TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

PULCHOWK CAMPUS

THESIS NO: T02/072

Development of Crash Frequency Prediction Model and Identification of Hazardous site locations: A case study of BP highway

by

Bidhan Dahal

A THESIS

SUBMITTED TO THE DEPARTMENT OF CIVIL ENGINEERING IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN TRANSPORTATION ENGINEERING

DEPARTMENT OF CIVIL ENGINEERING

LALITPUR, NEPAL

DECEMBER, 2019

COPYRIGHT

The author has agreed that the library, Department of Civil Engineering, Pulchowk Campus, Institute of Engineering may make this thesis freely available for inspection. Moreover, the author has agreed that permission for extensive copying of this thesis for scholarly purpose may be granted by the professor, who supervised the research work recorded herein or, in their absence, by the Head of the Department wherein the thesis was done. It is understood that the recognition will be given to the author of this thesis and to the Department of Civil Engineering, Pulchowk Campus, Institute of Engineering in any use of the material of this thesis. Copying or publication or the other use of this thesis for financial gain without approval of the Department of Civil Engineering, Pulchowk Campus, Institute of Engineering and author's written permission is prohibited.

Request for permission to copy or to make any other use of the material in this thesis in whole or in part should be addressed to:

.....
Dr. Bharat Mandal
Head of Department of Civil Engineering
Pulchowk Campus, Institute of Engineering
Lalitpur, Nepal

TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

PULCHOWK CAMPUS

DEPARTMENT OF CIVIL ENGINEERING

APPROVAL PAGE

The undersigned certify that they have read, and recommended to the Institute of Engineering for acceptance, a thesis entitled "Development of Crash Frequency Prediction Model and Identification of Hazardous site locations : A case study of BP highway" submitted by Bidhan Dahal (072/MST/252) in partial fulfilment of the requirements for the degree of Master of Science in Transportation Engineering.

_____

Supervisor: Anil Marsani

Program Coordinator

M.Sc. in Transportation Engineering

Department of Civil Engineering

_____

External Examiner: Saroj Kumar Pradhan

Deputy Director General

Department of Roads

_____

Committee Chairperson: Anil Marsani

Program Coordinator

M.Sc. in Transportation Engineering

Department of Civil Engineering

Date: _____

ABSTRACT

Crash prediction models (CPMs) have been used in many countries as a useful tool for road safety analysis and design. Each model is different in terms of methodology, data accuracy, variability in highway geometry and predictor variables used to predict crashes.

This research focuses on developing a relationship between crash counts and roadway attributes, namely curve density, length of horizontal curves, maximum length of continuous tangent, maximum longitudinal grade, average longitudinal grade, access density, minimum sight distance within a segment, minimum radius of curvature and average lane width. Generalized Linear Modelling Technique based on Poisson distribution was selected for the development of model.

The model was developed using the crash and road attribute data of Section II of BP highway. Out of the predictor variables, access density, minimum horizontal sight distance, maximum length of continuous tangent and minimum radius of curvature were found to be the most significant predictors. The proposed model was validated using crash and road attribute data from Section III of BP highway. The R^2 values obtained for the initial developed model was 0.509 whereas the one obtained during model validation was 0.4308. R^2 value obtained for the final model using both the core data-set and the data used for validation was obtained as 0.516.

Keywords

Crash prediction model, Generalized Linear Modelling Technique, Poisson distribution

ACKNOWLEDGEMENT

Firstly, I would like to express my sincere gratitude to Anil Marsani, Co-ordinator, M.Sc. in Transportation Engineering and my supervisor for this thesis assignment, for his continued guidance and motivation throughout my research work.

I am equally thankful to Saroj Kumar Pradhan (Deputy Director General, DOR) and Dr. Pradeep Shrestha (Assistant Professor, IOE) for their valuable opinions, ideas and suggestions during the course of my thesis.

I would also like to extend my deep gratitude to Bindu Shumsher Rana for his valuable insight on the situation of road safety in BP highway and helping me in the acquirement of road crash data. I am equally thankful to Dhulikhel-Sindhuli-Bardibas Road Project Office, Minbhawan for providing me with the as-built drawings which formed an integral basis for the research work.

Lastly, I would like to thank Er. Amit Shrestha and Sub Er. Santosh Sah for assisting me in data collection for my thesis.

Bidhan Dahal

072/MST/252

December, 2019

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS AND ABBREVIATIONS

CPM	- Crash Prediction Model
HSM	- Highway Safety Manual
WHO	-World Health Organization
AASHTO	- American Association of State Highway and Transportation Officials
SPF	- Safety Performance Function
GLM	- Generalized Linear Model
NB	-Negative Binomial
WHO	-World Health Organization
RTA	-Road Traffic Crashes
SDG	-Sustainable Development Goals
MLR	-Multiple linear regression
AADT	-Average Annual Daily Traffic
NB	-Negative Binomial
CPW	-Crash Point Weightage

CHAPTER I
INTRODUCTION

1.1 Background

Road safety, as we know, is an issue of global concern, leading to high number of injuries and fatalities each year throughout the world, and therefore a comprehensive understanding of traffic safety and ways to maintain traffic safety are always emphasized in transportation engineering. Road crashes and Crashes are generally used interchangeably. They are the incidences of injuries and fatalities resulting from a combination of four contributing elements – the driver, the road, the vehicle, and the environment.

Crash-prediction models are decision-making tools for transportation engineers to provide an estimate of expected crash frequency as a function of various Predictor variables depending on the scope of study. Modeling of crash count data is considered as an important task in road safety. The number of crash occurrences within a given time frame is called the crash frequency, which is used as an indicator of the crash occurrence at highways or certain segments of the roads.

CPMs have been developed for various kinds of roads in the past in different countries. The most prominent of the ones developed is the Safety Performance Function (SPF) suggested by Highway Safety Manual (HSM) to be used after applying calibration factor for local conditions. As the manual is applicable only to road segments of homogenous characteristics, researchers have recommended developing indigenous models to predict crash frequencies in developing countries where heterogeneity in traffic composition is observed (Shah and Basu, 2017).

As road crash is a rare event, typically, generalized linear models (GLMs) have been used to model crash outcomes based on Predictor variables(average annual daily traffic, lane width, segment length, presence of shoulders, access density etc). Generally Poisson and Negative binomial models have been extensively used for the purpose.

1.2 Problem Statement

Although initially envisioned as a bypass road, due to shorter travel time, BP highway has been exposed to traffic overload. The road crash data in all four sections of the highway from 2008 to 2016 indicate that there have been 1308 casualties; out of which 241 have been fatal injuries. In the study, a total of (70+22) road segments in the highway starting from Sindhulimadhi (Chainage 0+000 of Section II) to Purano Jhagajholi (Chainage 20+000 of Section III) have been considered as these critical sections have not been used in previous studies even though they have multiple accident prone-locations with varying geometric features.

1.3 Research objective

The main objective of this thesis is to develop a crash prediction model for the study area.

The specific objectives of this research study are:

1) To explore the relationship between the crash frequency and predictor variables related to roadway geometry.
2) To identify hazardous segments within the study area.

1.4 Scope and limitations of the study

1) The research is based upon the road crash data maintained by Dhulikhel-Sindhuli-Bardibas-Road Project and the missing data was found out from inquiry with local people and area police offices.
2) The crash data for the last 5 years was used for analysis.
3) Section II and parts of Section III were used for analysis.
4) Types of crashes were not considered as the records were missing in a number of cases.
5) The prediction is based upon the relationships between the crash data and highway geometry. Human factors and speed compliance have not been considered for the purpose of this research because of time limitation.

1.5 Organization of study

The thesis is divided into five chapters. Chapter One provides the background of the thesis, problem statement, objectives, scope and limitations of the thesis work. Chapter Two provides a review of the relevant literature associated with crash prediction model. Chapter Three consists of the methodology used for the purpose of the research. In Chapter Four, the model development and validation processes are elaborated. In Chapter Five, the results are analyzed and interpreted. Chapter Six contains the conclusion and recommendations.

CHAPTER II
LITERATURE REVIEW

2.1 General Overview on Road Crashes

As of 2018, road traffic injuries are the eighth leading cause of death, first among children aged 5-14 and young adults aged 15-29. 54% of deaths caused by road traffic Crashes (RTAs) are pedestrians, cyclists and motorcyclists. Low-income countries like Nepal have been hit hardest as 13% of all deaths occur in low-income countries even though their percentage share of vehicles is just 1%. Although the issue has been gained some attention internationally, for example, in United Nations Sustainable Development Goals (SDG), "halving the number of global deaths and injuries from road traffic Crashes by 2020" has been set as a Target under Goal 3: Ensure healthy lives and promote well-being for all at all ages, the actual achievements and milestones in reaching the goal have been met (WHO, 2018). If current trend persists, more than two million people are expected to die in road crashes per year by 2030 (WHO, 2018). Currently, road crashes are ranked as the ninth most, and without new initiatives to improve road safety. Fatal crashes are likely to rise to from the ninth place to the third place in the most serious cause of death in the world by the year 2020 (WHO, 2018). Road traffic injuries cost 2.0 percent to 3.0 percent of the Gross National Product of developing countries, which is twice the total amount of development aid provided to developing countries (World Bank, 2015).

Nepal has seen a continuous rise in road crash occurrences and fatalities in the past few years. According to Nepal Traffic Police data, road traffic accident incidents in Nepal have increased from 4,637 in 2007/08 to 10,965 in 2017/18 with the number of fatalities increasing from 1,131 to 2,541.

2.2 Factors Affecting Road Traffic Crashes

The contributing factors that lead to an actual event of crash occurrence are multi-dimensional. They have been generally classified in relevant literature into behavioral factors related to driver behavior and non-behavioral factors related to highway geometry, vehicle and traffic conditions, road side environment, etc (Caliendo et al., 2007). Risk factors associated with crash occurrences are further classified into the following groups (Greibe, 2003 ; Abdulhafedh, 2016) :

Table 2.1 Risk factors affecting road crashes

S.N.	Risk factors	Description
1	Driver behavior	Alcohol and drug abuse, psychological factors, use of electronic devices while driving
2	Vehicle factors	Type of vehicle and design, operating efficiency of mechanical parts
3	Roadway characteristic	Road geometry, shoulder width and type, sight distance, road safety barriers, traffic signals
4	Traffic volumes	AADT (vehicle flow over a road section on an average day) or VKT (vehicles kilometers travelled)
5	Environmental factors	Weather and light conditions
6	Time factors	Season, month, hour

2.3 Crash Prediction Models

Crash prediction models are widely used to estimate the frequency of crashes for a given spatial unit over a certain period of time using various factors related to traffic characteristics, road user characteristics, highway geometry, etc.

