the control of engineering system to artificial intelligence. Practical applications of fuzzy logic pose a unique set of problems. The design of systems, which apply fuzzy logic to make use of human knowledge and experience, is a daunting task without facing engineering problems of real-world systems. Fuzzy logic is a set of mathematical principles for knowledge representation based on degrees of membership (Unahabhokha et al., 2007). Fuzzy logic is a form of many-valued logic; it deals with reasoning that is approximate rather than fixed and exact. Compared to traditional binary sets (where variables may take on true or false values), fuzzy logic variables may have a truth value that ranges in degree between 0 and 1. Fuzzy logic has been extended to handle the concept of partial truth, where the truth value may range between completely true and completely false. When linguistic variables are used, these degrees may be managed by specific functions. Fuzzy logics provide the basis for logical systems dealing with vagueness, e.g. for formalizing common natural language predicates such as “tall” or “fast”. Design choices in this framework are made as to which real numbers to take as truth values, and which properties connectives should have. In fact, logics based on real numbers occur in a number of areas in logic. Fuzzy logic is based on the theory of fuzzy sets, which a generalization of the classical is set theory. Saying that the theory of fuzzy sets is a generalization of the classical set theory means that the latter is a special case of fuzzy sets theory. To make a metaphor in set theory speaking, the classical set theory is a subset of the theory of fuzzy sets (Zadeh, 2016).

A fuzzy set is a set without a crisp, not clearly defined boundary. It can contain elements with a partial degree of membership with multi-valued logic. Fuzzification comprises the process of transforming discrete values into grades of membership (continuous) for linguistic terms of fuzzy sets. The membership function is used to associate a grade to each linguistic term. Defuzzify evaluate several membership sets established by the system designer for a fuzzy logic-based control system, such as "speed too fast," "speed too slow" and "speed about right" at a specific input value (Zadeh, 2016).

Degree of membership is a specific value that defines how each point in the input space is mapped to the specific environment being studied lying between 0 and 1. Linguistic Variable means relating to language, (plain language words and statements). While variables in mathematics usually take numerical values, in fuzzy logic, the non-numeric linguistic variables are often used to facilitate the expression of rules and facts (Sirigiri *et al*., 2012).

A Fuzzy Logic System consists of four main parts: fuzzier, rules, inference engine, and defuzzifier (Unahabhokha *et al*., 2007). These components and the general architecture of a Fuzzy Logic System are shown in Figure 2.4

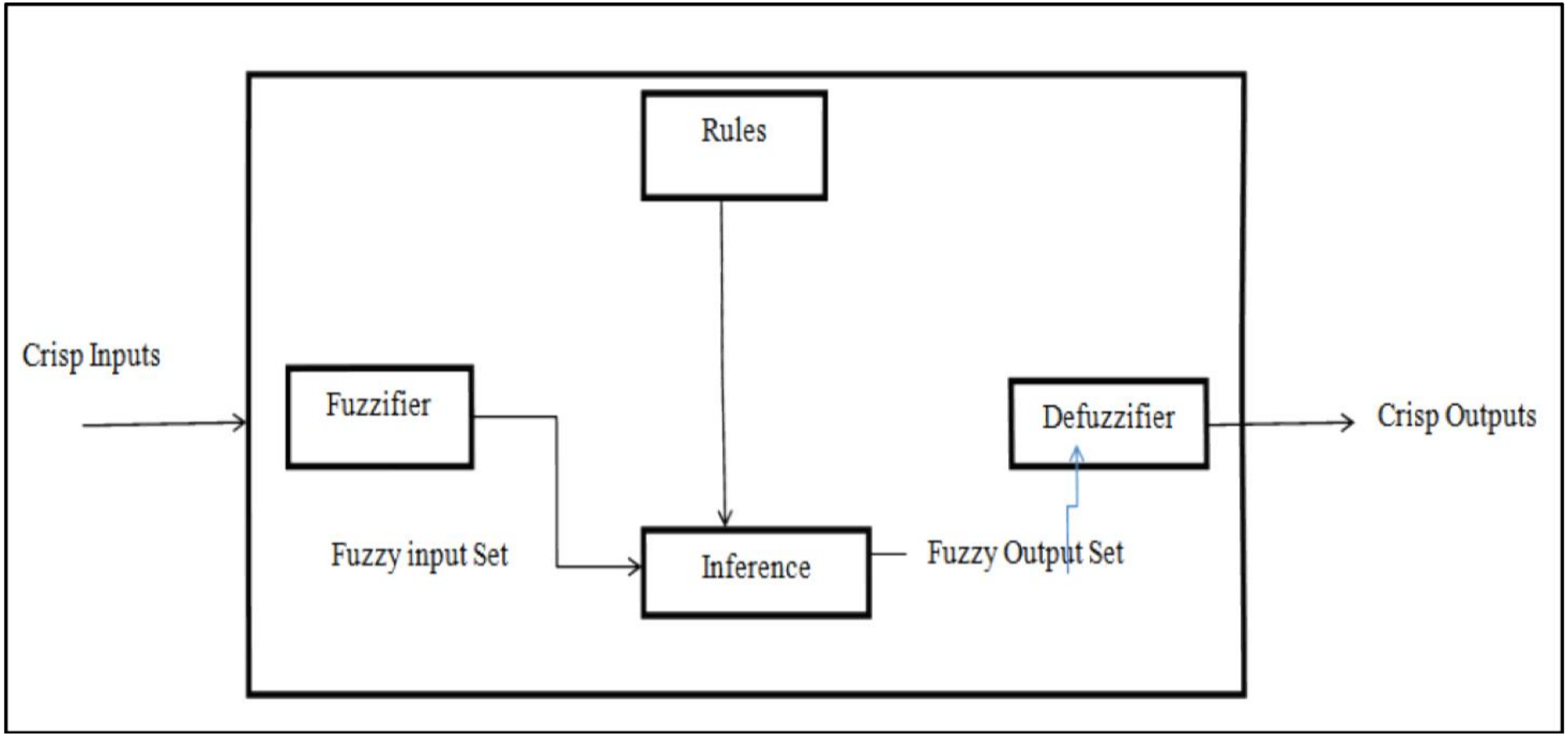


Figure 2.4: Fuzzy Logic System (Unahabhokhaeal, 2007)

2.10.6.1 Fuzzy logic modelling

To determine crisp variables used in the fuzzy logic systems, the formula was adopted from Qiu *et al*., 2014 as as shown in equation 2.14;

$$\mathrm{A} = \left(a_{ij}\right)_{nxn} = \begin{bmatrix} (1,1,1) & (l_{12}, m_{12}, u_{12}) & \ldots\ldots\ldots. & (l_{1n}, m_{1n}, u_{1n}) \\ (l_{21}, m_{21}, u_{12}) & (1,1,1) & \ldots\ldots\ldots.. & (l_{2n}, m_{2n}, u_{2n}) \\ \vdots & \vdots & \vdots & \vdots \\ (l_{n1}, m_{n1}, u_{n1}) & l_{n2}, m_{n2}, u_{n2} & \ldots\ldots\ldots.. & (1,1,1) \end{bmatrix} \qquad (2.14)$$

Where $a_{ij} = \left(l_{ij}, m_{ij}, u_{ij}\right) = aij = (1/u_{ji}, \ 1/m_{ji} \ 1/l_{ji}$) for I, j = 1----------, n and I $\neq$ j.

To calculate a priority vector of the above triangular fuzzy comparison matrix, the following five equations were used:

First, a fuzzy comparison operation was employed to sum up each row of the fuzzy comparison matrix as presented in Equation 2.15

$$RS_i = \sum_{j=1}^{n} a_{ij} = \left(\sum_{j=1}^{n} l_{ij}, \sum_{j=1}^{n} m_{ij}, \sum_{j=1}^{n} u_{ij} \right), \qquad i = 1 \ldots n \qquad (2.15)$$

Secondly, the row sums were normalized using Equation (2.15) as shown in Equation 2.16;

$$S_i = \frac{RS_i}{\sum_{j=1}^{n} RS_j} = \left(\frac{\sum_{j=1}^{n} l_{ij}}{\sum_{j=1}^{n} l_{ij} + \sum_{k=1,k\neq i}^{n} + \sum_{j=1}^{n} u_{kj}}, \frac{\sum_{j=1}^{n} m_{ij}}{\sum_{k=1}^{n}\sum_{j=1}^{n} m_{kj}}, \frac{\sum_{j=1}^{n} u_{ij}}{\sum_{j=1}^{n} ui_j + \sum_{k=1,k\neq i}^{n} 1\sum_{j=1}^{n} l_{kj}} \right), \quad i = 1 \ldots n \qquad (2.16)$$

Thirdly, the degree of possibility of $S_i \leq S_j$ was computed with the aid of Equation (2.16) as shown in Equation 2.17;

$$V(Si \geq Sj) = \begin{cases} 1 & mi \geq mj \quad 1 \\ \dfrac{ui - lj}{(ui - mi)(mj - lj)}, & lj \leq uj \quad 1 \\ 0 & others \quad 1 \end{cases} \quad \text{I, j = 1…n, j} \neq i \qquad (2.17)$$

Where the possibility degree is $S_i = (l_i, m_i, u_i)$ and $S_j = (l_j, m_j, u_j)$

Fourthly, the degree of possibility of S_i over all the other (n – 1) fuzzy member was computed using Eqution 2.18:

$$V\left(S_i \geq S_j\, j = 1.......,n; j \neq i\right) = V\left(S_i \geq S_j\right), i = 1,........,n \qquad (2.18)$$

Finally, the priority vector $W = \left(w_1........., w_n\right)^T$ of the fuzzy comparison matrix A was defined as shown in Equation 2.19

$$wi = \frac{V\left(S_i \geq S_j\, j = 1,..........,n; j \neq i\right)}{\sum_{k=1}^{n} V\left(S_k \geq S_j\, j = 1,..........,n; j \neq k\right)}, \quad i=1\ldots n \qquad (2.19)$$

2.10.6.2 Methodology of Fuzzy Logic

The basic steps involved in the application of fuzzy logic for the modelling and prediction of suitable site for landfill application are as follows;

i. Definition of input and output variables
ii. Conversion of crisp variables into fuzzy sets
iii. Definition of membership functions for each inputs and output
iv. Creation of fuzzy rules
v. Simulation

2.11 Geometry of Road Design

The geometric design of road is the branch of highway engineering concerned with the positioning of the physical elements of the roadway according to standards and constraints. The basic objectives in geometric designs are to optimize efficiency and safety while minimizing cost and environmental damage. Geometric design also affects an emerging fifth object also called "livability", which is defined as designing roads to a foster broader community goals, including providing access to employment, schools business and residence, accommodate a range of travel modes such as walking, bicycling, transit, and automobiles, and minimizing fuel use, emissions and environmental damage.

Geometric roadway design can be broken into three main parts; alignment, profile, and cross-section. Combined they provide a three-dimensional layout for a roadway.

2.11.1 Alignment:

This is the route of the road, defined as a series of horizontal tangent and curves.

2.11.2 Profile:

This is the vertical aspect of the road, including crest and curves, and the straight grades lines connecting them.

2.11.3 Cross-section:

The cross-section shows the position and number of vehicles lanes and sidewalks, along with their cross slope or banking. Cross section also shows drainage features, pavement structure and other items outside the category of geometric designs.

2.12 Review of Previous Related Work Done

In a research conducted by Mohammed et al., 2018 on classification of traffic accident prediction models; A review paper, the author affirms that; accident prediction models (APMs) are extremely important tools for estimating the expected number of accidents on entities such as intersections and street segments. The authors reaffirm; that the estimates from accident prediction models are typically used in the identification of sites for possible safety treatment and in the evaluation of such treatments. They simply define an accident prediction model as a mathematical equation that expresses the average accident frequency of a site as a function of traffic flow and other site characteristics. They concluded that; the credibility of an APM is enhanced if the APM is based on data collected for many years as possible especially if data for those same years are utilized in the safety analysis of a site. The paper covered a review of many papers as possible and various gaps in the research along with future possibility of study in this area were indicated. Several traffic accident prediction models were identified and discussed by the author's, namely; multiple linear regressions, Poisson regression, Conway-Maxwell Poisson regression models, artificial neural networks and fuzzy logics.

In a research conducted by Abdullah et al., 2012 on road accident models with two threshold levels of Fuzzy Linear regression; the author hypothesized that number of road accidents and road casualties are increasing in line with the raise in the variables of registered vehicles, population and road length. The authors stated that the effects of these variables toward road accidents are still inconclusive. The reviewed paper developed models based on the variables which can be used to determine road accidents in Malaysia. Abdullah *et al*., explained the effects of these variables to number of road accidents by testing fuzzy linear regression models with threshold level 0.5 and 0.9. The results showed that by applying a multi-variable approach of fuzzy linear regression, the models provide not only crisp output but also output range for number of road accidents in Malaysia. They concluded that the variables of registered vehicles and population were notable predictors to number of road accidents in Malaysia.

Osime et al., (2006) observed in their study that about 285,699 cases of road traffic accident (RTA) occurred between 1970 and 1979 with 57,136 deaths, which amounts to 20%. Again, 188,012 cases of RTA occurred between 1990 and 1999, where about 76,870 deaths were recorded, amounting to 41%. Reasons for this include the oil explosion in Nigeria, which occurred in the 1970s. This empowered many Nigerians financially to afford cars of their choices. This also aided in the repair of roads and the construction of new roads in Nigeria. But given that people were not yet exposed to high traffic of cars in Nigeria roads, this led to many cases of RTAs. Also, due to the fact that people are new to the development, cases of speeding could still be assumed to be relatively small. With the availability of good roads and due to the fact that some drivers drive with reasonable speed, which amount to some cases of RTAs, the deaths that occurred were about 20%. However, there was a sharp change observed between 1990 and 1999. The mortality rate increased to 41%, while the number of RTAs decreased. This is the time economic recession was observed in Nigeria, which led to the inability of acquiring new cars by most people; instead there was an increase in the purchase of used cars, which in turn led to an increase in RTAs. The economic recession also affected most roads, which left them in a bad state, making it more likely to cause fatal accidents. The country equally experienced neglect in the health sector, with little attention paid to RTAs cases.

In a research conducted by Gaber et al., 2016 on traffic accident prediction model using fuzzy logic: Aswan desert road case study; The author affirms that; transportation system plays an important role in human life and is one of the main indicators of the standard of living. They simply described traffic accidents as a major problem threatening people's lives, health, property and the management of transportation system. The authors affirmed that traffic accidents prediction models may help in understanding accident causes and the number of their occurrence under certain circumstances. The study was aimed at developing a prediction model for Aswan western desert road by using fuzzy logic which is known for its benefits in dealing with uncertainty problems. The study was carried out by the use of actual accident data obtained from the Egyptian General Authority for Roads, Bridges, and Land Transport (GARBLT) with survey data for pavement conditions, traffic flow presented as average hourly traffic per lane (AHTL), speed, minor access, traffic signs conditions and road width which are the inputs of the model. Several traffic accident prediction models were developed by the authors using the Poisson regression model, negative binomial regression model and negative multinomial model based on generalized linear regression technique. The authors reaffirm that, the relationship between an accident and the influencing factors is nonlinear and complicated and the using of fuzzy is preferable because fuzzy logic system is good for dealing with nonlinear input and output relationship. The overall results of the study revealed that the predicted results obtained by using the proposed fuzzy logic system produced accurate and stable traffic accident predictions.

