UHASSELT
KNOWLEDGE IN ACTION

School of Transportation Sciences

Master of Transportation Sciences

Master's thesis

Vehicle Control Behavior of Drivers as Proxy Indicator for Crashes

Dessalegn Alemu Bekele

Thesis presented in fulfillment of the requirements for the degree of Master of Transportation Sciences, specialization Traffic Safety

SUPERVISOR :

Prof. dr. Tom BRIJS

MENTOR :

Mevrouw Jana HOREMANS

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

2023
2024

School of Transportation Sciences

Master of Transportation Sciences

Master's thesis

Vehicle Control Behavior of Drivers as Proxy Indicator for Crashes

Dessalegn Alemu Bekele

Thesis presented in fulfillment of the requirements for the degree of Master of Transportation Sciences, specialization Traffic Safety

SUPERVISOR :

Prof. dr. Tom BRIJS

MENTOR :

Mevrouw Jana HOREMANS

ABSTRACT

Over the past decade, various researchers have conducted road safety studies. Traditional approaches in road safety research predominantly focused on analyzing actual traffic crashes, a process that demands considerable time to accumulate sufficient crash data for comprehensive analysis. Recognizing the limitations inherent in traditional methods, a paradigm shift occurred, giving rise to a new approach in traffic safety studies. This new method shifts the focus from infrequent traffic crashes to examining risky events or near crash events. Despite the recognized importance of studying near-crash events, the relationship between such events and actual crashes remains uncertain, and the spatial dynamics between them are undefined.

In this study, experimental research is carried out to investigate the spatial and statistical relationship of historical traffic crashes and vehicle control behaviors for Flanders region from i-DREAMS data set, a project funded by the European Union. Three years of historical traffic crash data (2017-2019) were analyzed.

The spatial analysis identifies concentration of historical traffic crashes around the major cities like Antwerp, Ghent, Bruges, Hasselt, Mechelen and in the motorways. The same risky events like acceleration, deceleration, speeding, steering and tailgating, concentrated on the Limburg, Antwerp and Flemish Brabant provinces. The hot spot analysis revealed significant clusters of traffic locations at a 95% confidence level. The local bivariate relationship and multivariate clustering show a positive relationship between vehicle control events and traffic crash in different places of the study area.

Statistical tests, including the chi-square test of association and Spearman's rank correlation test, showed significant associations and positive moderate correlations between risky events and historical traffic crashes. From Negative Binomial Model acceleration, speed and tailgating events significantly affect the expected count of traffic crashes. The spatial model conducted by spatial lag negative binomial to address the effect of spatial autocorrelation of historical crashes. The spatial lag model shows that the lag count of crash significantly affects the expected traffic crash count in addition to acceleration, speed and tailgating events. These results illustrate that close following and aggressive speeding is common risky driving behavior in the study area. The findings contribute to the body of research on driving behavior and can inform evidence-based interventions and policies to reduce crashes and promote safer road environments.

Key Words: Traffic Crash, Vehicle Control Behavior, Crash prediction, Spatial Analysis, Naturalistic Driving Study

ACKNOWLEDGEMENTS

First of all, I would like to thank Almighty God for his endless mercy, giving me excellence and efficiency in time of difficulty and a moment of weakness. Second in line, it is my great pleasure to forward my deepest gratitude to my research supervisor Prof. dr. Tom Brijs for his genuine guidance, valuable comments and help me achieve my goal in this research. Thirdly I would like to thank my mentor Ms. Jana Horemans for her valuable support, comment and guidance in the course of this research and become a good mentor. I am also indebted to the i-DREAMS project research team for providing indispensable data for this study.

I have a heart full appreciation to my colleagues and fellow students at Hasselt University and also it is my pleasure to express my gratitude to my family for their unreserved support and courage.

The last but not the least, I would like to acknowledge the Flemish Inter-University Council (Vlaamse Interuniversitaire Raad/ VLIR-UOS) for awarding me a scholarship that enabled me to pursue my master's studies at Hasselt University.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 General Background

Road Traffic crash is one of the leading causes of death worldwide. The United Nations Decade of Action for Road Safety targets worldwide road death and injury as a major public health issue with a wide variety of social and economic effects, and it is strongly tied to UN development goals. Every year, over 1.35 million people are killed in road traffic deaths worldwide, and around 26,000 of these occur in the European Union Countries (WHO, 2018).

According to the World Health Organization 2018 report, there is more progress in reducing road traffic death on middle-income and high-income countries than in low-income countries. Europe and other middle-income countries have more progress on reducing road traffic death (WHO, 2018).

Europe is the world's safest region and achieved incredible results in road safety, with 45 road fatalities per one million inhabitants in 2021. This is the lowest fatality rate out of all world regions (ETSC, 2022). Even if, the European Union (EU) had made significant success in reducing road safety problem, till now 70 people daily lost their lives in road accidents, which indicates it requires further efforts to significantly reduce road fatalities and serious injuries (European Commission, 2020).

In the same manner to European Union (EU), Belgium had made progress in reducing road traffic death in the last decades and recorded 43% reduction of traffic death from 2010 to 2020. In the past, Belgium's road traffic death is above to the EU average road traffic deaths. Recently, Belgium reached the average EU road traffic death which is 45 road deaths per million inhabitants. Still, this road traffic death constitutes a significant societal problem (Adminaité-Fodor et al., 2021).

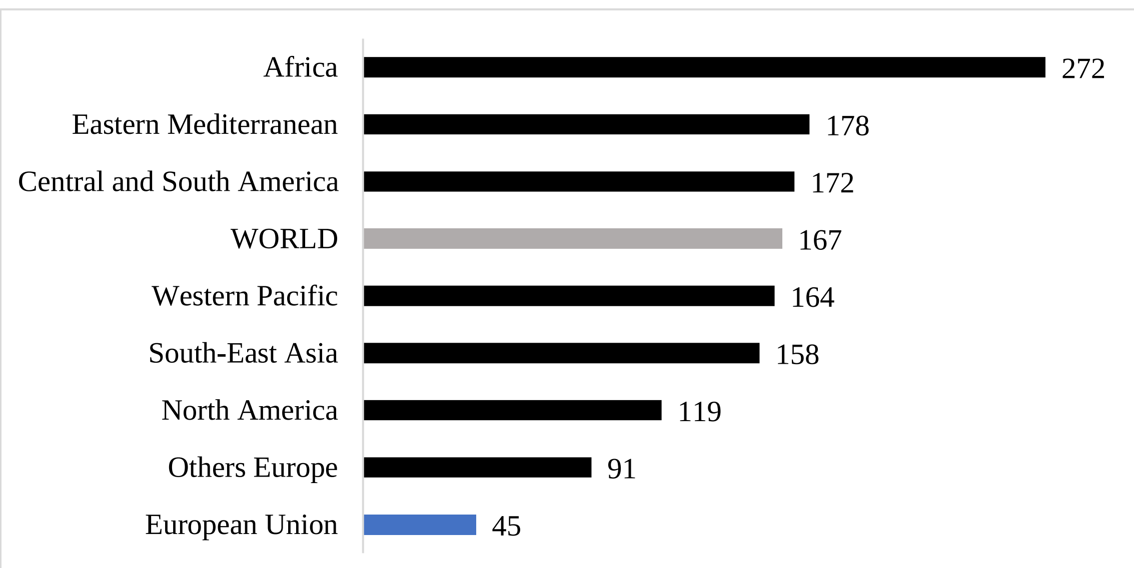

Figure 1.1: Road fatalities per million inhabitants (ETSC, 2022)

To overcome this health problem, World Health Organization (WHO) and United Nations Regional Commissions set a global decade of action plan for road safety with a 50% reduction of road traffic death and injury by 2030. Every Country worldwide set long term and short-term goal inline to the global decade of action plan. The same to this Flemish region of Belgium planned to reduce fatalities by 50% in 2030 and zero fatalities in 2050 (DANIELS, 2022). To achieve these goals and initiatives, all actors need to work their part in a coordinated manner. Innovative methods are required to ease road safety problems. Worldwide researches of road safety indicates that there are many contributing factors for road traffic crashes.

The Haddon matrix, which was published in 1960, is one of the studies that demonstrates the epidemiological idea in traffic safety. It shows the impact of human, vehicle, physical environment, and socioeconomic factors on the pre-crash, crash, and post-crash periods (Haddon, 1968). Each cell of the matrix is used to list the countermeasures to control the damage from such incidents.

Table 1.1: The Haddon matrix (Haddon, 1968)

		Factors			
		Human	Vehicle	Physical Environment	Socio- Economic Environment
Phases	Pre-crash				
	Crash				
	Post-crash				

Human Factors are one of the core factors in road traffic crashes. In all human factors, the driver behavior is an engine for safe movement on the roadway (Zhang et al., 2022). Driving is a complex task which requires execution of multiple tasks at a time. Several factors contribute for risky driving behavior of the driver, such as distraction, drowsiness, inattention, drunk driving, and speeding are the contributing factors for loss of control, close following of another vehicle as well as hard braking behaviors of drivers (NHTSA, 2017) & (VIAS, 2017). In relation to driver behavior study, researchers did enormous road safety research tasks, including new ways of researching road safety problems such as Naturalistic Driving study. Naturalistic Driving Study is one of the approaches to study road safety problems, especially to understand the interaction of the vehicle, driver behavior and the infrastructure. This naturalistic study provides an opportunity for researchers to conduct in depth studies on the association of crash with driver behaviors and vehicle kinematics (Schagen et al., 2012).

Understanding the behavior of drivers at the crash proxy location is essential to provide proactive measures based on the driver's reaction and the context of the road (Dozza, 2020). As cognizant that, these naturalistic studies are expensive and study for long time requires a lot of resources. In naturalistic studies, a lot of researchers mainly consider near-crash events as a crash event. Near crash event is any situation that requires the participant vehicle to make a quick, evasive maneuver in order to avoid a collision. A rapid, evasive maneuver are steering, braking, accelerating, or any combination of control inputs that limit the vehicle capability (Guo et al., 2010). Understanding the relationship between near-crash events and crash events is necessary (Dozza, 2020).

In previous research, enormous crash prediction models were performed using road infrastructure parameters like geometric design. However, consideration of only infrastructure parameter doesn't give effective crash prediction models (Shaon, 2019). Thus, considering driving behavior in crash prediction models is crucial.

This research is based on the data collected from i-DREAMS project and investigates the relationship of risky driving behavior events (vehicle control events) and historical crashes at crash spot locations. The vehicle control events include acceleration, deceleration, steering, speeding and tailgating events. i-DREAMS project is one of the naturalistic driving studies which was conducted on five European countries including Flanders region of Belgium (Talbot et al., 2020).

1.2 Problem Statement

Even if driver behavior is the major crash contributing factor, most road safety studies consider roadway geometric characteristics in developing crash prediction models. The driver behavior related variables are rarely considered in the crash prediction models. Such way of crash modeling results in biased models and inaccurate estimation of the crashes (Shaon, 2019).

In the past, unavailability of driver behavior-related parameters in the crash dataset was challenging for many researchers. Currently, Naturalistic Driving study opens a solution to such challenges. The naturalistic driving study improves road safety research from the perspective of normal driving behavior of drivers. Drivers under naturalistic driving conditions are evaluated based on safety related events such as acceleration, deceleration, tailgating, speeding and steering events. Association of such safety related events to crashes is necessary to predict the crashes on the perspective of driving behaviors (Guo et al., 2010). However, getting sufficient crashes for modeling is challenging in naturalistic study.

Owing to this limited number of crashes, previous researchers consider certain near crash events as crashes. Nevertheless, the relation of crash and near crash is unknown, and the debate to use near crash as crash is still open (Dozza, 2020). To this effect, understanding and validating the crash with safety-related events from the perspective of local and spatial context is essential. Comprehending the relationship between historical traffic crashes and safety-related events is vital to take proactive safety measures. Hence, based on this motivation, the researcher investigated the relationship between historical traffic crashes and vehicle control behavior of the drivers from the i-DREAMS naturalistic driving study data set.

Naturalistic Driving Study was conducted on Flanders region of Belgium by i-DREAMS project with the fund of European Union 2020 research and innovation program which aims to define, develop, test, and validate safety tolerance zone and prevent drivers from unsafe operation by mitigating risks in real time and after the trip.

1.3 Objective of the Research

1.3.1 General Objective

To assess the vehicle control behavior of Flanders drivers on the crash spot locations through Naturalistic Driving study and to predict crash from risky vehicle control behaviors at crash spot locations. To add the combined effort towards proactive traffic safety by understanding the causal relationship of crash and risky vehicle control behaviors.

1.3.2 Specific Objectives

The specific objective of the study are the followings: -

- ✓ To determine the frequency distribution of Vehicle control behaviors on crash spot locations
- ✓ To assess the effect of vehicle control behaviors on crash spot locations

- ✓ To determine the strength correlation between vehicle control behavior and crash
- ✓ To predict crashes based on vehicle control behavior events from naturalistic study
- ✓ To develop risk map of predicted traffic crashes from risky vehicle control behaviors in GIS

1.4 Hypothesis and Research Questions

1.4.1 Hypothesis

By statistical test the null and alternative hypothesis of this study will be tested. The null hypothesis for this study will be crash isn't frequent at risky vehicle control behavior locations and on the other hand the alternative hypothesis will be crash is frequent at risky vehicle control behavior locations.

Ho: crash is not frequent at risky vehicle control behavior locations
H1: crash is frequent at risky vehicle control behavior locations.

1.4.2 Research Questions

- ✓ How it looks the spatial distribution of crash and vehicle control behaviors in Flanders?
- ✓ How is the frequency of risky vehicle control behaviors on crash spots?
- ✓ What is the association of vehicle control behaviors and historical crashes?
- ✓ Which parameter of vehicle control behavior is significant for historical crashes?
- ✓ How can crash be predicted based on the significant vehicle control behaviors?

1.5 Scope and Limitation of the Study

The scope of the study is to assess the acceleration, deceleration, speeding, steering and tailgating behavior of Flanders drivers on the crash spot locations and to determine the relationship between vehicle control behaviors and historical crashes on the same locations. In this study, the effect of weather, time, driver socio-demographic character, traffic condition, geometric condition of the road, vehicle model difference between the crash data and naturalistic driving data are not considered.

1.6 Further Researches that can be made

The researcher can encourage others to work on the stated limitation with a wider study to arrive efficient model.

1.7 Significance of the Study

The study plays a vital role in proactive safety measures through naturalistic study. In naturalistic study, getting crash data is rear and waiting till getting sufficient crashes is not ethical because daily human lives are lost with road traffic accidents. Thus, these naturalistic studies need to forecast the crashes from drivers' unsafe or risky driving behavior. In order to forecast, understanding the relationship between crashes and this risky driving behavior of drivers is mandatory for the researchers, and an applicable crash prediction model from the risky driving behavior is necessary. Hence, this study is crucial for proactive crash prediction from risky driving and also it needs to be investigated everywhere because traffic crash factors vary from place to place.