Schneider et al. (2009) developed a crash prediction model for truck crashes on horizontal curves using truck ADT, passenger vehicle ADT, and degree of curvature and segment length. Other studies have developed crash prediction models for horizontal curves using limited variables. Bonneson et al. (2005) developed horizontal curve crash prediction models for multilane highways using radius and speed limit data. Similarly, Fitzpatrick et al. (2009) developed a crash prediction model for freeways using single in response variable: degree of curvature and assuming zero degree as the base condition. Likewise, there have been other studies on significant variables affecting crash frequency. 500-ft radius curve was found to be 200% more likely to produce a crash than an equivalent tangent section, and a 1,000-ft radius curve is 50% more likely to produce a crash than an equivalent tangent section (Zegeer et al.,1991).

Although crash prediction models were initially based on MLR (Multiple linear regression) models, but as the data was found to be better fitted with the Poisson distribution, it was started to be used using an advanced modeling technique called the Generalized Linear Models (GLM), instead of the conventional multiple linear

regression technique (Caliendo et al., 2007).

Multivariate regression models specifically Poisson regression model and Negative Binomial model have been widely used in the crash prediction models (Lord, D. and Mannering, F., 2010). Negative Binomial (NB) distribution (or Poisson-Gamma) overcomes the problem of mean equal to variance in Poisson distribution, and is considered more accurate for over-dispersed data (Geedipally et al.,2012).

2.4 Segmentation

Various segmentation approaches have been used to segregate the crash data based on their location. The Highway Safety Manual has prescribed the use of homogeneous segments with respect to AADT, lane width, , curvature, number of lanes, driveway density, shoulder width, shoulder type, roadside hazard rating, median width and clear zone width. The manual has suggested the minimum segment length to be no less than 0.10 miles to ensure ease of calculation and consistency in results (AASHTO, 2010).

As those variables may not be always be available, some researchers have questioned the practicality of such methods (Koorey, 2009; Fitzpatrick et al.,2006). Koorey (2009) has further inferred that variable and fixed- length segments both have their pros and cons as variable-length road segments seem intuitively more useful than fixed-length segments as the latter consists of multiple attributes but as the segments get shorter, the advantage is almost non-existent. He has also suggested that fixed length segments are computationally easier to create from constant-interval raw data.

Although it may seem intuitive to use short segments for a better meaningful interpretation of the results for localized safety interventions, transportation researchers have suggested that shorter segments, when used for the purpose of development of crash prediction models, are prone to high variation leading to uncertainty in the models (Souleyrette et al., 2007, Green,2018; Srinivasan et al., 2011, Lu et al., 2013). D' Agostino (2013) has indicated that short segments as well as those that are too long may not allow for proper statistical inference that can be drawn from the model to be used in identification of sites with safety problem.

Recent research (Cafiso et al.,2018; Green, 2018) have gone to great depths on investigating the statistical implications of various segmentation strategies on the performance of the crash prediction models. Cafiso et al. (2018) has discussed that while crash-based segmentation is likely to identify optimal segments for safety

analysis, it is less practical than a fixed segment based on roadway data. After comparative analysis of various segmentation approaches based on goodness of fit, Green (2018) found out that the segmentation approach with fixed length of 650 m, coinciding with the maximum length of an interchange area, and selected to be just longer than the longest horizontal curve, gave the best results.

2.5 Identification of hazardous locations

Various methods have been proposed for the identification and ranking of hazardous locations based on crash frequency and severity. Zegeer et al.(1974) published a set of methods that were used for identification of accident-prone locations by various transportation agencies in United States based on accident data using critical accident indicator. Fayaz et al. (2018) have used crash weightage formula as an alternative to using number of crashes for blackspot identification and ranking in Kerala City with weightage values of 6, 3 and 1 assigned to Fatal, Severe and Minor Crashes.

Mustakim et al. (2011) have suggested the use of Crash Point Weightage including the property-damage only crashes in the formula. The formula is as follows:

Crash Point Weightage= F*6+S*3+M*0.8+D*0.2 Equation 1.1

Where F=Number of fatal crashes

S= Number of severe crashes

M=Number of minor crashes

D= Number of property damage only crashes

Each crash severity level has its own weightage. For crashes involving fatalities, the numbers are multiplied by 6.0. For serious crashes, minor crashes and damage only crashes, the numbers are multiplied by 3.0, 0.8 and 0.2 respectively. The weights assigned to each level of severity is based on empirical judgement. This weightage formula is widely used in South-east Asia and has also been used recently in safety ranking for Slovenian roads (Zanne et al., 2018)

CHAPTER III
METHODOLOGY

3.1 Introduction

This chapter elaborates on the methodology used for the research work. The methodological framework is summarized in Figure 3.1.

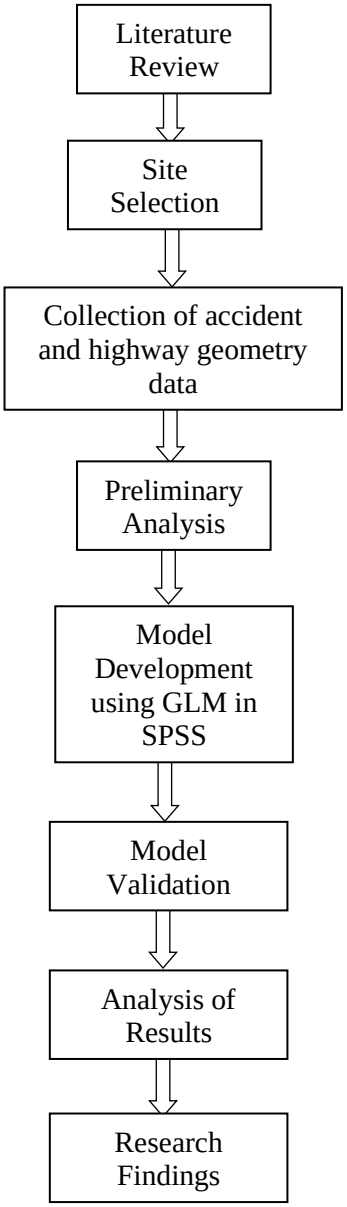

Figure 3.1 Methodological Framework

3.2 Site Selection

BP Highway (Banepa-Sindhuli-Bardibas Road) is the shortest linking road between Kathmandu valley and the Terai region of Nepal. The highway is divided into the following 4 sections:

Section I : Bardibas- Sindhulibazar section (37 km)

Section II : Sindhulibazar- Khurkot section (39.7 km)

Section III : Khurkot - Nepalthok section (32.9 km)

Section IV :Nepalthok-Dhulikhel section (50 km)

Section II (Sindhulibazar-Khurkot) of BP highway was chosen for the model development. As this critical section has not been used in previous research even though the trend of crash incidences in increasing in this section, it was thus selected for analysis. This section is 39.7 km long. 20 km of Section III (Khurkot-Nepalthok) was used for model validation. The sections used for the purpose of the thesis are shown in Figure 3.2.

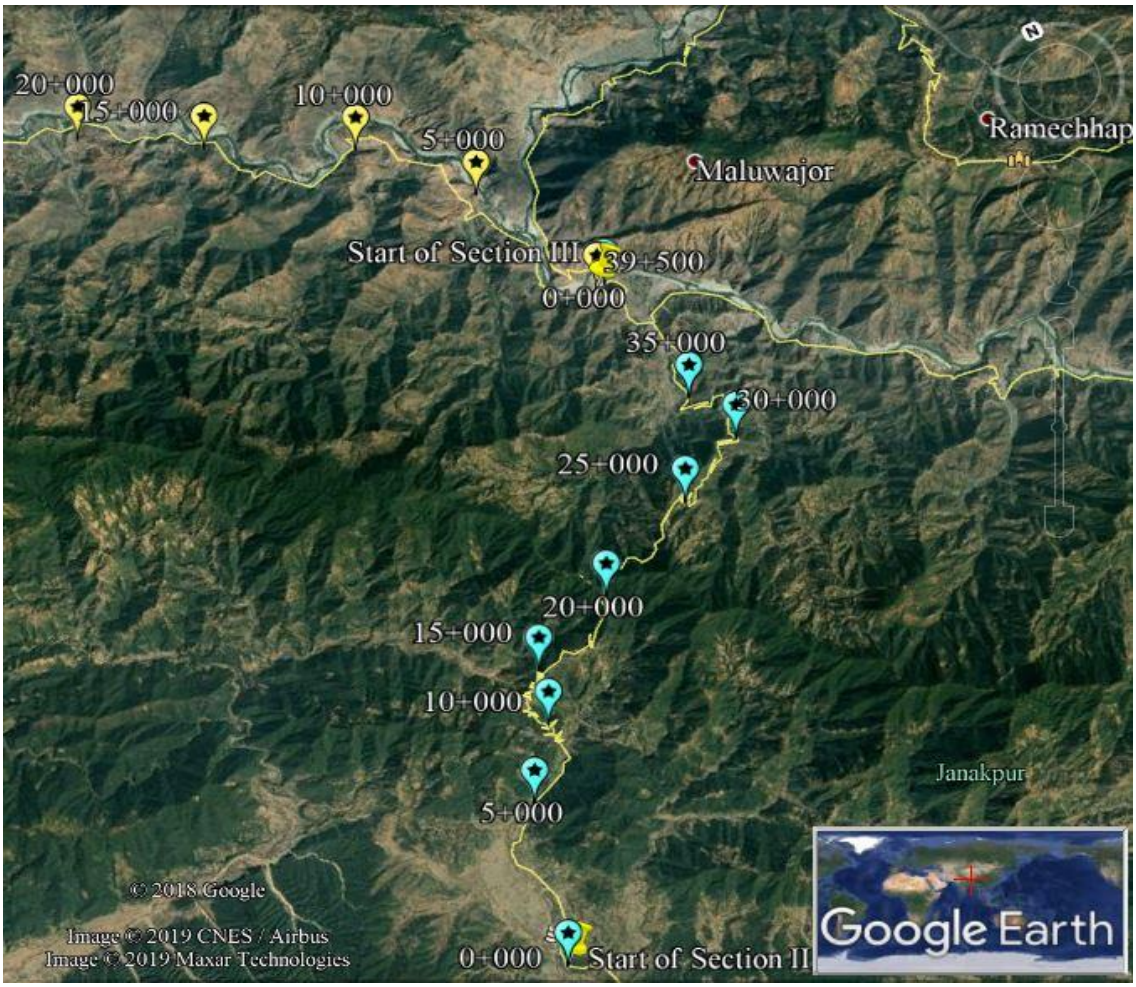

Figure 3.2 Location Map

These sections have multiple accident prone-locations with varying geometric features. The sections have been chosen because they possess a combination of horizontal curves and straight segments which is expected to aid in a more comprehensive analysis.

