ADDIS ABABA UNIVERSITY

ADDIS ABABA INSTITUTE OF TECHNOLOGY

SCHOOL OF CIVIL AND ENVIRONMENTAL ENGINEERING

PREDICTION MODELLING OF ROAD CRASH FREQUENCY: A REVIEW AND ASSESSMENT ON DIFFERENT APPROACHES

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for Degree of Master of Science in Civil Engineering (Road and Transport Engineering)

Advisor: Bikila Teklu (PhD)

By: Mohammed Assefa Chifraw

February, 2022

Addis Ababa, Ethiopia

CERTIFICATE

The undersigned have examined the thesis entitled "PREDICTION MODELING OF ROAD CRASH FREQUENCY: A REVIEW AND ASSESSMENT ON DIFFERENT APPROACHES" presented by Mohammed Assefa Chifraw, a candidate for the degree of Master of Science and herby certify, that it is worthy of acceptance.

Bikila Teklu (PhD)		28/03/22
Advisor	Signature	Date
Getu Segne (PhD)		21/03/22
Internal Examiner	Signature	Date
Wondwossen Taddesse		28-03-22
External Examiner	Signature	Date
Chair Person	Signature	Date

Mebruk Mohammed (Dr.-Ing.)
Dean, School of Civil &
Environmental Engineering

Addis Ababa University
Addis Ababa Institute of Technology
Director of Postgraduate Program

UNDERTAKING

I certify that research work titled **"Prediction Modeling of Road Crash Frequency: A Review and Assessment on Different Approaches"** is my own work. The work has not been presented elsewhere for assessment. Where material has been used from other sources it has been properly acknowledged / referred.

Signature of Student

Name of Student

ABSTRACT

Crash prediction version has often been utilized in dual carriageway protection studies. It may be utilized in figuring out foremost contributing elements or set up dating among crashes and explanatory twist of fate variables. Crash frequency refers back to the prediction of the range of crashes that might arise on a particular street section or intersection in a given time period. On the opposite hand predicted crash frequency is an anticipated street site visitors crashes which can be calculated the use of a calibrated SPF (Safety Performance Function) and weighted with the found crash frequency using Empirical Bayes (EB) method.

The objective of this report is to review and make an assessment on the road crash prediction models done previously, to identify which variables have a significant effect on crash prediction modeling and the methods employed in modeling of highway crash frequency, which can be used as a base line for future researches. Five articles with different modeling approaches, parameters and methods have been reviewed here.

The first study was "A Parametric Model for Accident Prediction along Ado Ekiti-Ikole Ekiti Road, Ekiti State, Nigeria" by Olumuyiwa S.Aderinola. "Statistical Model of Road Traffic Crashes Data in Anambra State, Nigeria was the second reviewed paper in this study by Nwankwo Chike H. and Nwaigwe Godwin I.The third study was "Accident Prediction Models for Two-Lane Rural Highways "by K.R.Kalokota. The fourth study was "An Artificial Neural Network Model for Road Accident Prediction: Case Study of a Developing Country" by F.N. Ogwueleka et al. The fifth study was "Development of Models for Crash Prediction and Collision Estimation-A case Study for Hyderabad City" by V.Niveditha, A. Ramesh and M. Kumar.

At the end of this study,AADT,85th percentile spot speed, length of road segments, number of lanes, surface condition of road, number of road access and number of road access controls, environmental and traffic characteristics were the main road traffic crash contributing factors considered in most researcher.

Keywords: Crash Frequency, Expected Crash Frequency, Prediction model, Review

ACKNOWLEDGMENT

Foremost, I am grateful to God Almighty ALLAH for helping me in every aspect. I would like to express my deepest gratitude to my advisor Dr. Bikila Teklu and co-advisor Mrs.Emnete T. for guiding and supervising my research work. They had been devoting their precious time and providing all necessary and relevant information to carry out the research. Also I would like to thank Dr. Getu S. for his precious and valuable comments and advising after presenting my paper as an examiner.

I acknowledge Ethiopian Road Authority (ERA) for the sponsorship and Addis Ababa Institute of Technology (AAiT) in writing a cooperation letter for concerned organizations for data collection and other information gathering purposes.

Finally, I take this opportunity to express gratitude to all my families and friends who have been a constant source of encouragement and support to accomplish this thesis.

ABBREVIATIONS

CART	Classification and Regression Tree
FRSC	Federal road Safety Commission
GPR	Generalized Poisson Regression
NBR	Negative Binomial Regression
AIC	Discrete and Akaike Information Criterion
AADT	Annual Average Daily Traffic
EJER	European Journal of Engineering Research and Science
VIF	Variance Inflation Factors
HSIS	Highway Safety Information System
FHWA	Federal Highway Administration
RSW	Right Shoulder Width
ANN	Artificial Neural Network
SOM	Self Organizing Map
RTA	Road Traffic Accident
MLPNN	Multi-Layer Perception Neural Network
BP	Back Propagation
IJTE	International Journal of Transportation Engineering
ERA	Ethiopian Road Authority
AAiT	Addis Ababa Institute of Technology
SPF	Safety Performance Function
EB	Empirical Bayes

TABLE OF CONTENTS

TABLE OF TABLES

TABLE OF FIGURES

CHAPTER ONE: INTRODUCTION

1.1 Background

Accident Prediction Models (APMs) are used for an entire lot of features typically used to estimate the expected twist of fate frequencies from splendid roadway entities consisting of, highways, intersections, interstates, etc. and moreover to turn out to be privy to geometric, environmental and operational factors which are associated with the superiority of injuries. It is crucial to look at the character of relationships among roadway, environmental and operational elements and, injuries to recognize the causal mechanisms concerned in injuries on the only hand and to higher are expecting their incidence on the opposite hand. APMs are one direction of inquiry regularly used to benefit those insights (Reurings et al., 2005).

In this report, cognizance is on APMs for street crash injuries on the street segments in exclusive evolved and growing countries. Since injuries happened each at intersections and on the mid-block segment of the street, those situations have to be studied separately.

Intersections are locations for extra injuries, because of the truth that there are numerous conflicting movements at those places in addition to many precise intersection layout trends. Intersections furthermore will be predisposed to experience immoderate injuries because of the fact that a number of the damage crashes together with left flip and way of thinking collisions generally rise up at intersection. Therefore, there is a need to understand the methodologies to assess the outcomes that geometric, web website online traffic flow, traffic control, environmental and operational trends have on the safety of intersections (Abdel-Aty and Keller, 2005). Since extraordinary roads meet at an intersection, one-of-a-type styles of accidents get up as a result. This calls for separate models to assess factors associated with one-of-a-type twist of fate types and the safety of diverse intersections sorts. Examples of varieties of intersections studied in this report include signalized intersections; prevent controlled intersections, intersections with cameras, etc.

Road phase additionally had a sizeable percentage of deadly and non-deadly crashes. The study carried out in Addis Ababa depicts that maximum crashes took place at mid-block places without junctions, because of the truth that, car pace is anticipated to be excessive and sufferers of those sorts of crashes are anticipated to be pedestrians (Jiregna, 2016). Also, the very best percent of street crashes severity inclusive of deadly damage, severe damage, minor damage and PDO passed off on directly roads, even as downward sloping roads shared decrease crash frequency (Tariku et al., 2017).

Several techniques had been evolved to pick out factors that have an effect on the protection of street Intersections and street section. These consist of the more than one logistic regression fashions, a couple of linear regression fashions, Poisson regression fashions, poor binomial regression Models, random outcomes fashions and the category and regression tree (CART) technique. This regression tactics attention on predicting overall injuries with fatalities, accidents, etc. for assessing the protection outcomes of diverse elements. However, powerful street protection control calls for understanding of the existing protection overall performance and what it's far possibly to be in destiny if proposed moves are taken. In effect, dependable techniques for estimating protection overall performance of a present or deliberate roadway are required (Harwood et al., 2000). As one of the most important strategies, APMs which describe relationships among the quantity of injuries and elements which are believed to be associated with coincidence prevalence were advanced through numerous authors to estimate modern-day or destiny avenue protection overall performance.

1.2 Statement of Problem

According to research made in extraordinary countries, it's miles feasible to alternate the trouble of street crash incidence between unique avenue users into a safer, more efficient and environmentally pleasant transportation device with the aid of using incorporating protection features. Investigate the street geometric variables, avenue surroundings and site visitors traits that reason street visitors crash occurrences at street segments and intersection can be very critical aspect. So modeling those variables with street site visitors crashes to estimate the predicted quantity of injuries on a selected street phase may be performed in specific procedures. Thus, know-how the street, surroundings and visitors traits thru distinctive research and studies can substantially assist in lowering the frequency of street visitor's crashes and make contributions to crashes while improperly implemented.

This look at closes the distance with the aid of using figuring out unique APMs the use of unique processes at the prevalence of street visitors crashes and develops regression fashions that may be implemented in new layout and renovation of street infrastructures for protection criteria. Therefore, the trouble isn't always to do a studies associated with APMs however the new thoughts evolved had been truly placed on the shelf with inside the shape of paper. It may be very critical that to study specific associated articles formerly carried out via way of means of exceptional researchers and integrates them to have a first-class answer for street crashes. The writer on this examine, particularly targeted at the elements taken into consideration in distinctive research, the techniques used for APMs each in advanced and growing countries.

1.3 General Objective

The main objective of this report is to review and made an assessment on prediction modeling of road crash frequency employed in literatures to predict the expected road crash frequency.

1.4 Specific Objectives

- To review Literatures on articles and papers for accident prediction models
- To identify the contribution of road geometric and other parameters for crash occurrence in engineering perspective for each study
- To make an assessment of the most critical factors for road crash occurrence based on the review of articles

CHAPTER TWO: LITERATURE REVIEW

2.1 Definition of Road Traffic Crashes

According to the Economic Commission for Europe (ECE), Road Traffic Accidents (RTAs) may be described as, the ones injuries which arise on a manner or avenue open to public traffic; resulted in a single or greater humans being killed or injured, and as a minimum one shifting automobile changed into involved”. It means that RTA is collisions among cars, motors and pedestrians, cars and animals and among cars and stuck obstacles (ECE, 2005).

The World Health Organization(WHO) defined road traffic crash as “any crash involving a device designed primarily for, or being used at the time primarily for, conveying persons or goods from one place to another”. It can also classified depending on different conditions; the 30 days crash death record of a person; the degree of injury to be treated in hospital; “the crash occurred on any road, street, or any place open to public, the crash involved one or more road vehicles which were in motion at the time of the accident” (WHO, 2004). Generally, road crash is an event which is unexpected, undersigned with an element or chance or probability or unfortunate result" and sometimes it is defined as “the occurrence which usually produces injury, death (fatal) or property damage (PDO)”. Therefore, it is important studying on the crash contributing factors to reduce traffic crashes by incorporating safety-conscious design and planning of road network.

Crash frequency which is one of the simplest forms of crash data analysis can be defined as the number of crashes occurring within a specific jurisdiction, on a roadway segment, or at an intersection. At the same location there may be multiple crashes which occurred over a period of time which indicate a safety issue and should be investigated and addressed appropriately. This can be achieved by a means of "clustering". Crashes can be clustered by route, specific location on that route, or by intersection.

On the other hand expected crash frequency is a predicted or calculated crash using a calibrated SPF (Safety Performance Function). This can be weighted with the observed crash frequency using Empirical Bayes (EB) method, weighted in terms of crash severity and the equivalent property damage only cost. With this method, the expected change in crash frequency of different design alternatives uses crash estimation procedures based on observed crash frequency. The EB method allows us to combine estimation from the statistical model with the observed crash.

2.2 Factors Contributing to Road Traffic Crashes

The most contributing elements for an excessive variety of avenue crashes can be recognized as follow, Tulu (2015).

- Deriver's deriving skills;
- Lack of know-how for visitors regulations and regulations ;
- Deriving over the velocity limits;
- Lack of punishment/enforcement;
- Vehicular performance/no maintenance;
- Pedestrian, Animal-carts and bicycle owner were now no longer separated from important highway;

Design and making plans of avenue community did now no longer recall protection issues; Lack of trendy protection attention with the aid of using pedestrians; and insufficient clinical facility for associated twist of fate severity.

Generally, those elements may be classified into three occasions as avenue surroundings deficiencies, car defects, and street person errors. In growing countries, those elements had been very crucial and without delay related to street visitors crashes. Therefore, those might also additionally want deep-rooted answers thru studies and investigations.

2.2.1 Driver Characteristics

As stated earlier than drivers, pedestrians, Vehicles, street and environmental associated and different elements have been the principle classes contributing to injuries. These elements extensively utilized to broaden statistical fashions forecasting anticipated range of injuries and casualties indicated. According to the examine motive force associated elements had been the maximum fantastically contributing elements, because of lack of manipulate of using wheel, speeding, stoppage (i.e. surprising stopping), Abbas (2004).

The incapacity of the deriver to manipulate each pace and course had been additionally fundamental additives concerned in street crashes. Particularly, run-off cars may also arise as a driving force is confronted uncommon information, Nicholas et al. (2001).

The motive force's behavior, distractions, the impact of alcohol, drugs (or medication), drowsiness, fatigue, illness, or blackout, speeding, and failure to obey signs, alerts or visitors police, which could be because of confusion or unfamiliarity with the roadway have been the primary classes contributing to injuries. According to the look at made via way of means of Chanyukong and Jikuang (2010), primarily based totally on actual world Accident-information the affiliation among the effect of velocity and hazard of pedestrian causalities in a passenger automobile collision in China can be investigated. In their take a look at they may enhance a more than one logistic regression version accepting effect velocity determined that the chance of pedestrian fatality at 50kmph, 58kmph and at 70kmph is 26%, 50% and 82% respectively. The pedestrians not often survived at an effect velocity of 80kmph.

2.2.2 Vehicle Characteristics

According to the observe made through Nicholas et al. (2001) any other essential aspect that contributes to avenue visitors crashes have been mechanical issues which include defective brakes, worn tires, and different automobile defects. These issues might also additionally have an effect on the controlling of a car, particularly at excessive speeds. Moreover, at excessive speeds, the tires may also blow out main to lack of control. So tire tread separation is some other element that results in lack of control. Since, heavy cars are plenty and large in size in comparison with passenger automobiles and that they enjoy instability and maneuverability issues. In addition, they've much less powerful acceleration talents than passenger automobiles and feature more issue keeping speeds on upgrades, and this pace version generates extra times of overtaking and the capability for head-on collisions with oncoming cars. Finally, they have got a decrease deceleration in reaction to braking than passenger automobiles, which will increase the capability for extreme rear-quit crashes.

The have a look at performed through Abbas (2004) indicates that tire burst, automobile flip off the street and car flip over, and crash because of mechanical failure are taken into consideration beneath car associated factors.

2.2.3 Road Geometric Characteristic

Based on an observation conducted in Addis Ababa using two specific statistical regression fashions i.e. bad binomial regression technique for black spot avenue segments and non-black spot street sections on the influence of street site visitors management and geometric characteristics on visitors protection. The studies observed that the principle influencing avenue geometric associated variables which considerably have an effect on visitors protection have been range of horizontal curve, quantity of a lane, variety of vertical curves, variety of get entry to and gradient according to kilometer (Tefera, 2015).

Nicholas et al. (2001) said that without the help of extra passing lanes, the overtaking maneuver on multi-lane roads is a complicated riding task. It requires crucial information-processing and decision-making skills, and a lengthy section of avenue to finish the maneuver.

The fee of overtaking crashes is associated with the availability and geometric layout of passing lanes. There is extended capacity for unstable or misjudged overtaking maneuvers, especially while sight distance is short, if passing lanes aren't supplied on lengthy sections of street lengths,. Also, evidently layout practices for passing lanes may not be appropriate for many drivers to pass slow visitors or more than one cars in a secure manner.

In addition to these, quantity of lanes, lane width, and presence of a median, median width, form of median, shoulder width, get admission to density, pace limit, vertical grade, horizontal curvature, and climate situation have been tested with crash occurrences in preceding studies. Knowing the connection among protection at the dual carriageway and elements indexed above is the number one awareness in crash discount and predictions (Deo, 2004). Moreover, carriageway, grade, horizontal curvature, shoulder, median, vertical curve have been the alternative number one geometric layout factors that may have an effect on street protection (Iyinamet et al., 1997).

Douglas et al. (2000) concluded from his look at that defining the visitors operational performance of any roadway and street visitor's injuries relies upon on geometric layout factors.

Based on the studies made, variety of lanes, widths of lanes, the presence of width of shoulders and dual carriageway medians, and the horizontal and vertical alignment of the highways were the essential factor geometric format elements that have an effect on site visitors operations and road traffic injuries.

The examine Conducted on the visitor's twist of destiny at 28 hazardous locations on Amman-(Jordon) town roads thru manner of manner of Obaidat and Ramadan (2012) found that the most critical and realistic models that can be used to count on the relationship some of the twist of fate characters and primarily based totally variables have been logarithmic and linear models. Based on their studies; huge sort of vertical curves, median width, type of road surface, not unusual place walking velocity, posted velocity, lighting, variety of vehicles consistent with hour, amount of crossing facilities, maximum and not unusual place degree of horizontal curves and percentage of automobiles were moreover the most crucial contributing factors to traffic twist of destiny in risky locations.

Tulu et al., (2013), made each different examine with in the town of Addis Ababa to signify land use have an effect on the crash occurrence, and concluded that most lethal and damage crashes occurred in CBD and domestic areas. Accordingly, 27.7% fatalities and 33.4% important injuries passed off in CBD and domestic areas respectively with within the six-yr. period beneath interest and 39.1% of PDO crashes occurred in CBD. Residential, commercial, industrial, and one of a kind land uses along roadways are associated with range amount of access the adjacent properties. However, the huge type of the road crashes could probably variety due to the degree of access the driveways that distinctive land uses faced.

2.3 Modeling Methods for Analyzing Crash-Frequency Data

A huge sort of techniques had been carried out over time so that it will cope with the statistics and methodological problems related to crash-frequency information. Many of that may compromise the statistical validity of an evaluation if now no longer nicely addressed. Table 1 and 2 below gives a list of techniques formerly implemented to crash-frequency evaluation in conjunction with their strengths and weaknesses.