1.8 Organization of the Study

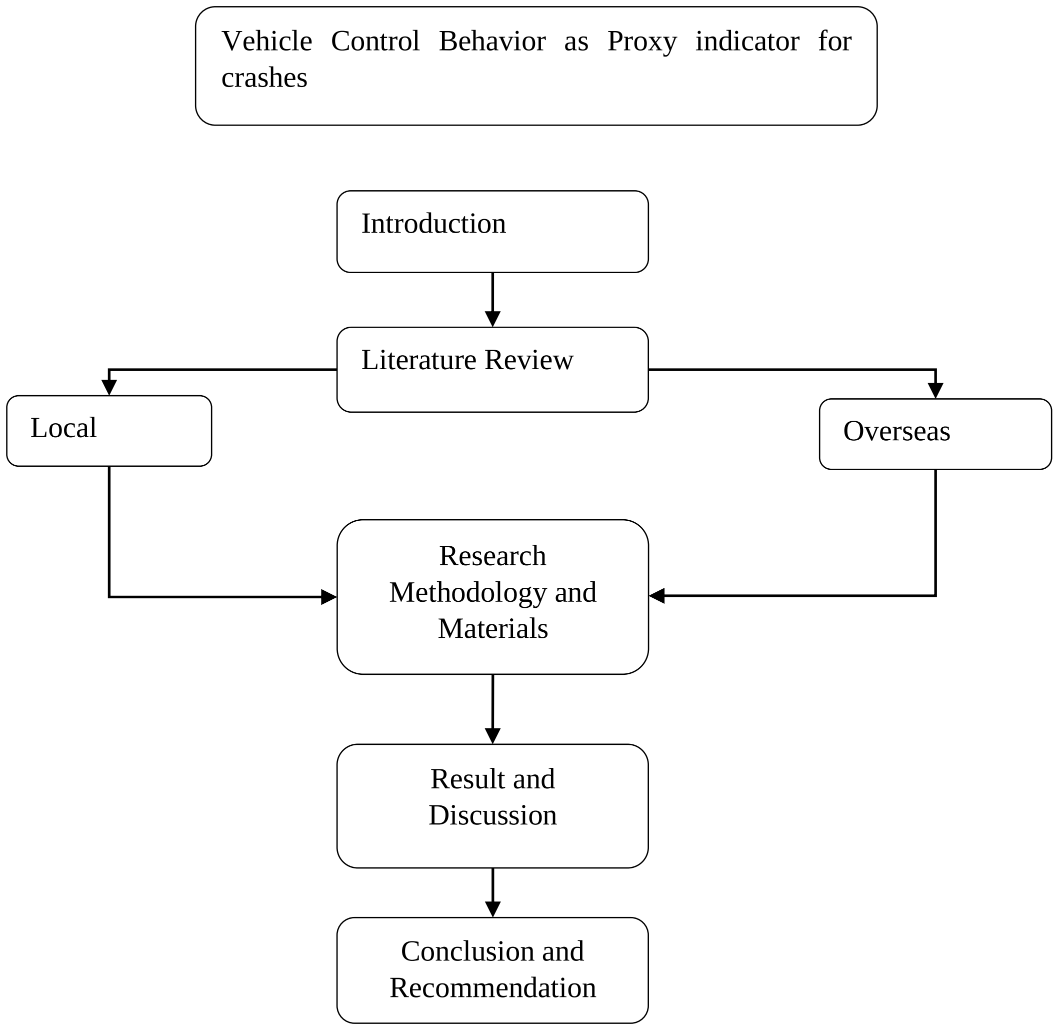

Figure 1.2: Study Structure

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

This chapter provides an overview of the current traffic safety literature on the subject topic and discuss the concepts traffic crash, naturalistic driving study, relationship of vehicle control behavior and traffic crash, spatial analysis of traffic crash, crash frequency and prediction models. It intends to look on several research written in the area of this study. Section 2.2 discussing road safety and their impacts. Section 2.3 Factors in Traffic Crash are discussed. Section 2.4 Road safety research approach including naturalistic driving study are presented. Section 2.5 Relationship of Vehicle Control Behaviors and Traffic Crash are presented. Section 2.6 Spatial Analysis of Traffic crash and Vehicle Control event are discussed. Section 2.7 Traffic Crash prediction models are presented. In the last section of this chapter, Section 2.7, Summary of Literature Review is discussed.

2.2 Road Traffic Safety

Traffic safety has become a major public health issue, and policymakers and researchers are increasingly focusing on strategies to reduce the number of road accidents and fatalities. The rapid growth of urbanization has led to an increase in the number of vehicles on the roads, which in turn has contributed to the rise in traffic accidents and fatalities. A study by the World Health Organization (WHO) estimates that road traffic crashes are the main cause of death for people aged 5-29 years, and that 1.35 million people lost their life each year as a result of road traffic crashes (WHO, Road traffic injuries., 2021).

Developed and middle-income countries had made significant reduction road traffic crash. However, human lives are still lost daily on the roads and will require further efforts to save precious human life. In the European Union (EU) 70 people daily lost their lives in road accidents, which indicates it requires further efforts to significantly reduce road fatalities and serious injuries (European Commission, 2020). In relation to this, Belgium had made significant progress in reduction of road traffic crash and reached the average EU road traffic death which is 45 road deaths per million inhabitants. Still, this road traffic death constitutes a significant societal problem (Adminaité-Fodor et al., 2021).

The same to the above on the Flemish region of Belgium still a lot of progress are made to reduce traffic crash whereas in 2021 the situation show reverse trend compared to the other regions of Belgium. According to (vrt NWS, 2022), The road traffic death on Flanders region increased by 18% in 2021 from the previous year. Whereas, in 2022 the road traffic fatalities show decrease compared to 2021 traffic fatality. However, traffic injury in Flanders region is significantly increased in 2022 compared to 2021 (VIAS institute, 2022). Thus, as the same to the above, this region of Belgium also requires active involvement of every stakeholder including researchers to save the human life.

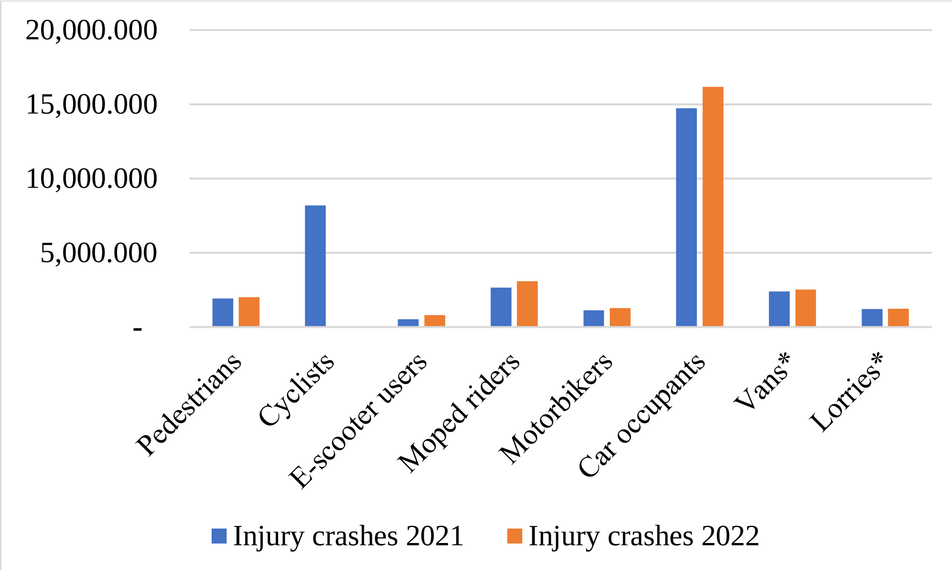

Figure 2.1: Injury crashes in Flanders region 2021 and 2022 (VIAS institute, 2022)

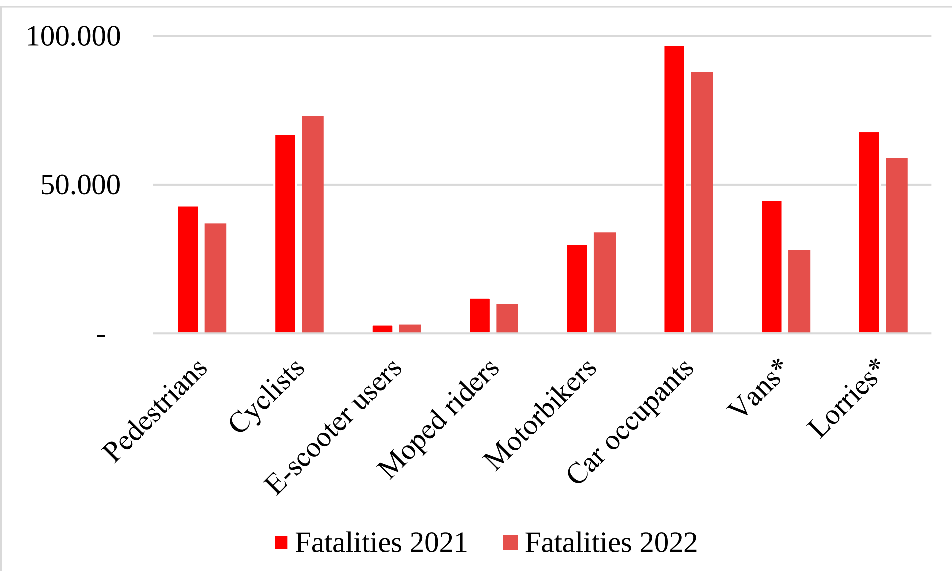

Figure 2.2: Traffic Fatalities in Flanders region on 2021 and 2022 (VIAS institute, 2022)

2.3 Road Traffic Crash Factors

The contributing factors for road traffic crashes are complex. One of the earliest studies on the road traffic crash factors through epidemiological concept is the Haddon matrix (Haddon, 1968). A lot of studies are conducted on the contributing factors of traffic crashes and revealed that the key factors are human factors, vehicle, road conditions, and environmental factors.

Crash occurs as a result of failure interaction between man, technology and organization environment at specific point in time and space. Based on Driver Reliability and Error Analysis Methodology (DREAM), crash model the failure classified into two as sharp end failure and blunt end failure (Ljung, 2002). The sharp end failures (critical events) are the observable consequences that lead to the crash and are expressed in time, space, or energy. The blunt end failures (causes) are the contributing factors for critical events (Talbot et al., 2013).

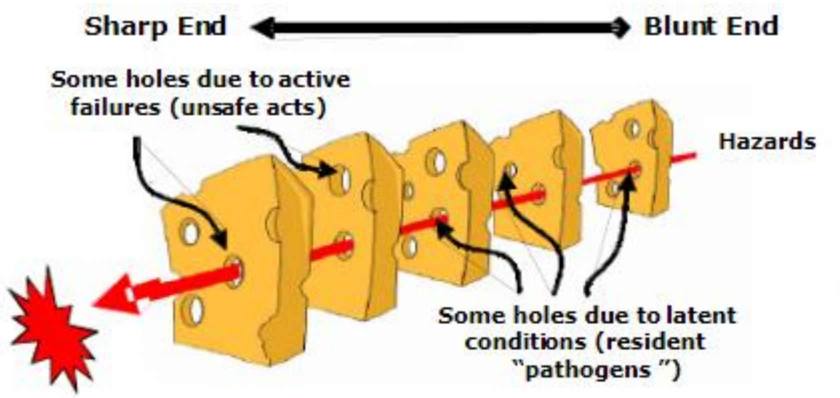

Figure 2.3: Swiss cheese model of accident causation (Reason J, 1997)

Driver Behavior is one of the engine in human factors and is a major contributing factor to traffic crashes. Studies have shown that driver error or driver-related factors, such as speeding, distraction, driving under the influence of alcohol or drugs, and driver fatigue, are contribute for a significant percentage of crashes. A study by the National Highway Traffic Safety Administration (NHTSA) found that significant crashes are caused by driver-related factors. This driver related factors include speeding, steering, and operating the vehicle in erratic, reckless, or neglect manner (NHTSA, 2015).

Vehicle design and safety features have made an enormous contribution to the traffic crash. Studies have shown that vehicle safety features can significantly reduce the risk of injury in traffic crashes (Bhalla & Gleason, 2016).

Road conditions are another important factor that can contribute to traffic crashes. Poor road design or maintenance, such as inadequate signage, poor lighting, or slippery road surfaces, can increase the risk of crashes. A study conducted by the World Health Organization (WHO) found that road traffic injuries are strongly associated with the quality of road infrastructure (World Health Organization, 2018).

Environmental factors, such as weather and visibility conditions, can also play a role in traffic crashes. Becker et al. (2022) conducted study to investigate the relationship between weather conditions and different types of road crashes. The study used data from road crashes that occurred in Queensland, Australia between 2001 and 2010. The researchers used generalized additive models (GAMs) to analyze the data and investigate the relationship between weather variables (such as rainfall, humidity, and temperature) and the incidence of various types of crashes (such as rear-end collisions, head-on collisions, and single-vehicle crashes).

Becker et al. (2022) revealed that weather conditions significantly impact the occurrence of different types of crashes. Such as, rain was found to have the largest impact on crashes, particularly for crashes involving loss of control of the vehicle. The study also found that humidity and temperature were important predictors of different types of crashes, with high humidity associated with an increased risk of rear-end collisions, and high temperature associated with an increased risk of single-vehicle crashes.

2.3.1 Driver Behavior as predominant factor for traffic Crash

Driver behavior is a critical factor in traffic crashes, and understanding driver behavior models is essential to improve road safety. These models need to comprehend the human driver by exploring driver intentions, states, maneuvers, vehicle condition, and environmental factors to identify contributing factors to crashes (Chan et al., 2020).

Research has shown that driver behavior significantly affects vehicle response and kinematics. NHTSA (2017) revealed that around 94% of traffic crashes are related to human behavior. Vehicle responses reflect driver behavior, and vehicle kinematics can be considered measurements of driver behavior. Many active safety technologies use these measurements to assess driving state and provide warnings or interventions to advance driver behavior or alleviate the consequences of deficient driver behavior in case of an imminent risk (Arvin et al., 2021).

Driver behavior includes various aspects, including driver inattention, distractions or impairments, and vehicle control behavior. Understanding how driver behavior modeling evolved and how it describes the dynamics between the driver, vehicle, and environment is crucial for future studies (Arvin et al., 2021).

2.4 Road Safety Research Approaches

The issue of road safety has been a global health problem for many years, and researchers have conducted numerous studies to provide reactive and proactive solutions to this issue. While reactive solutions have their benefits, proactive approaches have gained more attention in recent years. This is because proactive research allows for the prediction of future road safety issues, which can alert road users and organizations to provide countermeasures. As a result, many road safety agencies and organizations are now seeking innovative and proactive research approaches (Ziakopoulos et al., 2020).

The system thinking approach of the current road safety research also allows researchers to look an alternative way from the traditional road safety research (Ashleigh et al., 2016).

One such proactive research approach that has gained popularity in recent years is the Naturalistic Driving Study. This approach involves the observation of naturalistic driving behavior without any obstructions, allowing for a more accurate representation of actual driving behavior. Prior to the emergence of this approach, simulator driving studies were the most common method used to study driving behavior. However, these studies presented challenges in obtaining unobstructed observations and reliable data. Researchers sought a new method that could overcome these challenges, leading to the development of the Naturalistic Driving Study (Ziakopoulos et al., 2020).

The first large scale Naturalistic Driving Study, the 100 Cars Study, investigated driver behavior during risky events, providing valuable insights into the causes of accidents and potential interventions. This study paved the way for further Naturalistic Driving Studies, which have since become a popular and effective tool for studying driving behavior and improving road safety (NEALE et al, 2005).

2.4.1 Naturalistic Driving Study

Naturalistic driving studies collect real-world data to better understand driver behavior and road safety. The Naturalistic Driving Study utilizes sensors installed in the vehicle system to collect data on vehicle control behavior, driver distraction and fatigue, weather and road conditions, and other road users in the vicinity of the vehicle (Ziakopoulos et al., 2020).

Compared to laboratory-based experiments like driving simulator studies, Naturalistic studies provide a more realistic picture of the driver behavior due to the participant drivers not constrained by the artificial environment of the laboratory. Thus, the behavior of driver is not influenced by the observation (Ahmed et al., 2022). These studies provide valuable insights with unobstructed observation of drivers' responses to different road conditions. This information can help researchers develop more effective safety interventions and policies (Ahmed et al., 2022).