3.3 Collection of Accident Data and Highway Geometry Data

The crash data was collected from Department of Roads, Dhulikhel-Sindhuli-Bardibas Road Project Office and Area Police Office, Khurkot. The data in which the exact location of the crash site was not included was confirmed with the use of the accident form and public enquiry. Four of the crash locations of 2014 was not included in the analysis as the locations could not be confirmed as they were from 2014 and the crashes were 'Damage Only'. The final sorted accident data as tabulated in Annex IV was plotted in Google Earth as shown in Figure 3.3.

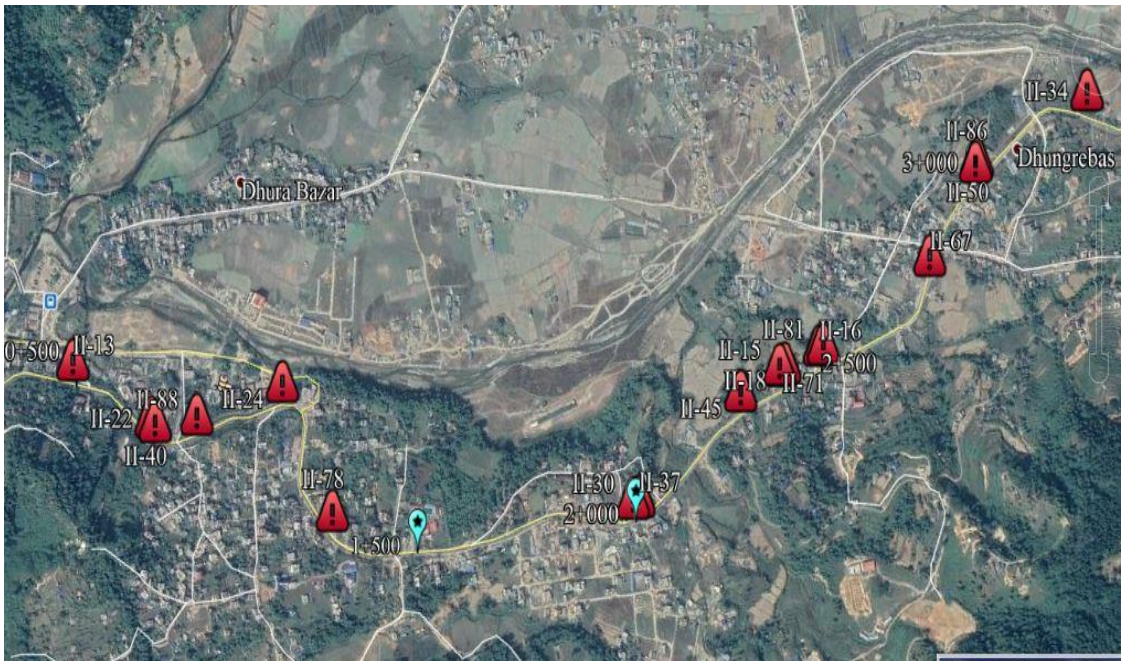

Fig 3.3 Sample Plotting of Accident Data

The highway geometry data of each sections were obtained from the as-built drawings of the sections as shown in Figure 3.4. The sight distance data and lane width were obtained from site by taping. The number of access points were counted during the site visit whereas the historical access point data was obtained from Google Earth.

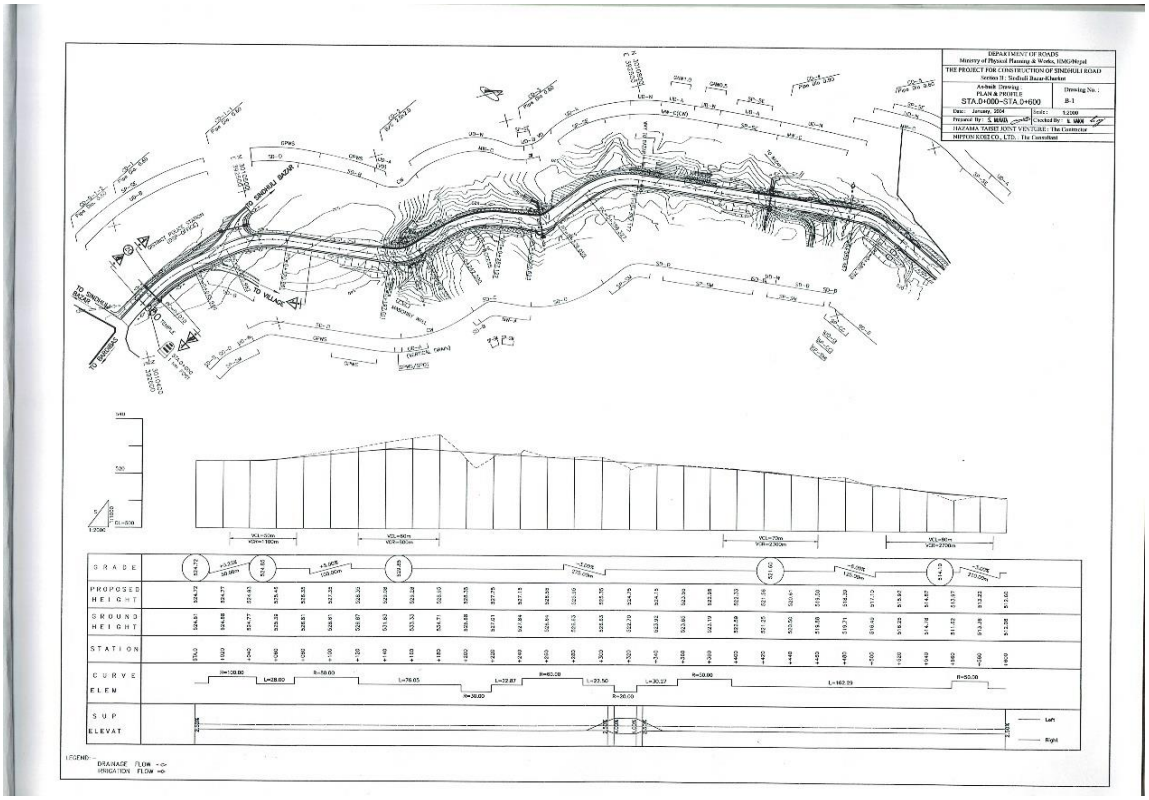

Fig 3.4 Sample of As-built drawing used to extract highway geometric data

3.4 Preliminary Analysis of crash data

3.4.1 Crash Data Summary

Table 3.1 show the total crash occurrences in the last five years in the four sections of BP Highway. The data suggests that crash incidents have been growing at a steady pace. Section II and Section IV have the highest number of crashes at 95 and 96 respectively while crash incidents in other two sections have also been on the rise. It is to be noted that Section II and III saw almost double the number of crashes in 2018 than in 2017.

Table 3.1 Total crash occurrences (Source: Sindhuli Road Maintenance Unit)

Year	Section I	Section II	Section III	Section IV	Total
2014	6	12	10	6	34
2015	12	13	17	18	60
2016	15	3	11	4	33
2017	18	23	15	32	88
2018	30	44	28	36	138
Total	81	95	81	96	353

3.4.2 Vehicle breakdown

Figure 3.5 shows the distribution of the type of vehicles involved in road crashes during the study period. Motorcycles have the biggest share of involvement in road crashes at 35% are the highest whereas car/jeep come in second at 24%.

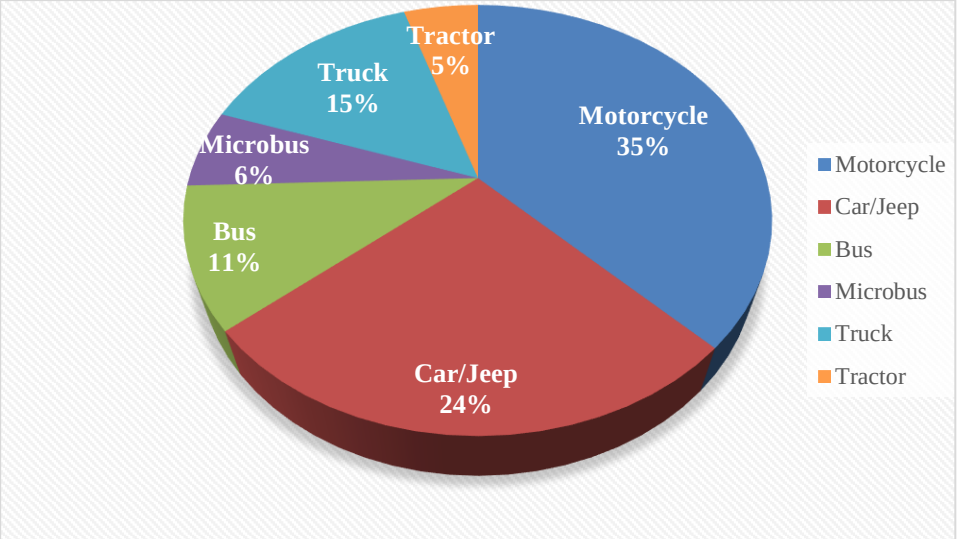

Fig 3.5 Vehicle-wise breakdown of Crashes (Source: Sindhuli Road Maintenance Unit)

3.4.3 Crash type distribution

Table 3.2 shows the distribution of the crashes based on type of collision. Out of the crashes of which the collision type was identifiable, head-on crashes were the most frequent followed by falling down of the vehicle. The data also suggests most of the crashes either don't fall in any of the categories or are unidentifiable.