Table 1 Listing of methods previously applied to crash-frequency analysis along with their strengths and weaknesses

Model Type	Advantages	Disadvantages
Negative binomial/Poisson-gamma	Easy to estimate can account for over dispersion	Cannot handle under-dispersion; can be adversely influenced by the low sample mean and small sample size bias
Poisson	Most basic model; easy to estimate	Cannot handle over- and under-dispersion; negatively influenced by the low sample mean and small sample size bias
Zero-inflated Poisson and negative binomial	Handles datasets that have a large number of zero-crash observations	Can create theoretical inconsistencies; zero- inflated negative binomial can be adversely influenced by the low sample mean and small sample size bias
Poisson-log Normal	More flexible than the Poisson-gamma to handle over-dispersion	Cannot handle under-dispersion; can be adversely influenced by the low sample mean and small sample size bias (less than the Poisson-gamma); cannot estimate a varying dispersion parameter
Conway-Maxwell-Poisson	Can handle under-and over-dispersion or combination of both using a variable dispersion (scaling) parameter	Could be negatively influenced by the low sample mean and small sample size bias; no multivariate extensions available to date
Gamma	Can handle under-dispersed Data	Dual state model with one state having a long term mean equal to zero
Generalized estimating equation models	Can handle temporal correlation	May need to determine or evaluate the type of temporal correlation a priori; results sensitive to missing values

Table 2 Listing of methods previously applied to crash-frequency analysis along with their strengths and weaknesses

Model Type	Previous Research
Negative binomial/Poisson-gamma	Maycock and Hall (1984); Hauer et al. (1988); Brude and Larsson (1993); BonnesonandMcCoy (1993); Miaou (1994); Persaud (1994); Kumala (1995); Shankar et al. (1995); Poch and Mannering (1996); Maher and Summersgill (1996); Mountain et al. (1996); Milton and Mannering (1998); Brude et al. (1998); Mountain et al. (1998); Karlaftis and Tarko (1998); Persaud and Nguyen, 1998; Turner and Nicholson (1998); Heydecker and Wu (2001); Carson and Mannering (2001); Miaou and Lord (2003); Amoros et al. (2003); Hirst et al. (2004);Abbas (2004); Lord et al. (2005a); El-Basyouny and Sayed (2006); Lord (2006); Kim and Washington (2006); Lord and Bonneson (2007); Lord et al. (2009); Malyshkina and Mannering (2010b); Daniels et al. (2010); Cafiso et al. (2010a)
Poisson	Jovanis and Chang (1986); Joshua and Garber (1990); Jones et. al. (1991); Miaou and Lum (1993); Miaou (1994)
Poisson-lognormal	Miaou et al. (2005); Lord and Miranda-Moreno (2008); Aquero-Valverde and Jovanis (2008)
Conway-Maxwell-Poisson	Lord et al. (2008); Sellers and Shmueli (2010)
Zero-inflated Poisson and negative binomial	Miaou (1994); Shankar et al. (1997); Carson and Mannering (2001);Lee and Mannering (2002); Kumara and Chin (2003); Shankar et al. (2003); Qin et al., 2004; Lord et al. (2005b); Lord et al. (2007); Malyshkina and Mannering (2010a)
Gamma	Oh et al. (2006); Daniels et al. (2010)
Generalize estimating equation models	Lord and Persaud (2000); Lord et al. (2005a); Halekoh et al. (2006); Wang and Abdel-Aty (2006); Lord and Mahlawat (2009)
Random-effects models	Johansson (1996); Shankar et al. (1998); Miaou and Lord (2003);Flahaut et al. (2003);MacNab (2004); Noland and Quddus (2004); Miaou et al., (2003); Miaou et al., (2005); Aquero-Valverde and Jovanis (2009); Li et al. (2008); Quddus (2008); Sittikariya and Shankar (2009); Wang et al. (2009); Guo et al. (2010)

The details of these methods are discussed below.

2.3.1 Poisson Regression Model

The utility of fashionable regular least-squares regression, which assumes a non-stop based variable, isn't always relevant due to the fact crash-frequency information are non-terrible integers.

But in latest wondering with inside the area has used the Poisson regression version as a place to begin for the reason that the structured variable is a non-bad integer. The opportunity of roadway entity along with segment, intersection, etc. i having y crashes in step with a while period, wherein y is a non-poor integer is given via way of means of a Poisson regression version as:

$$P(y_i) = \frac{EXP(-\lambda_i)\lambda_i{}^{y_i}}{y!} \qquad (2.1)$$

Where P (yi) is the chance of roadway entity i having yi. Crashes in keeping with term and λi is the Poisson parameter for roadway entity i, that's same to roadway entity i's anticipated range of crashes consistent with year, E[yi]. Specifying the Poisson parameter λi may be used to estimate Poisson regression fashions. The predicted wide variety of crashes consistent with length as a feature of explanatory variables, the maximum not unusual place useful shape being λi =EXP (βXi), wherein X, is a vector of explanatory variables and β is a vector of estimable parameters.

Even though the Poisson version has served as a place to begin for crash-frequency evaluation for numerous decades, researchers have regularly discovered that crash information showcase traits that make the software of the easy Poisson Regression as problematic. Since, Poisson fashions cannot deal with over and beneath dispersion and that they may be adversely laid low with low pattern method and may produce biased consequences in small samples.

2.3.2 Negative Binomial Regression Model

The bad binomial or Poisson-gamma version is an extension of the Poisson version to triumph over feasible over dispersion with inside the data. The bad binomial/Poisson- gamma version assumes that the Poisson parameter follows a gamma possibility distribution. The version effects in a closed-shape equation and the arithmetic to govern the connection among the imply and the variance systems is rather simple. The terrible binomial version is derived through rewriting the Poisson parameter for every commentary iλi =EXP (βXi+£i) in which EXP (£i) is a gamma-allotted mistakes time period with suggest 1 and variance α. The addition of this time period lets in the variance to vary from the imply as $\boldsymbol{VAR\ [y_i]} = E\,[y_i][1 + \boldsymbol{\alpha} E\,[y_i]] = E\,\boldsymbol{[y_i]} + \boldsymbol{\alpha} E\,[y_i]$.The Poisson regression version is a restricting version of the terrible binomial regression version as αapproaches zero, this means that that the choice among those fashions depends upon the price of α.The parameter α is regularly called the over dispersion parameter. The Poisson-gamma/terrible binomial version is the in all likelihood the maximum often used version in crash-frequency modeling. However, the version does have its limitations, maximum extensively its lack of ability to deal with beneath dispersed statistics, and dispersion-parameter estimation troubles whilst the facts are characterized through the low pattern imply values and small pattern sizes (Fred L. Mannering, 2010).

Poisson-Lognormal Model for modeling crash data negative binomial/Poisson-gamma model can be used as an alternative, but in recent time some researchers have proposed using the Poisson-lognormal model (Miaou et al., 2003; Lord and Miranda-Moreno, 2008; Aquero-Valverde and Jovanis, 2008). Both the Poisson-lognormal version and the poor binomial/Poisson-gamma version are similar, however the EXP (£i) time period used to compute the Poisson parameter is lognormal in place of gamma distributed. Although the Poisson-lognormal doubtlessly gives greater flexibility than the terrible binomial/Poisson-gamma, it does have its limitations. Among these, small pattern sizes and coffee pattern suggest values adversely have an effect on the Poisson-lognormal and version estimation is extra complicated due to the fact the Poisson-lognormal distribution does now no longer have a closed form (Fred L. Mannering, 2010).

2.3.3 Zero-inflated Poisson and Negative Binomial

These kinds of fashions were evolved to address information characterized via way of means of a large quantity of zeros or extra zeros than the only might anticipate in a conventional Poisson or bad binomial/Poisson-gamma version. Zero inflated fashions perform at the precept that the extra 0 density that cannot be accommodated via way of means of a conventional matter shape is accounted via way of means of a splitting regime that fashions as crash-loose as opposed to a crash susceptible propensity of a roadway segment. The chance of a roadway entity being in 0 or non-0 states may be decided through a binary logit or probit version (Fred L. Mannering, 2010).

Since its inception, the 0-inflated version (each for the Poisson and bad binomial fashions) has been famous amongst transportation protection analysts (Fred L. Mannering, 2010). Despite its huge applicability to a lot of conditions in which the discovered statistics are characterized with the aid of using massive 0 densities; others have criticized the software of this version in motorway protection.

2.3.4 Conway-Maxwell-Poisson Model

The Conway-Maxwell-Poisson distribution is a generalization of the Poisson distribution and changed into first delivered with the aid of using Conway and Maxwell (1962) for modeling queues and carrier rates. Shmueli et al.(2005) in addition explored the statistical homes of the Conway-Maxwell-Poisson distribution, and

Kadaneet al. (2006) evolved conjugate distributions for the parameters. The Conway-Maxwell-Poisson can deal with each beneath dispersed and over dispersed facts, and numerous not unusual place opportunity density features are unique instances of the Conway-Maxwell-Poisson (for example, the geometric distribution, the Bernoulli distribution, and the Poisson distribution). This flexibility significantly expands the sorts of troubles for which the Conway-Maxwell-Poisson distribution may be used to version crash frequency records.

This version has been lately implemented in highway-protection research, and has been observed to be similar to the Poisson-gamma version for information characterized via way of means of over dispersion. However, its predominant benefit is associated with information characterized with the aid of using below dispersion. On the down side, this version may be negatively encouraged via way of means of low pattern mean, small-pattern bias and, to date, there have now no longer been any multivariate programs of the approach. (Fred L. Mannering, 2010).

2.3.5 Gamma Model

The gamma version has been proposed via way of means of Oh et al. (2006) to research crash statistics showing beneath dispersion (see additionally Cameron and Trivedi, 1998). This version can deal with over dispersion and below dispersion and decreases to the Poisson version whilst the variance. Please see Sellers and Shmueli (2010), Guikema and Coffelt (2008) and Lord et al.(2009) for in addition information at the estimation residences and programs of this version.is kind of same to the imply of the quantity of crashes. Although this version plays nicely statistically, it's far nonetheless a dual-kingdom version, with one of the states having a long-time period suggest identical to zero (see the preceding dialogue of zero-inflated models). The gamma version has visible constrained use because it became first brought through Oh et al. (2006).

2.3.6 Generalized Estimating Equation Model

The generalized estimating equation version has been carried out to toll road protection evaluation with the aid of using Lord and Persaud (2000) to version crash information with repeated measurements. As mentioned previously, one often has statistics from roadway entities (roadway segments or intersections) over a couple of time durations which installation a serial correlation problem (see Liang and Zeger, 1986).

The generalized estimating equation isn't always definitely a regression version consistent, however a technique used to estimate fashions with information characterized through serial correlation. The generalized estimating equation version gives unique processes to address serial correlation consisting of independence, exchangeable, dependence, and autoregressive kind 1 correlation structures. Usually, the correlation shape with inside the estimation system has a minimum have an effect on at the modeling output whilst count-facts fashions are used with a whole dataset (few if any ignored variables), however, the choice of the correlation kind may be important whilst the database has ignored variables (Lord et al., 2005a; Halekoh et al., 2006; Lord and Mahlawat, 2009).

2.3.7 Generalized Additive Models

The generalized additive version has extra flexibility than the conventional remember-facts fashions (see Hastie and Tibshirani, 1990; Wood, 2006). According to the examine made through Xie and Zhang (2008), generalized additive fashions offer a extra bendy purposeful shape which entails smoothing capabilities for the explanatory variables of the version. Even though, it's far often utilized in conventional remember fashions, the smoothing characteristic represents a greater bendy dating in how explanatory variables are taken under consideration and this now no longer restricted to linear or logarithm relationships.

Although the generalized additive version may be greater bendy than conventional matter fashions, there are nevertheless limitations. First, due to the fact those fashions consist of extra parameters; the estimation procedure can turn out to be very complex, in particular whilst the default values aren't used. Secondly, it is more difficult to interpret the generalized additive models relative to traditional count models.Third, the modeling results between generalized additive model and traditional models are likely to be similar if the explanatory variables are exogenous and the dependent variable has a linear or exponential relationship with them. Thus far, applications of generalized additive models to crash-frequency analysis have been limited to a few papers including Xie and Zhang (2008) and Li et al. (2009).

2.3.8 Random–Effects Models

As discussed at length previously, there may be reason to expect correlation among observations. This correlation could arise from spatial considerations (data from the same geographic region may share unobserved effects), temporal considerations (such as in panel data - where data collected from the same observational unit over successive time periods could share unobserved effects), or a combination of the two.

To account for such correlation, random-effects models (where the common unobserved effects are assumed to be distributed over the spatial/temporal units according to some distribution and shared unobserved effects are assumed to be uncorrelated with explanatory variables) and fixed effects models (where common unobserved effects are accounted for by indicator variables and shared unobserved effects are assumed to be correlated with independent variables) models can be considered. Hausman et al. (1984) first examined random-effects and fixed-effects negative binomial models for panel data (which has temporal considerations) in their study of research and development patents In the context of count models. Random-effects models rework the Poisson parameter as $\lambda_{ij} = EXP(\beta X_{ij}) EXP(\eta_j)$ where λ_{ij} is the expected number of crashes for roadway entity i belonging to group j (for example, a spatial or temporal group expected to share unobserved effects), X_{ij} is a vector of explanatory variables, β is a vector of estimable parameters, and η_j is a random-effect for observation group j.

The most common model is derived by assuming η_j is randomly distributed across groups such that $EXP(\eta_j)$ is gamma-distributed with mean one and variance α(see Hausman et al., 1984). As mentioned previously, the Poisson model restricts the mean and variance to be equal which in this case would be $E[y_{ij}] = VAR[y_{ij}]$. However, with random-effects the Poisson variance to mean ratio is $1 + \lambda_{ij}/(1/\alpha)$.

Random-results with inside the context of crash-frequencies were studied with the aid of using some of researchers along with Johansson (1996) (who studied the impact of a diminished pace restriction at the quantity of crashes on roadways in Sweden), Shankar et al. (1998) (who as compared trendy poor binomial and random-consequences bad binomial fashions in a take a look at of crashes as a result of median crossovers in Washington State), Miaou et al.

2.3.9 Negative Multinomial Models

The problem of correlation among observations can also be addressed with a negative multinomial approach (Guo, 1996). This model is similar to the negative binomial in that it uses $\lambda i = EXP\ (\beta X_i + £_i)$ except now $EXP\ (£_i)$ is associated with a specific entity (roadway segment, intersection) as opposed to a specific observation. This is an important distinction because, for example, if one is considering annual crash frequencies and with 5 years of data, each roadway entity will produce 5 observations that would each have their own $EXP\ (£_i)$ in a standard negative binomial, which would create a potential correlation problem. For the negative multinomial model, at the segment/intersection level, the $EXP(£_i)$ is again assumed a gamma-distributed error term with mean 1 and variance a. Shankar and Ulfarsson (2003) presented an application of the negative multinomial model to crash frequency data and compare estimation results to standard negative binomial and random-effects negative binomial models (see also Hauer, 2004; Caliendo et al., 2007). Negative multinomial models cannot handle under dispersion and are susceptible to problems in the presence of low sample means and small sample sizes.

2.3.10 Random-Parameters Models

Random-parameter fashions may be regarded as an extension of random-results fashions. However, instead of successfully simplest influencing the intercept of the version, random-parameter fashions permit every anticipated parameter of the version to differ throughout every person commentary with inside the dataset. These fashions try to account for the unobserved heterogeneity from one roadway web page to another (Milton et al., 2008).To allow for such random-parameters in count-data models, estimable parameters can be written as $\beta = \beta + \emptyset_i$ where $\emptyset_i$ is a randomly distributed term (for example a normally distributed term with mean zero and variance a^2). With this equation, the Poisson parameter becomes $\lambda i \backslash \emptyset_i = EXP\ (\beta X_i)$ in the Poisson model and $\lambda i = EXP(\ \beta X_i + £_i)$in the negative binomial/Poisson-gamma with the corresponding probabilities for Poisson or negative binomial now $P(y_i \backslash \emptyset_i)$. These models have been applied to crash-frequency data by Anastasopoulos and Mannering (2009) and El- Basyouny and Sayed (2009b). Because each observation has its own parameters, the final model will often provide a statistical fit that is significantly better than a model with traditional fixed parameters. However, random-parameter models are very complex to estimate, they may not necessarily improve predictive capability, and model results may not be transferable to other data sets because the results are observation specific (see Shugan, 2006; Washington et al., 2010).

2.3.11 Bivariate/Multivariate Models

Bivariate/Multivariate models become necessary in crash-frequency modeling when, instead of total crash counts, one wishes to model specific types of crash counts (for example, the number of crashes resulting in fatalities, injuries, etc.). Modeling the counts of specific types of crashes (as opposed to total crashes) cannot be done with independent count models because the counts of specific crash Types are not independent (that is, the counts of crashes resulting in fatalities cannot increase or decrease without affecting the counts of crashes resulting in injuries and no injuries).

To solve this Problem, Bivariate/multivariate fashions are used due to the fact they explicitly keep in mind the correlation the various severity levels (for example) for every roadway entity (Miaou and Song, 2005; Bijleveld, 2005, Song et al., 2006).

Bivariate fashions are used for together modeling crash types (Subrahmainiam, 1973; Maher, 1990; N'Guessan et al., 2006; Geedipally and Lord, 2009; N'Guessan, 2010).Extensions to extra than crash types (multivariate-version formulations) were proposed which includes the multivariate Poisson version (Ma and Kockelman, 2006), the multivariate poor binomial version (Winkerlman, 2003), and the multivariate Poisson-lognormal version (Park and Lord, 2007; Ma et al.,2008; El- Basyouny and Sayed, 2009a; Park et al., 2010).   On the downside, bivariate/multivariate fashions are complicated to estimate in that they require a system of a correlation matrix.

2.3.13 Finite Mixture/Markov Switching Models

Finite aggregate/Markov switching fashions are a brand new sort of version that may be used to observe heterogeneous populations. Although this sort of version has been round for a few time, they have got lately turn out to be extra famous due to the development in computing energy and technology (Fruhwirth-Schnatter, 2006). For finite aggregate fashions, the idea is that the general facts are generated from numerous distributions which are combined collectively implying that man or woman observations are generated from an unknown quantity of subgroups. Markov Switching fashions additionally paintings on the idea that some of underlying distributions generate the records and that person observations can transfer amongst those distributions over time.

In current years, some researchers have tested the software of finite combination fashions (Park and Lord, 2009; Park et al., 2009) and Markov switching fashions (Malyshkina et al., 2009; Malyshkina and Mannering, 2010a) to motorway safety. Finite combination and Markov switching fashions provide substantial capability for offering vital new insights into the evaluation of crash statistics, however those fashions also are pretty complicated to estimate.

2.3.14 Duration Models

Another way of framing the crash-frequency problem is to consider the time between crashes, as opposed to the frequency of crashes over some time period. The frequency of crashes and the time between crashes are obviously interrelated. In fact, count-data models (such as the standard Poisson Model) imply an underlying distribution of time between crashes (for the standard Poisson the underlying time distribution is exponentially distributed), and a model of the duration of time between crashes can be aggregated to produce an expected frequency in any given period of time.

The maximum not unusual place period-version technique is a hazard-primarily based totally version that considers the conditional the chance of a crash going on at a while t +dt for the reason that it's been time t because the final crash occurred. Hazard-primarily based totally fashions may be predicted beneath an extensive type of distributional assumptions and non-parametric bureaucracy and permit critical inferences to be made on how the chances of getting a crash extrude over time (see Washington et al., 2003, 2010).