Although naturalistic driving studies have strong advantages, they also have some drawbacks, including the requirement for a huge database system, a huge investment, and a long period for study. Additionally, the ethical concerns regarding recording the data, such as video footage and participants' GPS locations, must be carefully considered and addressed (Ahmed et al., 2022).

Despite these challenges, naturalistic driving studies have successfully investigated a wide range of driving behaviors. One of the large-scale project is the Strategic Highway Research Program 2 (SHRP2) Naturalistic Driving Study, which collected data from over 3,000 participants across the United States (Ahmed et al., 2022). Strategic Highway Research Program 2 (SHRP2) Naturalistic Driving study provided valuable insights into the factors contributing to crashes and near-crashes, such as driver distraction and impairment.

The study on the recommendation for large scale European naturalistic driving revealed that Naturalistic driving studies give a better understanding of how and why crashes happen. In relation to analyzing crashes and crash-related events, there are two ways of analysis in naturalistic driving studies. First, the comparison of driver behavior based on the crash events and the baseline events to estimate the relative crash risk. The second is comparison of crashes and near crashes that can be used for validation of performance indicators (Fridulv Sagberg et al., 2011).

According to Fridulv Sagberg et al. (2011), Naturalistic studies are conducted in the following way.

- ✓ Driver-based measures: - measure fatigue and distraction of driver by face and eye recording, measures seat position by recording the top part of the body, measure the hand position and record the speech.
- ✓ Vehicle-based measures: - Vehicle-based measurement indirectly measures driver behavior through vehicle behavior. Speed, acceleration, steering wheel movement, gear position, brake force are mostly recorded measures.
- ✓ Environmental, situational, and infrastructural measures: - In addition to time, location and weather conditions, front, rear and side view of the road condition are measured through video.
- ✓ Different levels of measures: - recordings of the length of trip, purpose of trip and carrying passengers are some additional recordings from driver, vehicle, and environmental based measures.

In conclusion, naturalistic driving studies are a powerful tool for understanding driver behavior and improving road safety. While they present some challenges, the advantages of these studies, such as the realistic representation of driving behavior, outweigh the drawbacks. As technology continues to advance, naturalistic driving studies will play a significant role in improving road safety.

2.5 Relationship of Vehicle Control Behaviors and Traffic Crash

Vehicle control behaviors play a paramount role in one driver's involvement in crashes. Evaluating and assessing the vehicle control behavior through naturalistic data gives an unobstructed observation of driver’s involvement to the crash.

There are different vehicle control behaviors that lead the driver to crash involvement. In this study, the researcher focused on the common vehicle control behaviors: tailgating, speeding, steering, acceleration, and deceleration. A preliminary investigation relationship between historical traffic crash and naturalistic driving event was conducted by (Pande et al., 2017). The researcher sought to investigate how a driver's history of past crashes or traffic incidents relates to their subsequent driving behavior within quarter mile segment length.

Pande et al. (2017) focused on analyzing the driving behaviors of individuals with varying crash histories. It examined various parameters such as speed, acceleration, braking patterns, lane-keeping, and other critical driving factors recorded in everyday driving situations. Researchers aimed to determine if drivers with a history of crashes exhibit distinct driving behaviors compared to those without a crash history.

The findings from this preliminary investigation revealed significant differences in driving behaviors between individuals with a history of crashes and those without such incidents. Drivers with a history of crashes tended to display more aggressive driving behaviors, including higher instances of sudden accelerations or decelerations, increased variability in speed, and potentially riskier lane-changing patterns. This suggests a correlation between past crash involvement and certain risky driving behaviors observed in naturalistic driving scenarios.

Moreover, the study highlighted the potential implications for driver safety interventions and accident prevention strategies. Understanding the link between past crash history and driving behavior in real-world settings can aid in developing targeted training programs or interventions. By addressing and potentially modifying specific risky driving behaviors associated with crash histories, there may be an opportunity to reduce the likelihood of future accidents.

Alrassy et al. (2023) conducted assessment on frequency of risky driving behavior (speed, speed variation, hard braking rate, and hard acceleration rate) with frequency of crash. The study reveals a moderate correlation between risky driving behaviors and collision rate.

Alkaabi (2023) examined driver risky behavior and traffic crash through spatial and statistical analysis. The study revealed through logistics regression close following another vehicle and driving beyond the speed limit shows significant contribution on traffic crash at Abu Dhabi city. The same hot spot analysis of the study indicated that crashes concentrated around the city center.

There has been considerable research conducted by using traffic safety-related events to assess their relationship with traffic crashes. The study conducted by Wu et al. (2014) revealed that a Bayesian multivariate Poisson log-normal model is used to show the association between traffic safety-related events and crash risk.

The association of risky driving events and crash incidents conducted by different research. Cai et al. (2021) delves into the connection between safety-critical events (referred to as SCEs) and the incidence of crashes, injuries, and fatalities among commercial truck drivers. Data was gathered from a significant sample size of 31,828 truck drivers employed by a major commercial trucking company, covering a vast distance of approximately 2.3 billion miles traveled. Cai et al. (2021) findings revealed that an increase in SCEs was directly associated with an 8.4% increase in the rate of crashes per mile, and an 8.7% rise in the number of injuries occurring per mile. The research offers robust statistical evidence supporting the direct link between SCEs and the heightened occurrence of crashes and injuries among commercial truck drivers.

2.5.1 Tailgating

Tailgating is driving dangerously close to another vehicle without insufficient car following distance (Xu et al., 2021). It is a common aggressive driving and one of the causes for vehicle accident (Galovski, 2006). The study conducted by Joint (1997) revealed that about 62% of investigated drivers experience aggressive tailgating behavior, which results in traffic incidents.

Tailgating has been identified as one of the leading cause of rear end crash (Lee et al., 2002). Accordingly, National Highway Traffic Safety Administration (2015) declared that rear end crashes constitute one-third of the total crashes in 2015, which resulted 2,203 fatalities and about 556,000 injuries. The factors associated with rear-end crashes are inattention and tailgating.

The severity of crash is more dangerous when tailgating in highway driving compared to non-highway driving due to high speed and close following another vehicle makes the crash to become severe (Carter, C. J, May, A, Smith,F, & Fairclough,S, 1995).

By examining leading vehicle reaction using naturalistic driving data, highway tailgating behavior was investigated by (Xu et al., 2021). The study reveals that the leading vehicle experiences four types of behavior when there is close following another vehicle: changing lanes, slowing down, speeding up, and making no response. Through the random forest classification model, the study chosen changing lanes is the most prevalent reaction from leading vehicle during tailgating.

Distraction, speeding, and tailgating are the three leading causes of crash on the survey and simulation study on Rhode Island highways (Wang, 2011). To countermeasure tailgating, the presence of advisory signs showed promised result in modifying tailgating behavior and reducing rear end crash.

Tailgating behavior of drivers has been investigated by many researchers, and one of the earliest study conducted treatment of tailgating behavior to reduce the common traffic crash in Thailand by applying safe following distance based on car following theory and the study developed "dot" tailgating marking as a safety countermeasure for tailgating for highways of Thailand (Lertworawanich, 2006).

2.5.2 Acceleration and Deceleration

A study by Af Wåhlberg (2004) measured bus driver acceleration consistency and correlation with traffic crash in Sweden. The study examined a significant correlation between specific types of acceleration behavior and the likelihood of accidents. Drivers who exhibited more erratic or aggressive acceleration patterns were found to have a higher incidence of accidents compared to those who displayed smoother and more consistent acceleration styles.

Acceleration and Deceleration are commonly utilized to assess the driving performance and the frequent change in acceleration and deceleration lead to sever crashes. Comparisons were made between jerk occurrences and historical crash data, aiming to establish a correlation suggesting that road segments requiring frequent braking might be associated with long-term unsafe conditions, potentially leading to more crashes (Bagdadi, 2013).

Pande et al. (2017) utilizes GPS data from naturalistic driving to test the hypothesis that segments where drivers frequently apply hard braking might indicate crash-prone areas over time. It uses linear referencing in ArcMap to integrate GPS information with road characteristics along US Highway 101 in San Luis Obispo, California. The research aimed to link high-magnitude deceleration jerks observed in driving data with crash frequency. Negative binomial regression analysis results showed a significant relationship between high-magnitude deceleration jerks and long-term crash frequency in these segments. The study found that roadway curvature and auxiliary lane presence did not significantly affect crash frequency in the highway sections under consideration. The findings are promising as they suggest that explanatory variables derived from GPS data, accessible via off-the-shelf devices including smartphones, may be valuable for proactive safety assessments.

2.5.3 Speed

Speed-related crashes are common worldwide, and the study conducted by the European Union declared that 10-15% of all crashes and 30% of fatalities are in case of inappropriate speed (Van den Berghe, 2020). Speeding reduces the drivers ability to control the vehicle by increasing the stopping (Wagner et al., 2020). The research published by Elvik (2008) suggested that one-third to one-fourth of fatality related to speeding.

Driver instability is measured through the violation of the speed limit and the risk of crash becomes sever when the driver highly violates the speed limit of the road (Arvin et al., 2019). It is cognizant that the road geometry and topography significantly affect the speed variability because road design speeds depend on the terrain of the road. To avoid frequent speed variability, the road designs consider speed consistency on the road stretches by limiting the minimum distance between the two consecutive curves.

Charly & Mathew (2023) conducted a study to identify driving performance measures on freeway road segment through field measurement. The study revealed that higher speed variability on curved road segments makes higher crash frequency. Speeding, jerk events, and mean gradients are highly significant in estimating traffic crashes. Speeding and jerk events are identified as driving performance for proactive safety measurements.

The speed management of the driver is also associated with the speed limit of the road, traffic calming measures, road geometric conditions and the driver’s behavior. Some reckless drivers commute beyond the speed limit of the road. The higher speed the likelihood of crash will increase. Different researches indicated that speed management highly correlated with the driver characteristics such as age group and gender (Rahman et al., 2016) and (Kunnawee et al., 2012).

The degree of instability in a driver's behavior is gauged by assessing their speed and deceleration volatilities through certain parameters that capture the variations in their instantaneous driving patterns. The driving volatility of both speed and acceleration is measured using several indicators like standard deviation, time-varying stochastic volatility, coefficient of variation, and quartile coefficient of variation. Research suggests that volatility is a key factor that heightens the likelihood of serious accidents. Maintaining control over speed is often seen as a crucial step in preventing vehicular collisions, and losing focus on this can lead to distracted driving (Abd Rahman et al., 2020).

2.6 Spatial Analysis of risky events and traffic crash

The field of road safety used geographic information system (GIS) as a key tool for advancing research and to improve interventions. Understanding the spatial patterns of traffic crashes and risky events helps the traffic safety specialist to identify sections with high frequency of crashes and risky events. Exploring the risky locations through spatial analysis gives a better illustration for road safety management to tackle the undelaying factors in the risky location (Mohaymany et al., 2013).

2.6.1 Spatial units of analysis

Road safety studies considered different spatial units of analysis from road segment to zonal and regional level to investigate risky event counts, crash counts and crash rates (Chen et al., 2019), (Cheng et al., 2019), (Daniels et al., 2019). Some studies considered fixed distance and multiple grid spatial analysis for traffic crashes and risky events. The choice of spatial unit with different unit gives different insight of the study area (J. et al., 2010). Thus, it is recommended to consider the spatial unit that gives optimum insight into the safety of the study area.

2.6.2 Hot Spot Analysis

The hot spot analysis identifies the areas with the highest and lowest concentration of traffic crashes and risky events. Hot spot analysis is essential to seek the significant clusters high and low concentration (Alam & Tabassum, 2023). Arc GIS hot spot analysis tool in conjunction with Getis-Ord Gi* statistics identify significant high and low concentration of traffic crashes and risky events in the study area.

The Getis-Ord Gi* statistics is commonly used to identify significant high and low clusters in the data set. Getis-Ord Gi* statistics calculate the Z-score of each feature in the data set and this Z-score value identifies the feature surrounded by high or low values of the data set. A high positive Z-score indicates a higher value cluster (hot spot) and a low negative Z-score indicates low value cluster (cold spot). The results of hot spot and cold spot visualized in the map with showing the areas of statistically significant cluster (Alam & Tabassum, 2023).

2.6.3 Spatial Autocorrelation Analysis

The hot spot analysis by Getis-Ord Gi* statistics used to detect the hot spot and cold spot areas for traffic crashes and risky events, whereas it's necessary to evaluate the occurrence of crash at random or not. The spatial autocorrelation refers to the observation in the data set based on their spatial proximity.

The spatial connection between crashes measured in different ways such as inverse distance, inverse squared distance, fixed distance, K-nearest neighbors, the zone of indifference, the spacetime window technique, and contiguity edges and corners (Erdogan et al., 2015).

Alam & Tabassum (2023) used fixed distance metrics to measure the traffic incidents in space and time. The study used incremental spatial autocorrelation and evaluated the Global Moran's I test. The result of the study indicates that there is a positive Z-value in Global Moran's I test which indicates the accident data exhibits some sort of clustering and not random distribution.

Flahaut et al. (2003) used spatial autocorrelation and kernel estimation methods to identify black zones in Belgian roads. Similar study by Moons et al. (2009) used hot spot and hot zone analysis to identify high frequent crash locations in Limburg province and around Hasselt City. The study used a Moran Index to identify hot spot areas and succeeded in identifying dangerous locations requiring urgent intervention. In addition, the study investigated that on the top of hot spot area, hot spot zones or neighborhoods need to be identified as hot zone to address dangerous locations effectively.

2.7 Crash Frequency Modeling

The quantitative relationship of number of crash and contributing factor is used for modeling the crash frequency. In the past, several crash frequency models were conducted. The studies revealed that crash data have a variety of issues, including over-dispersion, under-dispersion, time and spatial variation, under-reporting, and rash-type correlation (Lord & Mannering, 2010).

Crash prediction is necessary especially to take corrective measures on the factors of crashes through engineering, coping driver behavior, vehicle assistive technologies, and Intelligent Transportation Systems. The development of safety performance functions like crash prediction models have been the subject of many researches in past decades (Miaou, 1994), (Lord et al., 2005), (Miaou & Lord, 2003) and (Khattak et al., 2021).

Currently, surrogate indicators have become the proactive approaches for the prediction of crashes. These surrogate methods are first proposed in medical science to evaluate the treatment beforehand and nowadays used by traffic engineers and researchers (Meng & Qu, 2012). As indicated by Tarko et al. (2009), surrogate measures should satisfy two conditions which are a surrogate event need to based observable non crash event and development of a method of converting surrogate outcome into meaningful outcome like to crash frequency and severity.

Various surrogate indicators are proposed like to Time to Collision (TTC), Deceleration Rate to Avoid Crash (DRAC), Crash Potential Index (CPI) and Proportion of Stopping Distance (PSD) are applied in safety evaluation (Tarko et al., 2009).

The Safety Tolerance Zone, a concept proposed by i-DREAM platform, is based on surrogate safety indicators. The Safety Tolerance Zone (STZ) is triggered when a certain value of the surrogate safety indicator exceeds the threshold value. The zone at which the demand of the driver task balanced with the driver capacity. According to i-DREAM platforms Safety Tolerance Zone has three phase which are normal driving phase (coping capacity balanced with the driving task complexity), danger phase (driving task complexity increases, and driver operates at risk of crash) and avoidable crash phase (driving task complexity beyond the capacity of the driver but there is time to avoid crashes) (Talbot et al., 2020).