Yong and Jung, (2012) conducted a research on automobile traffic accidents prediction models using by Artificial Neural Network. In their research, neural networks were utilized as useful modelling technique. Yong and Jong, (2012) proposed The use of neural network for a traffic accident in-depth analysis of scientific research through the pre-crash factors in order to reduce traffic accidents. For the prevention of traffic accident, the main factors are found associated with deaths. The authors affirm that; it is a difficult problem to present the perfect result while considering all circumstances if applied in actual problems, as multiple variables that affects the result. In the reviewed paper, two kinds of neural networks of MLP (Multi-Layer Perceptron) and RBFN (Radial-Basis Function Network) were experimented by XLMiner.

In a research conducted by Gajendran et al., 2015 on different methods of accident forecast based on real data; the authors affirmed that; loss of lives through road accidents are increasing day by day as there is increase in the number of motor vehicles on the road which has created a major problem. The paper discussed three types of accident prediction model namely; System Dynamic Model, Fuzzy logic and Bayesian Method. The authors reaffirmed that; investing in transportation sector leads to betterment of basic infrastructure at the development of a country. The Complex, Dynamic and Non-linear interaction can be understood using system dynamic model. Fuzzy logic deals with occurrence of sets and elements. Fuzzy model compresses of four sub process: Fuzzification, Rule Production, Composition or Aggregation and Defuzzification. Bayesian refer to methods in probability and Statics which has held to model the interaction between road geometry, traffic characteristics and accident frequencies by means of linear regression model.

In the research conducted by Rezaic et al., (2010) on prediction of accident severity using artificial neural networks; the author affirmed that; in spite of significant advances in highways safety, a lot of crashes in high severities still occur in highways. The reaffirm that; investigation of influential factors causing crashes enables engineers to carry out calculations in order to reduce crash severity. The paper deals with the models to illustrate the simultaneous influence of human factors, road, vehicle, weather conditions and traffic features including traffic volume and flow speed on the crash severity in urban highways. The study used a series of artificial neural networks to model and estimate crash severity and to identify significant crash-related factors in urban highways. According to Rezaic et al., (2010), artificial neural networks is capable to predict and present desired results in spite of limited data sets, which is the remarkable feature of the artificial neural networks models. The obtained results illustrated that the variables such as highway width, head-on collision, type of vehicle at fault, ignoring lateral clearance, following distance, inability to control the vehicle, violating the permissible velocity and deviation to left by drivers are most significant factors that increase crash severity in urban highways.

In the research conducted by Garda et al., 2018 on predicting road accidents using artificial neural network models; the authors presented a methodology for establishing an accident risk prediction model, which can be used as a decision-making tool in infrastructure management. The authors affirmed that; the methodology allowed for an appropriate handling of the available data, examines how it can be used to develop models using artificial neural networks (ANNs) and establishes a systematic ANN optimization process to determine the optimal architecture of the ANN model. The methodology was implemented using data for accident counts on the Swiss national roads from 2009 to 2012. It was found that ANNs can be used as a viable method to predict the frequency of road accidents. As accident occurrences are relatively rare events, the data are characterized by a large portion of zero observations. This poses a challenge for the training of the ANN. The results obtained showed that such models provide reliable results as indicated by the symmetric mean absolute percentage error, ranging from 17·5 to 32·7%.

In the research conducted by Mehmet et al., 2011, on prediction for traffic accident severity: comparing the artificial neural network, genetic algorithm, combined genetic algorithm and pattern search methods ; the authors focuses on predicting the severity of freeway traffic accidents by employing twelve accident related parameters in a genetic algorithm (GA), pattern search and artificial neural network (ANN) modelling methods. The models were developed using the input parameters of driver's age and gender, the use of a seat belt, the type and safety of a vehicle, weather conditions, road surface, speed ratio, crash time, crash type, collision type and traffic flow. The models were constructed based on 1000 of crashes in total that occurred during 2007 on the Tehran–Ghom Freeway due to the fact that the remaining records were not suitable for the study. From the reviewed work, the GA evaluated eleven equations to obtain the best one. The authors combined the GA and PS methods using the best GA equation. The neural network used multi-layer perceptron (MLP) architecture that consisted of a multi-layer feed-forward network with hidden sigmoid and linear output neurons that could also fit multi-dimensional mapping problems arbitrarily well. According to the investigation, the ANN was applied during training, testing and validation and had 12 inputs, 25 neurons in the hidden layers and 3 neurons in the output layer. The best-fit model was selected according to the *R*-value, root mean square errors (RMSE), mean absolute errors (MAE) and the sum of square error (SSE). They obtained the highest *R*-value for the ANN to be around 0.87, demonstrating that the ANN provided the best prediction. They concluded that the combination of GA and PS methods allowed for various prediction rankings ranging from linear relationships to complex equations. The advantage of these models is improving themselves adding new data.

In research conducted by Zianhu and Xiongbin (2014), on prediction of road traffic accidents using a combined model based on IOWGA operator; the authors affirm that; traffic accident prediction plays an important role in reducing the likelihood of traffic accidents and improving the management levels of traffic safety. Zianhu and Xiongbin, (2014) proposed a new combined prediction model based on the induced ordered weighted geometric average (IOWGA) operator. The new model combines the GM (1,1) model and the Verhulst model with changeable weight coefficients of each single model. A combined model based on the optimal weighted (OW) method was also presented for comparison. The author gave an example with the number of deaths by road traffic accidents in China from 2003 to 2008. The results of the investigation indicated that the proposed combined model is better than the other three models.

Murthy *et al*., 2015, investigated the development of models for road accidents based on intersection parameters using regression model. The author affirms that; the issue of road accidents is an increasing problem in developing countries. The authors reaffirm that this could be due to increasing road traffic /vehicle occupancy. This has been increasing over years. Regulating traffic on roads is an important task. There by reducing accidents in accident prone zones. The research showed that accident was drastically increased over a decade from 4% to 31%. This is an alarming issue. The analysis and identification of such road accident prone zones is essential to reduce the accidents. According to the paper, a model was developed based on intersection parameters and no. of accidents by regression analysis.

In research conducted by Mohd Zakwan, (2011) on development of accident prediction model by using artificial neural network (ANN), The author affirms that Statistical or crash prediction model can be used to identify major contributing factors or establish relationship between crashes and explanatory accident variables. The measures taken to prevent accident were from the speed reduction, widening the roads, speed enforcement, or construct the road divider, or other else. The purpose of the study was to develop an accident prediction model at federal road FT 050 Batu Pahat to Kluang. The study process involves the identification of accident blackspot locations, establishment of general patterns of accident, analysis of the factors involved, site studies, and development of accident prediction model using Artificial Neural Network (ANN) applied software which named NeuroShell2. According to the author, the significant of the variables that were selected from these accident factors were checked to ensure the developed model can give a good prediction result. The performance of neural network was evaluated by using the Mean Absolute Percentage Error (MAPE). The study result showed that the best neural network for accident prediction model at federal road FT 050 is 4-10-1 with 0.1 learning rate and 0.2 momentum rate. The network model contains the lowest value of MAPE and highest value of linear correlation, r which was 0.8986. The review study established the accident point weightage as the rank of the blackspot section by kilometer along the FT 050 road (km 1 – km 103). Several main accident factors also have been determined along this road, and after all the data.

In research conducted by Arun *et al*., 2015 on road crash frequency prediction for Indian national highways using soft-computing tools; the author affirms that crashes and road fatalities have risen quite drastically in India in the past decade. The authors reaffirm that there are several factors that affect the frequency of crashes on Indian roads ranging from deficiencies in geometric design, poor maintenance history and other environmental and human behavioural factors. The reviewed paper studies the effects of these factors in predicting road crash frequency on the National Highways of India. For this purpose, crash history of 4710 crashes was collected on total segment length of 889 kilometers falling on various National Highways in India. Additionally, data such as pavement roughness and road geometrics was also collected. Conventional Poisson-based Generalized Linear Regression

Models and modern soft-computing methods such as Multilayer Perceptron networks and a hybrid Adaptive Neuro-Fuzzy Inference System were configured to predict frequency of crashes. The results of the study indicated that the Multilayer Perceptron had the best prediction performance. A sensitivity analysis was also subsequently performed

In research conducted by Mehdi *et al*., 2012 on application of adaptive neuro-fuzzy inference system for road accident prediction; The author affirms that; several modeling approaches have been developed in road safety literature to establish the relationship between traffic accidents and road characteristics since the last two decades. The authors reaffirm that; no extensive research work has been published on application of Adaptive Neuro-fuzzy Inference System (ANFIS) on road accident modelling. The reviewed paper was aimed to develop an ANFIS technique for modelling traffic accidents as a function of road and roadside characteristics. To achieve the objective, accident data and road characteristics were collected over a two-year period along the Qazvin-Loshan intercity roadway in Iran. The candidate set of explanatory variables included the Mean Horizontal Curvature (MHC), Shoulder Width (SW), Road Width (RW), Land Use (LU), Access Points (AP), Longitudinal Grade (LG), and Horizontal Curve Density (HCD). The results showed that RW, SW, LU, and AP significantly affected accident frequencies. They used statistical performance indices to compare the ANFIS model with the Poisson, negative binomial, and non-linear exponential regression models. Based on the comparative results obtained, the proposed model had higher prediction performance than the other three traditional models which has been widely used in the literature. They concluded that; the proposed model could be used as a robust approach to handle uncertainty and complexity existed in accident data. They proposed that; the ANFIS model can be an effective tool for transportation agencies since intervention decisions and plans aiming at improving road safety depend on the prediction capabilities of a system.

In research work conducted by Bangaram *et al*., (2017) on a review of road crash prediction models for developed countries; The authors affirmed that; road crash losses have been on a growing trend for the preceding decade or so in India. consequently, traffic safety organization has emerged as a topic of argument for researchers all over the world. For this reason, Crash modeling on different factors causing them was conducted. Crash modelling helps anybody to recognize the real causative agents behind an accident to occur. The effect of one cause can be greater than the other. And those causes can only be known from Crash modelling. In the reviewed paper; the authors tried to divide this Crash modelling techniques into different categories based on the road geometrics characteristics, traffic characteristics and Environmental factors on urban roads and on rural roads of different developed countries. In both urban and rural road crash studies it was seen that for the most part regression techniques like linear, multi-linear, logit and poisons regression were used for modelling the road crashes. It was also noticeable that frequently authors have tried to research on one reason and go profound into it to a certain extent considering all factors at a time. From the study of different researches, the attention was paid to the safety effects of road environment such as traffic flow, lane width, number of accesses, speed and road connectors. The paper reviewed as much papers as possible and various gaps in research along with future possibility of study in the area was indicated. Starting from the basic models like Simple/Multiple regression model to the logistic and linear regressions to the new modeling techniques involving Negative Binomial/Zero inflated modelling, genetic mining and fuzzy logics were discussed in the paper.

In the research conducted by Tatiana *et al* (2014) on maximizing accuracy and efficiency of traffic accident prediction combining information mining with computational intelligence approaches and decision trees; The authors affirmed that the development of universal methodologies for the accurate, efficient, and timely prediction of traffic accident location and severity constitutes a crucial endeavour. The reviewed paper determined the best combinations of salient accident-related parameters and accurate accident severity prediction models for the 2005 accident dataset brought together by the Republic of Cyprus Police. The optimal methodology used involves: (a) information mining in the form of feature selection of the accident parameters that maximize prediction accuracy (implemented via scatter search), followed by feature extraction (implemented via principal component analysis) and selection of the minimal number of components that contain the salient information of the original parameters, which combined to bring about an overall 74.42% reduction in the dataset dimensionality; (b) accident severity prediction via probabilistic neural networks and random forests, both of which independently accomplished over 96% correct prediction and a balanced proportion of under- and over-estimations of accident severity. The authors gave an explanation of the superiority of the optimal combinations of parameters and models when compared with the existing accident classification/prediction approaches.

In research carried out by Nachimutu and Partheeban, (2013) on development of a road accident prediction model based on system dynamics approach; the authors researched on prediction of road accidents for Chennai city using system dynamics approach. In the research, the simulated road accident prediction model was developed from the base year 2010. Chennai City road accident data was collected from 2006 to 2010 from Chennai city traffic police. Nachimutu and Partheeban, 2013 attempted to identify the various factors causing the road accidents. The road accident prediction model was developed using factors of human behaviors, vehicle factors and road factors. The system dynamics road accident prediction model was developed using STELLA software. STELLA software is a powerful tool for making a simulation model instead of stock and flow diagram, graphical interface, table and graph view, causal relational diagrams and build in functions. The main objective for the studies was to establish simple, practicable simulation road accident models that can predict the expected number of accidents from 2010 to 2020. The predicted number of accidents in 2010 was 5255 and accident for the year 2020 will be 21612. The model was also validated by comparing the predicted accident values of the years 2010, 2011 and 2012 with actual accident values.

In research work done by Jerry *et al* (2019) on Development of Accident Prediction Model on Horizontal Curves; The author affirmed that; Nowadays accidents on horizontal curves increase daily. The authors reaffirmed that; the speed reduction affect the safety of the road. To reduce accidents, an accident prediction model has to be developed. According to the author, Accident Prediction Model is made to take remedial measures in advance by studying future trends, to take mitigation measures to minimize the accident rates to certain extent and to take other safety measures. The main objectives were to identify the factors influencing road crashes and to develop a accident prediction model using SPSS software.

In the research work done by Popoola et al. (2017), on the comparison of road traffic accident prediction models for two-lane highway integrating traffic and pavement condition parameters, focused on filling the gap between the integration of traffic of pavement condition and traffic characteristics in predicting road traffic accident frequency on 2-lane highways in Nigeria. The authors compared road traffic accident frequency prediction models on IIesha-Akure-Owo road based on data observed between 2012 and 2014, making use of Negative Binomial (NB), Ordered Logistic (OL) and Zero Inflated Negative Binomial (ZINB) models to model the frequency of road traffic accident occurrence using road traffic accident data from the Federal Road Safety Commission (FRSC) and pavement conditions parameters from pavement evaluation unit of the Federal Ministry of Works, Kaduna. The explanatory variables were: annual average daily traffic (aadt), shoulder factor (sf), rut depth (rd), pavement condition index (pci), and international roughness index (iri). The explanatory variables that were statistically significant for the three models were aadt, sf and iri with the estimated coefficients having the expected signs. The number of road traffic accident on the road increased with the traffic volume and the international roughness index while it decreased with shoulder factor. The systematic variation explained by the models amounted to 87.7, 78.1 and 74.4% for NB, ZINB and OL respectively. The research findings also suggested the accident prediction models that should be integrated into pavement rehabilitation.

In the research work done by Weihong et al. (2018), on the Analysis and comparison of traffic accident regression prediction model; The authors analyzed the relationship between the number of road traffic accidents and road length, traffic conditions and other factors, while taking note of the number of road traffic accidents subject to Poisson regression, negative binomial (NB) regression and Zero Inflated Negative Binomial (NINB) regression as response variables, was used to construct a generalized linear model by introducing a joint function. The Traffic Accident Prediction Model was constructed Based on Random Forest (RF) Regression. The reviewed paper compared the defected models, and based on the predictive model, selection of the significant factors and determination of the degree of influencing factors for road traffic accidents, reducing the number of traffic accidents and improving the overall security of the road was made.