Hazard-primarily based totally period fashions may be pretty state-of-the-art in phrases in their capacity to deal with records and not unusual place troubles related to crash facts (unobserved heterogeneity, etc.) and may provide insights referring to length results. As an example, rather than thinking about the range of crashes person drivers had over their lifetimes, Mannering (1993) used a hazard-primarily based totally period version to examine the elements affecting the time among their crashes. The consequences of this examine observed exciting period outcomes in that the longer men went while not having a crash, the much less possibly they have been to have a crash soon, however that the duration of time girls went without a crash turned into observed to don't have any statistically tremendous impact on their crash possibilities. There is considerable potential for the future application of duration models to crash frequency analysis, but the level of data required in terms of the timing of crashes and the values of explanatory variables and how they change over time can be prohibitive in many instances.

2.3.15 Hierarchical/Multilevel Models

Hierarchical models are used for analyzing data that are characterized by correlated responses within hierarchical clusters. In highway safety, crash data could be seen as exhibiting several levels of hierarchy. For instance, the lowest level of the hierarchy could be the crashes themselves. Then, the next level could be the type of vehicle (passenger cars, trucks, etc.). For the subsequent one, it could be the accident location on the transportation network, and so on. With this type of model, the primary assumption is that correlation may exists among crashes occurring for the same kind of vehicle and location, because they may share unobserved characteristics related to the vehicle type or location. One may lead to poorly estimated coefficients and associated standard errors, particularly when they are modeled using a traditional count-data modeling approaches, if Not considering the potential hierarchical structure of the data (the potential of a complex correlation structure) (Skinner et al., 1989; Goldstein, 1995). On the other hand, depending upon the study objectives, these models may not be warranted, even if correlations considered are not large (de Leeuw and Kreft, 1995) and the modeling output may be difficult to interpret, especially by non-statisticians (Pietz, 2003). There have been a number of applications of hierarchical models to crash data (Jones and Jorgenson, 2003; Kim et al., 2007).

2.3.16 Neural, Bayesian Neural Network, Support Vector Machine Models

Neural and Bayesian neural community fashions are capabilities which might be described the usage of multilevel community structures (Liang, 2005). The community shape includes a sequence of nodes and weight elements that hyperlink the diverse nodes collectively in hierarchical manner: enter layer, hidden layer, and output layer. Although each Neural and Bayesian neural community fashions have comparable modeling processes, they may be virtually unique with inside the manner they are expecting the final results variables. The weights are assumed constant for neural networks, while it follows a opportunity distribution for Bayesian neural networks, and the prediction procedure wishes to be incorporated over all of the chance weights.

These fashions were utilized in dual carriageway safety (Abdelwahab and Abdel-Aty, 2002; Chang, 2005; Riviere et al., 2006; Xie et al., 2007) especially as a predictive tool. Overall, those fashions have a tendency to showcase higher linear/non-linear approximation houses than conventional count-version approaches. However, those fashions regularly cannot be generalized to different information sets (Xie et al., 2007). Support vector system fashions that are primarily based totally on statistical studying idea are a brand new elegance of fashions that may be used for predicting count-frequencies (Kecman, 2005).

These fashions are a fixed of associated supervised gaining knowledge of strategies used for type and regression. For any preferred diploma of accuracy those fashions additionally own the famous capacity of being capable of approximate. Structural hazard minimization and Statistical studying concept are the theoretical foundations for the gaining knowledge of algorithms of aid vector system fashions. It has been discovered that those fashions display higher or similar effects to the results predicted through neural networks and different statistical fashions (Kecman, 2005).

Support vector device fashions have lately been brought for different transportation applications (see Zhang and Xie, 2007), inclusive of for predicting crashes (Li et al., 2008). However, they're complicated to estimate and, like neural and Bayesian neural community fashions, those fashions frequently can’t be generalized to different information sets. Another standard complaint of neural, Bayesian neural community, and guide vector device fashions is they all have a tendency to act as black-containers in that they do now no longer offer the interpretable parameters one receives whilst the usage of conventional crash-frequency fashions.

2.3.17 Parameter Estimation Methods

Maximum probability estimation and Bayesian techniques are the 2 maximum not unusual place techniques used for crash-frequency fashions. The predominant benefit of the most chance estimation is that closed shape capabilities frequently exist for the maximum not unusual place distributions used. On the opposite hand, most chance estimation cannot be used while the chance feature is hard to characterize.

Bayesian estimating techniques had been gaining in reputation because of advances in computing strategies (Gilks et al., 1996). Bayesian fashions have the benefit of being capable of deal with very complicated fashions, in particular the ones that don't have without problems calculable probability capabilities. Using Markov Chain Monte Carlo (MCMC) strategies, a sampling- primarily based totally technique to estimation this is properly suitable for Bayesian fashions; complicated practical version bureaucracy may be handled. For instance, random-parameter and Markov switching fashions are extra without difficulty envisioned the use of MCMC simulation. However, despite the first-rate computational advantages supplied with the aid of using Bayesian fashions, the simulation time for the Markov Chain Monte Carlo simulation can nevertheless be a barrier to complicated version bureaucracy. The simulation time, that's a feature of the scale of the pattern and the complexity of the version structure, can take numerous days and this time-difficulty can nevertheless be a proscribing element with inside the complexity of the version.

CHAPTER THREE: MATERIALS AND METHODOLOGY

3.1 Organization of the Report

This report was organized in to four sections according to the different classes of statistical methodologies applied in each study. For each study; 1st overview of the study's (article's) purpose were identified. Secondly the methods and procedures done by the author, how the data was collected, who was included in the sample, and any instruments used, sample size, demographic characteristics, or any interesting protocol were identified. In the third section the reader focuses on the findings in the study. Finally from the discussion part look for the author's critique of why the study did or did not produce results, the conclusion drawn, an unexpected influence in the findings, and the author's future line of research or "next steps" to improve the body of knowledge were assessed.

CHAPTER FOUR: RESULTS AND DISCUSSION

4.1 STUDY ONE

4.1.1 Overview of the Study (Article)

The study was "A Parametric Model for Accident Prediction along Ado Ekiti-Ikole Ekiti Road, Ekiti State, Nigeria" by Olumuyiwa S.Aderinola.The studies seemed in growing a parametric version for predicting injuries at particular places alongside Ado-Ekiti to Ikole-Ekiti road. The twist of fate inclined factors alongside the street and the elements inflicting to the incidence of injuries had been diagnosed via reconnaissance survey of the look at area.

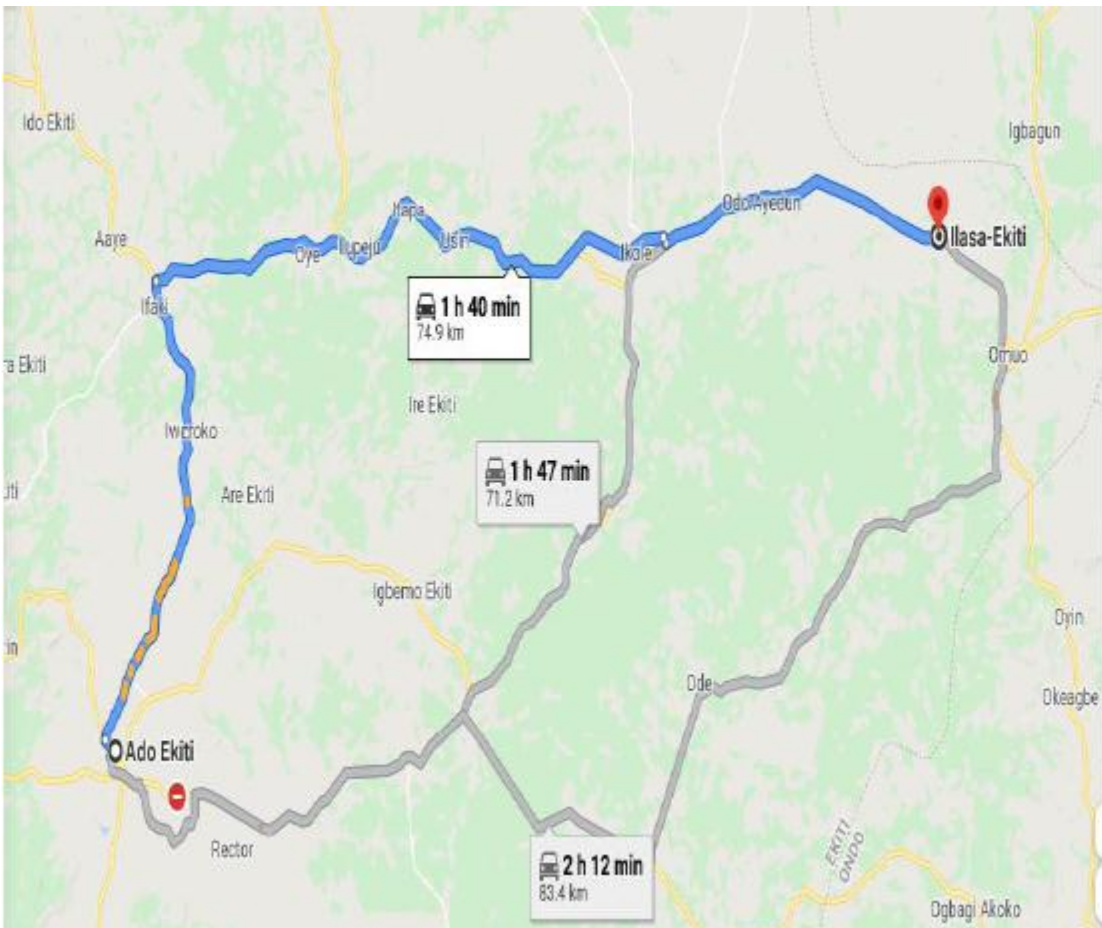

Figure 1 Ado-Ekiti to Ikole-Ekiti road Network

The factors considered in this study were Spot speed [S],Pavement condition[P], Condition of shoulder[C],Width of the Road[W],Elevation(super)/cambering[E],Gradient[G] and accident Vulnerability [AV] which form an acronym SPCWEG-AV. The paper was published on EJER, European Journal of Engineering Research and science on August 31, 2020. The author focused on five key words as parametric model, Ado Ekit to Ikole-Ekiti Road, Road accident and Total SPCWEG-AV.

4.1.2 Methods and Procedures in the Study

In this study the accident prediction modeling would be based on parametric more of geometric parameters were needed apart from the human factor, environmental factor and vehicle factor. The materials used here to determine those parameters were twine/rope, stop watch, plumb and car. The twine was used to determine the elevation (super/cambering across the road and the gradient along the road; the stop watch was used to determine the time taken from a particular vehicle to cover a given distance within the road; plumb was used to determine the perpendicularity (horizontal accuracy) of the twine in order to determine elevation (cross slope) or gradient along the road; and car was used to estimate the pavement and shoulder conditions.

Road traffic data was collected from Federal Road Safety Corps and the Nigeria Police Force, Ekiti State command. These data were used to identify the accident prone location the study area of Ado-Ikole highway. The five years recent data from 2014 to 2019 G.C were used by the author to model the accident prediction in the study area. Since accident prone locations are areas where accidents have occurred frequently, the data collected was analyzed to determine the accident pattern in the study area.

Once, the writer diagnosed the susceptible areas, then the spot velocity in every of the places become gotten through measuring a 60m duration and noting the time automobiles blanketed the distance. The pavement condition, width, extremely good elevation and gradient (horizontal) have been measured the use of car, tape, cord and plumb. Then the analyzed data from the investigated factors from the field were imputed into SPCWEG-AV Rating system and weights to get the index of each parameter. Then the summation of these indices gives what is called Total SPCWEG-AV index abbreviated as (T. SPCWEG-AV.I). This index defines the degree of accident vulnerability of the point in question.

Here for this study, the author designed a rating and weighting method for the ten locations. The rating system ranges from 1 to 5 for each variable and the weight system ranges from 1 to 6 in their order of significant contribution to road traffic crashes. Once, the author identified the prone areas, then the spot speed in each of the locations was gotten by measuring a 60m length and noting the time vehicles covered the distance. The pavement condition, width, super elevation and gradient (horizontal) were measured using car, tape, twine and plumb. Then the analyzed data from the investigated factors from the field were imputed into SPCWEG-AV Rating system and weights to get the index of each parameter. Then the summation of these indices gives what is called Total SPCWEG-AV index abbreviated as (T. SPCWEG-AV.I). This index defines the degree of accident vulnerability of the point in question.

Here for this study, the author designed a rating and weighting method for the ten locations. The rating system ranges from 1 to 5 for each variable and the weight system ranges from 1 to 6 in their order of significant contribution to road traffic crashes.

Finally, the mathematical model for SPCWEG-AV accident venerability evaluation could be expressed as

$$\mathrm{T.SPCWEG-AV.I} = SrSw + PrPw + CrCw + WrWw + ErEw + Grw \qquad (1)$$

Where,

Sr = Rating assigned to Spot speed

Sw =Weight assigned to Spot speed

Pr = Rating assigned to Pavement condition

Pw= Weight assigned to Pavement condition

Cr = Rating assigned to Condition of shoulder

Cw = Weight assigned to Condition of shoulder

Wr = Rating assigned to Width of the road and shoulder

Ww = Weight assigned to Width of the road and shoulder

Er =Rating assigned to Elevation (super)/cambering

Ew = Weight assigned to Elevation (super)/cambering

G = Rating assigned to Gradient

Gw = Weight assigned to Gradient

AV= Accident Vulnerability

The higher the SPCWEG-AV Index, the greater the accident proneness at a location. The SPCWEG-AV can be further divided into four categories: low, moderate, high and very high.

The weights assigned to these parameters as shown in (2):

$$\mathrm{T.SPCWEG} - \mathrm{AV.I} = 6\mathrm{Sw} + 5\mathrm{Pw} + 4\mathrm{Cw} + 3\mathrm{Ww} + 2\mathrm{Ew} + \mathrm{Gw} \qquad (2)$$

Where,

6 = weights of spot speed

5 = weights of pavement condition

4 = weights of condition of shoulder

3 = weights of width of the road

2 = weights of elevation (super)/cambering

1 = weights of gradient of the road

To check the validation and calibration of the model accurately the author used the collected and computed traffic accident on SPCWEG-AV index.

4.1.3 Data Analysis in the Study

Table 3 SPCWEG-AV Rating System and Weights

Parameters	Conditions	Classification	Range	Rate	Weight
Spot speed (S)	Slow	Very Good	0-30	1	6
	Moderate	Good	30-60	2	
	Average	Fair	60-90	3	
	Fast	Poor	90-120	4	
	Very Fast	Very Poor	120-150	5	
Pavement Condition (P)	Structurally ok	Very Good	0-20	1	5
	Crack/minor dent	Good	20-40	2	
	Isolated Potholes	Fair	40-60	3	
	Wavy/Heavy Surface	Poor	60-80	4	
	Shera/Massive Failure	Very Poor	80-100	5	
Condition of Shoulder (C)	Clean/Clear	Very Good	0-10	1	4
	Bushy	Good	10-20	2	
	Small Width	Fair	20-30	3	
	Eroded	Poor	30-40	4	
	Absent	Very poor	40-50	5	
Width of Pavement/Shoulder (W)	Too Small	Very poor	0.0-2.8	5	3
	Small	Poor	2.8-5.6	4	
	Normal	Fair	5.6-8.4	3	
	Wide	Good	8.4-11.2	2	
	Wider	Very Good	11.2-14.0	1	
Elevation(Super) /Cambering (E)	Very Bad	Very poor	0.00-0.75	5	2
	Bad	Poor	0.75-1.50	4	
	Fair	Fair	1.50-2.25	3	
	Good	Good	2.25-3.00	2	
	Very Good	Very Good	3.00-3.75	1	
Gradient of Pavement (G)	Normal	Very Good	0-3	1	1
	Moderate	Good	3-6	2	
	Fair	Fair	6-9	3	
	High	Poor	9-12	4	
	Very High	Very poor	12-15	5	

From the study area of about 64.8 km road the researcher identified ten prone locations from the data acquired and physical inspections. Table 1.2 shows the station and total number of accidents in each location from the data obtained with concerned body.

Table 4 Accident Vulnerable Locations along Ado-Ekiti to Ikole-Ekiti Road

Location	Chainage	Year	Total number of Crash
FGC,Ikole Ekiti	CH 0+000	2014-2019	21
NNPC, Ikole Ekiti	CH 3+200	2014-2019	16
Olokonia, Ikole Ekiti	CH 7+000	2014-2019	20
NPS,Oye Ekiti	CH 23+200	2014-2019	12
FUOYE,Oye Ekiti	CH 25+600	2014-2019	24
Ifaki Ekiti	CH 35+400	2014-2019	21
Iworoko Ekiti	CH 52+100	2014-2019	10
Iworoko Market	CH 53+100	2014-2019	13
EKSU,Iworoko Ekiti	CH 62+750	2014-2019	18
Irasa Ekiti	CH 64+800	2014-2019	22

Based on these locations and the assigned rating system the data were investigated for each location separately to get the SPCWEG-AV index from field data on the parametric model. Sample calculation shown below for location 1 was done here, in similar way the ten prone locations' SPCWEG-AV index was computed.

Table 5 SPCWEG-AV index Computation of Location 1 (FGC, Ikole-Ekiti)

Parameters	Field Data	Ra*	Wt	SPCWEG-AV.I
Spot speed (S)	48.39km/hr (moderate)	2	6	12
Pavement Condition (P)	60 (Wavy/Heavy Surface)	4	5	20
Condition of Shoulder (C)	30 (Eroded)	4	4	16
Width of Pavemen/Shoulder (W)	6.8m (normal)	3	3	9
Elevation(Super) /Cambering (E)	12% (Very Bad)	5	2	10
Gradient of Pavement (G)	5.83% (Good)	2	1	2
T.SPCWEG-AV I				69

4.1.4 Discussion of Findings in the Study

Accident vulnerability were categorized into six ranges based on SPCWEG-AV index as very low, low, moderate, high, very high and dangerously high with total SPCWEG-AV indices are between 0-21,22-31,32-41,42-51,52-1 and >62 respectively. According to this categorization, it is seen that locations 1, 2, 3, 5, 6, 9 and 10 had high possibility of accident occurrence with total SPCWEG-AV index greater than 62 and categorized as dangerously high. Whereas locations 4 and 7 were categorized as very high accident with total SPCWEG-AV indices 59 and 42 respectively.

In this study the author did a comparison between road traffic accidents data and computed total SPCWEG-AV index to check the parametric model was fit with accident data recorded in the study area. So location 5 had the highest accident occurrence i.e. 24 times each and also had highest T.SPCWEG-AV index of 71. Similar pattern runs through all the investigated sections of the road. It is therefore, reasonable to conclude that the parametric model can replicate and predict the occurrence of accidents along Ado-Ekiti to Ikole-Ekiti road and other roads where similar conditions of the highway occur.