A stakeholder survey was conducted to assess safety breaches and contributing issues on the i-DREAM project to aid system development. The survey was conducted on passenger car, bus/coach and truck modes of transport. The result revealed that the most important safety breach for passenger cars is loss of control, and the contributing factor is excessive speeding. The most safety breaches for bus/coach are loss of control and sudden brake and the contributing factors are fatigue/sleepiness and inattention/distraction respectively. However, the most important safety breach for trucks is close following another vehicle and the contributing factors are stress, time pressure and inattention or distraction. In general, loss of control, sudden brake and close following another vehicle are the most safety breaches identified in the survey and inattention, or distraction becomes the contributing factor (Talbot et al., 2020). Loss of control, abrupt braking, and close following are behaviors allied to vehicle control and expressed in terms of vehicle kinematics, such as acceleration, deceleration, tailgating, speeding, and driver steering actions.

Predicting crashes from safety breach behaviors is necessary to take proactive measures. In order to predict the cashes from such safety breach behaviors, it requires crash data. Most of the naturalistic studies consider near crash events as the crash in the lack of crash data. Thus, assess the relationship of near crashes and crash is necessary (Dozza, 2020).

Table 2.1: Observed Crash frequency modeling from risky events

Study	Statistical Method	Effect of risky event
(Cai et al., 2021)	Bayesian negative binomial regression	Increment of crash with the headway, hard braking, rolling stability and collision avoidance events
(Gitelman et al., 2018)	Negative binomial regression	Positive relationship between braking event and traffic crash while speed alert shows reverse relationship with crash
(Pande et al., 2017)	Negative binomial regression	High magnitude of jerks or deceleration events show positive relationship with traffic crash
(Wu et al., 2014)	Poisson regression	Positive and substantial relationship between aggressive driving and traffic crash
(Guo & Fang, 2013)	Negative binomial regression	Critical incident rates are an effective predictor for traffic crashes
(Guo, Klauer, et al., 2010)	Poisson regression	Near crash events shows significant positive relation with crash
(Gordon et al., 2011)	Poisson regression	Lane departure and time to edge crossings show significant positive relationship with traffic crash whereas lateral deviation doesn’t show significant relationship with traffic crash

2.7.1 Poisson Regression

Poisson regression is the most commonly used regression for count data, and many researchers have applied Poisson regression in crash frequency models. The assumption of Poisson regression is mean equal to variance which is not common on crash data. Hence, the limitation of consideration of overdispersion and under dispersion in the crash data researchers focused on other count models rather than Poisson regression (Jovanis & Chang, 1986), (Joshua & Garber, 1990) and (Jones et al., 1991).

2.7.2 Negative Binomial Regression

The Negative Binomial Regression is the most widely used method for crash prediction due to its consideration of over-dispersion data where the variance is greater than the mean (Lord & Bonneson, 2007), (Lord & Park, 2008) and (Lord et al., 2005). The main purpose of crash prediction models is to assess the relationship between crash frequency and the explanatory variables. The variables in the model and the statistically significant coefficients determined the maximum likelihood of predicted crash counts.

The Negative Binomial regression uses a linear combination of explanatory variables to explain over-dispersed count data with a log link function to ensure the predicted values are non-negative. The formula of Negative Binomial regression is:

$\log(E(Y|x)) = \beta 0 + \beta 1x1 + \beta 2x2 + ... + \beta kxk$ …………………………Eqn-2.1

Where,

- $E(Y|x)$ is the expected value of the count data Y given the values of explanatory variables x1, x2, ..., xk

- β0, β1, β2, ..., βk are the coefficients estimated for each explanatory variable

- x1, x2, ..., xk are the values of the explanatory variables for a particular observation.

Numerous Studies witnessed Negative Binomial Model to explore and analyze traffic safety issues.

The Negative Binomial model uses the gamma probability assumption to relax the mean equal to the variance. Through addition of an error term on Poisson model the negative binomial model estimates the expected number of crashes (Lord & Mannering, 2010).

$$\lambda i = EXP(\beta Xi + \varepsilon)$$ ………………………………. Eqn-2.2

Where EXP(ε) is the gamma distributed error, adding this error term allows the variance to differ from the mean.

$$VAR(ni) = E(ni)(1 + \alpha E(ni))$$ Eqn-2.3

When α equals zero the model become Poisson model and if α is significantly different from zero then negative binomial model will be used (Lord & Mannering, 2010).

On top of the better performance of Negative Binomial Model on over dispersed data, it has also limitation in handling under dispersion of data when the mean of the data is larger than the variance (Lord, 2006).

2.7.3 Goodness of Fit Measures

The goodness of fit evaluates how the model is fit to the data. Goodness of fit evaluates the discrepancies of observed and predicted values through different metrics. In Negative binomial model, the goodness of fit measured through Log likelihood (LL), Akaike's information criterion (AIC) and deviance statistics (Hilbe, 2011). The likelihood measures the occurrence probability of actual observed value under the given parameter estimates. The higher likelihood value, the better model. On the other hand, Akaike's information criterion (AIC) measures the trade-off between bias and variance. Akaike's information criterion (AIC) is computed based on the following equation.

$AIC = -2 \times LL + 2 \times (Number of Parameters)$………………………………Eqn-2.4

A lower value of Akaike's information criterion (AIC) is the better model because it discourages the overfitting of the data by penalizing adding parameters (Hilbe, 2011).

2.8 Summary of Literature Review

Based on the literature reviewed, the relationship between vehicle control behavior and historical crash is variable based on the local context and different factors. Past studies revealed that investigating the risky driver behavior with historical traffic crashes significantly contributes to proactive safety intervention and future autonomous transportation. Only limited studies are conducted to assess the relationship between vehicle control behavior from naturalistic driving study and historical traffic crash. The studies show that different vehicle control behaviors significantly predict traffic crashes and require local context analysis.

Integrating spatial analysis with quantitative analysis boosts the road safety research practicability to save precious human life. Hence, consideration of spatial and quantitative analysis to the local context is on demand.

The significant factors affecting traffic crash are variable based on the driver behavior and studies revealed that there is different argument on the effect of the factors due to the variability of drivers.

Owing to the above gaps and findings, study spatial and statistical relationship between vehicle control behavior from naturalistic driving study data set and historical traffic crash for Flanders region of Belgium is required.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Description of the Study Area

The selected location of the study area is Flanders region of Belgium on which the Naturalistic Driving Study was conducted. Flanders region is the northern part of Belgium with a land area coverage of 13,251 Km^2 and accounts approximately 6.35 million inhabitants. The population density of the region is four times higher than the average population density of the European Union. The Flemish region of Belgium is highly urbanized, with five provinces and 308 municipalities (POLIRURAL, 2022).

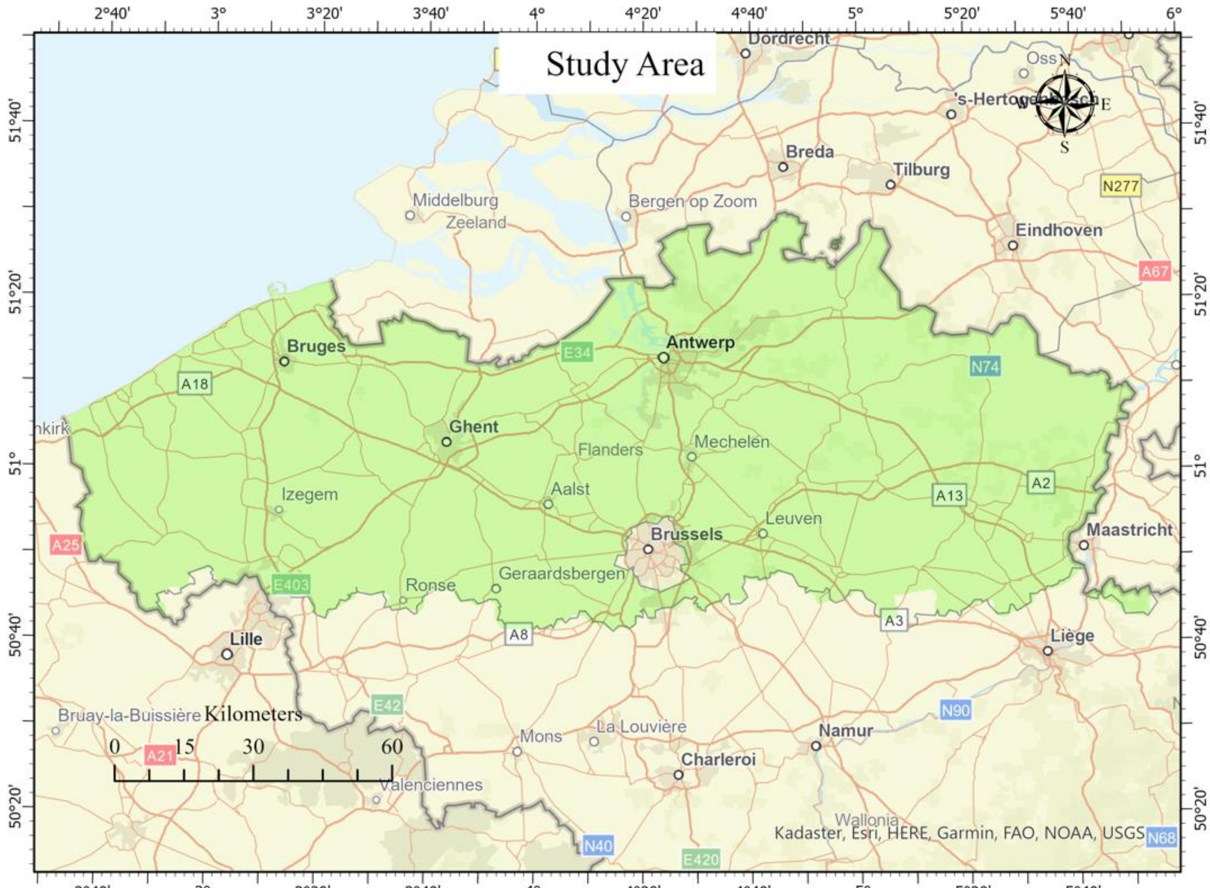

Figure 3.1: Map of study area

3.2 Study Area Selection Criteria

The study area of the research is selected based on the availability of Naturalistic driving data as well as publicized by (vrt NWS, 2022), the number of road deaths in Flanders are increased in 2021 compared to the other regions of Belgium. Hence, to save the precious human life everyone needs to be involved and researchers need to investigate the problems behind it and provide tangible mitigation measures.

3.3 Vehicle Control Behavior

The vehicle control behavior of the driver represented in naturalistic driving study through Acceleration, Deceleration, Tailgating, Speeding and Steering events. These events have three levels which are High, Medium, and Low levels. High levels of events are recorded on the naturalistic data when the driver performs sudden action due to distraction or sudden incidents.

Table 3.1: Vehicle Control Behaviors and Level of Events

No.	Vehicle Control Behavior Events	Level of Events		
		High	Medium	Low
1.	Deceleration	Hard Braking	Medium Braking	Low Braking
2.	Acceleration	High Acceleration	Medium Acceleration	Low acceleration
3.	Speed Management	High Speed Management	Medium Speed Management	Low Speed Management
3.	Steering	High Steering	Medium Steering	Low Steering
4.	Tailgating	High tailgating	Medium tailgating	Low tailgating

3.4 Historical Crash

Three years of historical Crash data from 2017 to 2019 was collected for Flanders region and the data was tagged on Geographic Information System and the corresponding vehicle control behavior of the driver also tagged on the GIS. The spatial distribution of crash data and vehicle control behaviors are assessed through GIS. Spatial join of naturalistic driving event and Historical crash was conducted through a distance cluster. Clustering of naturalistic driving events and historical traffic crashes was performed within an appropriate distance.

3.5 Research Data

As indicated in different studies, traffic crashes are affected by different factors. However, this study mainly focuses on the vehicle control behaviors effect on the historic traffic crash. In this study, ten factors are considered which are: acceleration high, acceleration medium, deceleration high, deceleration medium, speeding high, speeding medium, steering high, steering medium, tailgating high and tailgating medium events. These parameters are classified as acceleration, deceleration, speeding, steering and tailgating events.

3.6 Sampling

To come up with efficient analysis and model, determination of the sample size is a preliminary task of the researcher. For this study the researcher aims to sample the data from the selected area and for 95% level of confidence, Margin of Error E= 0.05 and standard deviation σ = 0.5 the minimum number of historical crashes required for the study will be:

$$n = [\frac{Zc\sigma}{E}] =(\frac{1.96*0.5}{0.05}) = 385 \text{ Historical Crash data} \ldots\ldots\ldots\ldots\ldots\ldots\ldots.. \text{Eqn-3.1}$$

According to VanVoorhis & Morgan (2007), the general rule of thumb for determining sample size for statistical analysis is 50+8*n where n is number of independent variables but the minimum number of samples should be 50.

For this study n=10, N = 50 + 8*10 = 130

A total of 43,448 historical crash data were collected for this study.

3.7 Sampling Eligibility Criteria

In this study, samples outside the spatial join of naturalistic driving event and traffic crash are rejected for the analysis. The study limited on the Flanders region of Belgium.

3.8 Data Collection

3.8.1 Secondary Data Collection

In this study, the naturalistic data's are taken from the data collected from i-DREAMS project which is European Union (EU) funded project to create safe travel zones. Historical traffic cash data was also collected from Flemish traffic agency and three years historical traffic crash data was collected for analysis.

3.8.2 Data Extraction and Mining

The major tasks in executing this study were data extraction and mining. Naturalistic driving events were collected from i-DREAMS project. The data include high, medium and low severity level of events. In this data, based on the scope of the study, extraction of high and medium severity events of acceleration, deceleration, speeding, steering and tailgating was performed for Flanders region of Belgium through QGIS and ARC GIS latest software.

This study extracted a total of 43,448 historical traffic crashes for the study area from 2017 to 2019. The total number of risky events extracted for the study area is shown below.

Table 3.2: Extracted number of events

Event Type	Severity	Number of Events	Total
Acceleration	High	10846	47061
	Medium	36215	
Deceleration	High	497	7137
	Medium	6640	
Speed	High	19107	26945
	Medium	7838	
Steer	High	4921	48255
	Medium	43334	
Tailgating	High	15554	77897
	Medium	62343	

3.9 Data Analysis

The manipulated data was used to analyze the relationship of historical traffic crash and naturalistic driving events. The data analysis shows the spatial relationship of traffic crashes and naturalistic driving events. Analyzing and predict traffic crash from significant risky events. The primary focus of the analysis was to determine the significant risky events in the Flanders region for predicting traffic crashes and develop a risk map based on the predicted traffic crashes. Spatial analysis and statistical analysis are used as basic tools in the analysis of the data.

- ✓ Historical crashes and vehicle control events are tagged in the Geographic Information System
- ✓ Spatial clustering of Traffic crash and Naturalistic driving event was conducted with 500m strip of road. 500 m distance for clustering was chosen through many trials to get adequate number of traffic crashes and risky events for the analysis. 200m and 300m clustering distance was tested and a limited number of crash and risky event obtained. Owing to the limited number of crashes on short distance clusters, the researcher diverted to increase the distance to include more crashes and to perform best on the overall data analysis.
- ✓ Strips with polygon format developed on GIS for study area
- ✓ Spatial join of traffic crash and vehicle control behaviors on the created polygon strips performed.
- ✓ Traffic crashes and vehicle control behaviors were counted on each segment through QGIS point count algorithm.
- ✓ After preparation of the data for analysis, different spatial and statistical analyses were performed.

3.9.1 Spatial Analysis

Spatial analysis plays a critical role in road safety research in achieving identification of risky zones and hot spot sections of the roads. After the introduction of spatial autocorrelation, spatial analysis application is promising for modeling of the relationship of variables in space through clustering. Spatial autocorrelation creates an association between different variables in space, and this method of analysis creates a more scalable outcome for road safety research.

Identification of hot spots, crash locations, high crash concentration locations is the common task for road safety organizations and research. Recent mapping software's advanced road safety research through spatial analysis tools. Geographic Information System (GIS) was first utilized in 1976 by Moellering for study of geospatial patterns of road accidents. GIS has been utilized for the past half-century to improve road safety (Mohaymany et al., 2013).