Table 3.2 Types of crashes based on collision type 2014-2018 (Source: Sindhuli Road Maintenance Unit)

Section	Head On	Over turned	Hit Object on road	Hit Object Off Road	Hit Pedestrian	Fall down	Other	Side Swipe	Total
I	23	3	4	1	12	14	17	7	81
II	34	1	1	0	7	20	26	6	95
III	28	3	0	0	10	0	30	1	81
IV	11	4	0	0	3	0	70	1	96
Total	93	11	5	1	30	31	143	15	353

3.4.4 Casualty Breakdown

Figure 3.6 shows the section-wise breakdown of casualties. The type of injuries have been classified in fatalities, serious injuries and minor injuries. Section I has seen highest number of casualties at 477 followed by Section II and Section I. The number of deaths due to road crashes were highest in Section II at 89 followed by Section I.

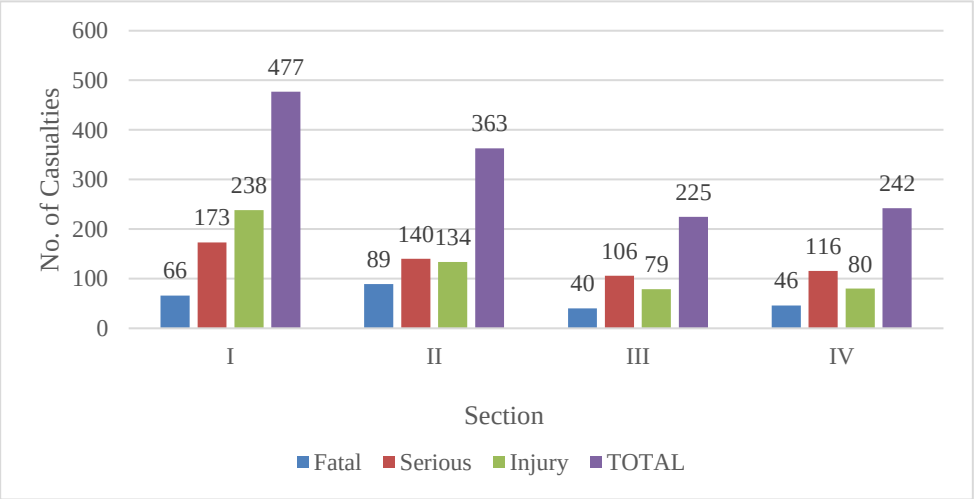

Fig 3.6 Section-wise Casualty Breakdown (Source: Sindhuli Road Maintenance Unit)

3.5 Model Development

3.5.1 Predictor Variables

From the literature review and observation of the crash patterns from Google Earth plot, the following predictor variables were considered for model development.

a. Minimum Radius of Curvature

It is the radius of the sharpest curve in a particular segment.

b. Curve Density

It is the number of horizontal curves per kilometer. The value is found out by counting the total number of curves in the segment and using the following formula:

$CD = N_c * 1000/L$ Equation 3.1

Where,

CD= Curve Density,

N_c=Number of horizontal curves in a segment,

L=Segment length in meters

c. Horizontal Curve Length %

It is the percentage ratio of the total length of horizontal curves in a segment to the total segment length. The total length of horizontal curves was obtained by adding up the individual curve lengths from the as-built drawings. The curves with degree of curvature lesser than 3.5 degrees is excluded from the analysis as they have been found to behave as a straight segment (Khan et al., 2012). As the segment length taken is not constant, the ratio of was used for analysis.

d. Access Density

It is the total number of access points per kilometer. The values are calculated using the following formula:

$$AD = AN * 1000 / L \qquad \text{Equation 3.2}$$

Where,

AD= Access Density

AN=Number of access points

L= Length of segment in meters

e. Minimum horizontal sight distance

It is the minimum value out of the sight distances of the horizontal curves in a particular segment. It was measured in site with measuring tape.

f. Maximum Grade

It is the maximum value of vertical grade within a segment. It is obtained from the as-built drawings.

g. Average Grade

It is the difference between the elevation of the two sides divided by the length of the segment. It is obtained from the as-built drawings.

$$AG = |EL1 - EL2| / L \qquad \text{Equation 3.3}$$

Where,

AG=Average Grade

EL1= Elevation of starting point of the segment in meters

EL2= Elevation of end point of the segment in meters

L=Length of the segment in meters

h. Difference between Maximum Grade and Average Grade

The above value was also included in the analysis to see if there is some abruptness in the gradient of the road which may lead to potential crashes.

i. Maximum length of continuous tangent

It is simply the maximum value of continuous tangent within a segment. As horizontal curves with degree of curvature lesser than 3.5 degrees are considered as straight segment, they are also included in the value if they follow a continuous tangent.

Although Average Annual Daily Traffic (AADT) is considered an essential exposure variable in crash prediction, AADT had to be excluded from the analysis because only the traffic count data of the start and end stations of Section II were available as there were no count stations in between. Since we are dealing with historical crash data of 2014-2018, it was impossible to calculate the AADT value for each of the segments for each year.

3.5.2 Response Variable

The response variable for the purpose of the study is taken as number of crashes. Even though fatal+severe crashes were initially considered for model development, as the number of minor and damage only crashes was less, the process was continued using only the total number of crashes.

The predictor and response variable data are tabulated in Annex I, II and III.

3.5.3 Model Form

GLM based Poisson and Negative Binomial Regression Models were used as predictive models.

a. The Poisson Regression Model

The Poisson model is expressed as:

$$P(n_i) = \frac{\lambda_i \mathrm{EXP}(-\lambda_i)}{n!} \qquad \text{Equation 3.4}$$

Where,

P(ni) : the probability of n crashes occurring on section i of a highway during a period of time,

λ_i: the expected crash frequency on section i of the highway.

Accordingly, the crash frequency can be estimated by the expression:

$$\lambda_i = EXP(\beta X_i)$$ Equation 3.5

where,

λ_i: the response variable (the expected number of crashes per time period), X_i : a vector of the independent (explanatory) variables,

β: a vector of the estimates (coefficients) of the predictor variables Xi.

b. The Negative Binomial (Poisson-Gamma) Regression Model (NB)

The Negative Binomial (or Poisson-Gamma) Regression Model was introduced as an alternative to the Poisson Model to consider the over-dispersion in the crash data counts. The NB model uses Gamma Probability Distribution, and helps in negating the assumption of mean equals the variance in the Poisson regression. The generalized form of negative binomial distribution thus becomes:

$$\lambda_i = EXP(\beta X_i + \varepsilon_i)$$ Equation 3.6

where:

EXP (εi): a gamma-distributed error with mean equals one and variance equals α. This error term which is called the over-dispersion parameter, allows the variance to differ from the mean.

3.5.4 Goodness of Fit Measures

The following goodness of fit measures are provided in SPSS V20 for the case of Generalized linear model which as used for the purpose of model choice.

- **Deviance**

The deviance of a model is based on the difference between the log-likelihood of the model of interest, L_M, and the log-likelihood of the most complex model that perfectly fits the data (i.e. saturated model), L_S. The deviance is represented by the following formula:

$$\text{Deviance} = -2(L_M - L_S)$$ Equation 3.7

In model comparison, the value is divided by Degree of Freedom (DoF). If the resulting value falls between 0.8 and 1.2, the hypothesis that the data will follow a particular type of distribution is considered true.

- **Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC)**

AIC and BIC are both penalized-likelihood criteria used for comparing non-nested models, which ordinary statistical tests cannot do. The AIC or BIC for a model is usually written in the form of:

$$AIC \text{ or } BIC = [-2\log L + kp] \qquad \text{Equation 3.8}$$

Where,

L = likelihood function,

p = the number of parameters in the model, and

k = 2 for AIC and log(n) for BIC.

The smaller the values of AIC or BIC, the better the model is considered.

- **Omnibus Test**

Omnibus Test, also called the Likelihood Ratio Chi-Squared Test is the test of whether all the independent variables collectively improve the model over the intercept-only model with no independent variables. In other words, it indicates the overall predictability of the model. The value of statistical Significance (Sig.) shall fall within $p < 0.05$ (i.e. 95% Confidence Interval) for a model to pass the Omnibus test.

3.5.5 Parameter Estimate

Maximum Likelihood method was used for estimation of the parameters. The values of parameter estimate or coefficients for each individual predictor variables are calculated by maximizing the log-likelihood function.

3.5.6 Wald Chi- Square test

Wald Chi-Square test is used to test the significance of the individual parameter estimates obtained by Maximum Likelihood Method. The Wald Value follows an asymptotic $\chi 2$-distribution under the null hypothesis. The following equation gives the Wald Value:

$$W = (\beta - \beta_0)^2 / \text{var}(\beta) \qquad \text{Equation 3.9}$$

Where,

W = Wald Value

β = coefficient which we are testing against the null hypothesis that it is 0 or the Maximum Likelihood Estimator (MLE)

Since the parameter of interest is usually 0 (i.e. $\beta_0=0$), the Wald statistic simplifies to

$W= \beta^2/$ var (β). The value of statistical Significance (Sig.) shall fall within 0.05 (i.e. 95% Confidence Interval) for a given predictor variable to be included in the model.

3.6 Model Validation

The developed model was validated using the crash data from 20 km of Section III. R-squared was used for model validation to compare the fit of values predicted by the developed model with respect to the observed values.

- R- squared or the coefficient of determination is the square of the correlation between observed values and predicted values.

$$R^2 = \frac{[\sum(O-O_m)(P-P_m)]}{\sum(O-O_m)^2\sum(P-P_m)^2}^2 \qquad \text{Equation 3.10}$$

Where,

O = Observed crash,

P = Predicted crash,

O_m = Mean of observed crashes,

P_m = Predicted mean of the predicted crashes.

CHAPTER IV

MODEL DEVELOPMENT AND VALIDATION

The predictor variables along with their respective codes are shown in Table 4.1.