According to Oyedepo *et al.*2010, on Accident Prediction Models For Akure – Ondo Carriageway, Ondo State Southwest Nigeria; Using Multiple Linear Regressions, Accident data on the 52km Akure-Ondo Carriageway and Spot Speed data were collected and analyzed. The analysis of the Spot Speed gave an average of 51.5km/hr. and 110.75km/hr. for the 15th and 85th percentile speed respectively, the 85th percentile speed was higher than the posted speed limit of 60km/hr. which indicated that the vehicle travelling at that speed was susceptible to accident. The analysis for the study area also showed that 759 people were involved in the accidents, 108 persons were killed and about 348 persons were injured between 2002-2007, 38% of the accident were fatal accident and 62% non-fatal accident, However, the regression analysis carried out on the accident data with number of accident as the dependent variables and number of people killed in the accident(X1), number of people injure(X2), number of people involved in the accident(X3) as independent variables; gave a coefficient of correlation "R" value of 70.70% and coefficient of determination "R2"of 49.70% respectively. Factors such as driver's behavior, poorly maintained vehicles, non-adherence to traffic rules, poorly maintained road and verges, and over-speeding were causatives factors to road accident. However, it suggested that the authorities concerned for mitigation of road accident could make use of road safety plans and road safety audit as effective strategies.

This study will help to improve/control and reduce drastically road accidents, because it helps to predict road accident using the two best fit models. These two models MLR (Multiple Linear Regression) and ANN (Artificial Neural Network) has been proven to be able to calculate and predict road accidents better than the FL (Fuzzy Logic). This study will help improve the safety of lives and properties on our highway.

CHAPTER THREE
RESEARCH METHODOLOGY

3.1 Description of study area

The study area is limited to 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 (Okhakhu, 2010).

The Benin City hydrological basin is partitioned into two main units. The first unit consists of the Ikpoba River Basin which drains the whole eastern part of the city while the second unit covers the Ogba River Basin which drains the western part. Although smaller rivers are found in some parts of the peripheral locations of the study area, in general, the hydrological basin clearly shows a north-south direction of flow owing to the high elevation of Nigeria from its northern part (Okhakhu, 2014). Fishing, irrigation, domestic consumption, industrial utilization, animal husbandry, recreation, environmental sanitation and research activities are some of the functions of rivers found in Benin City.

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 (Okhakhu, 2010). 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 showing the sampling zone is presented in Figure 3.1

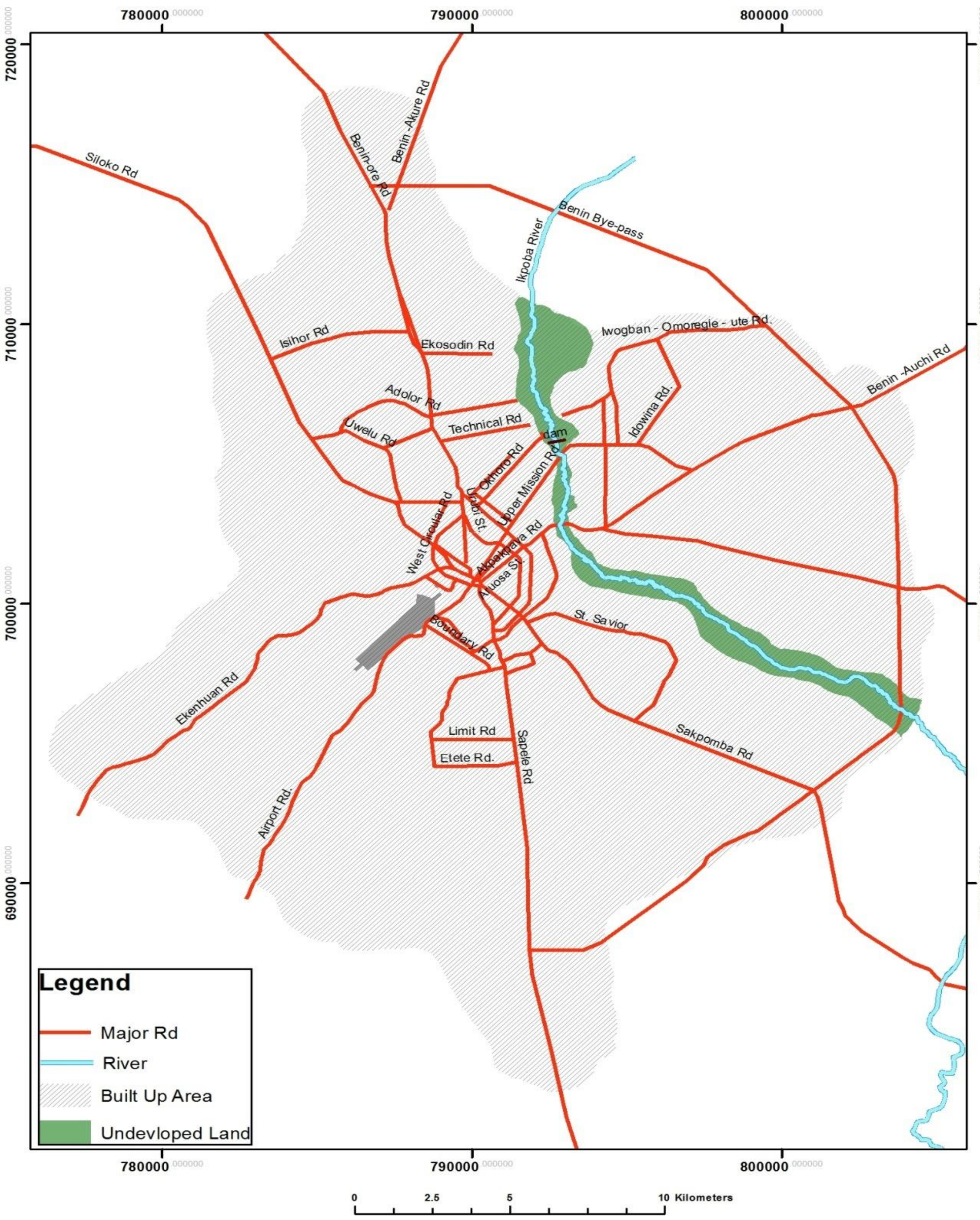


Figure 3.1: Base map of study area (Adapted from Google Earth)

3.2 Data collection

Two different types of data are employed in this study. The primary data which include; road accident data were collected from Federal Road Safety Office in Benin City while the secondary data which include; traffic volume and the geometric data were collected from the field. For a robust field data, a reconnaissance surveys was 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.

3.2.1 Prioritization of primary Data

Prioritization involves assigning suitable weights to different factors so as to achieve a desired result. In this model, the various factors, which tend to influence the occurrence of accidents on roads are assigned weighs on a scale of 0-10 in such a manner that the factors which tend to increase the probability of the accidents have lower weights. In order to prioritize roads for occurrence of accidents, the various factors are considered and the weights assigned to them.

3.2.1.1 Characterization of road geometry

To characterize the geometry of the road under study, the following information were collected.

1. Assessment of the presence of speed bumps.
2. Assessment of the presence of walkway.
3. Assessment of the presence of a shoulder.
4. Assessment of curves on the road.
5. Assessment of the presence of a median between the two carriage way.

To assess the presence of the above information concerning the road, a reconnaissance survey was done and relevant road geometry were measured and recorded.

3.3 Preliminary Analysis of Data

Preliminary analysis of the data includes:

i. Descriptive statistics of the data
ii. Detection of outliers
iii. Test of reliability of data
iv. Test of homogeneity
v. Test of normality
vi. Test of serial correlation

3.3.1 Descriptive statistics

Computed descriptive statistics include;

3.3.1.1 Skewness

Mathematically, the skewness of a distribution is defined as shown in equation 3.1:

$$y = \frac{E(x-\mu)^3}{\sigma^3} \tag{3.1}$$

where:

μ is the mean of x

σ is the standard deviation of x

E (t) is the expected value of the quantity x; SES is standard error of skewness

3.3.1.2 Kurtosis

A value of kurtosis significantly greater than 0 indicates that the variable has longer tails than those for a normal distribution; less than 0 indicates that the distribution is flatter than a normal distribution. A kurtosis coefficient is considered significant if the absolute value of (KURTOSIS / SEK) is greater than 2. Mathematically, the kurtosis of a distribution is defined as shown in equation 3.2 (Ramirez et al 2005):

$$k = \frac{E(x-\mu)^4}{\sigma^4} \tag{3.2}$$

Where:

μ is the mean of x

σ is the standard deviation of x

E (t) is the expected value of the quantity x

SEK is the standard error of kurtosis

3.3.1.3 Coefficient of Variation

The coefficient of variation is the standard deviation divided by the sample mean. Mathematically, the coefficient of variation (CV) of a distribution is defined as:

$$CV = \left(\frac{\sigma}{\mu}\right) \tag{3.3}$$

Where:

μ is the mean of x and σ is the standard deviation of x

3.3.2 Detection of outliers using the labeling rule

Although, the presence of outlier can be visualized using the histogram plot. In this study, the labeling rule method is employed to detect the presence of outliers. The labeling rule is the statistical method of detecting the presence of outliers in data sets using the 25th percentile (lower bound) and the 75th percentile (upper bound). The underlying mathematical equation based on the lower and the upper bound is presented as follows:

$$\textit{Lower Bound}\ \mathrm{Q}_1 - (2.2 \times (Q_3 - Q_1) \tag{3.4}$$

$$\textit{Upper}\ \text{Bound}\ \mathrm{Q}_3 + (2.2 \times (Q_3 - Q_1) \tag{3.5}$$

Q1 is the lower bound,

Q3 is the upper bound.

At 0.05 degree of freedom, any data lower than Q_1 or greater than Q_3 will be considered an outlier and needed to be removed before further analysis (Levi et al., 2009).

3.3.3 Reliability Analysis of the Data

Reliability analysis was done to ascertain the fitness of the data for the selected analysis. Figure 3.2 shows the statistical platform for reliability analysis.

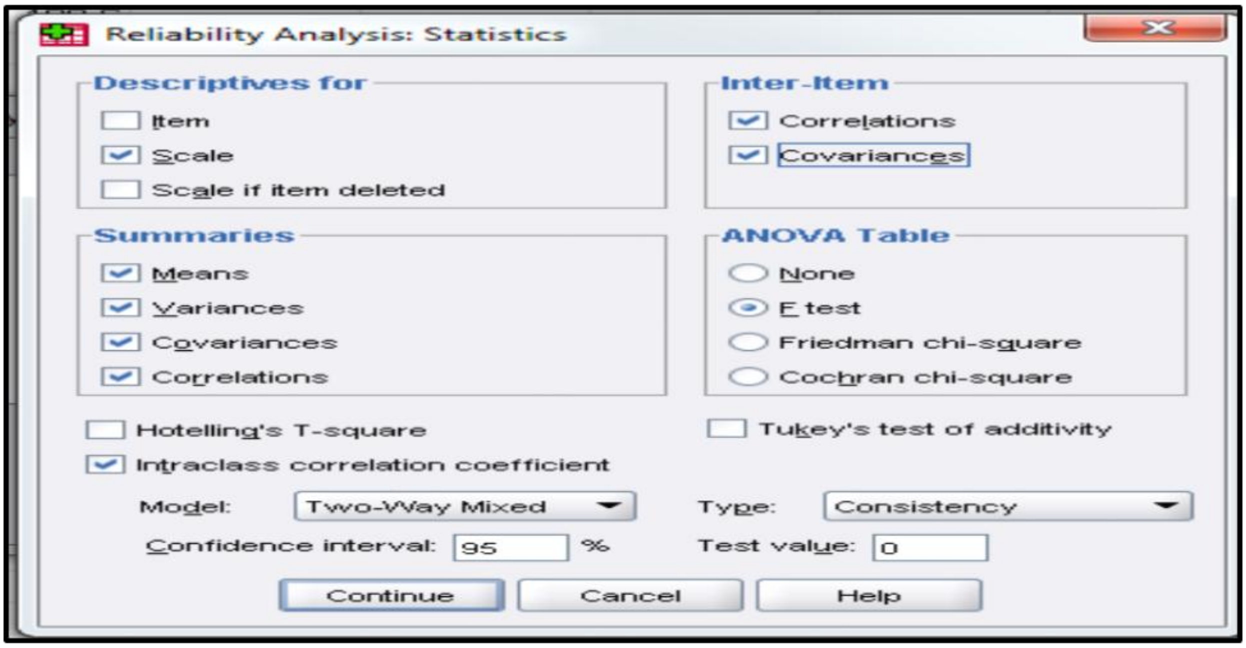


Figure 3.2: Reliability analysis platform

The null hypothesis of reliability is 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.

3.3.4 Test of Homogeneity

Frequency analysis of data requires that the data be homogeneous and independent. Homogeneity test was conducted to establish the fact that the data used for the analysis are from the same population. Homogeneity test is based on the cumulative deviation from the mean as expressed using the mathematical equation below (Raes et al., 2006).

$$S_k = \sum_{i=1}^{k} \left(X_i - \bar{X} \right) \quad k = 1, \text{----------} \; n \qquad (3.6)$$

where

X_i = The record for the series X_1 X_2 ----------- X_n

$\bar{X}$ = The mean

S_{ks} = the residual mass curve

For a homogeneous record, one may expect that the S_{ks} fluctuate around the zero-center line in the residual mass curve since there is no systematic pattern in the deviation X_i's from the average values $\bar{X}$. To perform the homogeneity test, a software package (Rainbow) for analyzing time series data was employed (Raes et al., 2006).

3.3.5: 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"

$$JB = n[(\sqrt{b_1})^2/6 + (b_2 - 3)^2/24] \qquad (3.7)$$

Where:

n is the sample size

√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.

3.4 Diagnostic Analysis of Data

It is pertinent to note that standard error estimation and computation of t-statistics are appropriate in calculating the probability (p-value) by which you test the significance of the regression model. In the presence of heteroskedasticity, it is assumed that the overall standard error of regression and the t-statistics computed for each independent variable may not be completely adequate to estimate the resulting probability (p-value) of regression. In addition, the presence of serial correlation can lead to a number of issues, namely;

i. Make reported standard error and t-statistics to be invalid

ii. Coefficient may be biased, though not necessarily inconsistent

Based on this argument, selected diagnostic statistics were conducted to verify the statistical properties of the overall regression model. The selected diagnostic statistics include;

i. Heteroskedasticity test using Breusch-Pagan Godfrey

ii. Serial Correlation test using Breusch Godfrey

iii. Variance Inflation Factor (VIF)

3.4.1 Heteroskedasticity Test

Heteroskedasticity is a diagnostic test statistic use to diagnose the adequacy of the probability (p-value) calculated for each individual variable. Hence it is important to know whether there is or there isn't heteroskedasticity in our data. The null and alternate hypothesis of heteroskedasticity was formulated as follows;

H0 = Presence of homoscedasticity

H1 = Absence of heteroskedasticity

H0 = Absence of homoscedasticity

H1 = Presence of heteroskedasticity

For p-value < 0.05 you reject the null hypothesis of homoskedasticity and conclude that there is no heteroskedasticity. For p-value > 0.05 you accept the null hypothesis of homoskedasticity and conclude that there is the presence of heteroskedasticity.

3.4.2 Serial Correlation Test

Serial correlation is a common occurrence in time series data because the data is ordered (overtime). It is therefore not surprising that neighbouring error terms turn out to be correlated. Serial correlation violates the standard assumption of regression theory that error terms are uncorrelated. The null and alternate hypothesis of serial correlation is formulated as follows;

H0 = Absence of serial correlation

H1 = Presence of serial correlation

Serial correlation analysis will be conducted using Breusch Godfrey test.

3.4.3 Variance inflation factor

Variance inflation factor (VIF) measures the correlation of the dependent variable with the independent variables. Ideal VIF is 1; VIF greater than 10 is cause for alarm showing the variables are uncorrelated due to multicollinearity.

3.5 Rate of Accident Prediction

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

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

3.5.1 Development of Multiple Linear Regression Equation

To apply multiple linear regression models, the independent variables that influences accident rate must first be known.