Recently a lot of road safety researches are conducted through spatial analysis. The spatial analysis by Rhee et al. (2016) and Hadayeghi et al. (2010) include Geographically Weighted Regression (GWR) in spatial analysis. The spatial analysis by Gomes et al. (2017) evaluates the performance of Geographically Weighted Regression (GWR) in a General Linear Model (GLM) context and revealed that for accounting the over-dispersion of crash data, it is necessary to consider Geographically Weighted Negative Binomial Regression (GWNBR) for spatially analyze traffic crash. However, (Xu & Huang, 2015) extended the Geographically Weighted Regression (GWR) to semiparametric by combine geographically varying parameters with geographically constant parameters.

Most commonly used Bayesian framework models also consider spatial autocorrelation in the spatial analysis of traffic crash (Wen et al., 2019). These models can also help to identify the spatial distribution of risk factors associated with traffic crashes and support the development of targeted interventions to improve road safety. Cai et al. (2019) explored the applicability of Random Forest models for ranking hotspots in Traffic Analysis Zones (TAZ) and investigated critical parameters for hot spots through Random Forest model using big data. Despite the limitation of interpretability, Machine Learning (ML) models have undeniable predictive power compared to conventional regression tools.

3.9.2 Hot Spot Analysis

Hot spot analysis of traffic crash and risky events was conducted to identify the significant hot and cold spots of the study area, which is essential in the analysis of traffic crash and risky events. The spatial distribution of traffic crashes and risky events observed through the latest ARC GIS software gives a clear insight into understanding the traffic crash and risky events concentrated locations. After conducting the spatial distribution, the hot spots of traffic crash are identified through Getis-Ord Gi* statistic. The Getis-Ord Gi* statistic, commonly called Gi * statistic, uses the spatial distance to determine clusters with high value of crash and clusters with low value of crash. The statistics calculate the Z score of the feature through the distance cluster and the Z-score high positive value is hot spot area whereas the Z-score with low negative value is cold spot. In addition, the statistics evaluate the significance of this hot spot and cold spot.

Gi * statistic is computed as follows (Soltani & Askari, 2017):

$$\text{Gi } *= \frac{\sum_{j=1}^{n} w_{i,j}X_j - \bar{X}\sum_{j=1}^{n} w_{i,j}}{S\sqrt{\frac{\left[n\sum_{j=1}^{n} w^2{}_{i,j} - \left(\sum_{j=1}^{n} w_{i,j}\right)^2\right]}{n-1}}} \quad \text{.....................................Eqn-3.2}$$

Where:

- $Gi*$ is the Getis-Ord Gi* statistic for location *i*
- *n* is the total number of locations.
- *xj* is the attribute value at location *j*.
- x^{-} is the mean of the attribute values.
- *wij* is the spatial weight between locations *i* and *j*

$$\bar{X} = \frac{\sum_{j=1}^{n} X_j}{n}$$

$$\text{S}= \sqrt{\frac{\sum_{j=1}^{n} Xj^2}{n} - (\bar{X})^2}$$

3.9.3 Spatial Autocorrelation

Further Hot spot analysis, spatial autocorrelation was conducted to test the occurrence of clusters in traffic crash at random effect or there is a certain underlying factor for the clustered areas to have hot spot and cold spot.

The spatial autocorrelation measures whether the features are random, clustered or dispersed based on the feature location and feature values. Moran's I developed to measure the spatial autocorrelation by Patrick Alfred Pierce Moran (Moran, 1950). Moran's I index value, P value, and Z score are used to evaluate spatial autocorrelation. Moran's I index statistics is computed as follows (Mitra, 2009).

$$\text{I} = \frac{N}{W}\frac{\sum_{i=1}^{N}\sum_{j=1}^{N} wij(xi-\bar{x})(xj-\bar{x})}{\sum_{i=1}^{N}(xi-\bar{x})2} \quad \text{........................Eqn-3.3}$$

Where: -

- N is the number of spatial units indexed by i and j
- x is the variable of the analysis
- $\bar{x}$ is mean of the variable interest
- wij is the spatial weight
- W is the sum of all wij

The spatial autocorrelation evaluates the null and alternative hypotheses through Moran's I index value, Z score and P-value. The hypothesis for spatial autocorrelation are the following;

Null Hypothesis (H_0): No spatial autocorrelation; the distribution of values is random across locations.

Alternative Hypothesis (H_1): Spatial autocorrelation exists; the distribution of values is not random and exhibits a spatial pattern.

The Moran I statistic ranges from -1 (perfect dispersion) to +1 (perfect clustering), with 0 indicated no spatial autocorrelation. The p-value evaluate the significance of Moran I statistics, if the p-value less than the chosen significance level ($P<0.05$), the null hypothesis is rejected and indicating the spatial autocorrelation. In this study, assessing the spatial autocorrelation of historical traffic crash data is essential to prove the hot spots.

3.9.4 Local Bivariate Relationship

Local bivariate relationship used to analyze the spatial relationship between two variables (i.e crash and risky events like acceleration, deceleration, steering, speed, and tailgating events). The local bivariate relationship of traffic crash and risky events was performed through the latest ARC GIS software to determine the significant bivariate spatial relationship between traffic crash and risky events in the study area.

3.9.5 Multivariate Clustering

In spatial analysis, multivariate clustering helps to identify locations with high frequency of crash and risky events. The multivariate clustering in ARC GIS used unsupervised machine learning to determine the natural clusters in the data set. The clustering tool specifies the number of optimum clusters. Multivariate clustering in ARC GIS uses k-means algorithm to partition the features into clusters (Selvi & Caglar, 2018). The K-means algorithm works first by identifying seeds to grow cluster. The first seed selected randomly, then the subsequent seeds will be selected. Once the seeds are identified, all features are assigned to the closest seed feature. The mean data center is computed for each cluster feature. This process of computing means data center and assigning features to the closest center continues until the cluster stabilize a maximum of 100 iteration. Then the tool identify clusters with the cluster id (Jain, 2010).

In this study two clusters are used to classify the data. The first cluster is the study areas with low crash and low risky driving events. The second cluster is the study areas with high crash and high risky driving events.

The locations with frequent crash and all risky events require top urgent action for road safety agencies to the study area and this locations with frequent number of traffic crash and risky events also show unknown relationship between traffic crash and risky events. Thus, it requires further analysis through statistical analysis and predictive models.

3.9.6 Statistical Analysis

Statistical analysis plays a vital role in road safety research and different crash frequency models and predictive models are developed through different conventional statistical models and machine learning models. Due to spatial and temporal variability of crash data, it requires analysis and predictive models with local context. Beyond the spatial and temporal variability of crash data, the risky factors for crash also vary based on the environment, road geometry, driver condition and vehicle factors. Owing to this, local crash frequency models based on the existing condition are favorable. Thus, this study considered different statistical analysis and crash frequency modeling from the risky driving events which observed from the i-DREAMS naturalistic driving study.

3.9.7 Chi-Squared test of association

The count data of historical crashes and risky events was processed in SPSS and Minitab software to conduct Chi-squared test. Chi-squared test of association is a tabular test to assess the significant relationship between two categorical or discrete variables (Lomax & Hahs-Vaughn, 2013). Historical traffic crash data with risky events assessed through Chi-squared test by using the following hypothesis.

Null Hypothesis (Ho): No association between crash and risky driving events within the study area.

Alternative Hypothesis (H1): there is association between crash and risky driving events within the study area.

Chi- squared value and test statistics computed based on the following formula (Mchugh, 2013).

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \text{ .... Eqn-3.4}$$

Where: - E_i is the expected frequency and O_i is the observed frequency. Based on the chi-squared value and degree of freedom the P-value will be computed and the computed P-value compared to the critical significance value which p value of 0.05.

3.9.8 Correlation test

A correlation test was performed after conducting the chi squared test to determine the correlation strength and direction between the risky event and the historical traffic crash. As cognizant that, the traffic crash and risky events are count data which are discrete data. Owing to the discrete nature of the data and lack of normal distribution in the count data, Pearson correlation isn't functional to this data set. Thus, other correlation tests are observed for this data set. Spearman's rank correlation is chosen because it is a non-parametric test and does not consider the data's normal distribution and linear relationship assumption. Spearman's rank correlation, denoted by ρ (rho) assesses the monotonic relationship of two variables. The correlation ranges from -1 to 1 where -1 indicates a perfect inverse relationship, 1 indicates a perfect positive relationship and o indicates no correlation between two variables.

Spearman's rank correlation is computed as follows for no ties in the data set (King & Eckersley, 2019).

$$r = 1 - \frac{6\sum d^2}{n(n^2-1)} \text{ ....Eqn-3.5}$$

Where: - d is the difference between ranks of observation, *n* is the number of observations.

Spearman's rank correlation with ties in data set (King & Eckersley, 2019)

$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)} - \frac{\sum d_i^2}{n(n^2-1)}\frac{1}{\sum y_i^2}$ …………………..Eqn-3.6

Where: - di is the difference between the ranks of pair of values, yi is the number of pairs that have the same rank as value i, and n is the number of pairs.

3.9.9 Crash Frequency Modeling

Researchers applied different crash frequency and predictive models including Poisson, Negative Binomial, Poisson-lognormal, Zero-inflated Poisson and negative binomial, Conway–Maxwell–Poisson, Gamma, Generalized estimating equation, Generalized additive, and Hierarchical/multilevel modeling based on the nature of data. Each of these models has pros and cons in predicting crash frequency. Negative Binomial Modeling is one of methodology in crash frequency modeling. This Negative Binomial Modeling handles the over -dispersion data by assuming gamma distribution of the exponential function of the disturbance term (Lord & Mannering, 2010).

In this study, Negative Binomial model used for the analysis by consideration of the overdispersion nature of the data and easily interpretability power of negative binomial model.

3.9.10 Negative Binomial Modeling

Negative Binomial regression is used to model traffic crashes from risky events using SPSS statistical software. Traffic crash model against below listed variables.

- Dependent variable = number of crashes
- Independent variable = Acceleration, Deceleration, Speed, Steer and Tailgating events with high and medium level of severity.

Note that the counts of risky events and traffic crashes represent the number of occurrences per half-kilometer distance within the study area.

The general mathematical formulation of the model is written as follows: -

$E[\,Y \mid X\,] = \exp(\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k)$ ………………………Eqn-3.7

Whereas: -

Y is the expected number of traffic crashes

$\beta 0$ is the intercept, representing the expected count of crashes when all risky driving event counts are zero

$\beta 1$, $\beta 2$, and $\beta 3$ are the coefficients associated with each risky driving event

X_1, X_2 and X_k are the predictor variables

Negative Binomial Model Assumptions

Before computing negative binomial regression, the following assumptions need to be fulfilled.

1. The predictor and response variable need to be count data or discrete variables.

2. The data need to show overdispersion which means the variance of the data larger than the mean. The over-dispersion nature of the data favors negative binomial model.
3. There should not be a high correlation within the predictor variables. Higher correlation within the predictor variables results in a biased and misleading model. Thus, the correlation between the predictors should be checked and those with high correlation need to be avoided from the model. Multicollinearity may be checked by:
    - Correlation matrix and when computing a matrix of Spearman's rank correlation among all independent variables, the magnitude of correlation should be less than .80.
4. Unusual observation on the data should be removed to avoid outliers in the model. Thus, in the data there should be no sign of outlier.
5. Homoscedasticity (residuals have a constant variance along the fitted value). A scatterplot of residuals is a good way to check homoscedasticity. If the data show homoscedasticity, the data points will be fairly randomly distributed with a fairly even spread of residuals at all predicted values.

Goodness of fit test

The developed model is tested with statistical goodness of fit test in this case the following metrics measurement are employed to evaluate the goodness of fit test.

Deviance: These measures how well the model fits the data and is computed based on the difference between expected and observed counts. A lower deviance indicates a better fit. When the deviance ratio to degree of freedom closes 1 it indicates the better fit of the model.

Akaike information criterion (AIC): This measure was developed by Hirotsugu Akaike in 1974 to evaluate the goodness of fit based on the tradeoff between model complexity and goodness of fit.

$AIC = -2 \times LL + 2 \times (Number of Parameters)$……………Eqn-3.8

This goodness of fit measure is a comparative metrics, and the lower value of AIC indicates the preferable model.

Bayesian information criterion (BIC): This metric is similar to AIC, but it uses a Bayesian approach to estimate the uncertainty of model parameters. BIC is also a comparative metrics to evaluate the goodness of fit and a lower value of BIC is a preferable model.

Pseudo R^2 Goodness of fit test: The count data uses pseudo R^2 goodness of fit test metrics to evaluate how much the full model is modified from the reduced or intercept-only model. The pseudo R^2 computed as follows:

$R^2 = 1\text{-}LL_F/LL_I$

Where: - R^2 is pseudo R^2, LL_F is Log likelihood of full model and LL_I is Log likelihood of intercept only model

In line with the above formula, the negative binomial model developed and the significant predictors for the response are determined based on their P-value.

CHAPTER 4: RESULTS AND DISCUSSIONS

4.1 Spatial Analysis

Spatial Analysis of historical traffic crashes and risky driving events performed through the latest GIS software packages ARC GIS PRO and QGIS. Accordingly, historical traffic crash distribution, risky driving events distribution, hot spot analysis of crashes, spatial autocorrelation of traffic crashes and different spatial analysis of the data was performed.

4.1.1 Historical Traffic Crash Distribution

The data of traffic crash was collected and deployed to latest ARC GIS software to visualize the distribution of traffic crash. Accordingly, the following map is developed based on the historical traffic crash in the Flanders region of Belgium. The map indicates that traffic crashes are mainly concentrated in major cities like Antwerp, Ghent, Bruges, and Mechelen. Three years of traffic crashes from 2017 to 2019 were deployed to the map.