Table 4.1 Predictor variables and codes

Variable	Coding	Variable Type
Access density	Access_Density	Continuous
Minimum Radius of Curvature	Min_Radius	Continuous
Horizontal Curve Length %	Curve_Length_Per	Continuous
Curve Density (Number of Curves per km)	Curve_Density	Continuous
Minimum horizontal sight distance	Sight_Distance	Continuous
Maximum Grade within a segment	Max_Grade	Continuous
Average Grade within a segment	Avg_Grade	Continuous
Difference between Maximum and Average Gradient	Grade_Diff	Continuous
Maximum length of continuous tangent	Tangent	Continuous
Average Lane Width	Lane_Width	Continuous
Segment Length	Length	Continuous

Table 4.2 Continuous Variable Information

		N	Minimum	Maximum	Mean
Dependent Variable	Crash_No	70	.0	5.0	1.329
Covariate	Min_Radius	70	13.0	150.0	27.16
	Tangent	70	9.1	169.2	63.02
	Sight_Distance	70	15.280	145.37	35.16
	Access_Density	70	.0	18.7	3.261
	Curve_Density	70	1.7	35.0	18.52
	Lane_Width	70	4.5	8.4	5.696
	Max_Grade	70	2.5	10.0	8.030
	Avg_Grade	70	.0	6.6	2.736
	Grade_Diff	70	.0	10.0	5.303
	Curve_Length_Per	70	19.3	75.9	53.066
	Length	70	451.4	707.9	560.493

Table 4.2 provides the information about the maximum, minimum and average values

of the variables. All of the variables used in the analysis are continuous. The maximum number of crashes in the segments is 5 whereas the average number of crashes is 1.329. As the segments are not chosen based upon the occurrence of crashes, there are a number of segments where there have been no crash occurrence in the past five years, which are also included in the model.

Table 4.3 Goodness of fit: Poisson VS Negative Binomial

	Poisson Regression			Negative Binomial Regression		
	Value	df	Value/df	Value	df	Value/df
Deviance	68.887	59	1.168	37.513	58	.647
Scaled Deviance	68.887	59		37.513	58	
Pearson Chi-Square	66.608	59	1.129	33.333	58	.575
Scaled Pearson Chi-Square	66.608	59		33.333	58	
Log Likelihood	-83.896			-91.258		
Akaike's Information Criterion (AIC)	189.792			206.516		
Finite Sample Corrected AIC (AICC)	194.344			211.990		
Bayesian Information Criterion (BIC)	214.526			233.498		
Consistent AIC (CAIC)	225.526			245.498		

Table 4.3 shows the various goodness of fit metrics which are used in the model selection process. The value of deviance / degree of freedom falls within the range between 0.8 and 1.2 in case of Poisson Model whereas it falls short in case of Negative Binomial Regression as it has the value of 0.647 only. Similarly, the Pearson Chi-Square value / degree of freedom is also within the acceptance limit in case of Poisson Regression. Below that, the values of various information criterion (smaller value better) also suggest that the Poisson Model should be chosen for model development.

Table 4.4 shows the omnibus test results. The third column indicates that the independent variables collectively improve the model over the intercept-only model.

Table 4.4 Omnibus Test

Likelihood Ratio Chi-Square	df	Sig.
46.355	10	.000

Table 4.5 Parameter Estimation: Poisson Regression

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	-1.534	1.9815	-5.418	2.349	.600	1	.439	.216	.004	10.478
Min_Radius	.013	.0057	.002	.024	5.044	1	.025	1.013	1.002	1.024
Sight_Distance	-.027	.0088	-.044	-.009	9.272	1	.002	.974	.957	.991
Access_Density	.073	.0328	.009	.138	5.027	1	.025	1.076	1.009	1.148
Tangent	.009	.0036	.002	.016	6.527	1	.011	1.009	1.002	1.016
Curve_Density	.006	.0244	-.041	.054	.067	1	.795	1.006	.959	1.056
Lane_Width	.103	.1717	-.234	.439	.358	1	.549	1.108	.792	1.552
Max_Grade	1.173	2.9644	-4.637	6.983	.157	1	.692	3.231	.010	1078.094
Avg_Grade	-1.162	2.9728	-6.988	4.665	.153	1	.696	.313	.001	106.159
Grade_Diff	-1.120	2.9654	-6.932	4.693	.143	1	.706	.326	.001	109.130
Curve_Length_Per	-.013	.0122	-.037	.011	1.165	1	.281	.987	.964	1.011
Length	.002	.0021	-.002	.006	.738	1	.390	1.002	.998	1.006

Table 4.5 shows the parameter estimates for each of the variables. Out of the variables considered for model development, only minimum horizontal sight distance, access density and maximum length of continuous tangent turned out to be significant variables (Sig <0.05).

Table 4.6 Correlation Matrix

	(Intercept)	Min_ Radius	Sight_ Distance	Access_ Density	Tangent	Curve_ Density	Lane_ Width	Max_ Grade	Avg_ Grade	Grade_ Diff	Curve_ Length_ Per	Length
(Intercept)	1.000	-.145	-.046	-.205	-.118	-.405	-.497	.040	-.045	-.046	-.210	-.783
Min_Radius	-.145	1.000	-.526	.172	-.172	.304	.012	-.022	.010	.019	-.043	.259
Sight_ Distance	-.046	-.526	1.000	-.142	.104	-.021	-.357	.083	-.079	-.084	.274	.045
Access_ Density	-.205	.172	-.142	1.000	-.125	.271	-.155	-.005	.002	.007	-.006	.276
Tangent	-.118	-.172	.104	-.125	1.000	.524	-.281	-.114	.113	.112	-.194	.142
Curve_ Density	-.405	.304	-.021	.271	.524	1.000	-.127	-.085	.080	.083	-.366	.472
Lane_Width	-.497	.012	-.357	-.155	-.281	-.127	1.000	-.029	.037	.035	.045	.043
Max_Grade	.040	-.022	.083	-.005	-.114	-.085	-.029	1.000	-.999	-1.000	.029	.008
Avg_Grade	-.045	.010	-.079	.002	.113	.080	.037	-.999	1.000	1.000	-.036	-.014
Grade_Diff	-.046	.019	-.084	.007	.112	.083	.035	-1.000	1.000	1.000	-.035	-.009
Curve_ Length_Per	-.210	-.043	.274	-.006	-.194	-.366	.045	.029	-.036	-.035	1.000	.032
Length	-.783	.259	.045	.276	.142	.472	.043	.008	-.014	-.009	.032	1.000

Table 4.6 is the correlation matrix showing the collinearity between the independent variables. Correlation of significant variables is only studied. As we can see from the table, the correlation of access density with minimum horizontal sight distance is -.142, access density with maximum length of continuous tangent is -.125 and maximum length of continuous tangent with horizontal distance distance is .104 which are considered weak correlations, so the process is continued with the three selected variables.

Table 4.7 Goodness of fit for revised model

	Value	df	Value/df
Deviance	74.411	64	1.163
Scaled Deviance	74.411	64	
Pearson Chi-Square	65.756	64	1.027
Scaled Pearson Chi-Square	65.756	64	
Log Likelihood[b]	-88.658		
Akaike's Information Criterion (AIC)	185.316		
Finite Sample Corrected AIC (AICC)	185.951		
Bayesian Information Criterion (BIC)	194.194		
Consistent AIC (CAIC)	198.194		

As seen in Table 4.7, both the values of deviance by degree of freedom and pearson chi-square by degree of freedom fall within the acceptance limit of 0.8 to 1.2.

Table 4.8 Omnibus Test for revised model

Likelihood Ratio Chi-Square	df	Sig.
31.130	3	.000

The third column of Table 4.8 indicates that the independent variables collectively improve the model over the intercept-only model.

Table 4.9 shows the parameter estimate for the revised model. As we can see, all three selected variables are statistically significant within the confidence interval. But the variables have to be checked for collinearity. As shown in the correlation matrix of Table 4.10, the correlation between the variables are within the permissible limit. So, the model is considered final.

Table 4.9 Parameter Estimate for revised model

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	-.310	.2359	-.772	.152	1.726	1	.189	.734	.462	1.165
Access_Density	.066	.0289	.010	.123	5.275	1	.022	1.069	1.010	1.131
Sight Distance	-.010	.0058	-.022	.001	3.172	1	.048	.990	.978	1.001
Tangent	.010	.0026	.005	.015	15.439	1	.000	1.010	1.005	1.015

Table 4.10 Correlation matrix for revised model

	(Intercept)	Access_Density	Sight_Distance	Tangent
(Intercept)	1.000	.145	-.428	-.529
Access_Density	.145	1.000	-.178	-.183
Sight_Distance	-.428	-.178	1.000	-.159
Tangent	-.529	-.183	-.159	1.000

The final model obtained is:

Total five-year crashes = EXP (-0.310 + 0.066*Access Density - 0.01*Min Horizontal Sight Distance + 0.01*Maximum Length of Continuous Tangent)

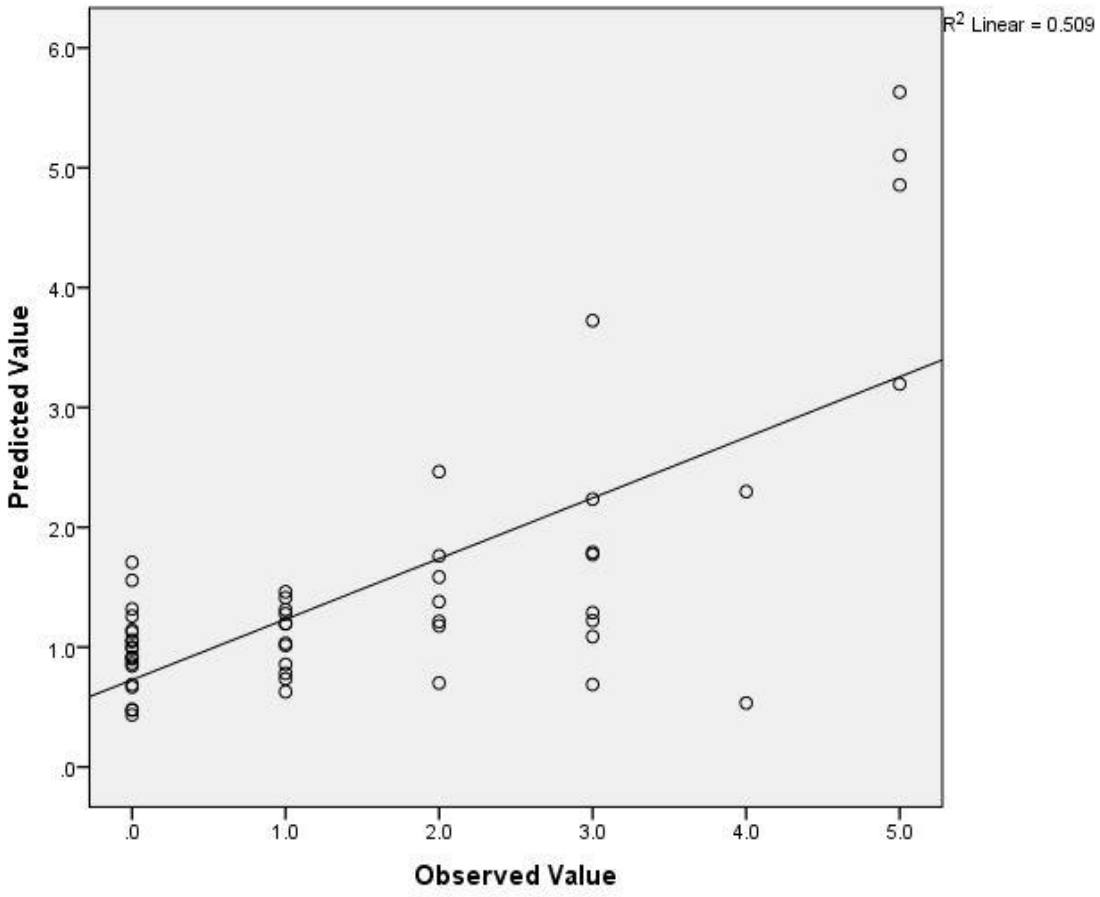

Fig 4.1 Predicted Vs Observed Crashes Plot of model data

The observed VS predicted plot of crashes obtained from SPSS Scatterplot option shows that the R^2 of the model is 0.509 which shows that a good percentage (50.9%) of the variation in the data is explained by the variation in the independent variables.