To ascertain the dependence of the selected independent variables on the dependent variable, multiple linear regression models will be applied to generate a regression equation of the form:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ---------- \beta_5 x_5 + \varepsilon \qquad (3.8)$$

where;

X_1, X_2 ---------- X_n = the selected independent variables

Y = the dependent variable (Rate of accident),

β_0 , β_1 are the regression constant;

Ɛ is the deviation.

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

3.5.1.1 Application of multiple linear regression equation to accident prediction

To apply the regression equation for the prediction of accident rate, a simple to use Microsoft Excel program is developed using the regression equation. Thereafter, arbitrary values were assigned to all the independent variables in order to predict the dependent variable.

3.5.2 Prediction of Accident Rate 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 is 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

3.5.2.1 Normalization of Input and Output Data

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 Sinan *et al*., 2011.

$$x_i = \frac{x - x_{\min}}{x_{\max} - x_{\min}} + 0.1 \qquad (3.9)$$

Where

x_i is the normalized value of the input and output data

x_{min} and x_{max} = the minimum and maximum value of the input and output data

3.5.2.2 Selection of training algorithm and hidden neurons

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) (Kin *et al*. 2001).

3.5.2.3 Network training/performance of ANN

To train the network, 3 runs of 1000 epochs, each with a precision rate of 0.00001 and a learning rate of 0.05 is used. In addition, cross validation data representing about 25% of the total input data was introduced to monitor the progress of training and prevent the network from memorizing the input data instead of leaning which is a common problem associated with overtraining (Kin et al 2001). The progress of the training was monitored using the mean square error (MSE) graph for training and cross validation

3.5.2.4 Network testing

To test the efficiency of the trained network, 15% of the input data was introduced to the network to generate the predicted accident rate

3.5.2.5 Reliability of trained Network

To test the reliability of the network and acertain the prediction accuracy, a reliability plot of output using the predicted accident rate as the vertical axis and the observed accident rate as the horizontal axis was obtained with the aid of Microsoft Excel Spreadsheet. Reliability of the network is evaluated using the value of the coefficient of determination (r^2) between the predicted and the observed accident rate.

3.6 Comparism of ANN, MLR .

To compare the performance of artficial neural network (ANN), 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.

CHAPTER FOUR
RESULTS AND DISCUSSION

4.1 Geometry features of Ugbowo Benin-Ore road

Result of the geometry features of the road under study is presented in the table 4.1a

Table 4.1a: The Geometric Features Along Ugbowo Benin-Ore Road

Chainage KM	Vertical curve %	Horizontal curve (M)	Super elevation %
11.5-13.0	12.37	2440.54	4.29
24.7-39.3	8.67	0	0
59.5-62.3	2.57	3642.45	1.49
74.0-76.6	0	3290.24	1.19
84.0-85.0	0	1022.94	0.39
86.0-87.0	0	5087.87	1.29
90.0-90.5	0	904.26	0.49

The geometry features of Ugbowo Benin-Ore road such as design speed, road width, median and shoulder were 100km/hr, 10.5m, and 1.5m respectively. The length of the road is 94 km and the AADT (annual average daily traffic) is 1850. In order to attain the primary goal of road transportation, road designers and the 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 behavior, which is expressed by the 85th percentile speed of passenger cars under free-flow conditions. Road curve radius is a primary element of road geometric design that is associated with horizontal curve design and it is related to trustier accidents as the smaller curve radius, the higher the possibility of accident to occur on the roads. Based on the theory of vehicles steering, Vehicles transverse stability which includes the slippage and overturns determines the curve radius values to be selected when designing road horizontal curves. Super elevation is defined as the transverse slope which is designed as higher on the outer side and lower on the inner side. The super elevation is for the purpose of aching against a centrifugal force that influences the vehicles running on a curve and also to enhance the stability of traveling vehicles and comfort of drivers. According to Zhang Yingxue, (2009) the super elevation traverse slope should be between 2.0% to 3.0%. Super elevation is a function of roads horizontal alignment, design speed, natural conditions and curve radius. The use of proper super elevation value can mitigate incidents which could contribute to accident severity. Some studies have indicated that super elevation and horizontal alignment have an influence on trashier safety on roads.

Generally speaking, freeway flow speeds have higher values on lanes with ideal widths, therefore, incidents are less possible, which is also stated by (HCM) especially for multilane freeway (Jerry et al., 2009). While, it might be rationally presumed that wider lanes reduce the effect of incident generating from drivers moistures, it can be argued that high operating speeds can oppose this effects. The use of an optimum lane width value of typically 3.5m to 3.6m was suggested by most researchers. Shoulders are used as a free space to allow vehicles on the road to stop out of the main road traffic flow whether for vehicle break-down incidences, emergences or as a section of road night of way. In case of control loss over the vehicle, shoulders are used as back up to take back vehicle control (US Department of Transportation 2007). More space provided by road shoulders allow for high free floe speed because drivers may have a perception that when they lose control over their vehicles there is a room for gaining control again. Several studies agreed on shoulder width impact on accident rates on roadway. Studies carried out had different findings on the effect of shoulder width on accident occurrence. Zeyeer et al., 1981), stated that wider shoulders result in lower accident rates and found that a decline up to a abut 20% of incidence occurring is attributed by shoulder width 0.9m-2.7m and theory suggested that the optimum road shoulder width should be 1.5m.

4.1 Comparison of Field Data with Previous Work

The field data obtained which include vertical curve (%), horizontal curve (M), and super elevation for Benin Ore Road was validated by comparing with the works of Osasu et al, 2020 and presented in Table 4.1b

Table 4.1b: comparison of geometric features along Ugbowo Benin-Ore road

Chainage	Vertical curve %		Horizontal curve (M)		Super Elevation %	
	Field Data	**Reference Data**	**Field Data**	**Reference Data**	**Field Data**	**Reference Data**
11.5-13.0	12.37	12.4	2440.54	2440.56	4.29	4.3
24.7-39.3	8.67	8.7	0	0	0	0
59.5-62.3	2.57	2.6	3642. 45	3642. 47	1. 49	1.5
74.0-76.6	0	0	3290.24	3290.26	1.19	1.2
84.0-85.0	0	0	1022.94	1022.96	0.39	0. 4
86.0-87.0	0	0	5087.87	5087.89	1.29	1.3
90.0-90.5	0	0	904.26	904.28	0. 49	0.5

4.2 Preliminary Analysis of Data

Accident data; which include the number of accident cases (NAC), number of vehicles involved (NVI), number of persons involved (NPI), number of persons injured (NPIJ) and number of persons killed (NPK) alongside the geometric features of the road which is the presence of a curve and shoulder are presented in appendix while the descriptive statistics of the data is presented in section 4.2.1

4.2.1 Descriptive Statistics

Descriptive statistics of the data employed for the analysis is presented in Table 4.1

Table 4.2: Descriptive statistics of accident data

Variable	Observations	Obs. with missing data	Obs. without missing data	Minimum	Maximum	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

Based on the results of Table 4.2, it was observed that the minimum cases of accident between 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.

4.2.2 Reliability Analysis

The correlation matrix of regression which shows how the individual variables relates to the others is presented in Table 4.3

Table 4.3 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 4.3 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 4.4

Table 4.4: 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 4.4, it was concluded that the model is significant, hence the Cronbach alpha value for assessing reliability was calculated and presented in Table 4.5

Table 4.5: 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 4.5, it was concluded that the accident data are reliable. Finally, the goodness of fit statistic of reliability were calculated and presented in Table 4.5

Table 4.6: 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 4.6 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.

4.2.3 Outlier Analysis

4.2.3.1 Dixon test for outliers / Two-tailed test (No. of accident cases):

Using 1000000 Monte Carlo Simulation at 99% confidence interval, the Dixon test for outlier was conducted on the number of accident cases (NAC) and the result obtained is presented in Table 4.7

Table 4.7: Dixon test for outlier on NAC

R10 (Observed value)	0.036
R10 (Critical value)	0.244
p-value (Two-tailed)	0.628
Alpha	0.05

Test interpretation:

H0: There is no outlier in the data

Ha: The minimum or maximum value is an outlier

As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0. Hence we accept the null hypothesis and concluded that there are no outliers in the number of accident cases. The plot of number of accident cases against the Dixon calculated Z-scores is presented in Figure 4.1

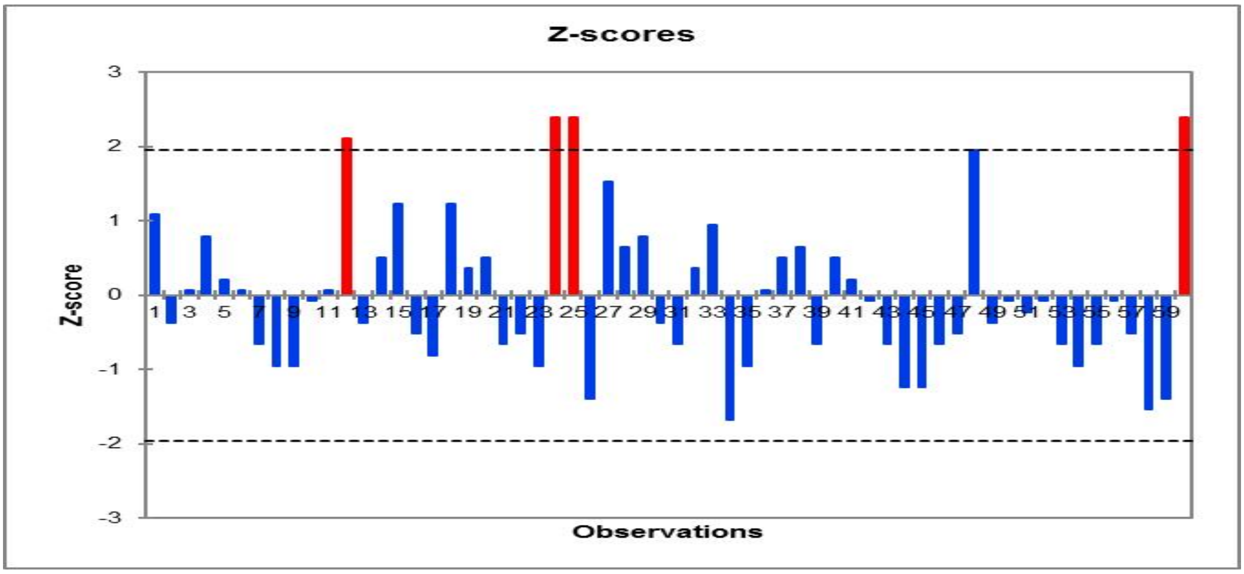


Figure 4.1: NAC versus computed Z-scores

The fluctuation of the data points around the zero center point of the Z-scores mass curve as observed in Figure 4.1 indicates that the data being tested are devoid of possible outliers.

4.2.3.2 Dixon test for outliers / Two-tailed test (No. of persons involved):

Using 1000000 Monte Carlo Simulation at 99% confidence interval, the Dixon test for outlier was conducted on the number of persons involved in accident cases (NPIV) and result obtained is presented in Table 4.8

Table 4.8: Dixon test for outlier on NPIV

R10 (Observed value)	0.039
R10 (Critical value)	0.244
p-value (Two-tailed)	0.673
Alpha	0.05

Test interpretation:

H0: There is no outlier in the data

Ha: The minimum or maximum value is an outlier

As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0. Hence we accept the null hypothesis and concluded that there are no outliers in the number of persons involved. The plot of number of persons involved in accident cases against the Dixon calculated Z-scores is presented in Figure 4.2

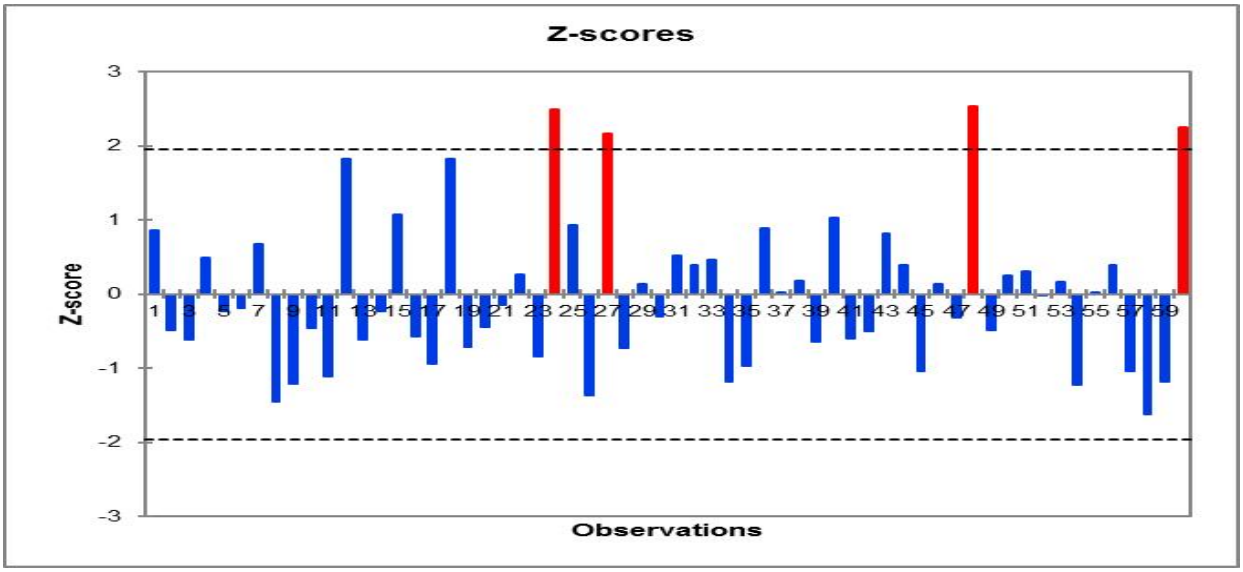


Figure 4.2: NPIV versus computed Z-scores

The fluctuation of the data points around the zero center point of the Z-scores mass curve as observed in Figures 4.2 indicates that the data being tested are devoid of possible outliers.

4.2.3.3 Dixon test for outliers / Two-tailed test (No. of persons injured):

Using 1000000 Monte Carlo Simulation at 99% confidence interval, the result of Dixon test for outlier which was conducted on the number of persons injured in the accident cases (NPIJ) is presented in Table 4.9

Table 4.9: Dixon test for outlier on NPIJ

R10 (Observed value)	0.274
R10 (Critical value)	0.244
p-value (Two-tailed)	0.725
Alpha	0.05

Test interpretation:

H0: There is no outlier in the data

Ha: The minimum or maximum value is an outlier

As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0. Hence we accept the null hypothesis and concluded that there are no outliers in the number of persons injured. The plot of number of persons injured in accident cases against the Dixon calculated Z-scores is presented in Figure 4.3

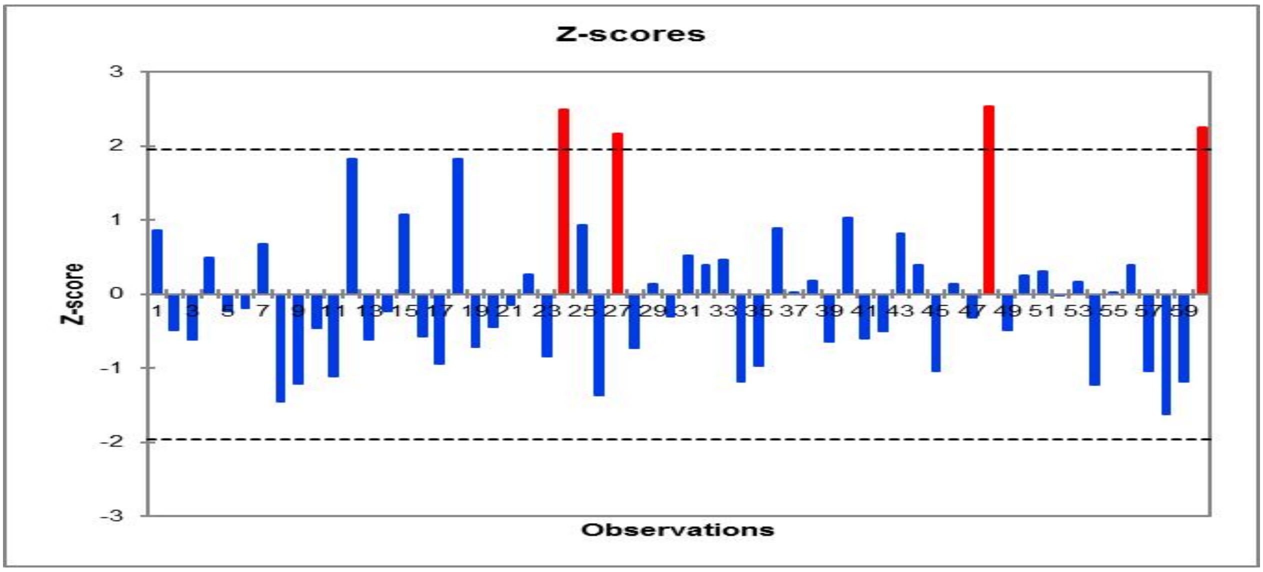


Figure 4.3: NPIJ versus computed Z-scores

The fluctuation of the data points around the zero center point of the Z-scores mass curve as observed in Figures 4.3 indicates that the data being tested are devoid of possible outliers.