Depending on the parameters considered in the study the analysis and results led to the development of total SPCWEG-AV index which shows the degree of accident vulnerability. The higher the index, the higher the index, the more vulnerable a section of the road is and thus the parametric model was developed. This parametric model showed very good correlation between the indices of the results of this research revealed that eight (8) out of the ten (10) locations have dangerously high accident vulnerability while one has very high accident vulnerability and another one has high accident vulnerability.

4.2 STUDY TWO

4.2.1 Overview of the Study (Article)

The study become "Statistical Model of Road Traffic Crashes Data in Anambra State, Nigeria: A Poisson Regression Approach "through Nwankwo Chike H. and Nwaigwe Godwin I., which gives the evaluation of statistics on avenue visitors crashes from the Anambra State Command of the Federal street Safety Commission (FRSC), Nigeria via way of means of the usage of Generalized Poisson Regression (GPR) and the Negative Binomial Regression (NBR) models. The paper become posted at the International Journal of Scientific & Technology Research in 2015 G.C with 5 key phrases. These 5 key phrases had been over dispersion, Road visitors, Crashes, Discrete and Akaike Information Criterion (AIC).

4.2.2 Methods and Procedures in the Study

In this study the road traffic crashes were considered as discrete or count data in nature. In order to modeling discrete data for characteristics and prediction of events, the researcher was used two approaches in this study. The data were collected over a fixed continuous space which is Anambra state, Nigeria and over a fixed time. So using Poisson distribution is appropriate in modeling of data in road traffic crashes.

The causes of accidents were determined based on the Federal Road Safety Commission in Nigeria as Speed Violation, Loss of Control, Dangerous Driving, Tire Burst, Brake Failure, Wrongful Overtaking, Route Violation, Mechanically Deficient Vehicle, Bad Road, Road Obstruction Violation, Dangerous Overtaking, Overloading, Sleeping on Street, Driving Under Alcohol, Use of Phone While Deriving, Fatigue, Poor Weather, Saign Light Violation and others. So the accident may be due to one or multiple causes as listed above. In this case the data were presented as 1 or ≥ 2 to indicate the cause of the accident is one and could be either of the above lists. And ≥ 2 means that there are two or more causes of accidents and so on as presented in table form. Once the data were organized in the manner as number of crashes, persons involved, season and number of causes in table column form it could be easy for analysis.

4.2.2.1 Negative Binomial Regression (NBR)

Negative binomial distribution certainly considered one among an critical software that is a combination and own circle of relatives of Poisson distributions with gamma blending weights. Thus this approach may be considered as a generalization of the Poisson distribution. Since the Poisson parameter is a random variable through itself, the poor binomial may be considered as a Poisson distribution in step with a Gamma distribution. Due to this the bad binomial distribution is called a Poisson-Gamma aggregate. The poor binomial regression addresses the difficulty of over-dispersion through together with a dispersion parameter to house the unobserved heterogeneity with inside the data, because the maximum not unusual place opportunity to Poisson regression. The bad binomial regression (Poisson-Gamma) also can be taken into consideration a generalization of Poisson regression. As its call implied, the bad binomial (Poisson-gamma) is a aggregate of distributions and became first derived via way of means of Greenwood and Yule (1920).

It have become very famous due to the fact the conjugate distribution (identical own circle of relatives of functions) has a closed shape and ends in the bad binomial distribution. As mentioned through Cook (2009), "the call of this distribution comes from making use of the binomial theorem with a bad exponent". This aggregate distribution become advanced to account for over-dispersion this is typically located in discrete or rely data (Lord et al (2005).

$$f\left(y_{i,}\lambda_{\mathrm{i},}\right) = \frac{\mathrm{e}^{-\lambda_{\mathrm{i}}}\lambda_{\mathrm{i}}{}^{y}}{y_i},\ \ \mathrm{y}\ \geq\ 0 \text{ and } \quad \lambda\ \geq\ 0 \qquad (3)$$

Where, $\mathrm{y_i}$ is the observed number of counts for =1, 2… and

λ_{i} is the mean of the Poisson distribution

4.2.2.2 The Generalized Poisson Regression Model (GPRM)

The advantage of using the generalized Poisson regression model is that it can be fitted for both over-dispersion as well as under-dispersion. The generalized Poisson distribution is a natural extension of the Poisson distribution.

$$\sum_{\mathrm{i=1}}^{\mathrm{n}} \frac{(\mathrm{yi}-\mu\mathrm{i})^2}{\mu\mathrm{i}(1+\alpha\mu\mathrm{i})^2} = \mathrm{n}-\mathrm{p} \qquad (4)$$

Where n is the number of values and P is the number of regression parameters.

4.2.2.3 Multi co-linearity Test

Multi co-linearity (additionally co-linearity) is a statistical phenomenon wherein or greater predictor variables in a a couple of regression version are particularly correlated, which means that one may be linearly expected from the others with a non-trivial degree. In different words, multi co-linearity is stated to have befallen if or extra unbiased variables are pretty correlated. The multi co-linearity check is carried out as a preliminary assumption for parameter estimation. One formal manner of detecting multi co-linearity is with the aid of using the usage of the variance inflation factors (VIF). The VIF is used to check for the presence of multi co-linearity, and is given via way of means of

$$\mathrm{VFI} = \frac{1}{1-{R_j}^2} \qquad (5)$$

Where ${R_j}^2$ is the coefficient of determination of a regression of an explanatory variable j on the other entire explanatory variable? A VIF value of 10 and above indicates a multicollinearity problem (Wikipedia.org).

4.2.2.4 Akaike Information Criterion (AIC)

The Akaike statistics criterion (AIC) is a degree of the relative excellent of a statistical version for a given set of data. That is, given a set of fashions for the data, AIC estimates the great of every version, relative to every of the opposite fashions. Hence, AIC presents a way for version selection. Given a fixed of candidate fashions for the data, the favored version is the only with the minimal AIC value (i.e., the smaller the AIC value, the higher the version). Hence, AIC does now no longer most effective praise goodness-of-healthy; however additionally consists of a penalty this is an growing characteristic of the quantity of expected parameters (Wikipedia.org). This degree additionally makes use of the log-probability characteristic; however upload a penalizing time period related to variety of variables. It is widely recognized that via way of means of including variables, you'll enhance the match of fashions. Thus, the AIC attempts to stability the goodness of-match as opposed to the inclusion of variables with inside the version. The AIC is computed as (Wikipedia.org).

$$\mathrm{AIC} = 2P - 2lnL \qquad (6)$$

Where P is the number of parameters in the model and L is the maximized value of the likelihood function for the model.

4.2.3 Data Analysis and Results

The data were analyzed using R Software and the results obtained are given below. Before performing the analysis on the two methods used, testing the data for multi co-linearity was conducted. The test results are shown in table 6 below:

Table 6 Co-linearity Statistics

Model	Tolerance	VIF
Number of Crashes	0.486	2.059
Season (weeks of the year)	0.995	1.005
Number of causes	0.485	2.063

Table 6 shows that all the variables have VIF values <10. Thus all the variables can be included in the subsequent analyses and modeling with the Generalized Poisson regression, and Negative Binomial Regression.

4.2.4 Modeling and Findings in the Study

Table 7 Negative Binomial Regression Model Parameter Estimation

Parameters	Estimate	Standard error	Z value	Pr(>|Z|)
Intercept	2.017437	0.128009	15.760	<2e-16
Number of Crashes	0.119212	0.025182	4.734	2.20e-06
Season (weeks of the year)	-0.002007	0.003494	-0.574	0.566
Number of Causes	0.155004	0.037553	4.128	3.67e-05

Table 8 AIC Value of Negative Binomial Regression Model Parameter Estimation

DEVIANCE	530.58
AIC	2742

Table 9 Generalized Poisson Regression Model Parameter Estimation

Parameters	Estimate	Standard error	Z value	Pr(>|Z|)
Intercept	2.1462560	0.0782676	27.4220	<2e-16
Number of Crashes	0.1005432	0.0178925	5.6193	<2e-16
Season (weeks of the year)	0.0033059	0.0028278	1.1691	0.00710
Number of Causes	0.1237426	0.0264850	4.6722	0.0841

Table 10 AIC Value of Generalized Poisson Regression Model Parameter Estimation

DEVIANCE	4.2710
AIC	3508.595

To conclude that the GPR and NBR were conducted to determine the better model to use in modeling data on number of road traffic crashes within the Anambra state Command of Federal Road Safety Commission, Nigeria. The criterion for selection of the best model used is the AIC. The best model is that with the smallest AIC value. This happened to be the NBR model. The deviance for the Generalized Poisson Regression models is larger than the deviance of the Negative Binomial Regression thus indicating the existence of significant over-dispersion if the Generalized Poisson Regression models Model were adopted. To test for over-dispersion, AIC of Poisson against Negative Binomial Regression model and Generalized Poisson Regression are obtained and compared. On the basis of the AIC values in tables 8 and 10 the estimated AIC for GPR model is 3508.595 whereas it is 2742 for the NBR model. The smallest AIC value is that of the negative binomial regression model. Therefore, the best model for the number of road traffic crashes in Anambra state road safety command is best modeled and described using the negative binomial regression model.

4.3 STUDY THREE

4.3.1 Overview of the Study (Article)

The study was "Accident Prediction Models for Two-Lane Rural Highways "by K.R.Kalokota, which aimed at modeling the influence of the geometric design variables on traffic accidents in a selected two lane rural highways of highway 89 and 91 in northern Utah, USA. The "Proc GLM" procedure in SAS (version 6.04) was employed to test the statistical significance of the variables in the study. The primary objective of this research was to establish statistically valid mathematical models that can explain the variations in accidents as a function of geometric variables. The second objective was to investigate the transferability of model parameters over time and space and develop a methodology to simplify the transfer process.

4.3.2 Methods and Procedures in the Study

In order to meet these objectives the author develops a methodology undertaken by dividing into seven sections. The first task was review previously identified geometric variables influencing the occurrence of accidents in literatures. So in this part the author conducted a thorough literature search and reviewed on several articles on accident prediction models and geometry collected here.

The second venture became gathering and analyzes statistics. Here the vital information for the have a look at location has been accumulated from the involved organization. The Highway Safety Information System (HSIS) facts for an 87-mile phase of dual carriageway 89 and ninety one in Utah had been acquired from the Federal Highway Administration (FHWA) and the information have been decoded in line with the look at needs. Therefore the writer recognized the importance of geometric variables recognized in challenge one through plotting injuries Observed with inside the check phase towards sure geometric variables. In the third project, discover statistically considerable geometric variables with inside the take a look at corridor. The "Proc GLM" methods in SAS (model 6.04) have been hired to check the statistical importance of the variables. GLM manner calculates the Student t-values for every variable and exams the importance at the specified level. The multi-co linearity of the variables became tested the use of SAS package. The fourth task in this study was focused in examine of previous accident prediction models on the basis of underlying statistical modeling principles. During the literature search the author looks that the forms of models and the statistical procedure for calibrating models differed from researcher to researcher. In this section the model form and the calibrating procedure best suited for this study corridor were determined.

Task five in the study was standing for calibrating and verifying the selected model forms. Here the forward stepwise regression models were calibrated using 1985-89 of five year data. In this study three separate models were calibrated i.e. for the curve sections, tangent sections, and the entire corridor respectively. And also these models were verified by using them to predict the accidents in 1990.

In the sixth section the study more focus on comparison of the calibrated model parameters/coefficients with parameters of the models from other studies to determine the potential for model transfer. The models of other researchers were employed to predict accidents at the study site so that the transferability and validity over space and time could be assessed.

In the last section of the methods in this study, the task was examine the procedures available for updating parameters and tests the accuracy of updated models. Different literatures search that several parameters updating techniques can be employed to enhance the predictive accuracy of transferred models. Here the Bayesian parameter updating approach was chosen and employed to update the parameters of the models calibrated in northern Utah by combining with sample data from southern Utah. The updated models were then tested for the spatial transferability. The temporal transfer was tested by updating the northern Utah models by combining with cross- sectional data from the same test section.

4.3.3 Data Collection and Analysis in the Study

4.3.3.1 Study Area

Among the five states that currently have HSIS for analysis, Utah is considered to be the state that has the most complete information on highway geometries. Therefore, for this study, an 87-mile segment of US Highway 89 and 91 from Brigham City to Bear Lake in northern Utah was selected by the author.

The study section from Brigham City to the city of Logan is a part of US 91 and it passes through the Sardine Canyon. The Logan to Bear Lake section of US 89 passes through the Logan Canyon. The study corridor of US 91 starts at mile point zero at Brigham City and ends at mile point 45.20 in the city of Logan. US 89 starts at mile point 374.24 in Logan and ends at mile point 415.84 at Bear Lake.

To limit the study to two-lane rural sections, the sections with four lanes, and sections in the City of Logan and to the immediate south were deleted. Only the sections from mile point 3.78 to mile point 15.75 on US 91 and sections from mile point 374.24 to mile point 376.95, mile point 377.31 to mile point 411.77, and mile point 411 to mile point 415.84 on US 89 were considered.

There were a few sections in the study corridor with climbing (auxiliary/truck) lanes. Since climbing lanes are an integral part of the two-lane highways in mountainous terrain, they were considered to be a variable design option and were included in the data set for further analysis.

The final database used in this study was 296 sections in all, made up of 143 tangent sections and 153 curved sections.

Data pertaining to 6 years (1985-1990) were made available by the FHWA from their Highway Safety Information System (HSIS) records for Utah. They were contained in four separate files: grade, road, curve, and accidents.

The grade file contained 1987 information on vertical grades in the form of section length in miles, beginning mile point, end mile point, percent grade, direction of grade, and route number. The road data file consisted of information on AADT, access, beginning point, end point, county, weighted design speed, federal aid system, functional classification, lane width, median type, number of lanes, percent of trucks in off peak, one-way or two-way factor, section length pavement condition, percent passing distance, present service rating, percent of commercial vehicles in peak, accident year, route number, type of sign posted, section length in miles, shoulder type (right side), speed limit, state code, and terrain type.

The curve data file contained 1987 information on length of curves in miles, route number, beginning and ending mile points of curve, direction of curve, and degree of curvature. The accident files consisted of information for the years 1985 through 1990. The information included the accident route code, accident route, accident mile point, accident year, accident month, accident day, time, number of vehicles involved in the accident, accident severity index, accident type in coded form, light condition in coded form, weather, collision type in coded form, object struck in coded form, road effect in coded form, number of injured in accident, accident severity, total people involved, traffic control, time of accident, accident recoded by, people treated by, and accident mile point. The coding was done as per accident report form standards.

4.3.3.2 Compilation of Data Files

For the purposes of this study, the curve file, road file, and grade files are combined manually into one file and named as the "roadway inventory" file. This file consisted of information on section length in feet, section length in miles, route number, degree of curvature, direction of the curve (to the left or right), beginning and ending mile points, AADT, number of lanes, pavement type, pavement condition, and percent grade.

The three accident files containing records from 1985-87, 1988-89, and 1990 were then combined with the roadway inventory file so that accidents in each section matched with the road characteristics in that section. This became a long and tedious process for several reasons.

First, because accidents were coded as separate entries in the HSIS, and a given section could have several accidents of different types in a given year. The number of accidents during each of the 6 years (1985-90) had to be aggregated and entered separately in a new column so that they correspond to the appropriate road section.

Second, some of the variables were at times defined by characters and other times defined by numbers or blank spaces. To overcome these inconsistencies, the data base was recoded and rearranged.

The programs that combined roadway inventory file and 1985-87, 1988-89, and 1990 accident file were done with some of the variables that are recoded as zero when information was missing.

4.3.4 Description of Variables in the Data Files

4.3.4.1 Section Length (L)

Table 11 Distribution of section lengths and accidents on tangents and curves in percentage

Tangent Section Length (feet)	% of Total Tangent Sections	% Total Accidents on Tangents	Curve Sections Length (feet)	% of Total Curve Sections	%Total Accidents on Curves
1000	70.63	15.64	1000	65.35	53.73
2000	11.89	7.50	2000	22.88	33.58
3000	8.39	8.80	3000	11.77	12.70
4000	1.40	0.70			
5000	0.70	0.84			
6000	2.80	10.61			
7000	1.40	17.04			
8000	0.70	1.82			
9000	0.00	0.00			
10,000	0.00	0.00			
11,000	0.70	10.75			
12,000	0.00	0.00			
13,000	0.00	0.00			
14,000	0.00	0.00			
15,000	0.00	0.00			
16,000	0.00	0.00			
17,000	0.00	0.00			
18,000	0.70	4.89			
19,000	0.70	21.37			

Section lengths were given in miles in the database. The distribution of the lengths of all the tangents and curves was shown in Table 11.

The test corridor consisted of 296 sections, of which 143 were tangent sections and the rest were curve sections. It can be seen from the Table above that over 70% of the tangent sections in the corridor were 1,000 feet or less. As for curves, over 60% were 1,000 feet or less, and the curved segments made up of approximately 32% of the corridor length. One important feature that was missing from the data was information on spirals. The spirals were contained in the tangents without being identified. But because terrain within the corridor does not permit long spirals, errors due to this aggregation could be regarded as minimal.

The 5-year (1985-89) accidents in percentages for each of the above roadway categories are also shown in Table 1. On the tangents, the largest share of accidents (approximately 30%) occurred in sections less than 3,000 feet. The next clustering was observed in the 5,000 to 7,000 feet sections.

On the curves, over 50% of the curve accidents occurred on sections less than 1,000 feet. However, in relation to the proportion of sections in each of the above categories, the propensity is for accidents to increase as section length increases. This can be seen from the ratio of the total number of accidents per million vehicles to the total number of sections in the interval of 0.5 miles as shown in Figure 1.

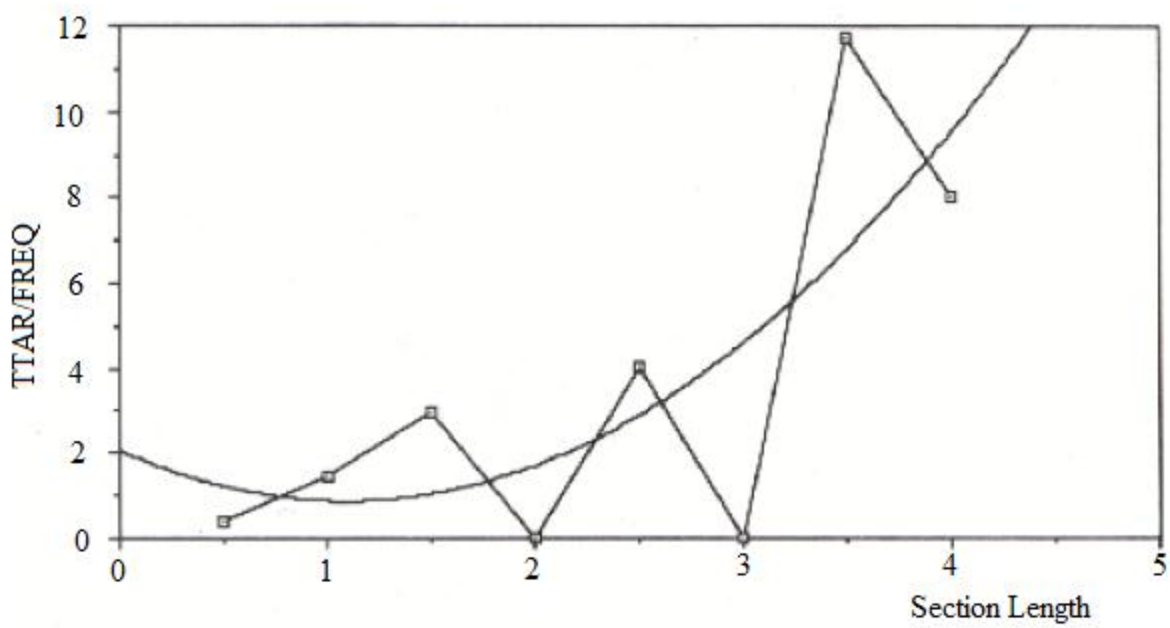

Figure 2 Distribution of accidents with section length

4.3.4.2 Degree of Curvature (D)

Since tangent sections in the HSIS base were identified as having a degree curvature equal to zero, it permitted separate analyses to be performed of accidents on curves and tangents. The degree of curvature in the selected corridor varied from 4^o to 30^o. The distribution of degrees of curvature and accidents on curves in the study corridor is depicted in Figure 2.