Model validation

The obtained model is validated using the crash data of 20 km of Section III. The validation results are tabulated in Table 4.11 and the R-Squared value is then obtained by plotting Observed Crashes VS Predicted Crashes from MS Excel 2019 as shown in Figure 4.2.

Table 4.11 Tabulation Chart for Model Validation

Segment No.	CH FROM (CH1)	CH TO (CH2)	Length	Access Points	Access Density	Sight Distance	Tangent	Observed Crash No.	Predicted Crash No.
1	00+095.96	00+600.52	504.56	-	-	42.68	74.890	3	1.012
2	00+600.52	01+127.93	527.41	1	2.49	25.38	23.990	1	0.853
3	01+728.53	02+264.71	536.18	2	3.73	28.34	23.480	2	0.894
4	02+574.17	03+173.58	599.41	3	5.00	24.82	79.760	2	1.767
5	03+673.57	04+225.72	552.15	4.00	7.24	19.31	60.238	3	1.781
6	04+225.72	04+753.64	527.92	1.00	1.89	34.95	51.230	1	0.978
7	04+753.64	05+314.83	561.19	3.00	5.34	24.32	38.640	1	1.204
8	05+706.27	06+257.81	551.54	2.00	3.62	27.56	77.334	6	1.532
9	06+257.81	06+793.13	535.32	6.00	11.20	42.02	89.632	5	2.473
10	06+793.13	07+352.68	559.55	3.00	5.36	47.58	68.964	1	1.294
11	08+209.91	08+755.69	545.78	3.00	5.49	28.68	56.234	1	1.388
12	09+776.22	10+309.02	532.80	1.00	1.87	80.13	104.940	1	1.063
13	10+309.02	10+818.06	509.04	2.00	3.92	29.1	61.270	1	1.311
14	11+338.54	11+907.36	568.82	1.00	1.75	35.27	33.731	1	0.811
15	11+907.36	12+448.57	541.21	1.00	1.84	23.64	89.520	4	1.600
16	13+102.64	13+691.32	588.68	1.00	1.69	26.5	34.101	2	0.885

Segment No.	CH FROM (CH1)	CH TO (CH2)	Length	Access Points	Access Density	Sight Distance	Tangent	Observed Crash No.	Predicted Crash No.
17	13+959.54	14+512.63	553.09	-	-	22.91	42.718	3	0.894
18	14+984.23	15+519.50	535.27	-	-	26.83	65.041	2	1.075
19	15+519.50	16+046.08	526.58	-	-	41.21	20.274	1	0.595
20	17+048.32	17+596.39	548.07	12.00	21.89	85.62	83.240	2	3.037
21	17+896.74	18+413.23	516.49	9.00	17.42	35	78.481	6	3.577
22	18+978.66	19+522.58	543.92	8.00	14.70	31.91	106.029	5	4.061

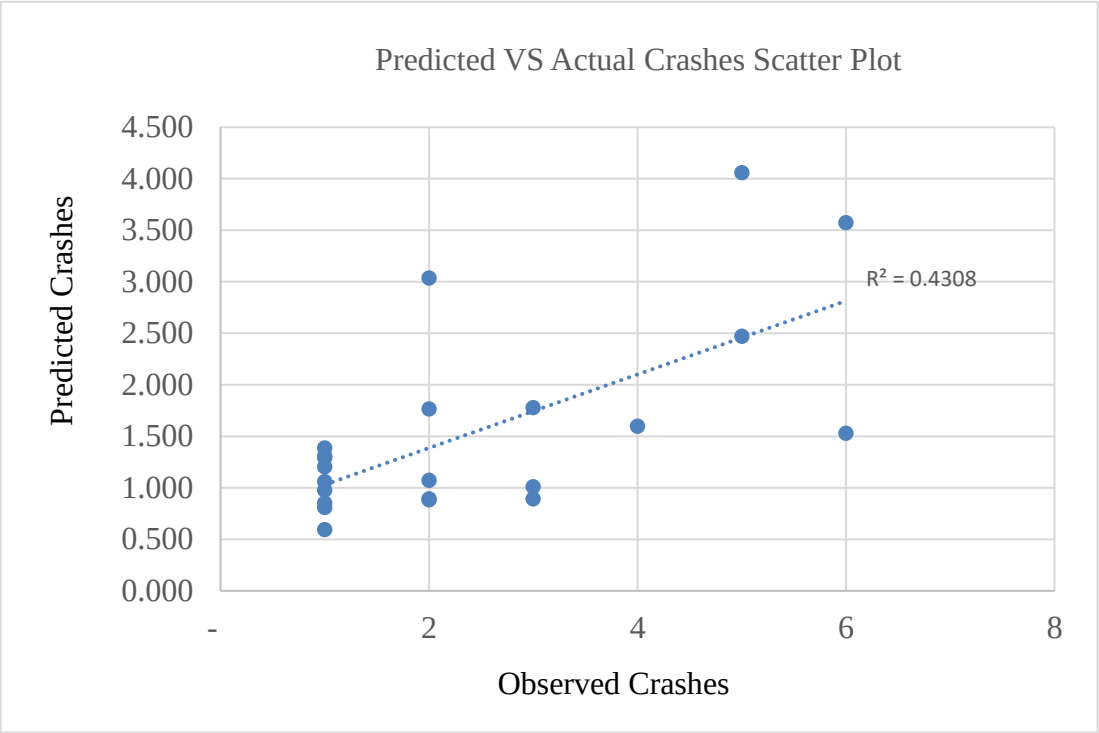

Fig 4.2 Predicted Vs Observed Crashes Plot for Model Validation

Final Model using all data

A final model was developed by using both the original data-set and the data used for model validation.

Table 4.12 Continuous Variable Information: Final Model

		N	Minimum	Maximum	Mean	Std. Deviation
Dependent Variable	Crash_No	92	0.0	6.0	1.598	1.611
Covariate	Length	92	451.4	707.9	556.517	57.930
	Access_Density	92	0.0	21.9	3.747	4.489
	Sight_ Distance	92	15.280	145.370	35.274	19.251
	Curve_Density	92	1.7	35.0	17.625	7.578
	Curve_Length_ Per	92	19.3	85.1	56.300	14.398
	Tangent	92	9.1	169.2	62.768	37.317
	Min_Radius	92	13.0	220.0	33.326	32.142
	Max_Grade	92	.5	10.0	7.562	2.170
	Avg_Grade	92	0.0	8.9	3.142	1.976
	Grade_Diff	92	0.0	10.0	4.426	2.656
	Lane_Width	92	4.5	8.4	5.564	0.763

Table 4.12 shows the information about the maximum, minimum and average values of the variables. All of the variables used in the analysis are continuous. The maximum number of crashes in the data-set is 6 whereas the average number of crashes is 1.598. Altogether 11 variables were selected for model development as earlier.

Table 4.13 Goodness of fit: Poisson VS Negative Binomial: Final Model

	Poisson Regression			Negative Binomial Regression		
	Value	df	Value/df	Value	df	Value/df
Deviance	84.202	80	1.053	36.755	55	.668
Scaled Deviance	84.202	80		36.755	55	
Pearson Chi-Square	77.223	80	.965	33.621	55	0.611
Scaled Pearson Chi-Square	77.223	80		33.621	55	
Log Likelihood	-117.296			-90.879		
Akaike's Information Criterion (AIC)	258.592			267.532		
Finite Sample Corrected AIC (AICC)	262.541			268.740		
Bayesian Information Criterion (BIC)	288.853			294.727		
Consistent AIC (CAIC)	300.853			310.282		

Table 4.13 shows the various goodness of fit metrics which are used in the model selection process. The value of deviance / degree of freedom falls within the range between 0.8 and 1.2 in case of Poisson Model whereas it falls short in case of Negative Binomial Regression as it has the value of 0.668 only. Similarly, the Pearson Chi-Square value / degree of freedom is also within the acceptance limit in case of Poisson Regression. Below that, the values of various information criterion (smaller value better) also suggest that the Poisson Model should be chosen for model development.

Table 4.14 Omnibus Test: Final Model

Likelihood Ratio Chi-Square	df	Sig.
61.345	11	.000

The third column of Table 4.14 indicates that the independent variables collectively improve the model over the intercept-only model.