4.2.3.4 Dixon test for outliers / Two-tailed test (No. of persons killed):

Using 1000000 Monte Carlo Simulation at 99% confidence interval, the Dixon test for outlier was conducted on the number of persons killed in the accident cases (NPK) and result obtained is presented in Table 4.9

Table 4.10: Dixon test for outlier on NPK

R10 (Observed value)	0.129
R10 (Critical value)	0.244
p-value (Two-tailed)	0.609
Alpha	0.05

Test interpretation:

H0: There is no outlier in the data

Ha: The minimum or maximum value is an outlier

As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0. Hence we accept the null hypothesis and concluded that there are no outliers in the number of persons killed. The plot of number of persons killed in accident cases against the Dixon calculated Z-scores is presented in Figure 4.4

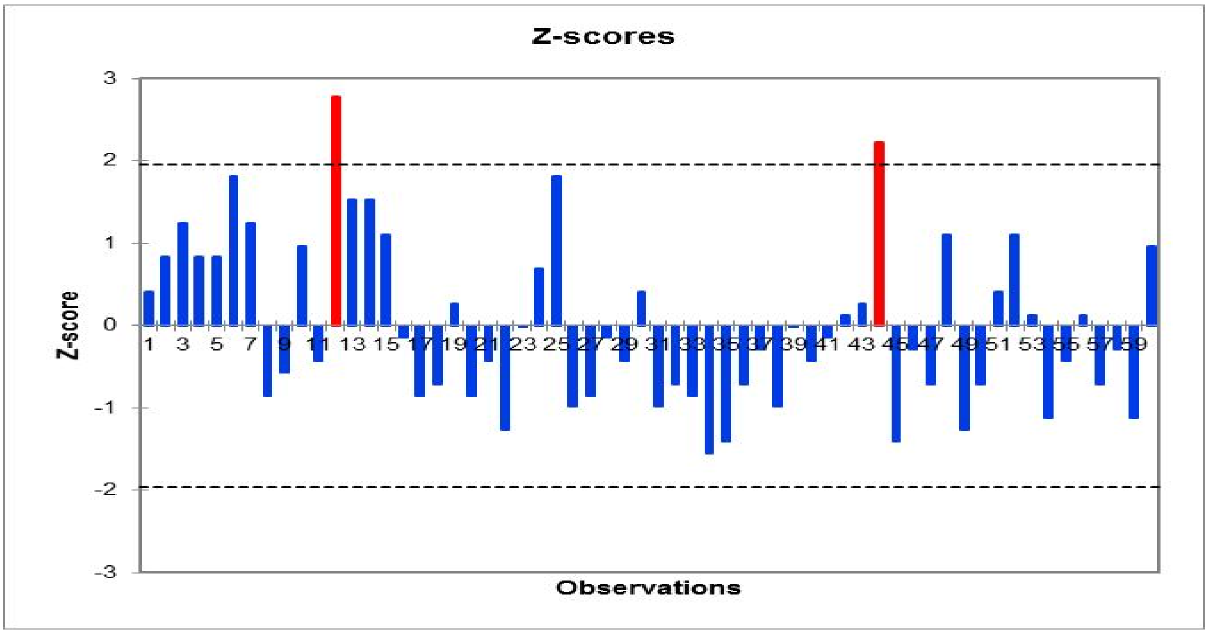


Figure 4.4: NPK versus computed Z-scores

The fluctuation of the data points around the zero center point of the Z-scores mass curve as observed in Figures 4.4 indicates that the data being tested are devoid of possible outliers.

4.2.3.5 Dixon test for outliers / Two-tailed test (No. of vehicles involved):

Using 1000000 Monte Carlo Simulation at 99% confidence interval, the Dixon test for outlier was conducted on the number of vehicle involved in accident cases (NVI) and result obtained is presented in Table 4.10

Table 4.11: Dixon test for outlier on NVI

R10 (Observed value)	0.077
R10 (Critical value)	0.244
p-value (Two-tailed)	0.840
Alpha	0.05

Test interpretation:

H0: There is no outlier in the data

Ha: The minimum or maximum value is an outlier

As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0. Hence we accept the null hypothesis and concluded that there are no outliers in the number of vehicles involved. The plot of number of vehicles involved against the Dixon calculated Z-scores is presented in Figure 4.5

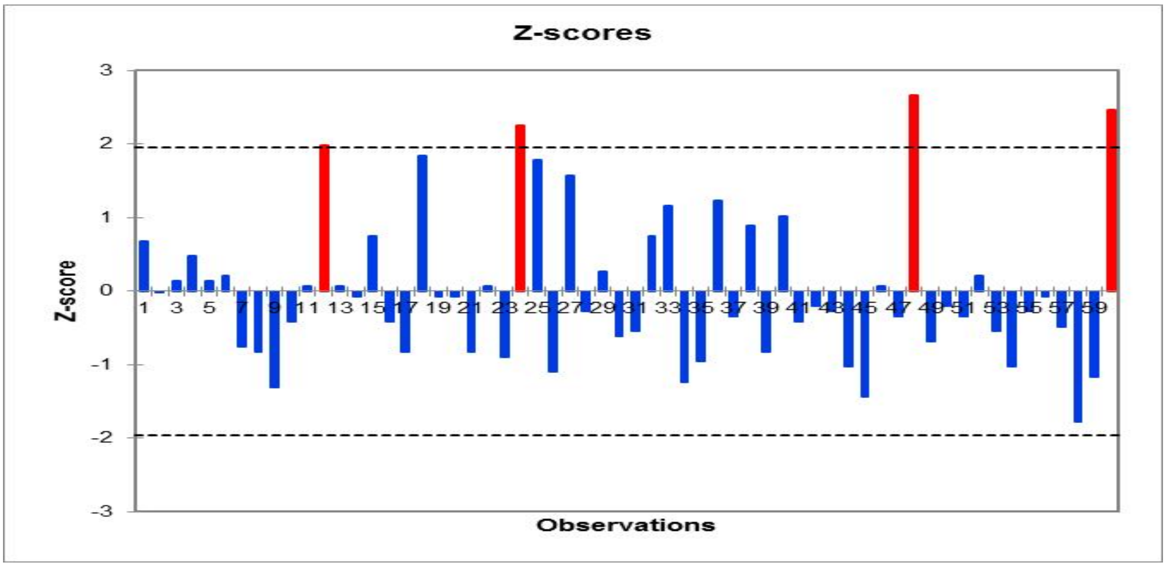


Figure 4.5: NVI versus computed Z-scores

The fluctuation of the data points around the zero center point of the Z-scores mass curve as observed in Figures 4.5 indicates that the data being tested are devoid of possible outliers.

4.2.4 Test of Normality

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, two methods were employed and they are; Lilliefors test and Jarque-Bera test. Results of the test are presented as follows;

4.2.4.1 Normality test for No. of accident cases

Table 4.12a: Lilliefors test (No. of accident cases):

D	0.112
D (standardized)	0.864
p-value (Two-tailed)	0.061
Alpha	0.05

Test interpretation:

H0: The variable from which the sample was extracted follows a Normal distribution.

H1: The variable from which the sample was extracted does not follow a Normal distribution.

As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0.

Table 4.12b: Jarque-Bera test (No. of accident cases):

JB (Observed value)	5.698
JB (Critical value)	5.991
DF	2
p-value (Two-tailed)	0.058
Alpha	0.05

Test interpretation:

H0: The variable from which the sample was extracted follows a Normal distribution.

H1: The variable from which the sample was extracted does not follow a Normal distribution.

As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0.

4.2.4.2 Normality test for No. of persons involved

Table 4.13a: Lilliefors test (No. of persons involved):

D	0.087
D (standardized)	0.674
p-value (Two-tailed)	0.312
Alpha	0.05

Test interpretation:

H0: The variable from which the sample was extracted follows a Normal distribution.

H1: The variable from which the sample was extracted does not follow a Normal distribution.

As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0.

Table 4.13b: Jarque-Bera test (No. of persons involved):

JB (Observed value)	6.312
JB (Critical value)	5.991
DF	2
p-value (Two-tailed)	0.063
Alpha	0.05

Test interpretation:

H0: The variable from which the sample was extracted follows a Normal distribution.

H1: The variable from which the sample was extracted does not follow a Normal distribution.

As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0.

4.2.4.3 Normality test for No. of persons injured

Table 4.14a: Lilliefors test (No. of person injured):

D	0.086
D (standardized)	0.668
p-value (Two-tailed)	0.325
Alpha	0.05

Test interpretation:

H0: The variable from which the sample was extracted follows a Normal distribution.

H1: The variable from which the sample was extracted does not follow a Normal distribution.

As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0.

Table 4.14b: Jarque-Bera test (No. of person injured):

JB (Observed value)	16.333
JB (Critical value)	5.991
DF	2
p-value (Two-tailed)	0.065
Alpha	0.05

Test interpretation:

H0: The variable from which the sample was extracted follows a Normal distribution.

H1: The variable from which the sample was extracted does not follow a Normal distribution.

As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0.

4.2.4.4 Normality test for No. of persons killed

Table 4.15a: Lilliefors test (No. of persons killed):

D	0.117
D (standardized)	0.904
p-value (Two-tailed)	0.071
Alpha	0.05

Test interpretation:

H0: The variable from which the sample was extracted follows a Normal distribution.

H1: The variable from which the sample was extracted does not follow a Normal distribution.

As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0.

Table 4.15b: Jarque-Bera test No. of persons killed):

JB (Observed value)	4.467
JB (Critical value)	5.991
DF	2
p-value (Two-tailed)	0.107
Alpha	0.05

Test interpretation:

H0: The variable from which the sample was extracted follows a Normal distribution.

H1: The variable from which the sample was extracted does not follow a Normal distribution.

As the computed p-value is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H0.

4.2.4.5 Normality test for No. of vehicles involved

Table 4.16a: Lilliefors test (No. of vehicles involved):

D	0.153
D (standardized)	1.188
p-value (Two-tailed)	0.001
Alpha	0.05

Test interpretation:

H0: The variable from which the sample was extracted follows a Normal distribution.

H1: The variable from which the sample was extracted does not follow a Normal distribution.

As the computed p-value is lower than the significance level alpha=0.05, one should reject the null hypothesis H0, and accept the alternative hypothesis H1.

Table 4.16b: Jarque-Bera test (No. of vehicles involved):

JB (Observed value)	8.079
JB (Critical value)	5.991
DF	2
p-value (Two-tailed)	0.018
Alpha	0.05

Test interpretation:

H0: The variable from which the sample was extracted follows a Normal distribution.

H1: The variable from which the sample was extracted does not follow a Normal distribution.

As the computed p-value is lower than the significance level alpha=0.05, one should reject the null hypothesis H0, and accept the alternative hypothesis H1.

4.2.5: Diagnostic Test

It is pertinent to note that standard error estimation and computation of t-statistics are appropriate in calculating the probability (p-value) by which you test the significance of the regression model. In the presence of heteroskedasticity, it is assumed that the overall standard error of regression and the t-statistics computed for each independent variable may not be completely adequate to estimate the resulting probability (p-value) of regression. In addition, the presence of serial correlation can lead to a number of issues, namely;

i. Make reported standard error and t-statistics to be invalid

ii. Coefficient may be biased, though not necessarily inconsistent

Based on this argument, selected diagnostic statistics were conducted to verify the statistical properties of the overall regression model. The selected diagnostic statistics include;

iii. Heteroskedasticity test using Breusch-Pagan Godfrey

iv. Serial Correlation test using Breusch Godfrey

v. Variance Inflation Factor (VIF)

4.2.5.1 Heteroskedasticity Test

Heteroskedasticity is a diagnostic test statistic use to diagnose the adequacy of the probability (p-value) calculated for each individual variable. Hence it is important to know whether there is or there isn't heteroskedasticity in our data. The null and alternate hypothesis of heteroskedasticity was formulated as follows;

For p-value < 0.05 reject H0

H0 = Presence of homoskedasticity

H1 = Absence of heteroskedasticity

For p-value > 0.05 accept H0

H0 = Absence of homoscedasticity

H1 = Presence of heteroskedasticity

For p-value < 0.05 you reject the null hypothesis of homoskedasticity and conclude that there is no heteroskedasticity. For p-value > 0.05 you accept the null hypothesis of homoskedasticity and conclude that there is the presence of heteroskedasticity.

Result of heteroskedasticity test using Breusch-Pagan Godfrey is presented in Table 4.16

Table 4.17: Result of Heteroskedasticity Test

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

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

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	1.215502	Prob. F(4,55)	0.3148
Obs*R-squared	4.873214	Prob. Chi-Square(4)	0.3006
Scaled explained SS	4.568656	Prob. Chi-Square(4)	0.3345

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 07/06/21 Time: 10:59
Sample: 1 60
Included observations: 60

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.723035	3.488870	0.207241	0.8366
NO__OF_VEHICLES_INVOLVED	0.274056	0.177194	1.546645	0.1277
NO__OF_PERSON_INJURED	0.018019	0.063115	0.285495	0.7763
NO__OF_PERSONS_INVOLVED	-0.030136	0.032840	-0.917647	0.3628
NO__OF_PERSONS_KILLED	-0.045960	0.184771	-0.248739	0.8045

R-squared	0.081220	Mean dependent var	6.184110
Adjusted R-squared	0.014400	S.D. dependent var	9.315730
S.E. of regression	9.248414	Akaike info criterion	7.366436
Sum squared resid	4704.324	Schwarz criterion	7.540965
Log likelihood	-215.9931	Hannan-Quinn criter.	7.434704
F-statistic	1.215502	Durbin-Watson stat	1.614077
Prob(F-statistic)	0.314771		

From the result of Table 4.17 it was observed that;

i. The calculated (p-value) based on the F-statistics is 0.3148

ii. The calculated (p-value) based on langrange multiplier (LM) is 0.3006

Since the computed (p-value) based on F-statistics and langrange multiplier is greater than 0.05 ($P > 0.05$), we accept the null hypothesis of homoskedasticity and conclude that there is heteroskedasticity in the data. The implication is that linear regression may not be the best model to assess the relationship between the accident data.