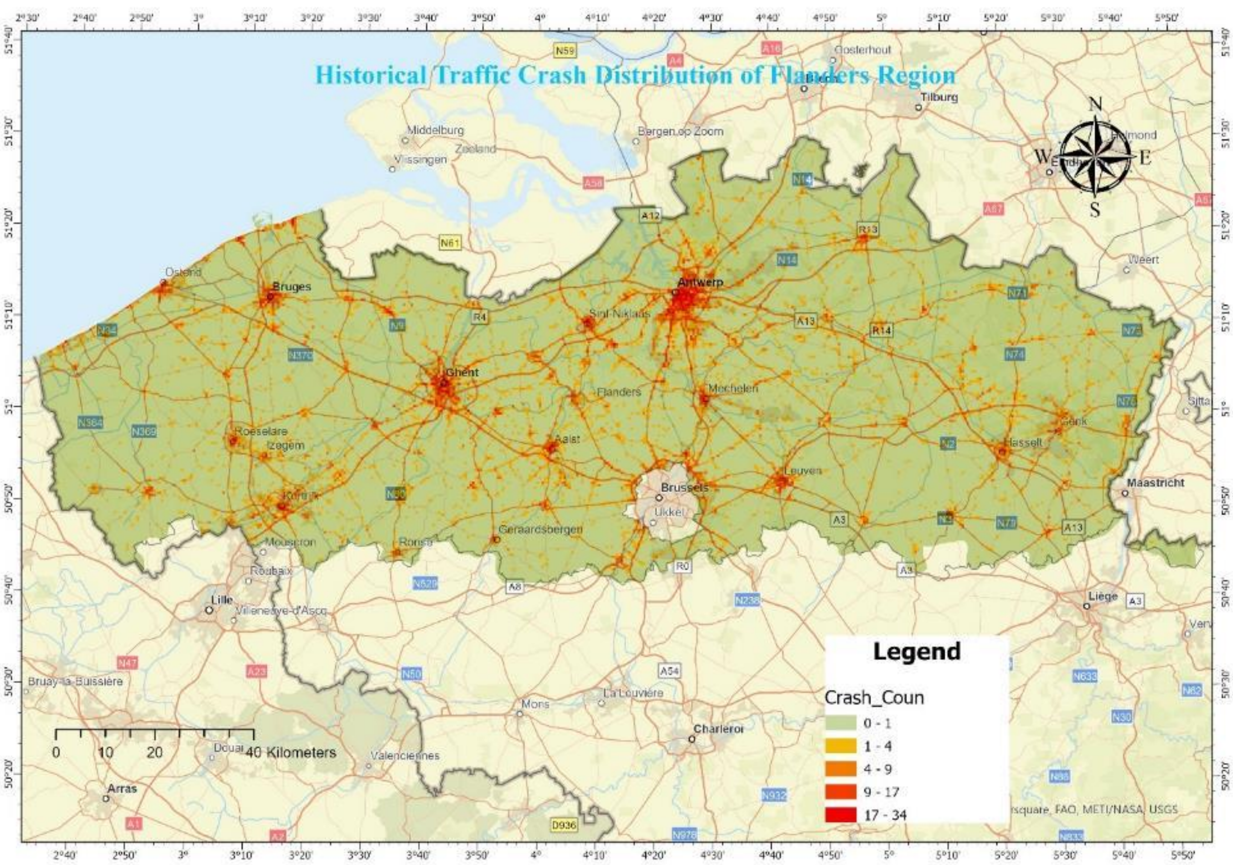

Figure 4.1: Historical Traffic Crash Distribution of Flanders Regions

4.1.2 Spatial Autocorrelation of Historical Traffic Crash

Spatial autocorrelation is conducted to evaluate the randomness and non-randomness of the data set. Accordingly, the test shows that historical traffic crashes have less than a 1% less likelihood to become randomness. This means that the historical traffic crash is not randomly distributed in the region, and the occurrence in the region has certain underlying factors. The test results of Spatial Autocorrelation are shown herein below: -

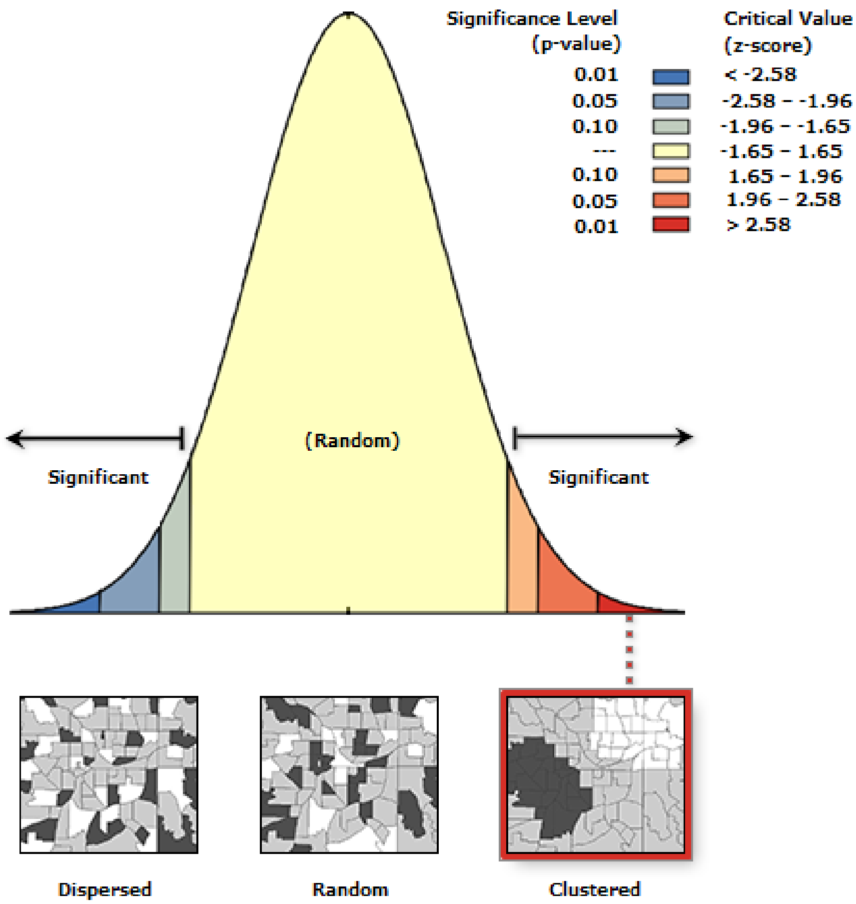

Figure 4.2: Spatial Autocorrelation of Traffic Crash

Moran's Index	0.560917
Expected Index	-0.000018
Variance	0.000009
z-score	185.886083
p-value	0.000000

Given the z-score of 185.886083, with a less than 1% likelihood that this traffic crash cluster could be the result of random chance.

Moran's (I) value : The Moran's I statistic measures the degree of spatial autocorrelation in the data. A value of 0.5609 indicates positive spatial autocorrelation, meaning that areas with similar values (in this case, crash counts) are clustered together.

Expected Index: This represents the expected Moran's I value under the assumption of complete spatial randomness. The expected index is very close to 0, indicating no spatial autocorrelation if the data were randomly distributed.

Variance: which is the measure of the dispersion of Moran's I values. In the data set, it's a small number, suggesting that Moran's I values are relatively consistent across the dataset.

Z-score: which indicates how many standard deviations observed Moran's I value is from the expected value. A Z-score of 185.8861 is high, indicating a high level of spatial autocorrelation.

P-value: The p-value is a measure of statistical significance. A p-value of 0.000000 (or very close to zero) indicates that the observed spatial autocorrelation is highly statistically significant. In other words, there is strong evidence of spatial clustering in the data.

Understanding the presence of spatial autocorrelation in this data is essential for making informed decisions in road safety, as it implies that crash counts are not distributed randomly, and there may be underlying factors contributing to this pattern. Usually, traffic crashes show spatial correlation due to the factor of traffic crashes affected by the road geometry, environment, weather condition, and traffic crashes show spillover effect to the neighborhood (Ahmed & Abdel-Aty, 2015) and (Guo, Wang, et al., 2010).

4.1.3 Hot Spot Analysis of Historical Traffic Crash

As illustrated in the historical traffic crash distribution of the region, it is easy to identify the areas with the highest number of traffic crashes. However, it’s necessary to identify the areas with significant hot spots and cold spot sections. Significant hot spot analysis was performed on ARC GIS with Getis-Ord Gi* statistic.

Getis-Ord Gi* statistic, also known as Gi* statistic is used to identify significant clusters with high values, low values and non-significant clusters. Getis-Ord Gi* statistic measures the spatial autocorrelation in the data set assess whether the feature is dependent on the values of the neighborhood feature. Gi* statistic results visualize on the map. A high positive Z score indicates the hot spot, and a low negative Z score indicates the cold spot.

Getis-Ord Gi* Formula:

$$(\mathrm{Gi}\,*)^{n} = \frac{\sum_{j=1}^{n} \mathrm{w\_ij} * \mathrm{z\_j}}{\sqrt{\sum_{j=1}^{n} \mathrm{w\_ij} * \sum_{j=1}^{n} \mathrm{z\_j\^{}2} \,/\, \mathrm{n}}} \text{……….................................Eqn-4.1}$$

Where:

Gi*: The Getis-Ord Gi* statistic for a specific feature (or location) in the dataset.

w_ij: Spatial weight between the feature of interest (i) and a neighboring feature (j). The weight is typically based on distance or contiguity.

z_j: The standardized value of the attribute for the neighboring feature (j). It is calculated as (x_j - μ) / σ, where x_j is the attribute value for feature j, μ is the mean of all attribute values, and σ is the standard deviation of all attribute values.

n: The total number of features (or locations) in the dataset.

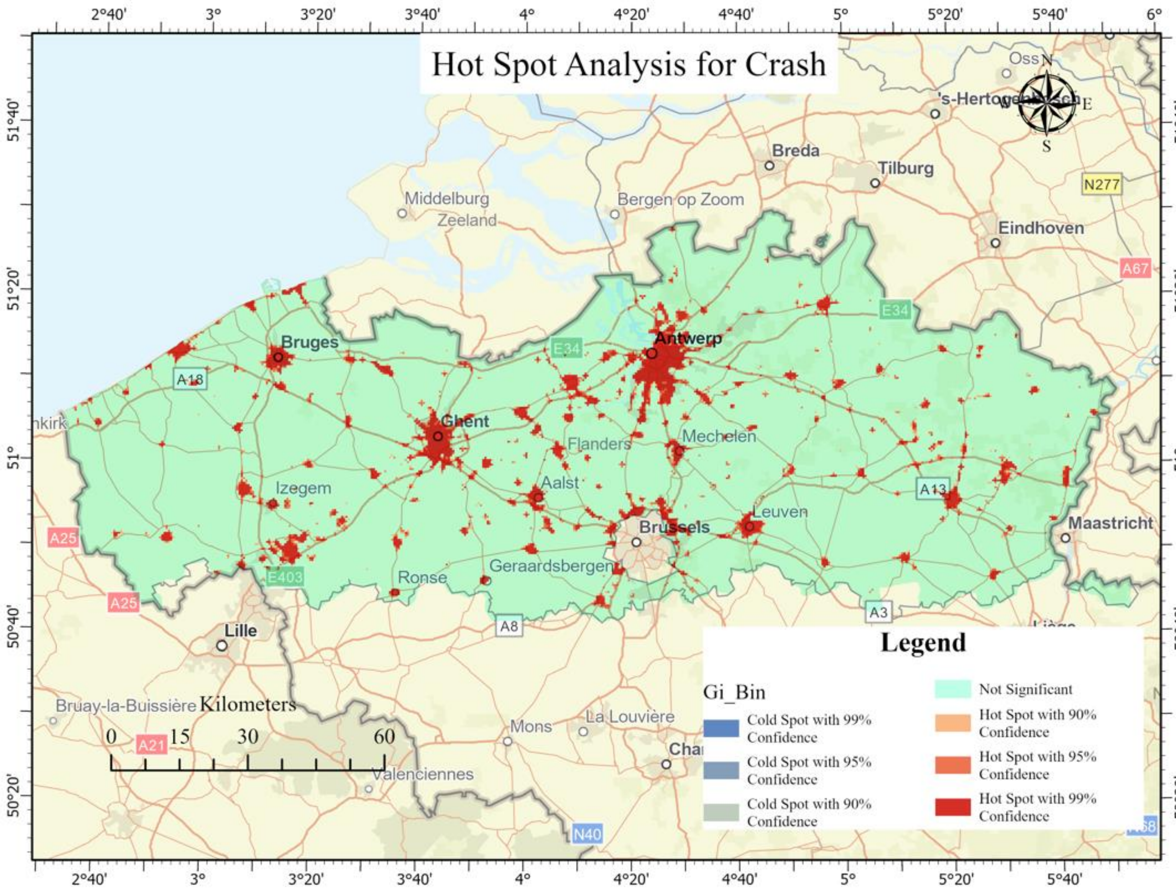

Figure 4.3: Hot spot Analysis of Traffic Crash

The hot spot analysis of traffic crashes indicates that on major cities like Antwerp, Ghent, Bruges, Hasselt, Mechelen and major motorways, the traffic crash distribution is a hot spot with more than 95% confidence. Thus, the road segments around these cities and locations show significant hot spots that require serious intervention from policymakers and decision-makers.

4.1.4 Distribution of Vehicle Control Events in Flanders Region

Distribution of vehicle control events assessed through the latest GIS software. The data collected from i-DREAMS project used for this analysis. The collected data includes vehicle control events with three risk levels: high, medium, and low. This analysis uses only high and medium risky driving events to assess their distribution on crash spot locations. Low-level risky driving events were excluded from the analysis. Accordingly, a spatial distribution map of vehicle control events developed.

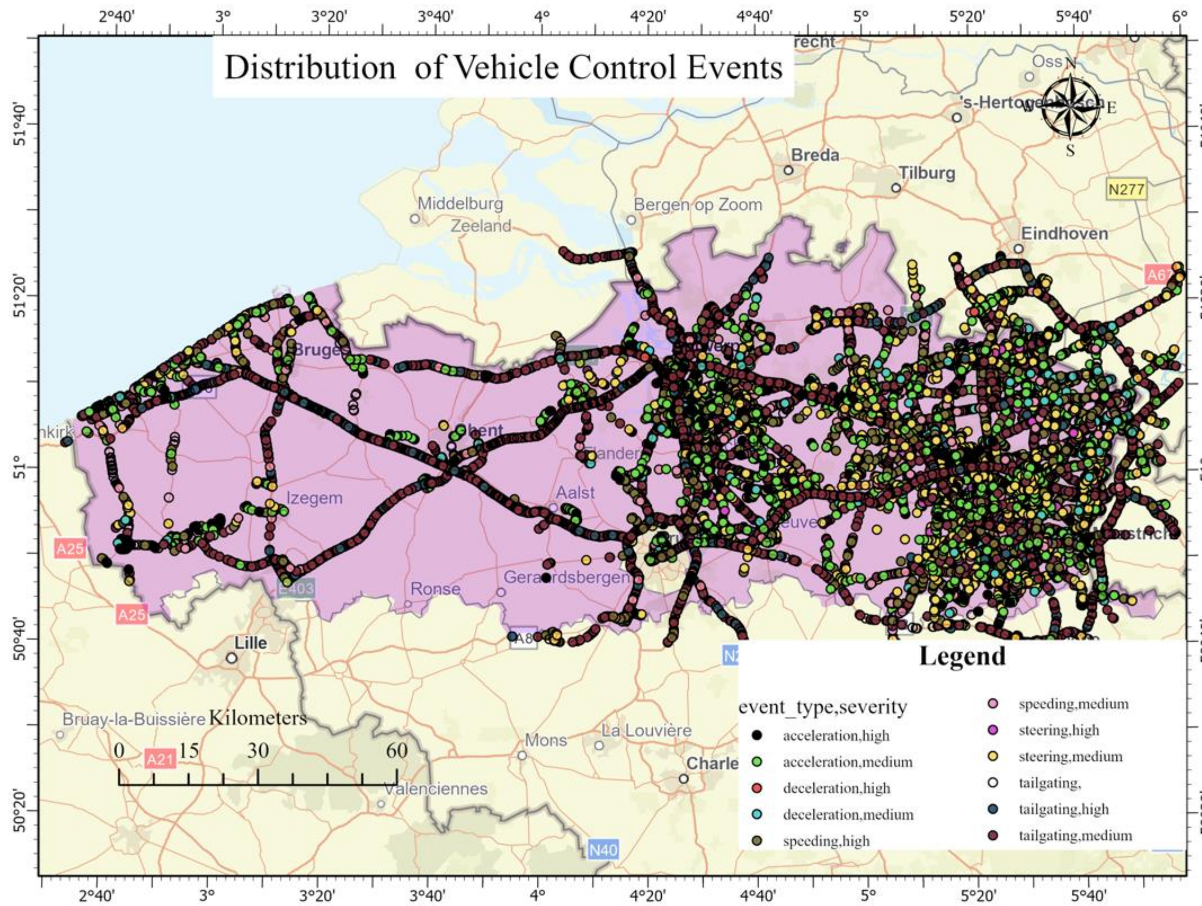

Figure 4.4: Distribution of high and medium vehicle control events in Flanders region

As shown in the above figure, the distribution of collected data by the i-DREAMS project indicated that there is a higher concentration of risky driving events in Limburg, Antwerp, Flemish Brabant provinces. The risky events also concentrated on cities like Hasselt, Genk, Antwerp, Mechelen and on the motorways. Tailgating high and medium are more frequent on motorways and entry and exit of cities. The same acceleration event is more concentrated on motorways and around the cities. Generally, the distribution of vehicle control events in the region shows high frequency at junctions, roundabouts, motorways, and entry as well as exit of cities. Junctions, roundabouts, entry and exit of cities, and motorways are usually known as crash-prone locations, and serious attention is recommended in these sections of the road. The distribution of vehicle control events of the i-DREAMS project also proves this general truth and it's necessary to investigate this unknown relationship of crash with vehicle control event to have proactive remedial measure in the future. On the next section, the study figures out this unknown relationship of traffic crash and vehicle control events with different spatial analysis and statistical analysis techniques.