Table 4.15 Parameter Estimates: Final Model

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	-.222	1.6581	-3.472	3.028	.018	1	.893	.801	.031	20.648
Length	2.984E-05	.0018	-.003	.004	.000	1	.987	1.000	.997	1.004
Access_Density	.059	.0216	.017	.102	7.562	1	.006	1.061	1.017	1.107
Sight_ Distance	-.018	.0060	-.030	-.006	9.055	1	.003	.982	.970	.994
Curve_Density	-.022	.0179	-.057	.013	1.550	1	.213	.978	.944	1.013
Curve_Length_ Per	.002	.0072	-.012	.016	.056	1	.813	1.002	.988	1.016
Tangent	.008	.0031	.002	.014	6.755	1	.009	1.008	1.002	1.014
Min_Radius	.005	.0027	.000	.011	4.058	1	.044	1.005	1.000	1.011
Max_Grade	1.219	2.8030	-4.275	6.712	.189	1	.664	3.383	.014	822.566
Avg_Grade	-1.139	2.8062	-6.639	4.361	.165	1	.685	.320	.001	78.361
Grade_Diff	-1.219	2.8014	-6.709	4.272	.189	1	.664	.296	.001	71.647
Lane_Width	.046	.1500	-.248	.340	.094	1	.759	1.047	.780	1.405

Table 4.15 shows the parameter estimates for each of the variables. Out of the variables considered for model development, only minimum horizontal sight distance, access density and Maximum length of continuous tangent and minimum radius of curvature turned out to be significant variables (Sig <0.05)

Table 4.16 Correlation Matrix: Final Model

	(Intercept)	Length	Access_ Density	Sight_ Distance	Curve_ Density	Curve_ Length_ Per	Tangent	Min_ Radius	Max_ Grade	Avg_ Grade	Grade_ Diff	Lane_Width
(Intercept)	1.000	-.768	-.097	-.007	-.330	-.467	.042	-.466	.017	-.020	-.019	-.588
Length	-.768	1.000	.155	.133	.338	.166	.050	.268	.034	-.037	-.037	.089
Access_Density	-.097	.155	1.000	-.183	.052	-.099	-.164	.189	.031	-.033	-.026	.018
Sight_ Distance	-.007	.133	-.183	1.000	.088	.119	.102	-.187	.081	-.082	-.082	-.404
Curve_Density	-.330	.338	.052	.088	1.000	-.203	.453	.352	-.066	.060	.061	-.057
Curve_Length_ Per	-.467	.166	-.099	.119	-.203	1.000	-.041	.036	-.025	.024	.026	.294
Tangent	.042	.050	-.164	.102	.453	-.041	1.000	-.161	-.092	.086	.086	-.391
Min_Radius	-.466	.268	.189	-.187	.352	.036	-.161	1.000	-.038	.043	.042	.295
Max_Grade	.017	.034	.031	.081	-.066	-.025	-.092	-.038	1.000	-1.000	-1.000	-.023
Avg_Grade	-.020	-.037	-.033	-.082	.060	.024	.086	.043	-1.000	1.000	1.000	.029
Grade_ Diff	-.019	-.037	-.026	-.082	.061	.026	.086	.042	-1.000	1.000	1.000	.026
Lane_Width	-.588	.089	.018	-.404	-.057	.294	-.391	.295	-.023	.029	.026	1.000

Table 4.16 is the correlation matrix showing the collinearity between the independent variable. Correlation of significant variables is only studied. As we can see from the table, the correlation of access density with minimum horizontal sight distance is -.183, access density with maximum length of continuous tangent is -.164 and maximum length of continuous tangent with minimum horizontal sight distance is .102, which are considered weak correlations. Similarly, the correlation of minimum radius of curvature with the other three significant variables is also weak, so the process is continued with the four selected variables.

Table 4.17 Goodness of fit: Final Model Revised

	Value	df	Value/df
Deviance	88.577	87	1.018
Scaled Deviance	88.577	87	
Pearson Chi-Square	81.818	87	.940
Scaled Pearson Chi-Square	81.818	87	
Log Likelihood	-119.483		
Akaike's Information Criterion (AIC)	248.967		
Finite Sample Corrected AIC (AICC)	249.664		
Bayesian Information Criterion (BIC)	261.576		
Consistent AIC (CAIC)	266.576		

Table 4.17 shows the various goodness of fit metrics which are used in the model selection process. The value of deviance / degree of freedom falls within the range between 0.8 and 1.2 in case of Poisson Model. Similarly, the Pearson Chi-Square value / degree of freedom is also within the acceptance limit.

Table 4.18 Omnibus Test: Final Model Revised

Likelihood Ratio Chi-Square	df	Sig.
56.970	4	.000

Table 4.18 indicates that the independent variables collectively improve the model over the intercept-only model.

Table 4.19 shows the parameter estimate for the final revised model after omission of insignificant variables. As we can see, all three selected variables are statistically significant within the confidence interval. But the variables have to be checked for collinearity. As shown in the correlation matrix of Table 4.20, the correlation between the variables are within the permissible limit. So, the model is considered final.

Table 4.19 Parameter Estimates: Final Revised Model

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	-.154	.2133	-.572	.264	.519	1	.471	.858	.565	1.303
Access_Density	.061	.0189	.036	.110	14.963	1	.000	1.062	1.037	1.116
Sight_Distance	-.019	.0052	-.029	-.008	12.832	1	.000	.982	.972	.992
Tangent	.010	.0024	.005	.014	16.748	1	.000	1.010	1.005	1.015
Min_Radius	.006	.0021	.002	.010	8.463	1	.004	1.006	1.002	1.010

Table 4.20 Correlations of Parameter Estimates: Final Revised Model

	(Intercept)	Access_Density	Sight_Distance	Tangent	Min_Radius
(Intercept)	1.000	.033	-.417	-.470	-.126
Access_Density	.033	1.000	-.217	-.249	.159
Sight_Distance	-.417	-.217	1.000	-.209	-.167
Tangent	-.470	-.249	-.209	1.000	-.260
Min_Radius	-.126	.159	-.167	-.260	1.000

The final model obtained is:

Total five-year crashes= EXP (-0.154 + 0.061*Access Density - 0.019*Minimum Horizontal Sight Distance + 0.01*Maximum Length of Continuous Tangent - 0.006* Minimum Radius of Curvature)

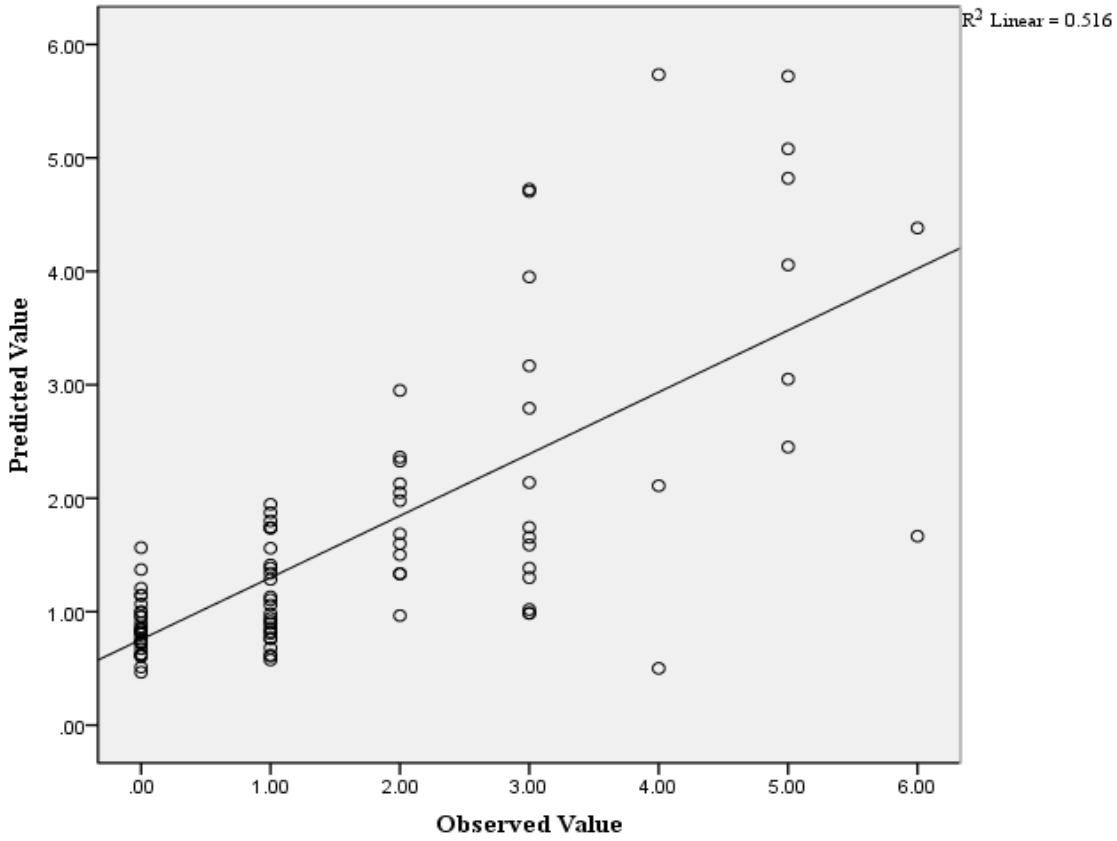

Fig 4.3 Predicted Vs Observed Crashes Plot for Final Model

The observed VS predicted plot of crashes obtained from SPSS Scatterplot option shows that the R^2 of the model is 0.516, which is slightly better than the R^2 value of the initial model.

Identification of Hazardous Locations

The Identification of Hazardous locations and ranking was done by using the formula:

Crash Point Weightage= F*6+S*3+M*0.8+D*0.2

The ranking of hazardous segments based on CPW value can be seen in Table 4.21

Table 4.21 Ranking of hazardous segments based on Crash Point Weightage Value

S.N.	Segment No.	CH FROM (CH1)	CH TO (CH2)	Crash No	Fatal	Severe	Minor	Damage Only	Crash Point Weightage	Rank
1	34	18+161.38	18+762.17	5	3	2	-	-	24.00	**1**
2	2	00+614.22	01+213.96	5	2	2	1	-	18.80	**2**
3	61	33+897.42	34+453.76	3	2	1	-	-	15.00	**3**
4	4	01+854.34	02+457.41	5	1	2	1	1	13.00	**4**
5	10	05+214.03	05+679.93	4	1	2	1	-	12.80	**5**
6	5	02+457.41	02+938.86	3	1	2	-	-	12.00	**6**
7	9	04+686.25	05+214.03	3	1	2	-	-	12.00	**6**
8	11	05+679.93	06+387.86	4	-	4	-	-	12.00	**6**
9	13	06+900.00	07+387.22	3	1	2	-	-	12.00	**6**
10	24	12+909.43	13+481.70	3	1	2	-	-	12.00	**6**
11	65	36+103.64	36+745.14	3	1	2	-	-	12.00	**6**
12	62	34+453.76	35+057.02	2	1	1	-	-	9.00	**7**
13	21	11+274.71	11+918.40	3	-	2	1	-	6.80	**8**
14	22	11+918.40	12+409.00	5	-	2	-	3	6.60	**9**
15	29	15+592.58	16+148.33	3	-	2	-	1	6.20	**10**
16	66	36+745.14	37+285.89	3	-	2	-	1	6.20	**10**
17	1	00+000.00	00+614.22	1	1	-	-	-	6.00	**11**
18	3	01+213.34	01+854.34	1	1	-	-	-	6.00	**11**
19	7	03+534.00	04+114.92	2	-	2	-	-	6.00	**11**
20	23	12+409.00	12+909.43	1	1	-	-	-	6.00	**11**
21	25	13+481.70	14+008.62	1	1	-	-	-	6.00	**11**