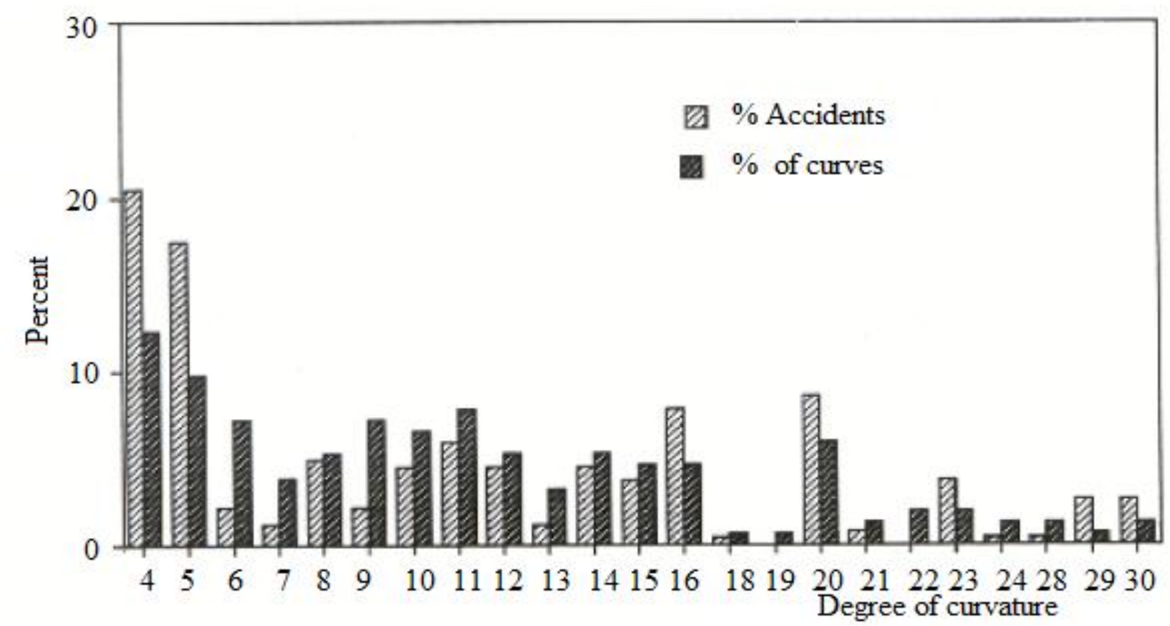

Figure 3 Distribution of degree of curvature and accidents

Contrary to previous findings, most curve accidents seem to have occurred on 4^o to 6^o curves. But, in terms of total accidents in the corridor, less than 15% of the accidents have occurred in these curves. Table 12 shows the number of accidents that occurred from 1985 through 1989 and the number of curves with a given degree of curvature. Figure 3 shows the distribution of accidents in relation to a relative safety index (explained later), which indicates that the accidents increase with the degree of curvature.

Table 12 Distribution of curves and accidents in percentage

Degree of Curvature	% of curve Sections	% Accidents In 1985	% Accidents In 1986	% Accidents In 1987	% Accidents In 1988	% Accidents In 1989
4	12.4	17.5	9.2	21.6	20.3	17.8
5	9.8	15.0	4.6	17.6	23.0	15.6
6	7.2	2.5	0.7	2.0	4.1	0.0
7	3.9	0.0	0.0	0.0	4.1	0.0
8	5.2	5.0	0.7	2.0	6.8	8.9
9	7.2	0.0	0.7	3.9	1.4	4.4
10	6.5	7.5	0.0	0.0	4.1	4.4
11	7.8	7.5	2.0	13.7	0.0	6.6
12	5.2	5.0	2.0	2.0	2.7	8.8
13	3.3	0.0	0.7	3.9	0.0	0.0
14	5.2	2.5	3.3	5.9	1.4	4.4
15	4.6	2.5	2.6	7.8	0.0	2.2
16	4.6	5.0	2.0	7.8	12.2	6.6
18	0.7	2.5	0.0	0.0	0.0	0.0
19	0.7	0.0	0.0	0.0	0.0	0.0
20	5.9	12.5	3.9	7.8	8.1	4.4
21	1.3	2.5	0.0	0.0	1.4	0.0
22	2.0	0.0	1.3	0.0	0.0	0.0
23	2.0	5.0	1.3	0.0	5.4	4.4
24	1.3	0.0	0.0	0.0	1.4	0.0
28	1.3	0.0	0.0	0.0	1.4	0.0
29	0.7	5.0	0.7	3.9	1.4	2.2
30	1.3	2.5	0.7	0.0	1.4	8.8

4.3.4.3 Vertical Grade (G)

The study corridor traverses two major canyons (Sardine and Logan) in northern Utah. Consequently, there were many segments where a horizontal curve is connected directly or through a short tangent section to a vertical curve. These sections were not identified in the HSIS records and were difficult to find without extensive field surveys or referring to state DOT inventories. Therefore, no distinction was made between sections on the basis of this condition. The distribution of vertical grades and accidents within the grades were shown in figure 4.

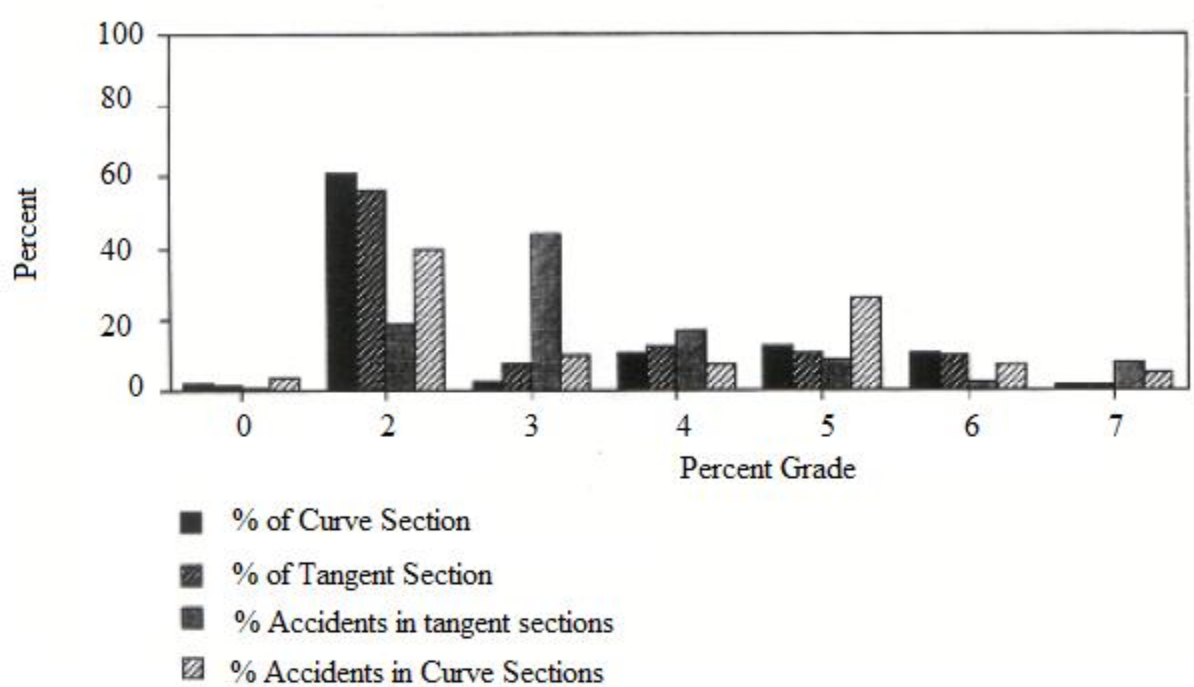

Figure 4 Distribution of vertical grade, frequency and accident rate

The number of accidents that occurred from 1985 through 1989 at sites with a given grade (in percentage) in tangents and curves, and the number of sections for each grade are given in Table 13. It can be seen that the grade varies from 0 to 7% and most of the sections are of 2% grade.

Table 13 number of accidents that occurred from 1985 through 1989 on two- and three-lane road sections in tangents and curves

Tangent % Grade	% Tangent Sections	% Accidents in 1985	in 1986	in 1987	in 1988	in 1989
0	1.4	0.9	1.9	0.6	0.6	0.0
2	55.6	17.5	22.6	19.5	18.5	19.3
2.5	0.7	0.0	0.0	12.7	0.0	0.7
3	7.7	41.2	45.3	44.8	45.1	50.0
4	12.7	21.9	19.8	13.0	18.5	15.0
5	10.6	6.1	15.1	11.7	6.4	7.9
6	9.9	5.3	4.7	2.6	1.2	0.7
7	1.4	7.0	8.5	5.8	9.8	6.4

Curves % Grade	% Curves Sections	% Accidents in 1985	in 1986	in 1987	in 1988	in 1989
0	2.0	7.5	1.9	7.5	1.4	2.2
2	60.8	22.5	41.5	50.9	39.2	44.4
3	2.6	7.5	13.2	13.2	13.5	4.4
4	10.5	10.0	9.4	0.0	9.5	8.8
5	12.4	30.0	24.5	20.8	23.0	28.9
6	10.5	17.5	5.7	1.9	8.1	6.6
7	1.3	5.0	3.8	5.7	5.4	4.4

4.3.4.4 Number of Lanes (N)

There were a few sections in the study corridor with climbing (auxiliary/truck) lanes (third lane). Since climbing lanes are an integral part of two-lane highways in mountainous terrain, it was considered to be a potential variable that could explain the occurrence of accidents, and was included in the analysis. It is also evident that 97.3% of the accidents between 1985 and 1989 occurred in two-lane sections and only 2.7% in three lane sections.

Table 14 shows the number of accidents that occurred from 1985 through 1989 on two- and three-lane road sections in tangents and curves separately.

Table 14 Distribution of lanes and accidents in percentage

Tangents # of Lanes	% Tangent Sections	% Accidents in 1985	in 1986	in 1987	in 1988	in 1989
2	96.5	98.2	96.0	97.5	96.5	100
3	3.5	1.8	4.0	3.5	3.5	0.0

Curves # of Lanes	% Curves Sections	% Accidents in 1985	in 1986	in 1987	in 1988	in 1989
2	96.1	97.5	94.3	98.0	95.9	97.83
3.9	2.5	5.7	2.0	4.1	2.2	

4.3.4.5 Right Shoulder Width (RSW)

The nature of the terrain had determined the width of the shoulder in many of the sections through the canyons. The distribution of the widths and accidents is shown in Figure 5.

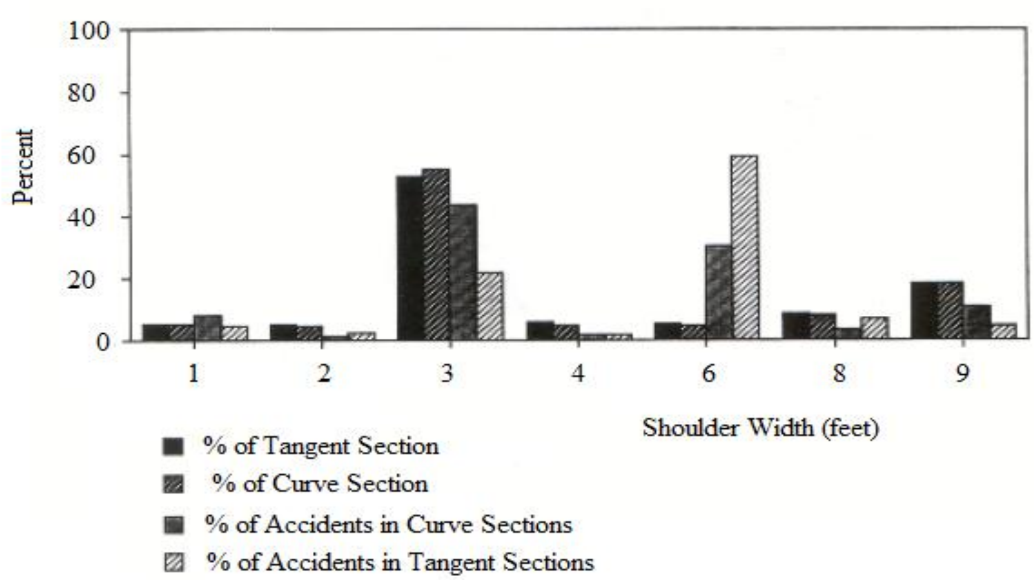

Figure 5 Distribution of shoulder width and accidents

Table 15 shows the number of accidents that occurred from 1985 through 1989 on curves and tangents and the number of sections with a given width of shoulder. It is also evident from Table 3.5 that in both tangents and curves, more accidents occurred on 6-foot shoulders, though not many sections had 6-foot shoulders.

Table 15 Distribution of right side shoulder width and accidents in percentage

Tangent R. Shoulder Width	% Tangent Sections	% Accidents in 1985	in 1986	in 1987	in 1988	in 1989
0	4.9	3.5	1.8	5.0	3.5	7.9
1	4.9	1.8	2.7	1.9	2.9	2.9
2	2.8	23.7	21.2	22.4	19.1	23.6
4	5.6	0.9	1.8	2.5	1.7	1.4
6	4.9	57.9	65.5	55.3	63.0	57.9
8	8.5	6.1	11.5	7.5	5.2	4.3
9	8.3	6.1	6.2	5.6	4.6	2.1

Curves R. Shoulder Width	% Curve Sections	% Accidents in 1985	in 1986	in 1987	in 1988	in 1989
0	5.2	5	7.5	11.8	8.1	8.9
1	4.6	0.0	1.9	0.0	2.7	0.0
2	54.9	47.5	45.3	39.2	41.9	51.1
4	4.6	2.5	0.0	0.0	0.0	0.0
6	4.6	30.0	30.2	31.4	35.1	24.4
8	7.8	2.5	5.7	3.9	4.1	2.2
9	18.3	12.5	9.4	13.7	8.1	13.3

4.3.4.6 Traffic Volume (AADT)

Section traffic volumes were recorded in terms of AADT in the data files. However, the AADT changed little within the corridor. The major difference was noted in the section through the City of Logan, which was eliminated to satisfy the rural condition. In view of this invariability and also the fact that it was more of an operational variable, AADT was incorporated into the dependent variable as the exposure variable.

Table 16 shows the number of sections and the number of accidents observed in each year from 1985 through 1989 for tangents and curves.

Table 16 Distribution of AADT and accidents in percentage

Tangents AADT	% Tangent Sections	% Accidents in 1985	in 1986	in 1987	in 1988	in 1989
1230	13.4	2.6	3.2	7.5	1.7	0.0
1280	11.3	6.1	0.0	2.5	1.7	19.3
1415	5.6	3.5	2.4	3.7	2.8	0.7
1635	3.5	3.5	4.8	6.8	9.7	50.0
1835	26.8	11.4	12.8	8.1	7.4	15.0
3140	32.4	10.5	12.0	11.2	10.8	7.9
10450	4.9	57.9	59.2	55.3	61.9	0.7
14465	2.1	4.4	5.6	5.0	4.0	6.4

Curves AADT	% Curves Sections	% Accidents in 1985	in 1986	in 1987	in 1988	in 1989
1230	13.7	7.5	3.8	9.8	10.8	6.7
1280	10.5	17.5	5.7	5.9	6.8	13.3
1415	4.6	0	1.9	0	2.7	0
1635	2.6	0	3.8	3.9	5.4	6.7
1835	28.1	22.5	32.1	23.5	21.6	26.7
3140	33.3	22.5	20.8	23.5	17.6	20
10450	4.6	30	30.2	31.4	35.1	24.4
14465	2.6	0	1.9	2	0	2.2

4.3.4.7 Data Analysis

Selection of Variables

The author used prior research and the present study objectives as a guide to select variables for significance testing. The task was not difficult because there were no other truly random geometric variables in the database except those mentioned above. Some other variables such as lane width and left shoulder width were either irrelevant or invariable. Sight distance was not included in the analysis since it was given in terms of percent and it could not have been included in our database so that it corresponds to the curve and tangent sections.

Independent Variables

The relationships between the independent variables and accidents were not clearly apparent from the individual plots. For example, it is seen from Figure 2 that accidents in the present case are not responding to the degree of curvature as it had been hypothesized by prior researchers. Moreover, the relationships found in previous studies were not exactly in concurrence with one another. Therefore, it was important to test the contribution of all forms and combinations of variables to accident occurrence.

The general linear model procedure (ProcGLM) of SAS was used to check the significance of the variables. ProcGLM procedure uses the method of least squares to fit general linear models (33). For this study, the factorial modeling (with interaction) procedure was adopted. Degree of curvature, percent grade, right and left shoulder widths, number of lanes, and section length were tested for significance by computing t-values and p-values for each of the parameters.

It can be seen from test result, the section length is the most significant variable. Degree of curvature, section length, right shoulder width, and number of lanes are significant at the ten percent level. No t-value could be estimated for left shoulder width, which means that it is highly insignificant. Percent grade could not meet the ten percent significance level.

To minimize the errors due to multi co-linearity, the variables were screened once more by examining the correlation coefficients between the variables in which all of them were less than 4%. This is a strong indication of the absence of multi co-linearity between the chosen variables. Even though percent grade could not fall within the chosen level of significance in terms of t-value, it was included in the final analysis since the study corridor is in mountainous terrain.

In summary, the independent variables selected for further analysis in this study were section length in miles, degree of curvature, shoulder width on right side of the road in feet, grade in percentage, and combinations of some of them.