4.2.5.2: Serial Correlation Test

Serial correlation is a common occurrence in time series data because the data is ordered (overtime). It is therefore not surprising that neighbouring error terms turn out to be correlated. Serial correlation violates the standard assumption of regression theory that error terms are uncorrelated. The null and alternate hypothesis of serial correlation was formulated as follows;

H0 = Absence of serial correlation

H1 = Presence of serial correlation

Result of serial correlation using Breusch Godfrey is presented in Table 4.18

Table 4.18: Result of Serial Correlation Test

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

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

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.020058	Prob. F(2,53)	0.9801
Obs*R-squared	0.045379	Prob. Chi-Square(2)	0.9776

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 07/06/21 Time: 11:09
Sample: 1 60
Included observations: 60
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.020344	1.015356	-0.020037	0.9841
NO__OF_VEHICLES_INVOLVED	0.003054	0.052948	0.057671	0.9542
NO__OF_PERSON_INJURED	0.000188	0.018904	0.009944	0.9921
NO__OF_PERSONS_INVOLVED	-0.000542	0.009839	-0.055040	0.9563
NO__OF_PERSONS_KILLED	-0.000819	0.053348	-0.015344	0.9878
RESID(-1)	-0.024849	0.141825	-0.175212	0.8616
RESID(-2)	0.014926	0.151505	0.098517	0.9219

R-squared	0.000756	Mean dependent var	2.04E-16
Adjusted R-squared	-0.112366	S.D. dependent var	2.507773
S.E. of regression	2.644917	Akaike info criterion	4.892437
Sum squared resid	370.7660	Schwarz criterion	5.136777
Log likelihood	-139.7731	Hannan-Quinn criter.	4.988012
F-statistic	0.006686	Durbin-Watson stat	1.988191
Prob(F-statistic)	0.999999		

From the result of Table 4.18, it was observed that;

i. The calculated (p-value) based on the F-statistics is 0.9801

ii. The calculated (p-value) based on langrange multiplier (LM) is 0.9776

Since the computed (p-value) based on F-statistics and langrange multiplier is greater than 0.05 ($P > 0.05$), we accept the null hypothesis of serial correlation and concluded that there is no serial correlation in the data

4.2.5.3: Variance Inflation Factor (VIF)

Variance inflation factor (VIF) measures the correlation of the dependent variable with the independent variables. Ideal VIF is 1; VIF greater than 10 is cause for alarm showing the variables are uncorrelated due to multicollinearity. Result of the calculated VIF for the selected variables is presented in Table 4.19

Table 4.19: Calculated variance inflation factors

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

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

Variance Inflation Factors
Date: 07/06/21 Time: 11:14
Sample: 1 60
Included observations: 60

Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	0.960066	8.538595	NA
NO__OF_VEHICLES...	0.002476	34.85694	4.623334
NO__OF_PERSON_I...	0.000314	15.29861	2.698573
NO__OF_PERSONS...	8.51E-05	27.55772	4.145661
NO__OF_PERSONS...	0.002693	5.305859	1.206455

Since the computed variance inflation factors (centered VIF) for the selected independent variables are less than 10, it was concluded that the variables are well correlated with the dependent variable, hence absence of multicollinearity.

4.3 Analysis of Accident Data using Linear Regression

The dependence of the dependent variable on the selected independent variables was evaluated using the coded least square regression equation presented as shown in Equation 4.1;

NAC C NVI NPI NPIJ NPK (4.1)

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 4.20

Table 4.20: 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 4.20, 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 4.20, the overall linear regression equation was thereafter generated and presented as follows;

$$NAC = 2.8958 + 0.4034(NVI) + 0.0346(NPIJ) - 0.0102(NPI) + 0.0897(NPK) \qquad (4.2)$$

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.

4.4 Analysis of Accident Data using Artificial Neural Network (ANN)

Five (5) years monthly accident data was employed for this analysis. The data range was between 2014 to 2018 resulting to sixty (60) data. The classification of the data into input and output variables is presented in Table 4.21

Table 4.21: 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

4.4.1 Normalization of data

Normalization of the input and output data was done to reduce the effect of weight variation that may subsequently result to overtraining. The whole idea was to reduce the weight of the input and output data to between zero to one (0 to 1).

4.3.2 Selection of training algorithm and hidden neurons

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 4.22 shows the performance of the different training algorithm tested.

Table 4.22 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 4.22 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 4.23 was selected to design the network architecture.

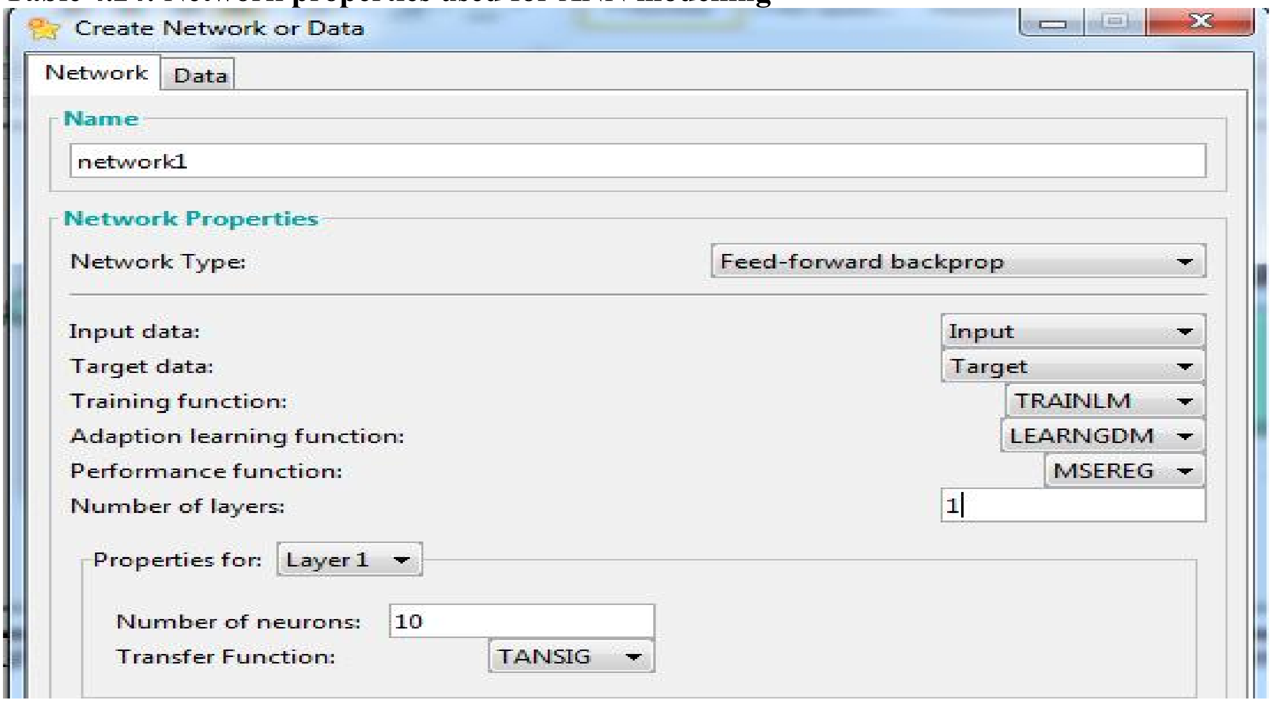
Table 4.23 Selection of optimum number of hidden neurons for ANN

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 4.22 and 4.23, 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). The network properties is presented in Table 4.24

Table 4.24: Network properties used for ANN modelling



A learning rate of 0.01, momentum coefficient of 0.1, target error of 0.01, analysis update interval of 500 and a maximum training cycle of 1000 epochs was used. The network generation process divides the input data into training data sets, validation and testing. For this study, 60% of the data was employed to perform the network training, 25% for validating the network while the remaining 15% was used to test the performance of the network. Using these parameters, an optimum neural network architecture was generated. The network training diagram generated for the prediction of number of persons killed using back propagation neural network is presented in Figure 4.6

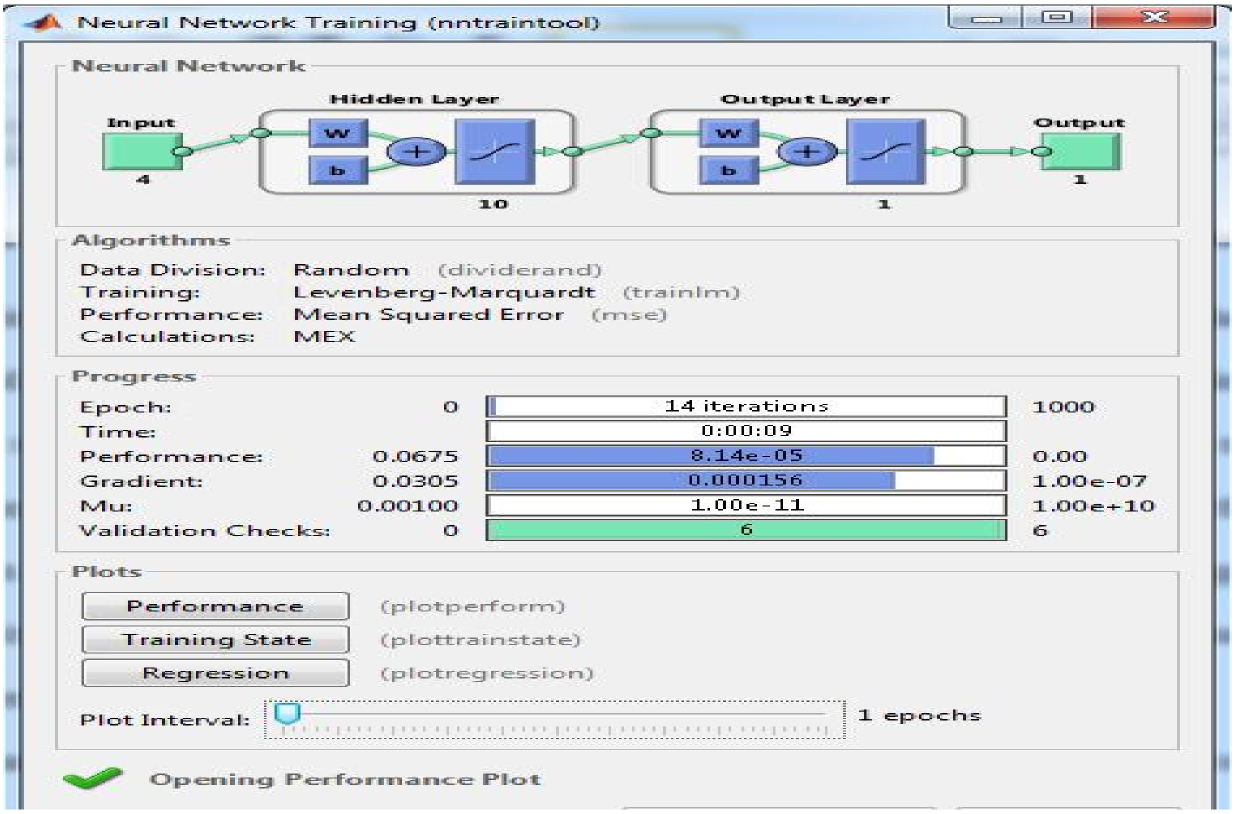


Figure 4.6: Network training diagram for predicting number of persons killed

From the network training diagram of Figure 4.6, it was observed that the network performance was significantly good with a performace error of 8.14e-05 which is far lesser than the set target error of 0.01. The maximum number of iteration needed for the network to reach this performance was observed to be 14 iterations which is also lesser than the initial 1000 epochs. The gradient function was calculated to be 0.000156 with a training gain (Mu) of 1.00e-11. Validation check of six (6) was recorded which is expected since the issue of wieght biased had been addressed via normalization of the raw data. A performance evaluation plot which shows the progress of training, validation and testing is presented in Figure 4.7

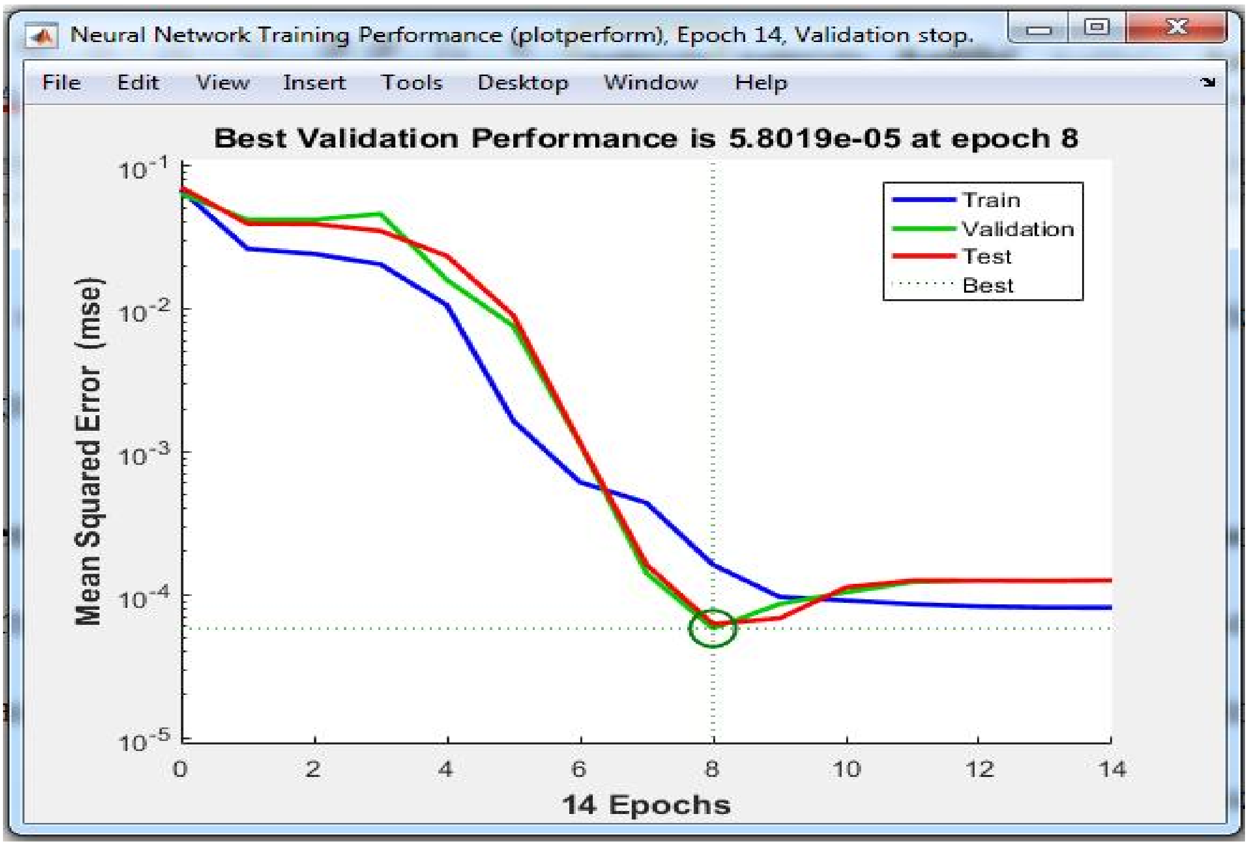


Figure 4.7: Performance curve of trained network for predicting number of persons killed

From the performance plot of Figure 4.7, 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 4.8