S.N.	Segment No.	CH FROM (CH1)	CH TO (CH2)	Crash No	Fatal	Severe	Minor	Damage Only	Crash Point Weightage	Rank
22	37	19+882.36	20+493.89	1	1	-	-	-	6.00	**11**
23	44	23+798.81	24+370.13	2	-	2	-	-	6.00	**11**
24	52	28+330.54	28+874.19	1	1	-	-	-	6.00	**11**
25	60	32+774.21	33+335.07	1	1	-	-	-	6.00	**11**
26	6	02+938.86	03+534.00	3	-	1	1	1	4.00	**12**
27	12	06+387.86	06+900.00	3	-	1	1	1	4.00	**13**
28	47	25+449.90	26+053.96	2	-	1	1	-	3.80	**14**
29	69	38+474.51	39+158.59	1	-	1	1	-	3.80	**14**
30	8	04+114.92	04+686.25	2	-	1	-	1	3.20	**15**
31	33	17+709.94	18+161.38	2	-	1	-	1	3.20	**15**
32	15	07+878.69	08+432.57	1	-	1	-	-	3.00	**16**
33	32	17+146.61	17+709.94	1	-	1	-	-	3.00	**16**
34	56	30+516.22	31+092.64	1	-	1	-	-	3.00	**16**
35	64	35+618.50	36+103.64	1	-	1	-	-	3.00	**16**
36	68	37+815.08	38+474.51	1	-	1	-	-	3.00	**16**
37	70	39+158.59	00+096.43	2	-	1	-	-	3.00	**16**
38	14	07+387.22	07+878.69	1	-	-	1	-	0.80	**17**
39	54	29+340.03	29+900.89	-	-	-	1	-	0.80	**17**
40	63	35+057.02	35+618.50	1	-	-	1	-	0.80	**17**
41	67	37+285.89	37+815.08	1	-	-	1	-	0.80	**17**
42	19	10+223.07	10+732.39	1	-	-	-	1	0.20	**18**
43	58	31+611.39	32+078.01	1	-	-	-	1	0.20	**18**

The crash point weightage values can’t be interpreted independently. It is just a metric to ensure the severity level of the crashes is considered which my result in a more complete understanding of the safety hazard present in a road segment. The significance of the values can be analyzed only by comparing them with the number of crashes and with each other. As suggested by Table 4.21, the segments from 18+161.38 to 18+762.17, 00+614.22 to 01+213.96 and 33+897.42 to 34+453.76 are highest in the ranking with weightage values of 24, 18.8 and 15 for crash numbers of 5,3 and 3 respectively. The fact that multiple segments with lower crash numbers having higher weightage values indicate the necessity for such analyses.

CHAPTER V

RESULTS AND DISCUSSION

1) Poisson Distribution was chosen for the development of the model based on all of the goodness of fit indicating parameters. That would imply that the data is not over dispersed enough for negative binomial distribution to be a better option.

2) The Poisson models were within the level of significance (p=0.05) for both the core-model and the model including the validation data-set based on the omnibus test which compares the fitted model against the intercept only model.

3) Out of the predictor variables, access density, maximum length of continuous tangent and horizontal sight distance fell within the confidence interval becoming statistically significant predictors.

4) The initial model developed using the core data-set was:

   Total five-year crashes= EXP (-0.310 + 0.066*Access Density - 0.01*Minimum Horizontal Sight Distance + 0.01*Maximum Length of Continuous Tangent)

5) The final model obtained from the complete data-set including the data used for validation was:

   Total five-year crashes= EXP (-0.154 + 0.061*Access Density - 0.019*Minimum Horizontal Sight Distance + 0.01*Maximum Length of Continuous Tangent - 0.006* Minimum Radius of Curvature)

6) For every unit increase in access density, the number of crashes in the particular segment increases by 6.1% whereas for every unit decrease in minimum horizontal sight distance, the number of crashes in the particular segment increases by 1.9%

7) Even though the impact of mimimum horizontal sight distance and maximum length of continuous tangent seems minimal based on the value of coefficient, they are still significant variables and removal of their values drastically impacts the predictive capacity of the model.

8) The R^2 values obtained are tabulated in Table 5.1

Table 5.1 R^2 Comparison

	Initial Model	Model Validation	Final Model
R^2	0.509	0.4308	0.516

9) The table indicates that the initial model developed using the core data-set is able to explain 50.9% of the variation in the data, while it only explains 43.08% of the variation in data used for validation. The value is considered moderate. The R^2 value of the final model using the complete data-set including the data used for validation of the earlier model was obtained as 0.516 i.e it was able to explain 51.6% of variation in the data.

10) From the Crash Point Weightage ranking, the segments from 18+161.38 to 18+762.17, 00+614.22 to 01+213.96 and 33+897.42 to 34+453.76 were considered to be the most hazardous locations with Weightage Value of 24, 18 and 15.5 respectively. Observation of explanatory variables in the respective segments indicate that these segments have either relatively higher number of access points or lower minimum horizontal sight distance or both.

CHAPTER VI
CONCLUSION AND RECOMMENDATION

Based on the results obtained, the access density is the most significant variable in crash frequency determination. The problem of unmanaged and haphazard access road opening around the highway has been a growing phenomenon in the last few years which should be controlled with the help of local authorities. Along with proper regulation, hill roads have to be designed predicting the fact that a number of access roads may prop up after the construction which may be difficult to manage after project completion.

As suggested by the model, minimum sight distance and maximum length of tangent within a segment also play a significant role in the occurrence of crashes, which points out the risk of crash occurrences in both straight and curved segments.

Nepal still lacks a proper accident database management system. The exact locations of crash points are very tough to find which makes prediction modelling a difficult task. Government funding has to be increased on providing traffic officials will all the essential equipment and trainings that they require for accurate record-keeping.

Predictive analysis of road crashes in developing countries like Nepal has a good scope and potential given the lack of readily applicable international models that suit indigenous local conditions. More predictive models need to be formulated using other sections of the road network to check the consistency of the relationships and to introduce other combinations of predictor variables which can be beneficial for road-safety decision making.

REFERENCES

1. Zegeer, C.V. and Deen, R.C., 1974, "Identification of Hazardous Locations in city streets".
2. Basu, S. and Saha, P., 2017, "Regression Models of Highway Traffic Crashes: A Review of Recent Research and Future Research Needs", *Procedia Engineering 187*, pp.59 – 66.
3. "Global status report on road safety 2018", Geneva: World Health Organization; 2018. License: CC BYNC-SA 3.0 IGO.
4. "Transport for Development", World Bank, 2015. <URL:http://blogs.worldbank.org/transport/why-vehicle-safety-matters-crash-related-deaths?cid=EXT_WBBlogSocialShare_D_EXT>
5. Caliendo, C1., Guida, M., and Parisi A., 2007, "A crash-prediction model for multilane roads", *Accident Analysis & Prevention* Volume 39, Issue 4, July 2007, pp. 657-670
6. Greibe, Paul. ,2003, "Accident prediction models for urban roads", *Accident Analysis & Prevention*, Volume 35, Issue 2, pp 273-285
7. Abdulhafedh, A., 2016, " Crash Frequency Analysis", *Journal of Transportation Technologies*, 6, pp 169-180.
8. Bonneson, J.A., K. Zimmerman, and K. Fitzpatrick, 2005, "Roadway Safety Design Synthesis", *Report No. FHWA/TX-05/04703*—1, Texas Department of Transportation.
9. Fitzpatrick, K., Lord, D., and Park, B, 2009, "Evaluating Safety Effects of Ramp Density and Horizontal Curve for Freeways Using Texas Data".
10. Zegeer, C., Stewart, R., 1992, "Safety Effects of geometric improvements on horizontal curves", *Transportation Research Board. No. 1356*, Washington D.C, pp. 11-19.
11. Geedipally, S.R., Lord, D., and Dhavala, S.S., 2012, "The Negative-Binomial Lindley Generalized Linear Model: Characteristics and Application Using Crash Data", *Accident Analysis and Prevention*, pp. 258-265.
12. "Highway Safety Manual", 2010, American Association of State Highway and

Transportation Officials, Washington D.C..

13. Koorey, G, 2009, “Road Data Aggregation and Sectioning Considerations for Crash Analysis” *Transportation Research Record*, 2103(1), pp. 61–68.
14. Cafiso, S., D’Agostino, C., Persaud, B., 2018, “Investigating the influence of segmentation in estimating safety performance functions for roadway sections”, *Journal of Traffic and Transportation Engineering.*
15. Green, Eric R., 2018, “Segmentation strategies for road safety analysis”, *P.Hd. Dissertation*, Department of Civil Engineering, University of Kentucky.
16. Souleyrette, R. R, Haas., R. P., and Maze., T.H., 2007, “Validation and implication of segmentation on Empirical Bayes for highway safety studies”.
17. Srinivasan, B., Lyon, Persaud, B., Martell, C., and Baek, J., 2011, “Methods for Identifying High Collision Concentration Locations (HCCL) for Potential Safety Improvements: Phase II, Evaluation of Alternative Methods for Identifying HCCL”.
18. Lu, J., Gan, A., Haleem, K., and Wu, W., 2013, “Clustering-based roadway segment division for the identification of high-crash locations”, *Journal of Transportation Safety & Security*, Vol. 5, No. 3, pp. 224-239.
19. Fayaz, M.M., Mrudula, S.P., George, S.J., Yoyak, S.P., Roy, S.S., 2018, “Black Spot Identification Using Accident Severity Index Method.”, *ISSN:2394-0697*, Volume-5, Issue-3.
20. F. Mustakim and M. Fujita, 2011, “Development of accident predictive model for rural roadway”, pp. 46–51.
21. Zanne, M., Bajec, P., 2018 “External Costs of Traffic Crashes on Slovenian Roads”.