Dependent Variable

The dependent variable was the expected accident rate (per million vehicles per year) estimated as follows:

$$AR = \frac{\text{Total Accidents}(1985 - 89) * 10^6}{(5 * 365 * \text{AADT})} \tag{7}$$

Where,

AR, is the expected accident rate (per million vehicles per year)

4.3.4.8 Selection of Model Form and Calibration

Many model forms have been tested in previous safety research. These can be broadly categorized as linear and nonlinear models. The following model forms were chosen and calibrated separately with each of the three data sets corresponding to tangents, horizontal curves, and the entire corridor:

$Y = \sum a_i \, X_{ii}{}^{b}$………………………………………...… (Model A)

Log$Y = \sum a_i \, Log \; X_i$…………………………………… (Model B)

$LogY = \sum a_i \, X_i{}^{bi}$……………………………………..… (Model C)

Where,

Y = accident rate

X = independent variable

a & b = calibrated parameters/coefficients

Since making a decision about which of the variables to include in a regression model is difficult, stepwise regression technique is sometimes used. This technique allows the computer to experiment with different combinations of the independent variable and give insight into the relations between the independent variables and the dependent variable. The problem with stepwise regression is that some of the variables which are not significant may enter the model, or variables which are significant may not enter the model. For the purpose of this study, it was decided to employ the “forward stepwise” GLM procedure to screen the variables and calibrate the traffic accident prediction models.

GLM also enables the reduction of confusion created by the inclusion of all the variables. The forward stepwise procedure begins with no variables in the model. For each of the independent variables, the ProcGLM can be used to calculate the F-statistic. The F-statistic signifies the variable's contribution to the explanatory capability of the model if included. The ahead stepwise feature compares the F-records to the extent of significance (in this example the extent particular is 0.1) this is certain withinside the version statements. If no F records have values more than the required level, the procedure ends. Otherwise, the system provides the variable that has the most important F-statistic to the version. It then calculates Statistics once more for the variables nevertheless last outdoor the version, and the system continues. Thus, unbiased variables are brought one at a time to the version till not one of the closing variables are significant. The outcomes of the ahead stepwise more than one regression evaluation for every of the facts units is given as below.

Corridor

Model A

$$AR = 0.0092 + 0.016\,(D) + 3.5\,(L) - 0.02\,(L)(S - 0.006(L)(D)(G)(SWR) \tag{8}$$

R2 = 0.74

Model B

$$Log\,(AR) = 1.02 + 0.77\,Log\,(L) - 0.1\,Log\,[(SWR)\,(D)(L)] \tag{9}$$

R2 = 0.38

Model C

$$Log\,(AR) = 0.004 + 0.0071\,(D) - 0.87\,(L)\,(SWR) - 0.003\,((G)(SWR)(L)(D) \tag{10}$$

R2 = 0.74

Tangents

Model A

$$AR = 0.1 + 3.4\,(L) - 0.2(L)\,(SWR) \tag{11}$$

R2 = 0.843

Model B

$$Log\,(AR) = 1 + 0.74\,Log\,(L) \tag{12}$$

R2 = 0.43

Model C

$$Log\,(AR) = 0.02 + 1.5\,(L) - 0.1(L)\,(SWR) \tag{13}$$

R2 = 0.84

Curves

Model A

$$AR = -0.3 + 3.8\,(D) + 0.37\,(D)\,(L) + 0.011(D)(G) + 0.004(D)(SWR) - 0.12(G)(D)(L) \tag{14}$$

R2 = 0.28

Model B

$$Log\,(AR) = 0.5 + 0.4\,Log\,(D)\,(L) - 2\,Log\,(SWR)\,(G)(D) \tag{15}$$

R2 = 0.18

Model C

$$Log\ (AR) = -\ 0.1 + 1.7\ (L) + 0.16\ (D)\ (L) + 0.005\ (D)\ (G) - 0.1((G)(D)(L) \quad (16)$$

R2 = 0.28

4.3.4.9 Model Verification

The models developed were verified by using them to predict accidents in 1990 and comparing the outcome with the observed accidents. The predicted values of the present models were then regressed with the observed accidents in 1990 to test the predictive accuracy of the model.

Table 17 Comparison of Model A predicted values and observed values

AR	Corridor Predicted	1990 Observed	Tangent Predicted	1990 Observed	Curve Predicted	1990 Observed
0	0	209	0	98	17	111
1	269	23	115	9	129	14
2	15	26	16	11	7	15
3	5	25	5	13	0	12
4	4	4	3	3	0	0
5	1	1	1	4	0	0
6	0	1	0	1	0	0
7	0	1	0	1	0	1
8	0	0	0	0		
9	1	0	1	1		
11	0	0	0	0		
12	0	0	0	0		
13	0	0	0	0		
14	0	1	0	1		

Table 18 Comparison of Model B predicted values and observed values

AR	Corridor Predicted	1990 Observed	Tangent Predicted	1990 Observed	Curve Predicted	1990 Observed
0	29	209	24	98	7	111
0.1	90	2	26	1	59	1
0.2	108	7	40	2	74	5
0.3	37	14	20	6	13	8
0.4	21	18	19	6	0	12
0.5	8	14	8	6	0	8
0.6	3	12	4	9	0	3
0.7	0	11	1	7	0	4
0.8	0	5	0	5	0	0
0.9	0	2	0	1	0	1
1.0	0	0	0	0		
1.1	0	1	0	1		
1.2	0	1	0	1		

Table 19 Comparison of Model C predicted values and observed values

AR	Corridor Predicted	1990 Observed	Tangent Predicted	1990 Observed	Curve Predicted	1990 Observed
0	2	209	0	98	7	111
0.1	132	2	54	1	70	1
0.2	118	7	60	2	59	5
0.3	31	14	14	6	15	8
0.4	2	18	4	6	2	12
0.5	4	14	4	6	0	8
0.6	2	12	2	9	0	3
0.7	3	11	3	7	0	4
0.8	0	5	0	5	0	0
0.9	0	2	0	1	0	1
1.0	0	0	0		0	
1.1	1	1	1		1	
1.2	0	1	0		1	
1.3	0	0	0		0	
1.4	1	0	1		0	

Corridor Models

From the result of forward stepwise multiple regression analysis result, it can be seen that Model A has an R2 value of 0.74. The predicted accidents by this model, when regressed with the observed accidents of 1990, could explain 56.31 (R2) percent of the variation in accidents. The total number of accidents predicted by this model for 1990 is 85.86% (predicted number of accidents is 164 and the observed number of accidents was 191) of the accidents observed in 1990. This model predicts a mean accident rate of 0.5524 accidents per million vehicles per year, whereas the observed mean accident rate was 0.6426 accidents per million vehicles per year.

However, from Table 17 it can be seen that this model predicts 269 sections to have an accident rate of one accident per million vehicles per year in the year 1990. But it was observed in 1990 that 209 sections had a zero accident rate and only 23 sections had an accident rate of one.

Model B has an R2 value of 0.38. The predicted accidents for 1990 by this model when regressed with observed accidents in 1990 gave an R2 value of 0.2424. This model predicted about 6% more accidents than accidents observed in 1990. The model predicts a mean accident rate of 0.141 for 1990, whereas the observed mean rate in 1990 was 0.133 accidents per million vehicles per year.

However, from Table 18 it can be seen that this model predicts 108 sections to have an accident rate of 0.2 accidents per million vehicles per year but 209 sections were observed to have an accident rate of zero in 1990 and only seven sections had an accident rate of 0.2.

Model C has an R2 value of 0.516. The number of predicted accidents by this model for 1990 when regressed with observed accidents in 1990 could explain 33.39% of the variation in accidents observed in 1990. The model predicts about 6% more accidents than observed in 1990 with an estimated mean accident rate of 0.14088 per million vehicles per year, whereas the observed accident rate in 1990 was 0.133 per million vehicles per year.

However, from Table 19 it can be seen that this model predicts only two sections to have an accident rate of zero when 209 sections were observed to have an accident rate of zero in 1990.

Tangent Models

From Table 17 it can be seen that Model A has an R2 value of 0.843. The predicted accident by this model for 1990, when regressed with the observed accidents of 1990, could explain 66.80% (R2) in the variation of accidents. This model predicts 17.89% (predicted number of accident =145 and observed number of accidents = 123) more accidents than those that occurred in 1990. The model predicts a mean accident rate of 0.733, whereas the observed mean accident rate in 1990 was 0.855 accidents per million vehicles per year.

However, from Table 18 it can be seen that Model A predicts 115 sections to have an accident rate of one accident per million vehicles per year in the year 1990. But it was observed in 1990 that only nine sections had an accident rate of one, and 98 sections had an accident rate of 0.

From Table 18 it can be seen that Model B has an R2 value of 0.43. The predicted accident values when regressed with the observed accidents in 1990 could explain about 34.74% of the variations in accidents observed in 1990. The model could predict about 6.15% more accidents than those observed in 1990. The predicted mean accident rate for 1990 was 0.1715 whereas the observed accident rate was 0.1612 per million vehicles per year.

However, from Table 19 it can be seen that this model predicts 24 sections to have an accident rate of zero, but it was observed in 1990 that 98 sections had an accident rate of zero.

Model C has an R2 value of 0.84 as seen from Table 9. The predicted accident values by this model when regressed with observed accidents in 1990 could explain 33.39% of the variation in accidents. The model could predict about 6.15% more accidents than those observed in 1990. The mean accident rate observed was 0.1715 whereas the observed rate was 0.1612 accidents per million vehicles per year. However, from Table 12 it can be seen that this model predicts no sections to have an accident rate of zero but 98 sections were observed.

Curve Models

From Table 17 it can be seen that Model A has an R2 value of 0.28. The predicted accident values by this model for 1990 when regressed with observed accident values in 1990 could explain 20.93% of the variation in accidents. This model predicts 13.24% (predicted number of accidents = 59, and observed number of accident = 68) fewer accidents than those that occurred in 1990. The predicted mean accident rate for 1990 was 0.3838, whereas the observed accident rate was 0.4439 accidents.

However, from Table 17 it can be seen that this model predicts 129 sections to have an accident rate of one accident per million vehicles per year for the year 1990 but it was observed that 111 sections had a zero accident rate in 1990, and only 14 sections had an accident rate of one.

Model B has an R2 value of 0.18. The predicted accident values by this model for 1990 when regressed with observed accidents in 1990 could explain about 24.24% of the variations in accidents. This model could predict about 5.22% more accidents than those observed in 1990 with a mean

Accident rate of 0.1122, whereas the observed accident rate was 0.1066 accidents per million vehicles per year.

However, from Table 18 it can be seen that this model predicts 74 sections to have accident rate of 0.2, but 111 sections were observed to have an accident rate of 0 in 1990 and only 5 sections had an accident rate of 0.2.

Model C has an R2 value of 0.28. The predicted accidents by this model for 1990 when regressed with observed accident in 1990 could explain only 16.5% of the variation in accidents. The model could predict about 5.22% more accidents than those were observed in 1990. The mean predicted accident rate was 0.1120, whereas the observed rate was 0.1066 accidents per million vehicles per year. However, from Table 19 it can be seen that this model predicts 70 sections to have an accident rate of 0.1, but 111 sections were observed to have an accident rate of zero in 1990 and only one section had an accident rate of 0.1.

It can be seen that the model form of Model A has more independent variables than the model form Model B. Even though Model C has more variables, its R2 value is less than the Model A form.

4.3.4.10 Modeling and Discussion of Findings in the Study

Among all the model forms, Model A has the highest predictability in terms of R2 value. Model A is also easier to understand and interpret. Hence, it was taken to be the best model form for the study corridor and was further examined.

Selection of best accident prediction models for northern Utah

Corridor

$$AR = 0.0092 + 0.016\,(D) + 3.5\,(L) - 0.02(L)(SWR) - 0.006(L)(D)(G)(SWR) \qquad (17)$$

R2 = 0.74

Tangents

$$AR = 0.1 + 3.4\,(L) - 0.2(L)\,(SWR) \qquad (19)$$

R2 = 0.843

Curve

$$AR = -0.3 + 3.8\,(D) + 0.37(D)(L) + 0.011(D)(G) + 0.004(D)(SWR) - 0.12(G)(D)(L) \qquad (20)$$

R2 = 0.28

4.3.4.11 Conclusion drawn from the Study

The study reported here confirmed that road geometry can explain a substantial portion of the variation in accidents on two-lane rural highways. It was also found that disaggregation of data does not always increase the predictive accuracy of a model. For instance, it was found that the curve (disaggregate) model was able to explain only 28% of the variation compared to 74% by the corridor (aggregate) model. The model for tangent sections (another disaggregate model), on the other hand, was more credible than all the aggregate models.

The other principal finding was that the aggregate and disaggregate models are transferable over short time spans, but their transferability over space is not clearly evident. For example, in spite of being widely cited as global models, Zegeer, et al.'s (3) models were able to explain less than 5% of the variation in accidents at the cross section or on the curves when employed to predict accidents in Utah. Of course, a few missing variables might have been a cause. But Zegeer, et al. (3) have demonstrated that curvature and section length are the most significant among the variables. Thus, the models should have been able to explain a larger part of the variation in accidents in Utah since information on the degree of curvature and section length was available.

The impact of horizontal curvature on accidents was found to be much less significant than what was noted in previous studies. Thus, if curve flattening results in a significant elongation of curve length and an increase in operating speed, the safety benefits noted in the previous literature (24, 27, 31, 33) may not be too realistic. This question calls for a further examination of the relationship between accident rate and degree of curvature.

In general, it could be said that there is a need for considerably more work in the accident modeling area. It may need to start with some standard definitions or criteria for disaggregating road sections so that models and results are easily transferable. If a reasonable level of uniformity can be achieved, parameter estimation and transferability would become practical. The use of modeling principles in transportation safety analysis will become commonplace. Ultimately, agencies responsible for transportation will be able to make informed decisions as opposed to educated guesses about the influence of geometric elements on traffic accidents.

4.4 STUDY FOUR

4.4.1 Overview of the Study (Article)

The examine was "An Artificial Neural Network Model for Road Accident Prediction: Case Study of a Developing Country" with the aid of using F.N. Ogwueleka et al, which aimed toward generating a layout of an Artificial Neural Network (ANN) version for the evaluation and prediction of coincidence prices in a growing country. The authors used the most recent 1998 up to 2010 data to the model design, by using the number of vehicles, accidents and populations were used as model parameters. The sigmoid and linear functions were used as activation functions with the feed forward-back propagation algorithm.

4.4.2 Methods and Procedures in the Study

. The first task in the study was obtained the road traffic accident data from Federal Road Safety Commission, Abuja-Nigeria and Nigerian police as shown in the table below, from 1998 to 2010.

Table 20 Data of road traffic accident cases, the injured persons with deaths

Year	Total Cases Reported	No. of Persons Killed	No. of Persons Injured
1998	17,117	6,578	17,547
1999	12,503	5,953	18,000
2000	12,325	6,336	20,555
2001	15,621	7,845	26,745
2002	16,452	8,452	27,102
2003	16,795	8,672	28,215
2004	14,279	5,351	16,897
2005	8,962	4,519	15,779
2006	9,114	4,944	17,390
2007	9,132	4,916	20,944
2008	11,341	6,661	27,980
2009	11,031	4,120	20,975
Jan-June 2010	5,560	3,183	14,349

Then the road accident was categorized using a self-organizing map (SOM) based clustering. The number of RTAs, vehicles, and populace were taken and used as model factors in the study. Road types and country are also used as parameters. The input variables are the number of vehicles per day and the road length in kilometers. The designed Multi-Layer Perceptron Neural Network (MLPNN) consists of the input layers, hidden layers and an output layer, as shown in Figure 6 below.

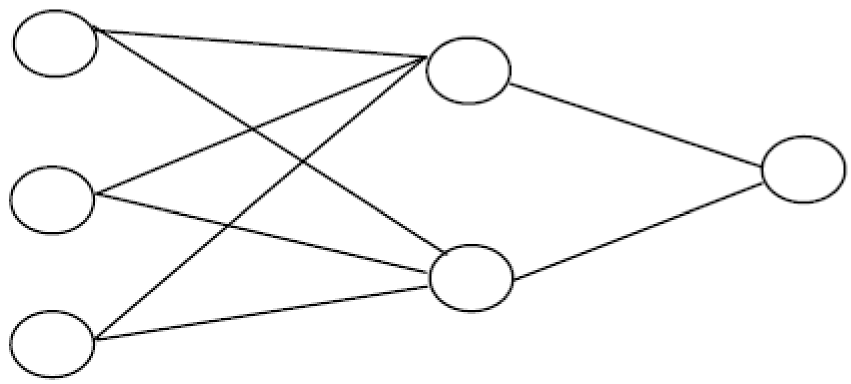

Figure 6 Neural Network

After clustering the entire datasets, (ANN) is used to get a model that is the best for predicting road traffic accident, as ANN always learns from past experience after it completes its first training, and so it becomes an appropriate methodology for prediction. The ANN technique shows some tolerance to a good extent over errors that may exist within the training set. It has the ability to show the veiled and dependencies that are not linear and still learn from its past experience after completing its first training, and this makes it appropriate method that is suitable for prediction. The Multi-Layer Perceptron Neural Network (MLPNN), which is also called the multilayer feed-forward neural network, was chosen and used in this study. Figure 7 is a graphical representation of the overall architecture of the proposed system.

The data that is present in the databases are obtained from Federal Road Safety Commission, Abuja-Nigeria. The data are preprocessed by removing the duplicates and providing the values that are missing. The data were first formatted to an acceptable form for clustering; this is because data that have similar factors are clustered together, while data that have less peculiar factors are clustered differently through a methodology called unsupervised grouping of similar datasets into a predefined groups.

To prepare the data that is best for mapping out the process, there is need to change data consequently. The formatted data are further preprocessed in order to put the data in the form best for clustering. Clustering was performed with a k-means algorithm. SOM was used as the basis for clustering of data items. The method proposed uses k-means measurement, which is used to measure distance or differences that exist between the sampled data, because it is the major factor for cluster scrutiny. After clustering all the datasets, with self-organizing map techniques, ANNs are used to obtained best pair of road accidents for the specified type of accident characteristics. The prediction was done using Multilayer Perceptron Neural Networks (MLPNN).