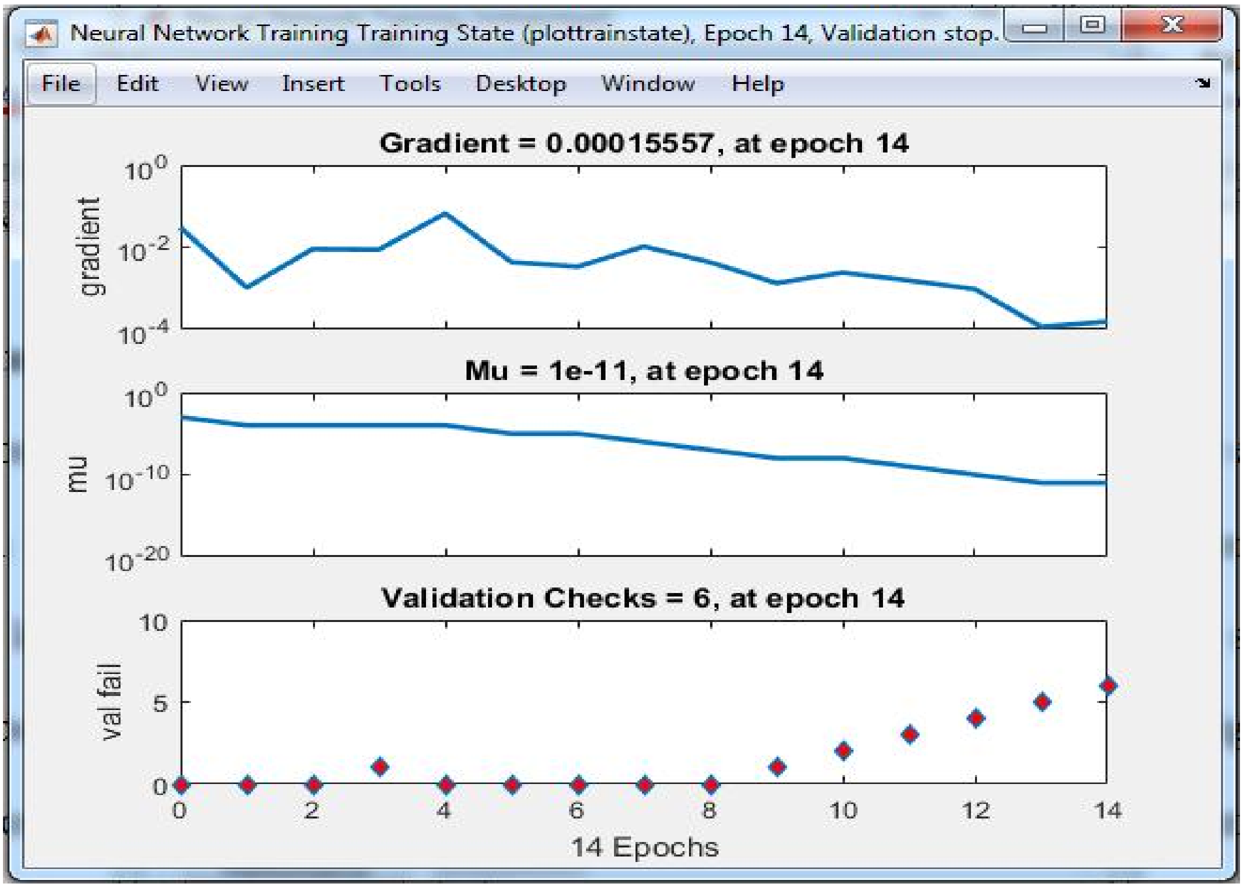


Figure 4.8: Neural network training state for predicting number of persons killed

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 4.8 indicates that the error contributions 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 4.9

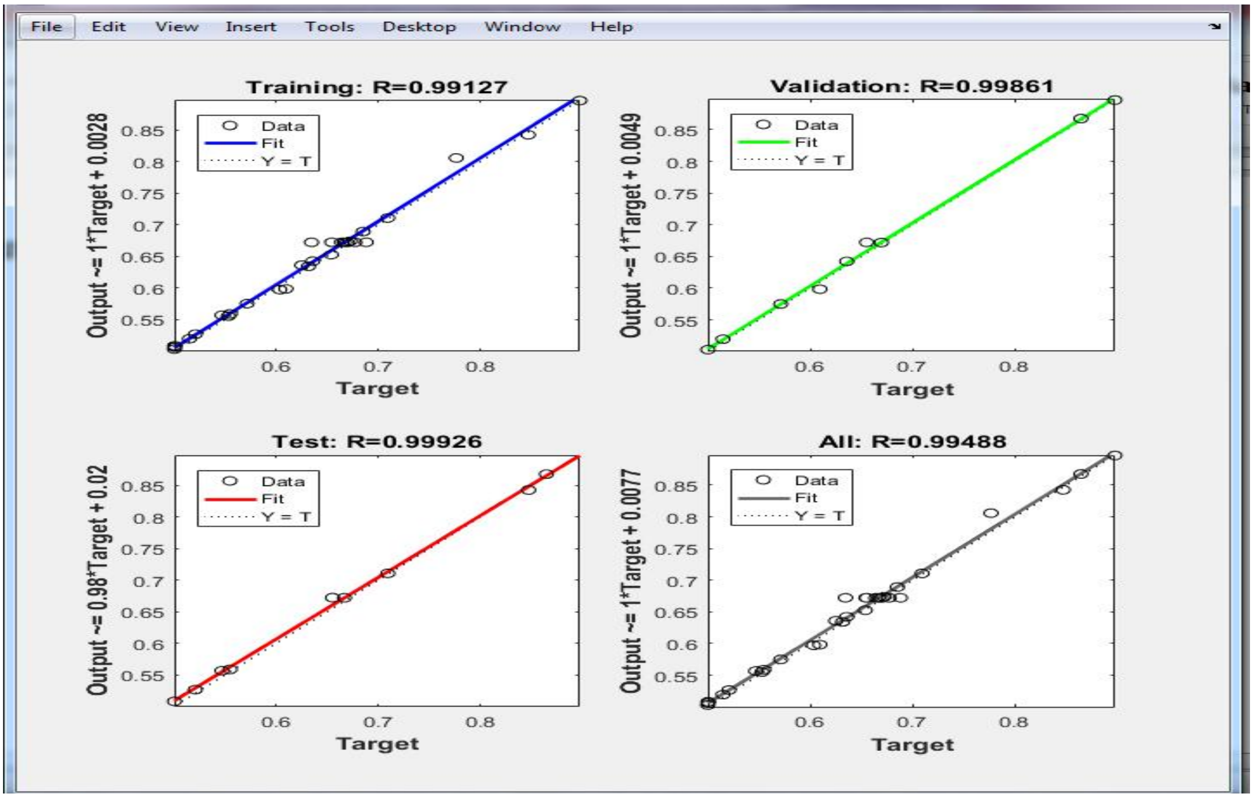


Figure 4.9: Regression plot showing the progress of training, validation and testing

Based on the computed values of the correlation coefficient (R) as observed in Figure 4.9, 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 4.10

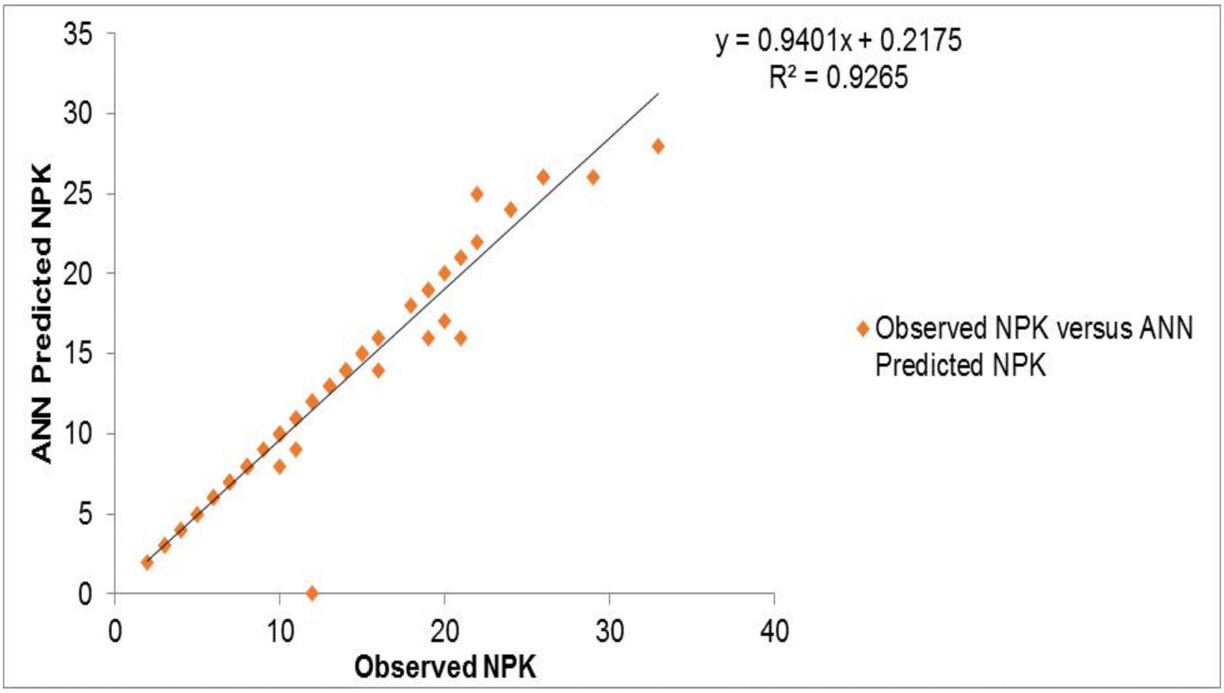


Figure 4.10: Regression plot of observed versus ANN predicted NPK

Based on the observed and the linear regression predicted values, a regression plot of outputs was also generated as presented in Figure 4.11

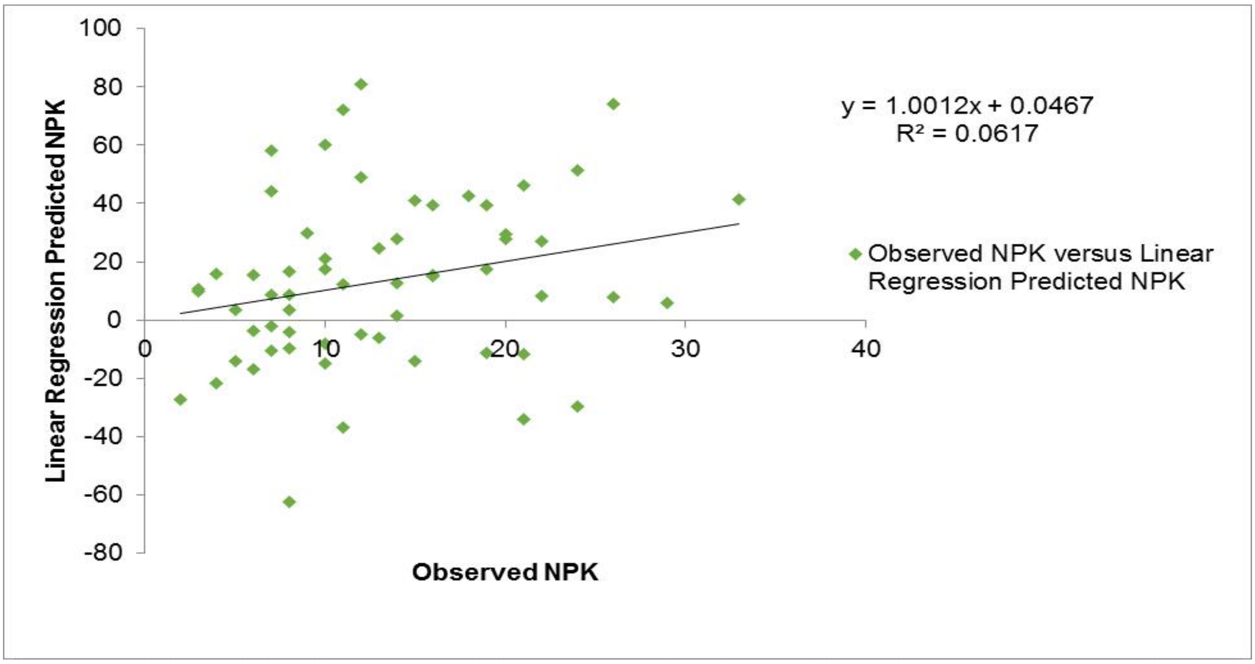


Figure 4.11: Regression plot of observed versus LRM predicted NPK

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.

CHAPTER FIVE
CONCLUSION AND RECOMMENDATION

5.1 Summary of findings from data analysis;

i. For reliability test, the cronbach alpha value must be greater than 0.65. for standardized cronbach alpha values of 0.900 was concluded that the accident data are all reliable.

ii. For the outlier analysis, 1000000 monte carlo simulation at 99% confidence interval, Dixon test for outlier was conducted on the number of accident cases (NAC)

iii. And the Dixon test for outlier was conducted on the number of persons involved in an accident cases (NPIV)

5.2 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 test of outliers, reliability test, test of normality and diagnostic 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.

5.3 Recommendations

From the study carried out, and after several observations, the following are recommended;

1. There is need to re-educate the Nigerian Police, FRSC, with respect to records keeping when it comes to road accident matters because one will discover that they always have very little or no knowledge about the various causes of accidents most especially when road condition, and vehicle defects are the causes.

2. There is need to have a thorough analysis of road accidents into different causes so that proper attention could be given to the problem in order to adopt a proper remedial measure and in the long run to access its effectiveness.
3. Effective campaign and enlightenment against bad driving culture/non maintenance of vehicles and taking of alcohol when driving should be intensified and encouraged by the media, law enforcement agents, federal Road Safety Corps and the Government.

References

Abdelwahab, H. T. and Abdel-Aty, M. A. (2002) Artificial Neural Networks and Logit Models for Traffic Safety Analysis of Toll Plazas. Transport. Res. Rec., 1784: pp. 115-125

Abdulhafedh, A. (2017); Road Crash Prediction Models: Different Statistical Modeling Approaches. Journal of Transportation Technologies, vol. 7, pp: 190-205.

Agbonkhese O. (2013) Road Traffic Accidents in Nigeria, Causes and Preventive Measures civil and environmental research, ISSN 2224-5790 (Paper) ISSN 2225-0514 (online) Vol 3(13), pp:1-14

Ali, A. M., Kamarudin, B. A., Ahmed, M. M., and Deprizon, S (2018), Classification Of Traffic Accident Prediction Models: A Review Paper, International Journal of Advances in Science Engineering and Technology, ISSN(p): 2321 –8991, ISSN(e): 2321 –9009 Volume-6, Issue-2, Apr.-2018.

Ali, A.M., Kamarudin, B. A., Ahmed, M. M and Deprizon, S, (2018), Classification of Traffic Accident Prediction Models: A Review Paper, International Journal of Advances In Science Engineering And Technology, Vol.6 (2), Pp.8991-9009.

Anyaoku, E. (2009), “Capital Waste on Nigeria, Roads” Inaugural FRSC Annual Lecture Series delivered in Abuja on 27th October.

Ashutosh, A. E and Madhu, S. V (2015), Road Crash Frequency Prediction for Indian National Highways using Soft Computing Tools, pp: 34.56

Aworemi, Joshua, R., Abdul, A., Ibraheem, A., Olabode, and Segun, O. (2010); Analytical study of the causal factors of road traffic crashes in southwestern Nigeria.