In addition to the above, after merging traffic crash and risky event data set, the spatial distribution of all events with traffic crash plotted with one map through dot density to visualize the spatial distribution of traffic crash with driving events in the study area. The spatial distribution of traffic crash and vehicle control events shows that high frequent occurrence of both crash and vehicle control events in Limburg Province, Flemish Brabant Province, Antwerp, Mechelen and Motorways.

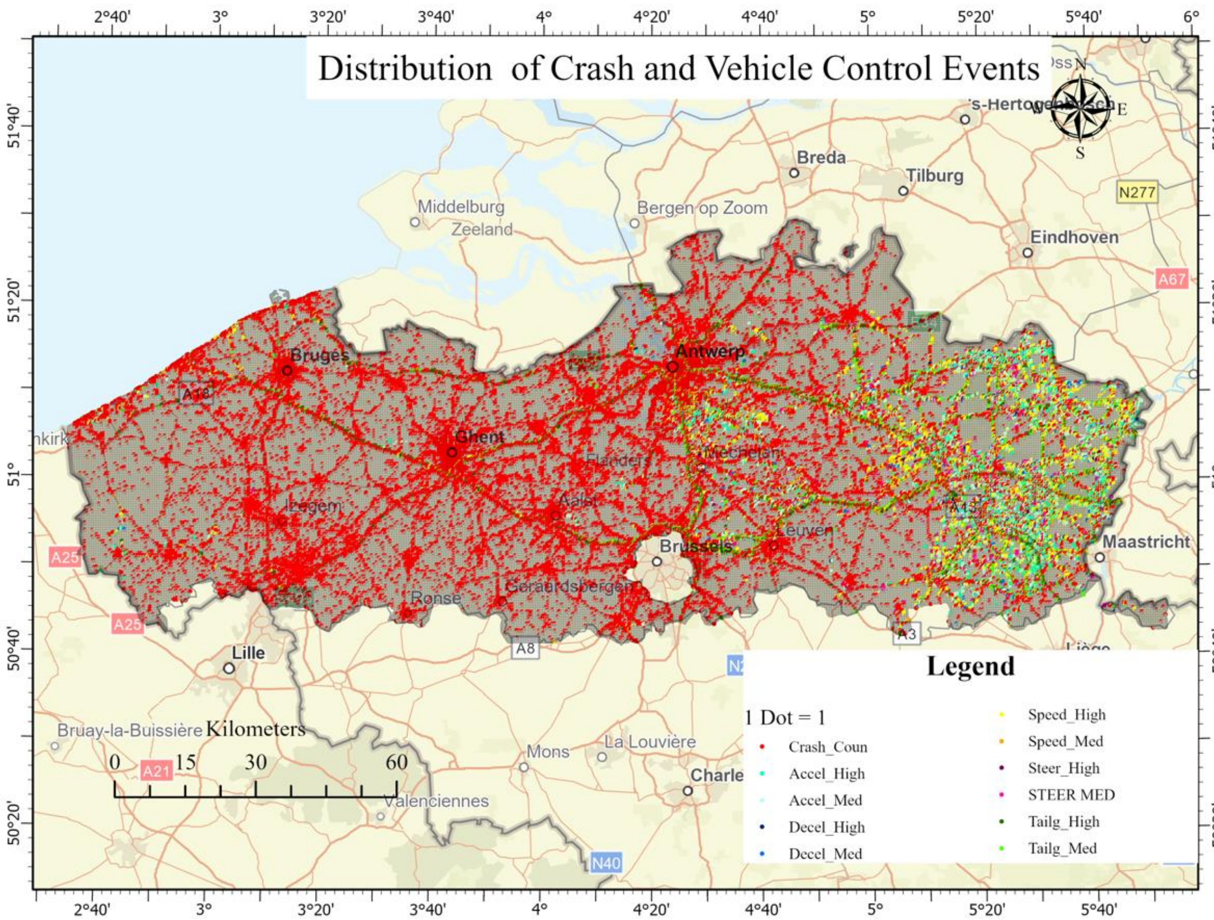

Figure 4.5: Spatial Distribution of Crash with vehicle control events

4.1.5 Local Bivariate Relationship of Crash with Vehicle control events

Local bivariate relationship is a way of analyzing the relationship between two variables in the spatial context. ARC GIS PRO is utilized to develop the relationship between traffic crash and vehicle control events. It can help to find out the relationship between traffic crashes and vehicle control events and to figure out how the relationship varies across the study area.

Local Bivariate relationship in Arc GIS PRO calculates the entropy statistics for each feature in the input data. Entropy is a measure of uncertainty or randomness of a variable. The entropy of each variable is compared to the entropy of both variables, then determines how much information is shared between them. The more information shared the stronger relationship. Local bivariate relationship classifies the relationship into six categories as positive linear, negative linear, concave, convex, undefined complex and not significant relationship.

Accordingly, the following local bivariate relationship between traffic crash and vehicle control events is developed for the study area.

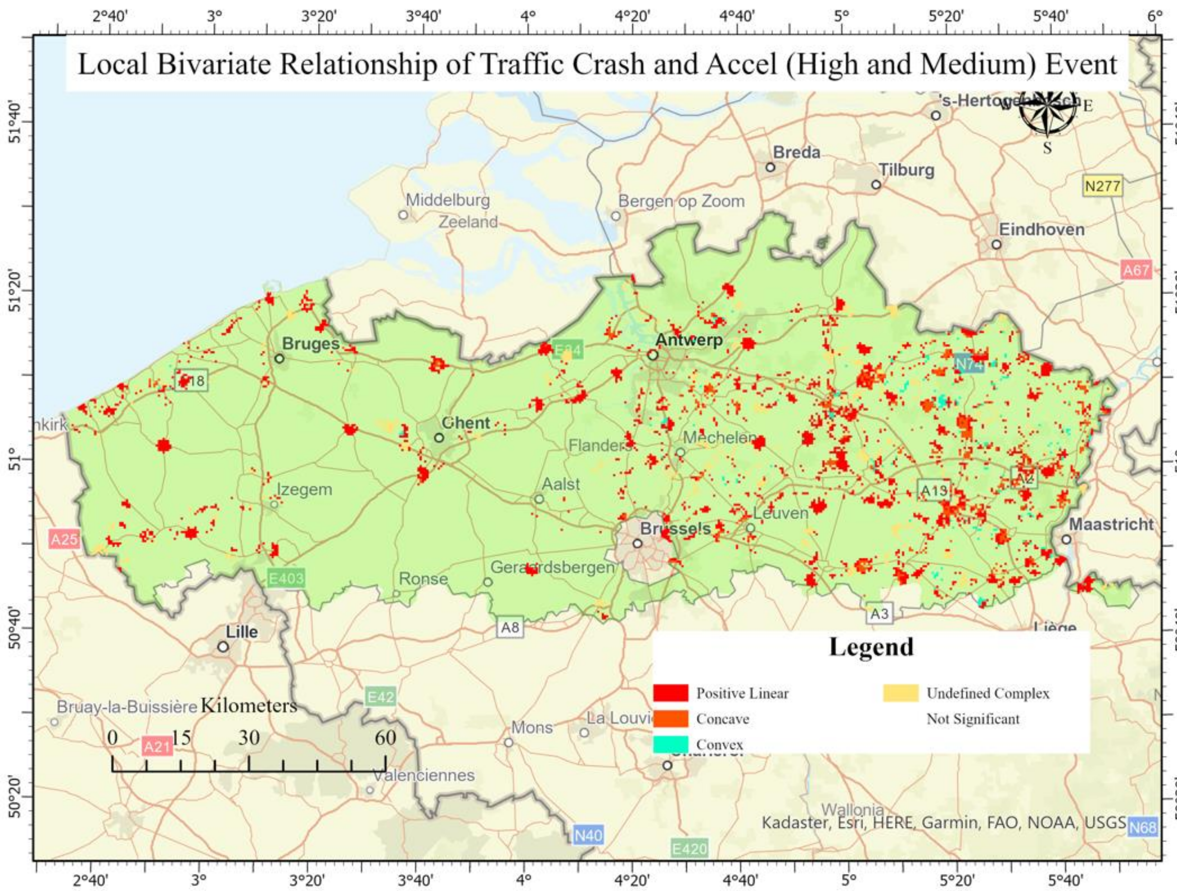

Figure 4.6: Local Bivariate relationship between traffic crash and acceleration event

The Local bivariate relationship between traffic crash and acceleration (high and medium) event shows that in many sections of the study area there is a positive relationship. This positive relationship signifies that these locations exhibit an increase in acceleration (high and medium) events is associated with high risk of traffic crash.

In particular, areas surrounding major cities and motorways within the study region demonstrate a positive relationship between traffic crashes and acceleration events. The identified locations underscore the critical need of targeted intervention to reduce and mitigate the future traffic crash in the region.

Further, certain sections in the study area unveil patterns characterized by both concave and convex relationship between traffic crash and acceleration events. The concave relationship indicates higher concentration of traffic crash towards the central region of the neighbors. The convex relationship indicates the higher concentration of traffic crashes towards the outer and peripheral areas of the neighbors.

In light of this local bivariate relationship, it is essential to implement proactive measures on the identified locations with positive relationship and concave and convex relationships.

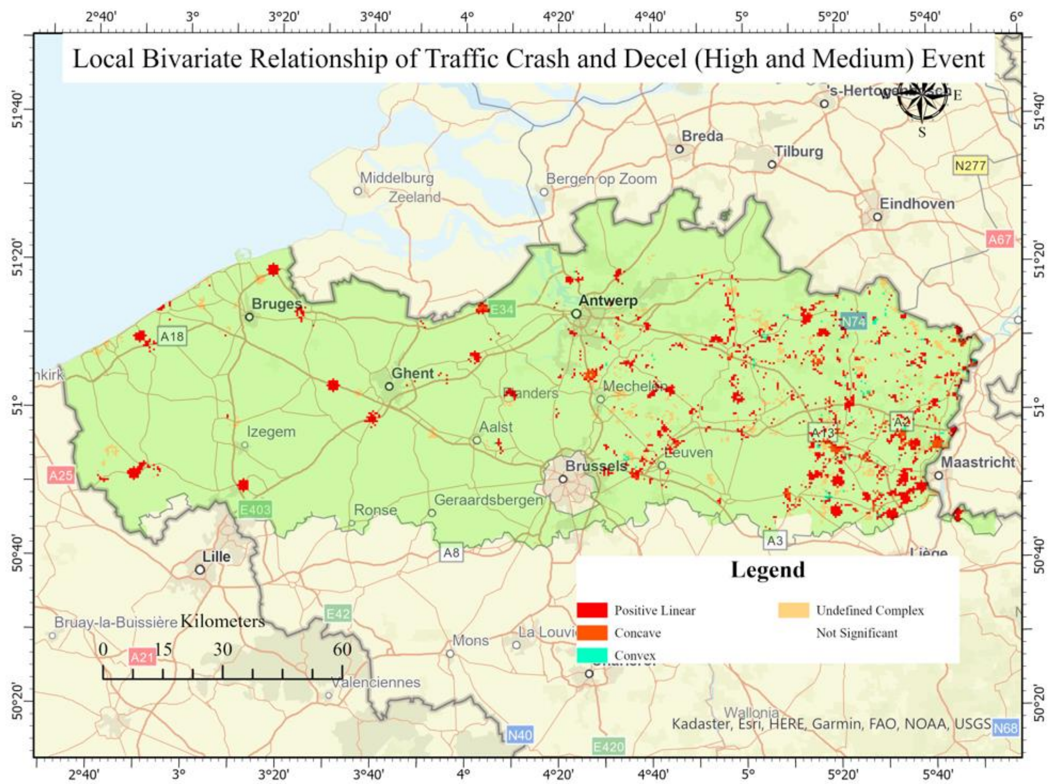

Figure 4.7: Local Bivariate relationship between traffic crash and deceleration event

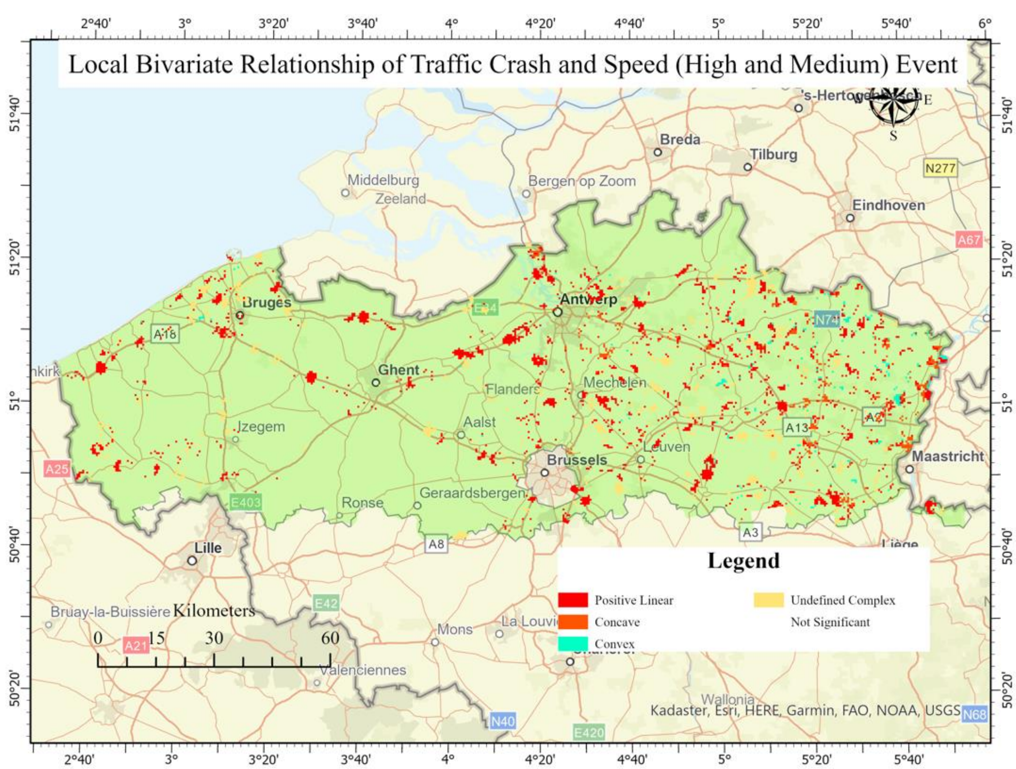

Figure 4.8: Local Bivariate relationship between traffic crash and speed (high and medium) event