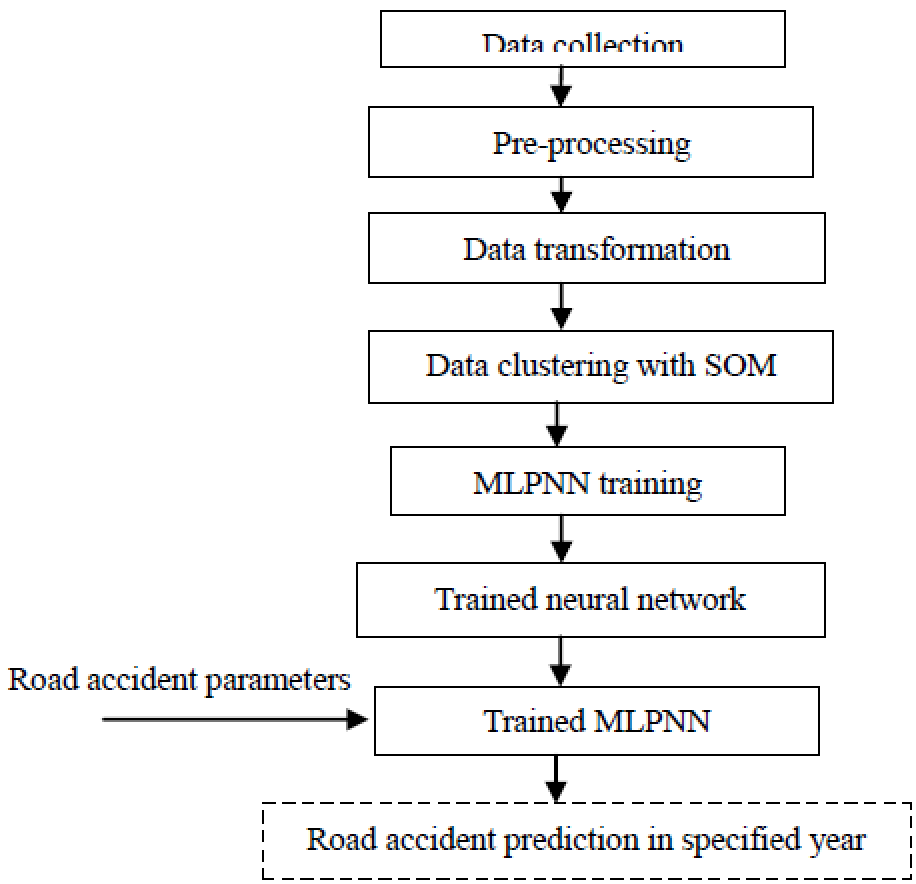

Figure 7 the training architecture

4.4.3 Data Analysis in the Study

One of the simple algorithms used in this research is:

Initialize map

For r from 0 to 1 (where r is the weight vector)

Randomly select a sample

Get best matching unit

Scale neighbors

Increase r with small amount

End for

Initialization of weight vectors is step one in building up of the SOM, after which a vector among the samples is selected randomly and SOM searched for weight vectors that best represent the sample. Each weight vector has a location, with neighboring weights very close to it. The chosen weight is compensated as it is more likely to be the randomly chosen sample vector. Also, the neighbors of that weight are also rewarded, as it is likely to be the selected sample vector. Here, r is increased a little with respect to amount of neighbors, and to what extent will each weight can learn to decrease with time. The steps are repeated a good number of times.

The major aim for carrying out training in a multilayer feed-forward network is to obtain what will make ANN output weight values to match the actual target values very closely. To design and train multilayer perceptron network involves several challenges, which include determining the number of hidden layers to be used in the network, determining the number of neurons to be used in each hidden layer, establishing a general acceptable solution that avoids local minima, converging to an optimal solution as and when due or in good time, and validating the neural network to test for over fitting. Though there exist errors and noise in the training set, ANN still possesses the capability to find the dependencies that are hidden and are not linear, and it also learns from past experience as it completes its training. ANN is still the best prediction tool.

One of the best training ANN algorithm for prediction is Back propagation (BP). During prediction using BP, errors found in the network are propagated backward to the appropriate nodes. BP carries out is process by adjusting the weight values along with the bias values in order to increase the square sum of the difference that exist among the given output and output values that is generated by the network.

The back propagation technique was used efficiently by these steps:

1) A sample for training was presented to the ANN.

2) The output of the ANN was compared with required output and the calculation of error in each output neuron is done.
3) The calculation of the local error, which is estimated from each neuron by establishing what the output should have been with a scaling factor and to what extent either low or high the output, must be adjusted in order to match the needed output.
4) Lowering the local error by adjusting of the weights of each neuron.
5) Fault is assigned to neurons for the local error found at the previous level; this assigns a higher task to neurons that are strongly connected by weights.
6) Iterate algorithm starting from step 3 using neurons found at the previous level; use each neuron fault as its error.

In this study, the adjustment of the load cost turned into executed the usage of conjugated gradient set of rules with assist of gradient at some stage in backward propagation of mistakes withinside the network. The conjugate gradients set of rules makes use of extra paths which are direct to satisfactory organization of weight values while as compared with the gradient descent. It is also faster and more robust. It does not require explicit specification of learning rate and momentum factors.

The steps in the proposed approach are given as follows; 1) Calculate the amount of data in the dataset 2) Generate a group of clusters and establish the centroid of the clusters. 3) Establish the Euclidean distance for each data using the centroids the existing groups of clusters. 4) Allocate data to the group of clusters using minimum distance 5) Iterate steps three and step four until all changes in the clusters disappears. 6) This step deals with the generation of the standard deviation for the group of clusters that are formed. Put aside all the clusters with generated standard deviation that are less than 0. 7) The above steps are repeated until the generated standard deviation for the whole clusters attains a value that is less than 0. The following pseudo-code shows the algorithm for clustering: Set NC, No (where NC = 1 and No = 1) Calculate NC (where NC = NC – No + 1 Form Cif clusters of NC size compute centroid Iterate For every value starting from one to N and NC Compute the Euclidean distance Find the list distance Apply data to a group of cluster until all no change is found in the clusters For every cluster value which is equal from one to NC Compute the standard deviation If standard deviation is less than Φ

Remove Cr from Cif where Cr € Cif and add to Ca Find the rest of the clusters until its values reaches zero Where N = Amount of data in the dataset, NC = Amount or a group of clusters No = Amount or a group of clusters that has zero standard deviation Ca = Original clusters, Cif = Intermediately formed clusters Cr = Cluster that will be removed.

Figure 6 is the flow chart diagram depicting the process of clustering T. The first value for NC = 1 and No = 0. The minimum distance among the clusters is obtained by after computing the centroid and Euclidean distance of the data items.

4.4.4 Result and Discussion in the Study

The designed MLPNN contains three input layers with two hidden layers and one output layer. The output layer carries out the prediction of the RTA rate when presented with the factors.

Comparing 2004 and 2005 RTA summary from Table 21, there is an observed 37% reduction in total RTA cases; the amount of persons killed was reduced by 16%; and the amount of persons injured was reduced by 7%.

Table 21 Summary of 2004 and 2005 RTA

Year	2004	2005	Remarks
Total reported RTA cases	14279	8962	37% reduction in RTA
Persons killed	5351	4519	16% reduction in the number of persons killed
Persons injured	16897	15779	7% reduction in the number of persons injured

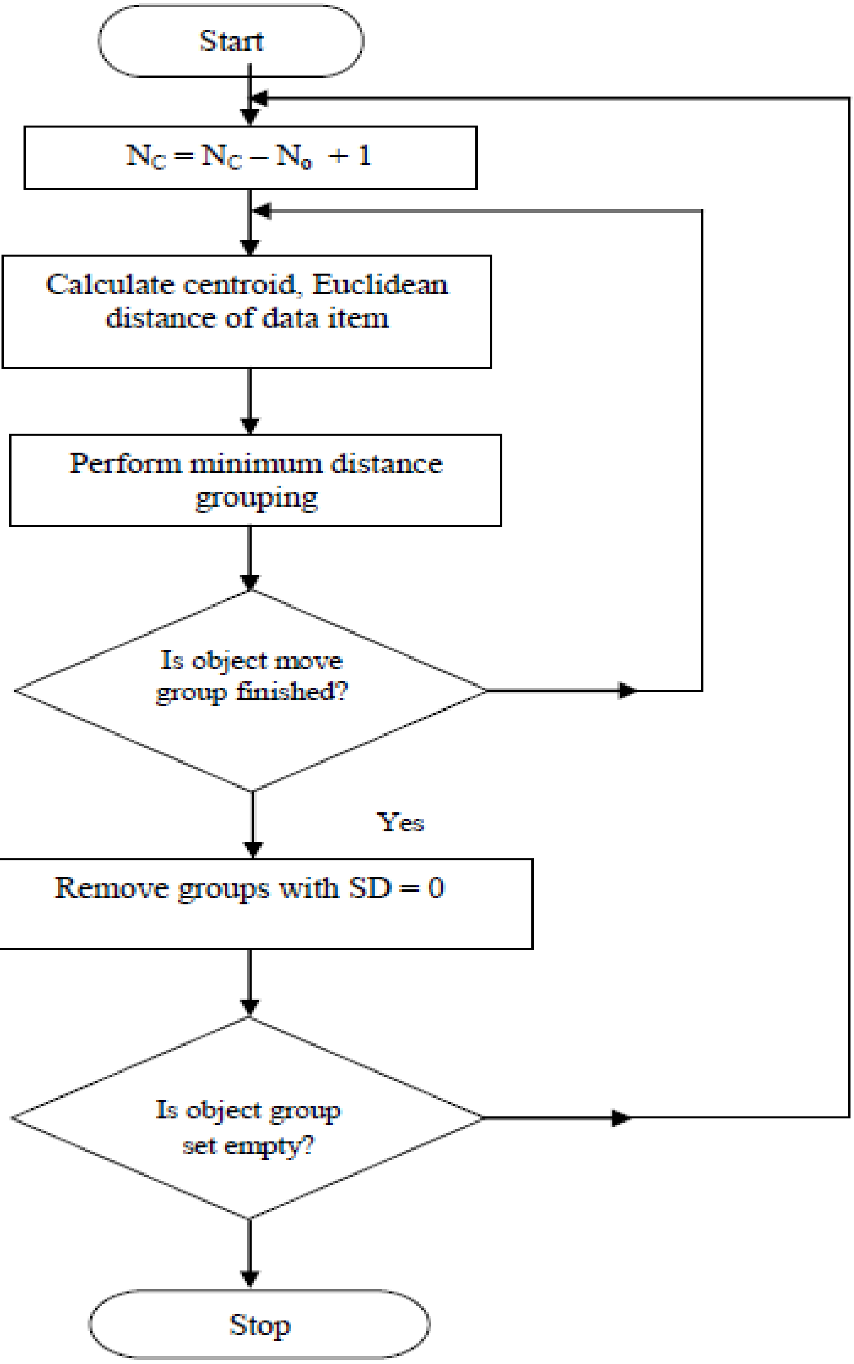

Figure 8 Clustering process flowcharts

The road traffic accident summary of 2004 and 2005 are shown in Figures 9 and 10, respectively.

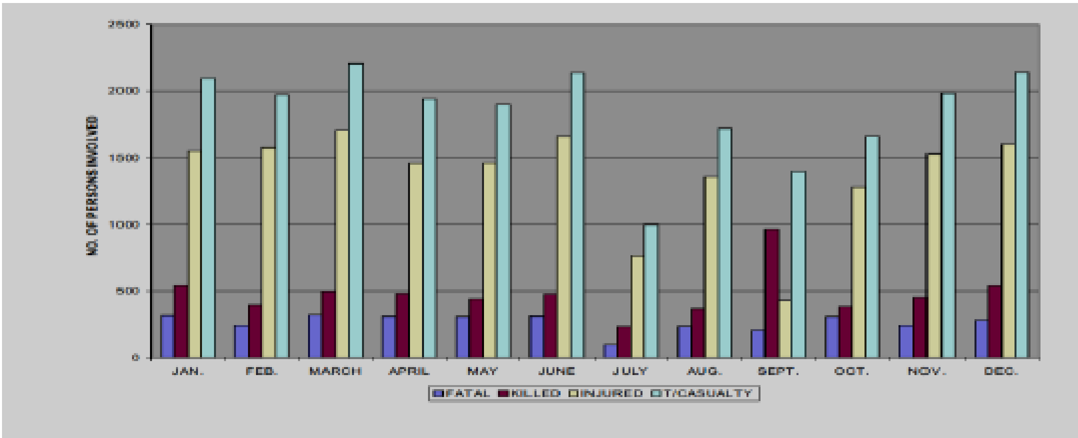

Figure 9 Road Traffic Accident summary 2004

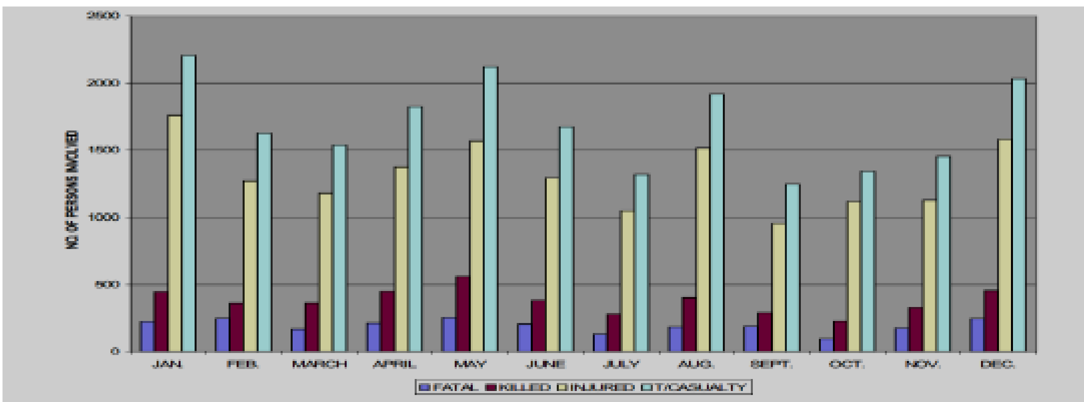

Figure 10 Road Traffic Accident summary 2005

Table 22 shows the data of RTA year summary of cases reported, total persons killed and persons injured for 2004 and 2005.

Table 22 the data of RTA year summary of cases reported

Month	Total Cases Reported	No. of Persons Killed	No. of Persons Injured	Month	Total Cases Reported	No. of Persons Killed	No. of Persons Injured
Jan	980	447	1,760	Jan	1333	544	1550
Feb	1040	358	1267	Feb	1039	397	1572
Mar	591	359	1178	Mar	1312	498	1705
April	679	450	1373	April	1498	481	1458
May	1118	556	1567	May	1206	440	1459
June	841	381	1291	June	1144	476	1661
July	438	276	1045	July	487	233	761
Aug	956	399	1518	Aug	1136	365	1354
Sept	630	289	955	Sept	829	963	428
Oct	376	224	1117	Oct	1925	383	1279
Nov	624	325	1128	Nov	1036	455	1526
Dec	689	4,120	20,975	Dec	1104	542	1598
Total	8,962	4,519	15,779	Total	14,049	5,777	16,355

From the assessment of 2004 and 2005, it turned into determined that 2004 has the better stated RTA cases, variety of men and women killed and additionally quantity of men and women injured.

Figure eleven indicates the waft in RTAs beginning from the year 2000 to the year 2007. The quantity of RTAs with inside the year 2007 equals to 12,038, which suggests a growth of 3.2% more that of the year 2006 and any other growth of 2.8% more than that of the 12 months 2000.

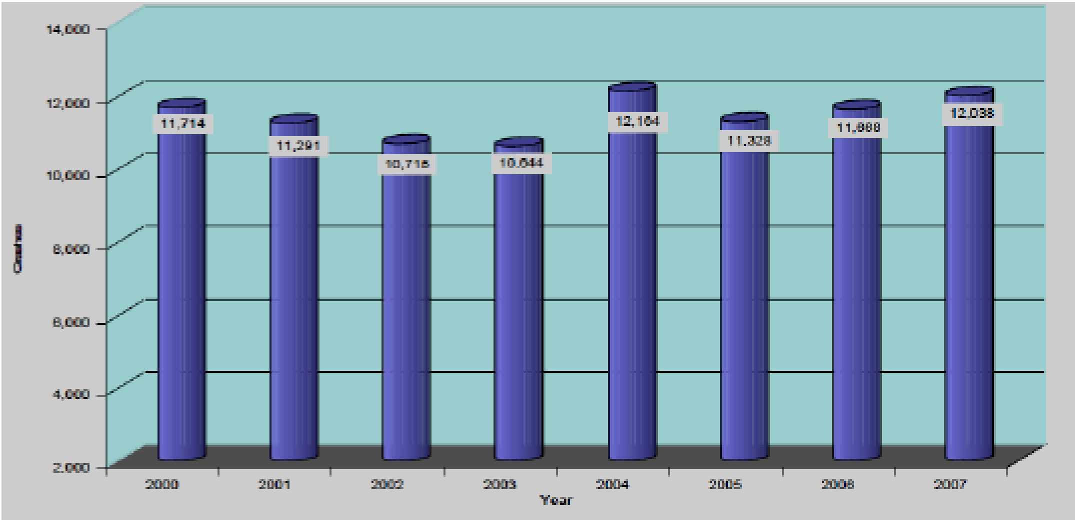

Figure 11 RTA 2000 – 2007

From the analysis of 2004 against 2005 and also from 2000 to 2007, the RTA target verses actual from 2001-2015 were obtained through prediction from collected data. Figure 12 shows the predicted RTA for 2010 to 2015.

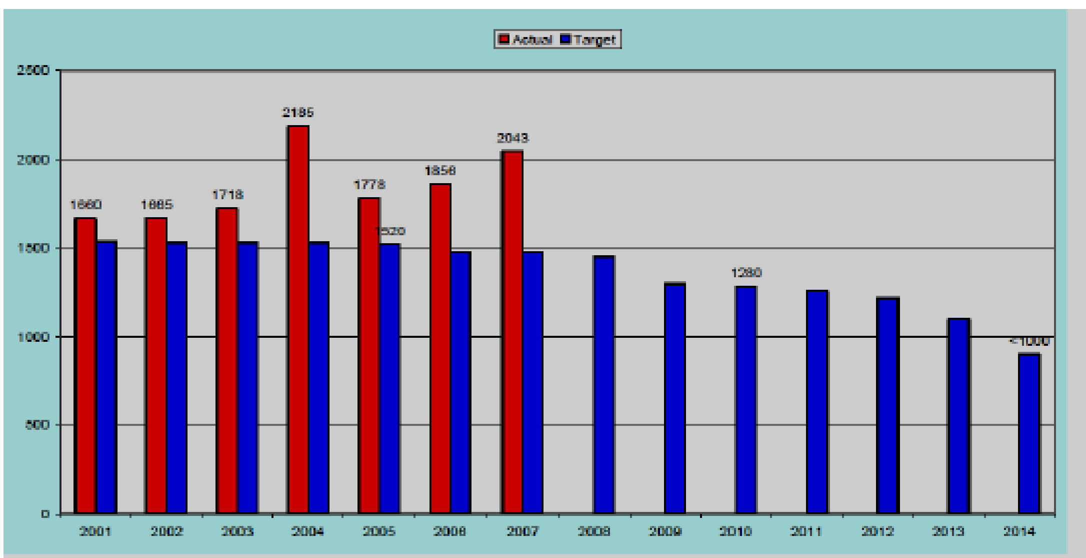

Figure 12 RTA fatalities target vs. actual 2001-2015

4.4.5 Discussion

The comparative analysis of 2004 against 2005 and 2000 to 2007 made the prediction of RTA for 2010 to 2015 using primary source collected data capable of allowing an accurate and good datamodel. The multilayer feed-forward neural network with its learning technique worked through the output value comparison with the accurate answer and also performed the computation of this already established error function. The error is inputted back to the ANN algorithm and it adjusts the weight values of every connection to bring down the values of the error function to minimal.