Abdullah et al., 2012 on road accident models with two threshold levels of Fuzzy Linear regression; the author hypothesized that number of road accidents and road casualties are increasing in line with the raise in the variables of registered vehicles, population and road length. Pp 38

Abdullah et al., 2012 on road accident models with two threshold levels of Fuzzy Linear regression; the author hypothesized that number of road accidents and road casualties are increasing in line with the raise in the variables of registered vehicles, population and road length. Pp 38

Abdullah et al., 2012 on road accident models with two threshold levels of Fuzzy Linear regression; the author hypothesized that number of road accidents and road casualties are increasing in line with the raise in the variables of registered vehicles, population and road length. Pp 38

Carlos, G. (1999); Artificial Neural Networks for Beginners. Retrieved on September 27, 2010, from http://www.cogs.susx.ac.uk/users/davidy/fcs.

Celikoglua, H. B. and Cigizoglub, H. K. (2007); Modelling public transport trips by radial basis function neural Networks, Mathematical and Computer Modelling. Vol. 45, pp. 480-489.

Chang, L.Y. (2005). “Analysis of freeway accident frequencies: negative binomial regression versus artificial neural network”. Safety Science vol.43, pp: 541-557.

Chen, G. (2010) Road Traffic safety in African countries- status, Tread, contributing factors counter measures and challenges international journal of injury control and safety promotion, vol.17 (4): pp: 241-255.

Abdullah et al., 2012 on road accident models with two threshold levels of Fuzzy Linear regression; the author hypothesized that number of road accidents and road casualties are increasing in line with the raise in the variables of registered vehicles, population and road length. Pp 38

Chris, B. (1995); Interim guide on identifying, prioritizing and treating hazardous locations on roads in Malaysia. Malaysia: Public Work Department.

Dahee, H. (2005); Development of Traffic Accident Prediction Models by Traffic and Road Characteristics in Urban Areas. Proc. of the Eastern Asia Society for Transportation Studies, vol. 5. Kyunggi, Korea: EASTS.

Delen, D., Sharada, R., and Bessonov, M. (2006). “Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks”. Accident Analysis and Prevention, vol. 38, pp: 434-444.

Department of Transportation (2000); Federal Highway Administration. United States of America.

Eleni I. V. and Matthew G. K (2013), Fuzzy-Entropy Neural Network Freeway Incident Duration Modeling with Single and Competing Uncertainties, National Technical University of Athens, 5 Iroon Polytechniou Str, Zografou Campus, 157 73, Athens, Greece.

Esnizah, M. S. (2008). A study on Road Safety at High Accident Sites and the FRSC Reports of 51693 fatal accidents in 2009 (FRSC 2019)

Esnizah, M. S. (2008); A study on Road Safety at High Accident Sites and the Development of Accident Prediction Model at Federal Route FT 50 Batu Pahat to Ayer Hitam. Universiti Tun Hussein Onn Malaysia: Bachelor’s Degree Thesis.

Fajaruddin, M., Ismail, Y., Ismail, R., Aziz, A. S., and Esnizah, M. S. (2008); Blackspot Study and Accident Prediction Model Using Multiple Liner Regression. Proc. of the First International Conference on Construction in Developing Countries. Karachi, Pakistan: Universiti Tun Hussein Onn Malaysia.

Garda et al., 2018 on predicting road accidents using artificial neural network models; the authors presented a methodology for establishing an accident risk prediction model, which can be used as a decision-making tool in infrastructure management. The authors affirmed that; the methodology allowed for an appropriate handling of the available data, examines how it can be used to develop models using artificial neural networks (ANNs) and establishes a systematic ANN optimization process to determine the optimal architecture of the ANN model.pp 41.

Gaber et al., 2016 on traffic accident prediction model using fuzzy logic: Aswan desert road case study; The author affirms that; transportation system plays an important role in human life and is one of the main indicators of the standard of living. They simply described traffic accidents as a major problem threatening people's lives, health, property and the management of transportation system. Pp 39

Gajendran, C., Serin, V. K., Seenu, S. G and Swati, P (2015); Different Methods of Accident Forecast Based on Real Data. J. Civil Environ Eng, vol.5, pp:180. doi:10.4172/2165-784X.1000180.

García de Soto B, Bumbacher A, Deublein M and Adey B. T (2018); Predicting road traffic accidents using artificial neural network models. Infrastructure Asset Management, vol. 5(4): pp:132–144.

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 Modul. 2nd Ed. Johor: Penerbit UTHM.

Google Copyright (2011); Google Earth. Retrieved on May 2011 from Google website: http://maps.google.com

Gwynn, D.W. (1967); Relationship between Road Accident and Hourly Volumes, Traffic Quartely. Vol. 14, no 1, pp. 407-418.

Haykin, S. (1999); Neural Neworks: A Comprehensive Foundation. 2nd Edition. Prentice Hall. pp 135-155.

Hung et al., 2008. In regression problems only a single output is estimated, and so the second hidden layer usually has two units.

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

Hizal, H. H. and Sharifah, A. S. M. R. (2009); The Construction of Road Accident Analysis and Database System in Malaysia. Proc. of IRTAD Conference, vol.4. Seoul, Korea: IRTAD Institute

Jerry, S and Jisha, A (2019); Development of Accident Prediction Model on Horizontal Curves, International Research Journal of Engineering and Technology (IRJET), Vol.6 (3), Pp. 56-72.

Jianhu, Z and Xiongbin, W. (2014), Prediction of Road Traffic Accidents Using a Combined Model Based on IOWGA Operator, Periodica Polytechnica Transportation Engineering 43(3), pp. 146-153.

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

Lazim, A. and Nurnadiah, Z. (2012), Road Accident Models with Two Threshold Levels of Fuzzy Linear Regression, Journal of Emerging Trends in Computing and Information Sciences, vol.3(2), pp:23.34

Levi, D. B.; Julie, E. K.; Olsen, J. R.; Pulwarty, R. S.; Raff, D. A.; Turnipseed, D. P.; Webb, R. S and Kathleen D. W (2009); Climate Change and Water Resources Management: A Federal Perspective, circular 1331, pp: 1 – 72

Lester, A. H., Garber, N. J., and Adel, W. S. (2010); Transportation Infrastructure Engineering. SI Edition. Virginia: Thomson Learning.

Lugarde, G. (2007) Road Traffic injury is an escalating Burden in African and Deserves proportionate research efforts, plosmed, vol.4(6): pp: 967-971

Zadeh, 2016 The membership function is used to associate a grade to each linguistic term. Defuzzify evaluate several membership sets established by the system designer for a fuzzy logic-based control system, such as "speed too fast," "speed too slow" and "speed about right" at a specific input value.

Mehdi, H., Ahmad, S. Y., Seyed, M. G, and Joewono, P (2013), Application of Adaptive Neuro-Fuzzy Inference System for Road Accident Prediction, KSCE Journal of Civil Engineering, vol.17(7); pp: 1761-1772.

Mehmet, M., Kunt, I and Aghayan, N. N (2011); Prediction for Traffic Accident Severity: Comparing the Artificial Neural Network, Genetic Algorithm, Combined Genetic Algorithm and Pattern Search Methods, Transport, vol. 26(4): pp: 353–366.

Meng, X.H., Zheng, L., and Qing, G.M. (2009). “Traffic accidents prediction and prominent influencing factors analysis based on fuzzy logic”. Journal of Transportation Systems and Information Technology, vol. 9(2), pp: 87-92.

Mohammed, G., Amr, M. W., Ayman, M. O and Aboelkasim, D (2018), Traffic Accidents Prediction Model Using Fuzzy Logic: Aswan Desert Road Case Study, Journal of Engineering Sciences Assiut University Faculty of Engineering Vol. 45 (1), PP. 28 – 44

Mohd, Z. B. R., (2011), Development of Accident Prediction Model by Using Artificial Neural Network (Ann), A project report submitted in partial fulfillment of the requirement for the award of the Degree of Master of Civil Engineering Faculty of Civil and Environmental Engineering University, Tun Hussien Onn Malaysia.

Nachimuthu, K. and Partheeban, P. (2013), Development of Road Accident Prediction Model Based on System Dynamics Approach, Indian Journal of Applied Research, Vol.3 (9), pp: 26-31

NSSR, M. and Srinivasa, R. R. (2015); Development of model for road accidents based on intersection parameters using regression models, International Journal of Scientific and Research Publications, vol. 5 (1), pp: 34.42 ISSN 2250-3153.

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

Ogwueleka, T. C and Ogwueleka, F. N. (2010) Prediction of Solid Waste Generation Using Self Organizing Map and Multilayer Perceptron Neural Networks. IUP University Journal of Science and Technology, Vol. 6, No. 1, pp. 39-50

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

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.

Okhakhu, P.A. (2014): Environmental and Human Challenges in the Niger-Delta Region of Nigeria, Journal of Environment and Earth Science, Vol. 4(23), pp: 112 – 134

Oladehinde, M. K, Adeoye, A. O, Adegbehingbe, B. O, and Onakoya A. O. (2007) Visual Functions of Commercial Drivers in Relation to Road Accidents in Nigeria. Indian Journal of Occup Environ Med. Vol. 11, pp. 71-75

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

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.

Murthy et al., 2015, investigated the development of models for road accidents based on intersection parameters using regression model. The author affirms that; the issue of road accidents is an increasing problem in developing countries. Pp43

Olugbenga et al.2010, on Accident Prediction Models For Akure – Ondo Carriageway, Ondo State Southwest Nigeria; Using Multiple Linear Regressions, Accident data on the 52km Akure-Ondo Carriageway and Spot Speed data were collected and analyzed. The analysis of the Spot Speed gave an average of 51.5km/hr. and 110.75km/hr. for the 15th and 85th percentile speed respectively, the 85th percentile speed was higher than the posted speed limit of 60km/hr. which indicated that the vehicle travelling at that speed was susceptible to accident.pp49

Ozgan, E. and Demirci, R. (2008) Neural Networks-based Modeling of Traffic Accidents in Interurban Rural Highways, Duzce Sampling. J. Applied Sci., vol. 8, pp. 146-151

Persson and Berndtsson, 200.1Artificial Neural Networks (ANNs) have been motivated right from their inception by the recognition that the human brain computes in an entirely different way from the conventional digital computer.

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.

Raes, D; Willens, P and Gbaguidi, F. (2006), Rainbow – A software package for analyzing data and testing the homogeneity of historical data sets, vol. 1, pp: 1-15

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

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

Sirigiri et al., 2012 While variables in mathematics usually take numerical values, in fuzzy logic, the non-numeric linguistic variables are often used to facilitate the expression of rules and facts.

Tatiana, T., Dora, S., Miltiadis, C and Andreas, G. (2014), Maximizing Accuracy and Efficiency of Traffic Accident Prediction Combining Information Mining with Computational Intelligence Approaches and Decision Trees, Jaiscr, Vol.4 (1), Pp.31-42.

Weihong, M. A. and Zhenzhou, Y. (2018), Analysis and Comparison of Traffic Accident Regression Prediction Model, MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing 100044, China, Pp. 364-369.

Yong, G. J. and Jong, H. L. (2012); Automobile Traffic Accident Prediction Model using by

Artificial Neural Networks, pp. 713-719.

(Jacob,1995) Use of seat belt reduces the risk of death or services injury by about 45per cent

pp 12

Central African Journal of Medicine, Vol. 28, No. 7 pp 173

Hung et al, (2008) an artificial neural network model for rainfall forecast in bankok.

(Gbadmosi, 1994).Extensive remedial measures aimed at improving road safety have been

developed in the field of engineering, education, and enforcement “Three Es”.pp 12

Road Traffic Injuries, W.H.O (World Health Organization), report, pp 3.

Geometric Design of Roads, Google assisted and Wikipedia Reports. Pp, 37.

Super Elevation Transverse Slope, Horizontal Alignment, Mitigating Incidents which contributes to Accident Severity. (Zhang Yingxue, 2009). PP 66.

General spacing, Freeway flow speeds and (HCM) Indicating multilane freeway (Jerry Et Al, 2009). PP 66. Ozgan and Demirci, (2008) used series of artificial neural network to model the potential non-linear relationship between injury severity levels and crash related factors.

Median type Study, Impact of Median on Accident, and how it prevents errant maneuvers on the highway (Hadi Et Al, 1995) PP 67.

Xiao et al., (1999) developed two fuzzy logic models for predicting the risk of accident that occurred on wet pavement.

Zadeh, 2016. Saying that the theory of fuzzy sets is a generalization of the classical set theory means that the latter is a special case of fuzzy sets theory. To make a metaphor in set theory speaking, the classical set theory is a subset of the theory of fuzzy sets

Unahabhokha et al., 2007. Fuzzy logic is a set of mathematical principles for knowledge representation based on degrees of membership.

Zadeh, 2016 The membership function is used to associate a grade to each linguistic term.

Zianhu and Xiongbin (2014), on prediction of road traffic accidents using a combined model based on IOWGA operator; the authors affirm that; traffic accident prediction plays an important role in reducing the likelihood of traffic accidents and improving the management levels of traffic safety. Pp43

Zianhu and Xiongbin (2014), on prediction of road traffic accidents using a combined model based on IOWGA operator; the authors affirm that; traffic accident prediction plays an important role in reducing the likelihood of traffic accidents and improving the management levels of traffic safety. Pp43

APPENDIX

Appendix 1: Accident data for 2014

Months	2014				
	No. of accident cases	No. of persons involved	No. of person injured	No. of persons killed	No. of vehicles involved
Jan	27	240	117	16	47
Feb	17	139	47	19	37
March	20	130	57	22	39
April	25	213	87	19	44
May	21	159	70	19	39
June	20	162	55	26	40
July	15	226	43	22	26
August	13	67	26	7	25
September	13	85	30	9	18
October	19	142	74	20	31
November	20	93	34	10	38
December	34	312	114	33	66

Appendix 2: Accident data for 2015

Months	2015				
	No. of accident cases	No. of persons involved	No. of person injured	No. of persons killed	No. of vehicles involved
Jan	17	130	80	24	38
Feb	23	159	75	24	36
March	28	256	122	21	48
April	16	133	69	12	31
May	14	106	88	7	25
June	28	312	97	8	64
July	22	123	62	15	36
August	23	143	53	7	36
September	15	165	53	10	25
October	16	196	50	4	38
November	13	113	61	13	24
December	36	362	137	18	70

Appendix 3: Accident data for 2016

Months	2016				
	No. of accident cases	No. of persons involved	No. of person injured	No. of persons killed	No. of vehicles involved
Jan	36	245	102	26	63
Feb	10	74	26	6	21
March	30	338	69	7	60
April	24	121	51	12	33
May	25	186	60	10	41
June	17	153	86	16	28
July	15	215	35	6	29
August	22	205	44	8	48
September	26	210	78	7	54
October	8	88	23	2	19
November	13	104	29	3	23
December	20	242	86	8	55

Appendix 4: Accident data for 2017

Months	2017				
	No. of accident cases	No. of persons involved	No. of person injured	No. of persons killed	No. of vehicles involved
Jan	23	178	73	11	32
Feb	24	189	92	6	50
March	15	128	32	13	25
April	23	253	88	10	52
May	21	131	74	12	31
June	19	138	38	14	34
July	15	237	72	15	33
August	11	205	23	29	22
September	11	98	49	3	16
October	15	186	57	11	38
November	16	152	61	8	32
December	33	365	180	21	76

Appendix 5: Accident data for 2018

Months	2018				
	No. of accident cases	No. of persons involved	No. of person injured	No. of persons killed	No. of vehicles involved
Jan	17	140	93	4	27
Feb	19	195	83	8	34
March	18	199	83	16	32
April	19	174	81	21	40
May	15	188	63	14	29
June	13	84	51	5	22
July	15	178	39	10	33
August	19	205	73	14	36
September	16	98	35	8	30
October	9	55	33	11	11
November	10	88	35	5	20
December	36	344	131	20	73