4.4.6 Conclusion drawn from the Study

Some of these designed solutions will go a long way in reducing RTAs in Nigeria when implemented. The institution of a high level National Road Safety Council will complement the effort of FRSC, and also a special sub-group for road safety information and campaigns. Revisiting the activities and duties of all the bodies that are charged with the national responsibility for road safety management activities will equally go a long way to reduce RTAs in Nigeria; enhancing the collaboration among the Ministry of Works, FRSC, VIO, hospitals, transporters, NCDC, Army, Police Force, and insurance companies with respect to RTAs coming together to develop one inclusive and comprehensive annual national statistical yearbook on all RTAs; the development of strong a policy to regulate driving speed limits; an urgent plan to develop means for black spot management, and creating a special funding for black spots; and revisiting the rules for commercial traffic, and fortify the full implementation of the rules, together with the rules governing the transportation of human beings, their luggage and goods.

ANN showed its advantage over conventional programming in this study. This is due to its capability to provide solutions to non-algorithmic problems and can learn how to deal with the new and unexpected situations by the help of past experience. Neural networks are able to relate input with output, allow large number of variables and are error tolerant.

4.5 STUDY FIVE

4.5.1 Overview of the Study (Article)

The study was “Development of Models for Crash Prediction and Collision Estimation-A case Study for Hyderabad City” by V.Niveditha, A. Ramesh and M. Kumar. The research looked in developing a model which is suitable for road crash prediction and estimation of collision type that influencing road crash in Hyderabad City. A detailed study was conducted in the city of considering factors as roadway geometrics, traffic data, and type of collision. So Regression models like Multiple Linear Regration, Poisson Regression, Logit Model and Multinomial Logistic Models were considered in the study. Then the suitable model for the study area was selected based on R^2 and chi-square test. The objectives undertaken in this study were ascertained which type of mathematical model is suitable for prediction of road crashes and its influencing variables and estimation of the type of collision significance on road crashes.The paper was published on IJTE, International Journal of Transportation Engineering, and February 15, 2016. The author focused on three key words as Road crash, collision and regression models.

4.5.2 Methods and Procedures in the Study

In this study three years road crash data was collected for the city and it is considered in the analysis. Here the analysis was divided into two parts as analysis carried using Multiple Linear Regression, Poisson regression and Logit models in the first part and in the second part the analysis traffic and crash data were used for collision analysis. In the study the authors done a related literature review on crash causing factors and key concepts in regression modeling to safely accomplish the job.

Once the crash data were collected from the recognized organization, the authors made crash type data analysis and collision type data analysis to identify the location where more number of crashes was occurred. So based on the crash type data analysis njara Hills and Jublee Hills locations are selected for model development as more number of crashes is occurred long these stretches. The data includes road geometrics, crash data (as fatal non-fatal and total number of crashes) and traffic volume.

Table 23 Selected Variables and method of data collection in Study Five

Road Data	Road crashes data Police Records	Traffic Data
Inventory Survey	Road Crash Location	
Shoulder Width	Age of Driver	Manual Method
Number of curves	Type of Vehicle Involved	Volume count conducted at
Number of bus stops	Number of Person injured or dead	midblock section
Subjective Rating	Type of collision	
Pavement Condition		
Shoulder Condition		

4.5.3 Data Analysis in the Study

Table 24 Road crash areas

Year	Location	Total no of crashes	Total no of fatal crashes	Total no of non fatal crashes
	Bowenpally	77	24	67
	Osmania University	54	10	47
2012	Banjarahills	130	19	122
	Punjagutta	97	20	83
	Jubleehills	56	15	48
	Bowenpally	103	22	90
	Osmania University	64	10	57
2013	Banjarahills	115	19	107
	Punjagutta	129	18	120
	Jubleehills	83	17	74
	Bowenpally	110	30	92
	Osmania University	95	19	72
2014	Banjarahills	135	20	117
	Punjagutta	112	22	102
	Jubleehills	135	15	112

Table 25 Parametric considered

	Fatal	763
Total number of accidents	Non-fatal	6496
Type of vehicles	2-Wheeler	3325
	3-Wheeler	478
	Pedestrian	3456
Type of collision	Side Swipe	4929
	Rear-end	713
	Head-on	617

Table 26 Descriptive statistics

Variable Code	Variable Description (per kilometer data)	Minimum	Maximum
PC	Pavement Condition*	1	5
SC	Shoulder condition*	1	5
SW	Shoulder Width	0	2.5
AADT	Traffic Volume (100 vehicles)	2	100
AGE	Age	2	75
NFC	Number of Non-Fatal Crashes	0	53
FC	Number of Fatal Crashes	0	15
TC	Number of Total Crashes	0	68

*Subjective rating: 1-Very Good, 2-Good, 3-Fair, 4-Poor, 5-Very Poor

Based on this analysis the correlation analysis gives a quantitative assessment of association between two variables. So positive or negative values could be obtained with varying strengths. Here a positive Correlation coefficient means that as the value of one variable increases, the value of other variable also increases. And also as one variable decreases the other also decreases.

On the other hand negative correlation coefficient indicates that as one variable increases the other variable decreases. For example in the analysis of the study, it is observed that as pavement condition deteriorates more number of non-fatal crashes was increased.

4.5.4 Result and Discussions of the Study

In this study three types of regression models were used for developing crash models. The author considers non-fatal crashes, fatal crashes and Total Crashes as dependent variables and PC, SC, SW, AADT and AGE were independent variables. The modeling parameters were estimated as shown below in table….

Model Interpretation for Crash Prediction in Table 23 all the variables are significant for model predication and it can attribute on better suitability of model for crash prediction. Condition of the pavement is also one of the affecting factors for prediction of model development.

4.5.4.1 Multiple Linear Regression Model

Multiple Linear Regression Model R^2 value and coefficient obtained from model analysis explains that there is relationship between dependent and independent variables. Fatal crashes are less frequent when compared to Non-Fatal crashes which approximates as linear in relation. It is observed that no variable coefficients are zero and is an indication of existence of model. All the statistical values are within the limit at 1 degree of freedom for 0.05 level of significance. The developed model of linear regression has lower R^2 value (0.204) and shows poor performance in prediction of crashes. Hence it is not fit for prediction of crashes. The models emphasize that there is influence of pavement condition on non-fatal crashes. This value is higher than shoulder condition. Improvement in pavement condition and increase in shoulder width may reduce non-fatal crashes.

4.5.4.2 Poisson Regression

Poisson Regression: R^2 value of Poisson is very low than linear and logit. It is not applicable for model development. The variable coefficients of Poisson are zero. The chi-square values also indicates zero.
As chi-square values are not within the limit of this model and are not suitable for development of crash prediction. There is no relationship between the dependent and independent variables. Due to the improper values of R^2 and chi-square parameters of this model is not possible.

4.5.4.3 Logit Model

Logit Model, The R^2 value of logit is higher than that of Linear and Poisson models. The performance of logit model is analyzed using two parameters - chi-square and R^2 values. The variable coefficients are non-zero. The choice of Logit model depends on dispersion in the data. For all types of crashes chi-square values are within the limit at 0.05 level of significance. The chi-square values are significantly different from zero. The occurrences of non-fatal crashes are more due to improper shoulder provided along pavement edge and condition of pavement. It is observed from the model constants that as the pavement condition deteriorates which is also influenced by width of the shoulder more numbers of non-fatal crashes are occurred.

4.5.4.4 Collision Type Analysis

Using multinomial logistic regression, number of crashes as dependent variable, factors as collision and covariates as vehicles are considered. Collision analysis carried through SPSS package and its values are provided in table 27 Chi-square values are within the limit for vehicles and collision type. Two-wheelers and pedestrians are more effected when road crashes takes place.

From the analysis it is observed that side swipe type of collision is occurring more when compared to other type of collision. Swipe and rear-end collisions are significant for non-fatal type of crashes.

Table 27 Chi-square values

	FC	NFC	TC	PC	SW	AADT	SC	Age
FC	1							
NFC	-0.077	1						
TC	0.121	0.883	1					
PC	-0.136	0.305	0.232	1				
SW	-0.216	-0.190	-0.267	0.125	1			
AADT	-0.210	0.176	0.142	0.146	0.225	1		
SC	0.019	0.015	-0.033	-0.004	0.191	0.142	1	
Age	0.083	0.133	0.177	-0.109	-0.110	-0.012	0.119	1

Model	Variable	Non-Fatal crashes			Fatal crashes			Total crashes		
		B	Std. error	t-statistic	B	Std. error	t-statistic	B	Std. error	t-statistic
Multiple Linear Regression	Constant	-0.21	0.57	-0.36	0.702	0.55	1.271	0.25	0.485	0.5
	PC	0.288	0.12	2.37	-0.07	0.12	-0.579	0.199	0.103	1.93
	SW	-1.45	0.78	-1.85	-0.88	0.75	1.164	-1.48	0.663	-2.2
	SC	0.02	0.11	0.17	0.049	0.10	0.481	-0.01	0.09	-0.1
	Age	0.007	0.01	1.01	0.002	0.01	0.298	0.007	0.006	1.2
	AADT	0.001	0.01	1.33	-0.001	0.01	-1.122	0.001	0.001	1.2
	R^2	0.202			0.09			0.204		
		B	Std. error	χ^2	B	Std. error	χ^2	B	Std. error	χ^2
Poisson Regression	Constant	-0.59	0.64	0.86	-0.796	0.38	4.38	0.251	0.218	1.3
	PC	0.256	0.24	1.13	0	0	-	0	0	-
	SW	0	0	-	-3.6	2.84	1.608	-1.11	1.45	0.59
	SC	0	0	-	0	0	-	0	0	-
	Age	0	0	-	0	0	-	0	0	-
	AADT	0	0	-	0	0	-	0	0	-
	R^2	0.025			0.044			0.013		
Logit Model	Constant	-4.20	5.04	0.69	0.98	2.71	0.130	-12.5	8.508	2.16
	PC	2.00	1.63	1.50	-0.193	0.61	0.099	2.37	2.125	1.24
	SW	2.49	8.04	0.09	-4.627	3.75	1.519	-10.2	12.03	0.72
	SC	0.95	1.14	0.70	0.288	0.53	0.294	-0.43	1.59	0.07
	Age	-0.04	0.07	0.28	0.015	0.03	0.189	0.06	0.103	0.34
	AADT	0.209	0.01	0.01	-0.004	0.01	1.373	0.006	0.012	0.21
	R^2	0.57			0.132			0.7		

Effect		Significance	Obtained Chi-Square Value	Table Chi-Square Values	Degree Of Freedom
Vehicles		0.095	2.431	3.841	1
Collision	Side Swipe	0.154	3.867	5.991	2
	Rear-End	0.060			
	Head-On	0			

CHAPTER FIVE: CONCLUSIONS AND RECOMMENDATIONS

5.1 Summary and Conclusions

The study has reviewed crash prediction models on road traffic crash occurrences using different approaches which have been studied in different areas. Five Selected articles were reviewed and assessed in the study, particularly the method used, the variables considered in the study and the result of each study were the focus of the author during review.

Since it is difficult to consider all cash contributing factors during modeling, most researchers consider the most contributing variables for a given study area.

Generally,AADT,85th percentile spot speed, length of road segments, number of lanes, surface condition of road, number of road access and number of road access controls, environmental and traffic characteristics were the main factors considered in most studies.

Similarly the approaches used to prediction of road crash accidents were very unique for different study area. Researchers have developed a number of statistical analytical tools for analyzing crash data in the past. For example, crash count modeling techniques are used in the highway safety analysis. Lao et al. (2011) stated that different types of explanatory variables can affect road crash frequency such as road geometric, driver behaviors, vehicle, and environment. The researches indicated that both behavioral factors related to the driver's errors, and non-behavioral factors related to road geometry, vehicle, and environment can significantly affect road traffic crashes.

Road traffic accident prediction models were at the beginning based on the simple multiple linear regression models that accepts normally distributed errors (Caliendo et al., 2007). Researchers concluded that crash occurrences more fitted with the Poisson distribution, which was developed by an advanced modeling technique (i.e. generalized linear models) (GLM), parameter Estimation Methods, and Neural Network techniques were recently used to model the relationships between road geometry, site characteristics, traffic variables and the expected number of resulting crashes on roadway segments or intersections.

5.2 Recommendations

Based on the review and assessment made, as well as from the perspectives of the main causes of road crashes, the following recommendations were given to facilitate road safety at the road segments.

•The Road Traffic Management Agencies in different countries should expend considerable resources and effective monitoring in an effort to improve safety by implementing countermeasures that include improving highway geometrics, highway signing, and other road safety considerations.

- High raised pedestrian crossing structures should be provided specially in front of markets, schools and church areas considering disabled persons.
- Minimizing road access (or driveways), placing warning signs at the roads access, and also removing obstacles from road layouts.
- Providing well-constructed sidewalk on the both sides of road segments for pedestrians.
- Provide access control (or fences) to guide both vehicles and the pedestrians.
- Providing properly channelized traffics.
- Banning street shops (or not to sell) some goods on the sidewalks of road segments.
- Trimming trees from the road medians that limits the visibility of the driver during the turning maneuvers.
- Providing Posted speeds on each direction of road segments and taking measures on the drivers that break speed limits.
- The transport authority should consider proportionating a number of lanes with annual average daily traffics as safety criteria, and as well as number of horizontal and vertical curves, sidewalks and driveways while designing the new street and the rehabilitations of the existing road geometric design elements.
- Providing appropriate crash location information.

REFERENCE

Abbas, K. A. (2004). Traffic safety assessment and development of predictive models for accidents on rural roads in Egypt. Accident Analysis and Prevention, Vol. 36,149-163

Abdulhafedh, A. (2017). Road Crash Prediction Models : Different Statistical Modeling Approaches. 190–205. https://doi.org/10.4236/jtts.2017.72014

Aderinola, O. S. (2020). A Parametric Model for Accident Prediction Along Ado Ekiti – Ikole Ekiti Road , Ekiti State , Nigeria. 5(8), 980–985.

Brijs, T. (2015). A review of accident prediction models for road intersections. June.2015

Deo, Chimba. (2004). Evaluation of Geometric and Traffic Characteristics Affecting The Safety of Six-Lane Divided Roadways. M.Sc, The Florida State University, College of Engineering.

Douglas, W., Joseoh, E., and Kelth, K.,. (2000). Operational and Safety Effects of Highway Geometrics at the Turn of the Millennium and Beyond. TRB, National Research Council,Washington, D.C,.ECE. (2005). Mobility and Transport.

Fred L. Mannering, 2010. Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives. Article in Transportation Research Part A Policy and Practice· June, University of South Florida.

Hassouna, F. M. A., & Al-sahili, K. (2020). "Practical Minimum Sample Size for Road Crash Time-Series Prediction Models. 2020".

H, N. C., & I, N. G. (2015). Statistical Model Of Road Traffic Crashes Data In Anambra State , Nigeria : A Poisson Regression Approach. 4(09), 226–233.

http://www.itm.su.se/documents/poster/Road_sweepers_20120308. pdf(last accessed on 30.10.2012)

Iyinam, Af ,Iyinam, S and Ergun, M. (1997). Analysis of relationship between highway safety and road geometric design elements: Turkish case. Faculty of Civil Engineering, Turkey, 1-8.

Kalokota, K. R., & Seneviratne, P. N. (1994). TWO-LANE RURAL HIGHWAYS. May.1994

Miaou, S. P. (1994). "The relationship between truck accidents and geometric design of road section: Poisson versus negative binomial regressions. Accident Analysis and Prevention, 1-482."

Milton and Mannering. (1998). The relationship among highway geometrics, traffic related elements and motor vehicle accident frequencies. Journal of Transportation, Vol.25(4), 395-413.

Mohammed, H. A. (2013). The Influence of Road Geometric Design Elements on Highway Safety . International Journal of Civil Engineering and Technology (IJCIET), 1-18.

Mohita Mohan Garnaik. (2014). Effects of Highway Geometric Elements on Accident Modelling.Msc. Thesis, 1-125.

Niveditha, V., Ramesh, A., & Kumar, M. (2016). Development of Models for Crash Prediction and Collision Estimation- A Case Study for Hyderabad City. 3(2), 143–150.

Ogwueleka, F. N. (2014). An Artificial Neural Network Model for Road Accident Prediction : A Case Study of a Developing Country. 11(5), 177–197.

Obaidat, M. T., and Ramadan, T. M. .(2012). Traffic accidents at hazardous locations of urban roads. . Jordan Journal of Civil Engineering, Vol.6(4), 436-447.

Repository, I. (1990). Institutional Repository Predicting accident frequency at their severity levels and its application in site ranking using a two-stage mixed multivariate model. 1979–1990.

Road Injuries and Fatalities - The Health Impacts of Urban Sprawl. An Information Series from Ontario College of Family Physician. Vo. 2. www.ocfp.on.ca

Road Safety Country Profile (2005) Magyarország, Hungary, pp. 1-2

Sarbaz Othman & Robert Thomson.(2009). Influence of Road Characteristics on Traffic Safety, Sweden.1-10.

Subramanian, R., (2003) Motor Vehicle Traffic Crashes as a Leading Cause of Death in the United States, 2001. U.S. Department of Transportation, National Highway Traffic Safety Administration, document DOT HS 809 695

Tefera Bahiru. (2015). Influence of Road Traffic Management and Geometric Characteristics on Traffic Safety Case Study Addis Ababa, Ethiopia, Master Thesis. Addis Ababa.

Timur, M. (2010). Analysis and evaluation of safety impacts of median types and midblock left turn treatments for urban arterials. Phd Dissertation, University of Nevada, Las vegas.

Tobey, N. H., Shunamen,E. M.,& Knoblaunch, R.(1983). Pedestrian Trip Making Characteristics and Exposure Measures. Federal Highway Administration.

Tom Brijs, ElkeHermans. (2014). A review of accident prediction models for road intersections. Research Gate, 1-72.

Thomson, R. W. (2009). Identifying Critical Road Geometry Parameters Affecting Crash Rate and Crash Type; IDENTIFYING CRITICAL ROAD GEOMETRY PARAMETERS. October.2009

Tulu, G. S. (2015). Pedestrian Crashes in Ethiopia: Identification of Contributing Factors through Modeling of Exposure and Road Environment Variables. PhD Dissertation, Research Gate, 1-314

Tulu, GS, Washington, S and King, MJ. (2013). Characteristics of police-reported road traffic crashes in Ethiopia over a six year period.roceedings of the 2013 , 1-13.

WHO.(2004). Road Traffic Accident.

Zhang, Z., Yang, W., & Wushour, S. (2020). Traffic Accident Prediction Based on LSTM-GBRT Model. 2020.