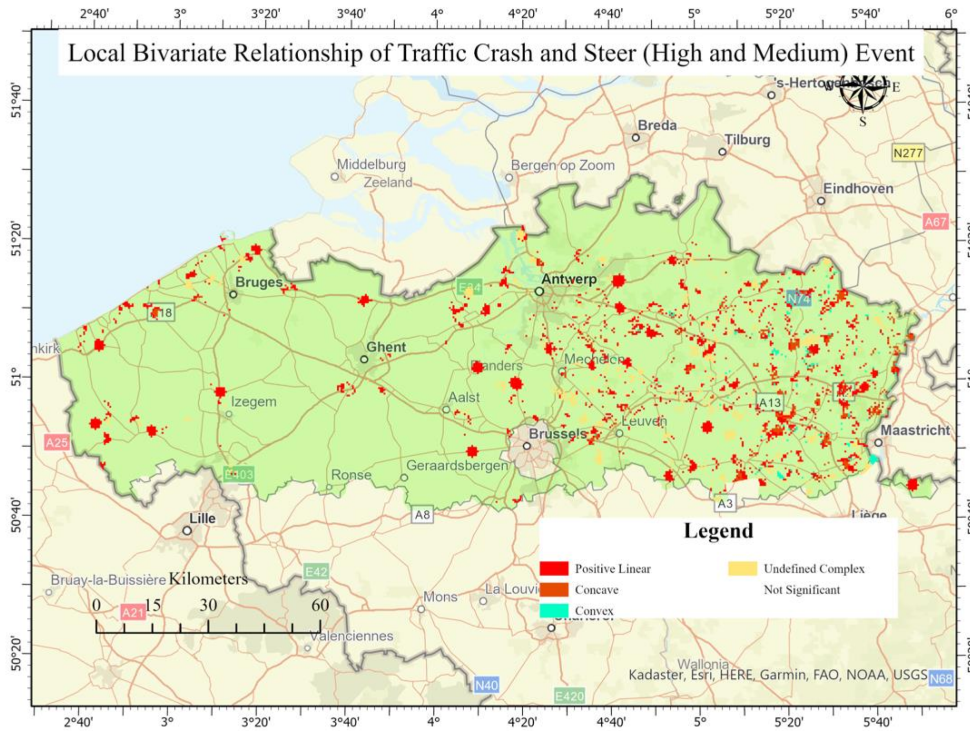

Figure 4.9: Local Bivariate relationship between traffic crash and steer (high and medium) event

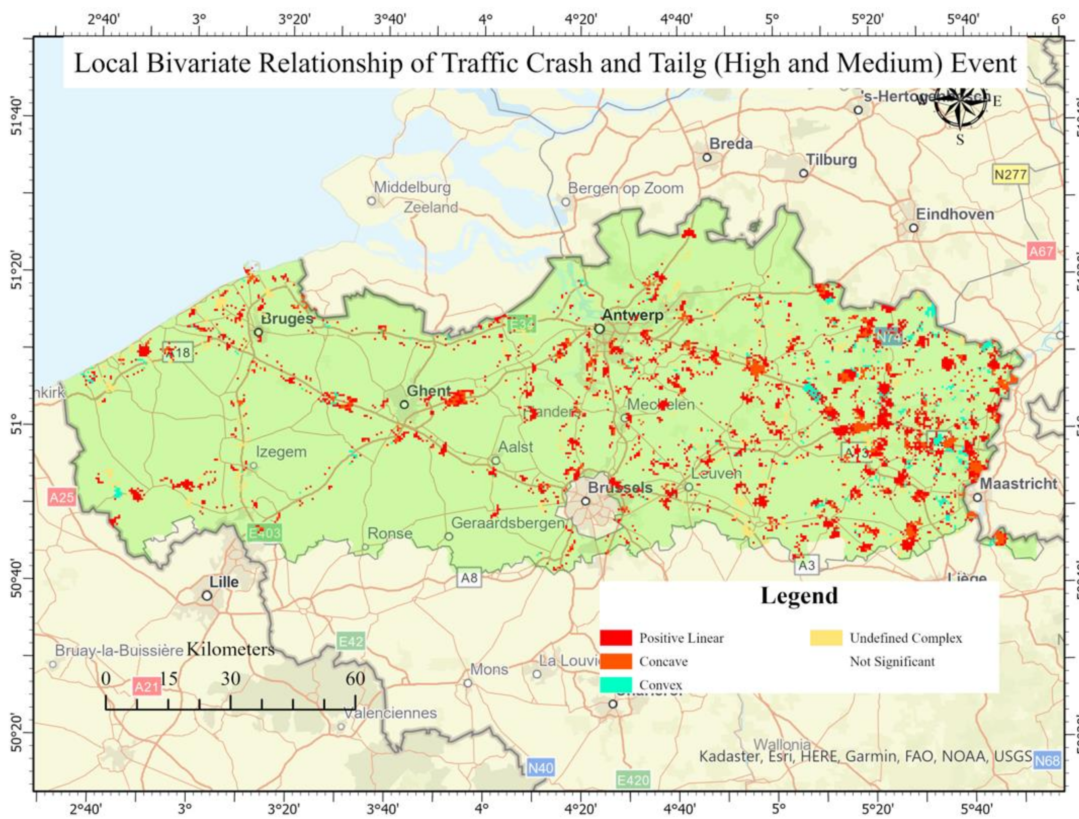

Figure 4.10: Local Bivariate relationship between traffic crash and tailgating (high and medium) event

As shown in the above local bivariate maps, traffic crashes and vehicle control event shows a positive linear relationship on many study area locations. This positive relationship indicated the heightened risk of traffic crashes on frequent risky event locations. In a few study area locations, it shows convex and concave type of relationships.

4.1.6 Multivariate Clustering

Multivariate clustering is a spatial analysis technique to figure out the most prone locations with high incident rates of all the variables. The above section notes that crash has unknown bivariate spatial relationship with each risky driving event. Identifying the most common high frequent locations for all variables gives some picture of the association of crashes with all other risky driving events. Multivariate clustering reveals the areas with similar attribute values concentrated together. In the Multivariate clustering technique, two clusters are assigned to classify the study region. The first cluster is the study areas with low crash and low risky driving events. The second cluster is the study areas with high crash and high risky driving events.

The following multivariate clustering developed by using two clusters.

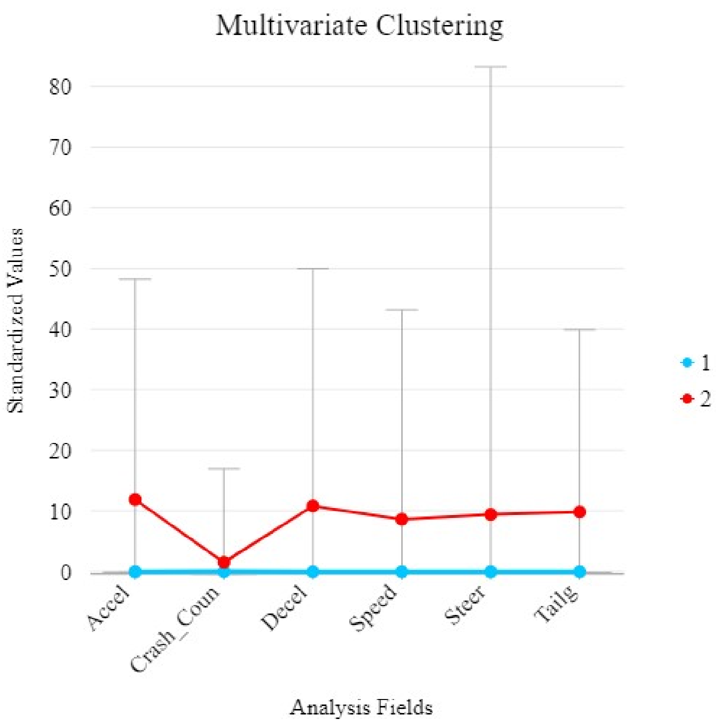

Figure 4.11: Multivariate clustering

As indicated on figure 4.11, the multivariate clustering shows two clusters, whereas cluster 1 is a cluster with low crash and low risky driving events. Cluster 2 shows high crashes with high risky driving events. The figure also illustrates that the number of crashes increased with frequent risky driving events. Based on the above figure, the two clusters develop a multivariate clustering map.

As indicated on the following multivariate clustering map, clustering 1 covers a large part of the study area since the probability of low crashes and low number of events at same location is high. Furthermore, the location under cluster 2 are too much necessary for policymakers and road safety institutions of the study area to take proactive measures. The locations identified on cluster 2 are around Hasselt, Genk city, along motorways.

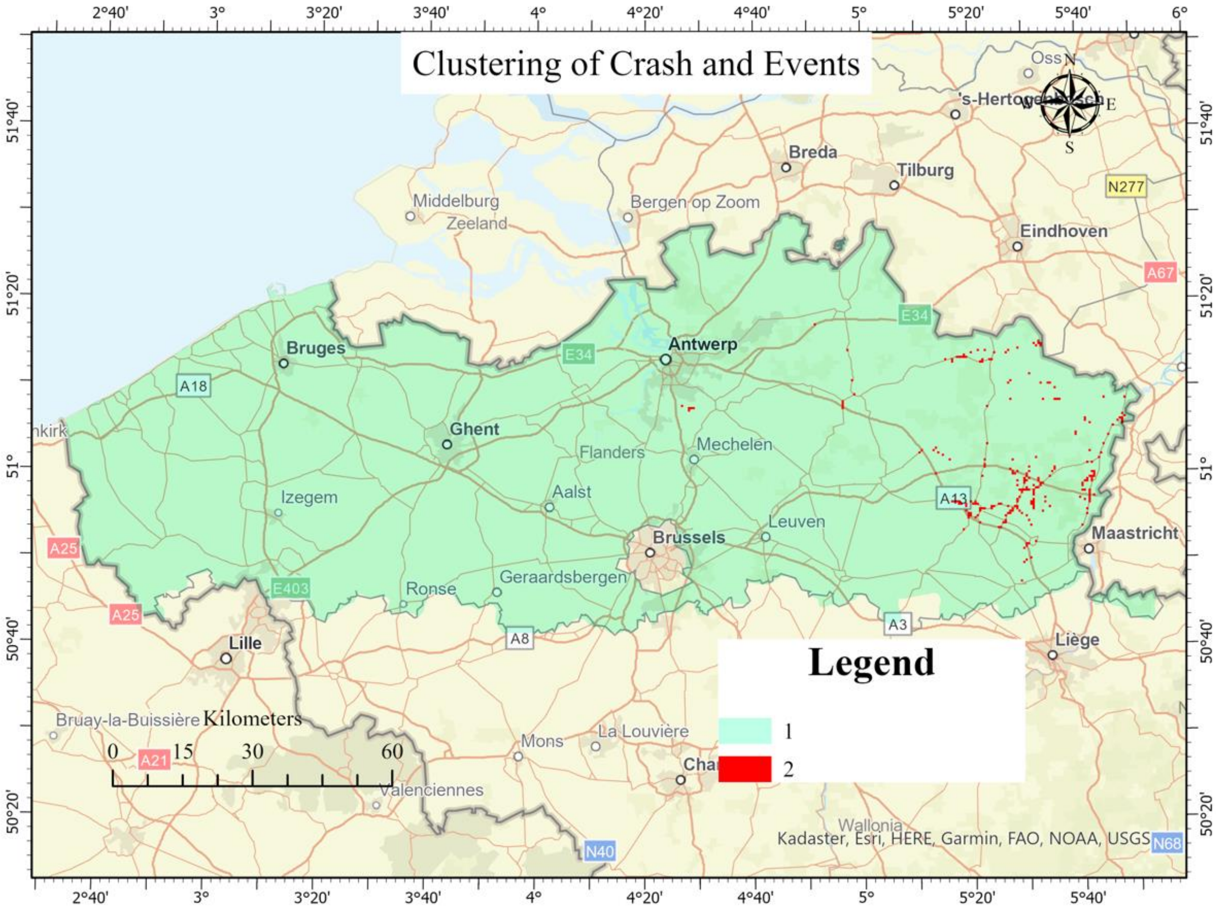

Figure 4.12: Multivariate Clustering of Crash and risky driving events

4.2 Statistical Analysis

As indicated above, the count data of events and traffic crashes within 500m distance of road segment performed through latest GIS software. Accordingly, different statistical analysis are performed through this count data information.

4.2.1 Historical Traffic Crash Distribution

The following traffic crash frequency graph and scatter plots are developed based on the count within 500m distance of road segments.

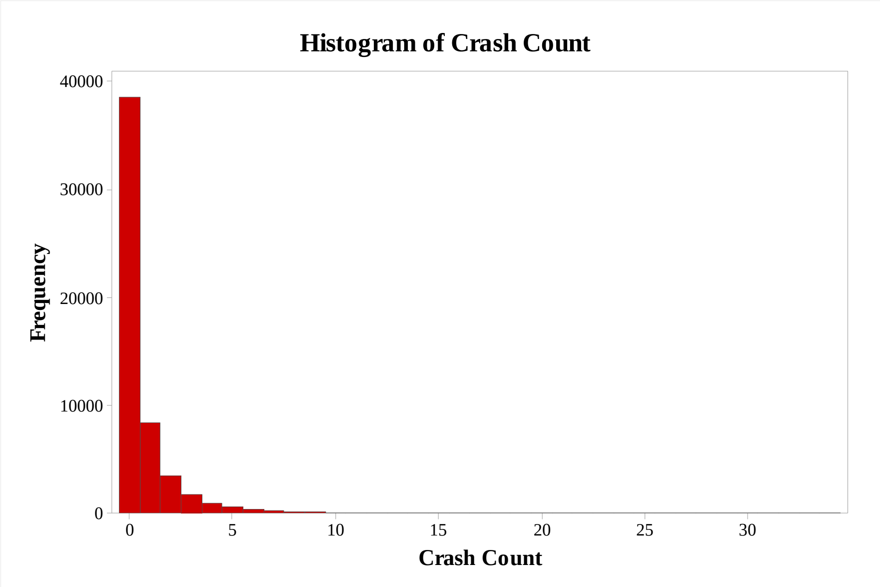

Figure 4.13: Frequency distribution of traffic crash

Table 4.1:Descriptive Statistics of Traffic Crash

Variable	Mean	StDev	Variance	Minimum	Maximum
Crash Coun	0.7798	1.9574	3.8315	0.00000	34.0000

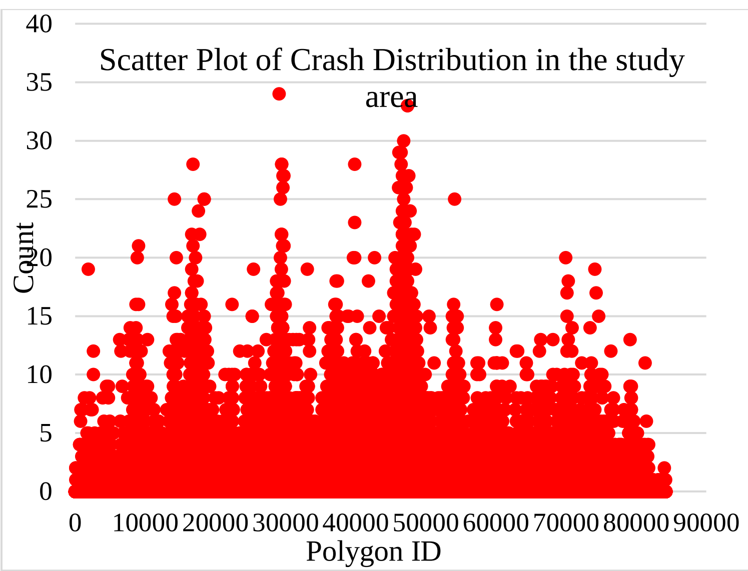

Figure 4.14: Scatter plot of Traffic Crash

The scatter plot of traffic crashes is represented by the polygon ID, and polygons represent each 500m road segment. The count of traffic crashes performed within 500m intervals in the study area. The traffic crash is plotted against with this polygon. The ID is used to designate the strips of the road sections which used for the analysis with 500m distance. Accordingly, the provinces of Flanders region designated with the following polygon ID's in ARC GIS for the analysis.

- ✓ ID 0 to 21000 for West Flanders
- ✓ ID 21000 to 44000 to East Flanders
- ✓ ID 44000 to 57000 to Antwerp
- ✓ ID 57000 to 65000 to Flemish Brabant
- ✓ ID 650000 to 85000 to Limburg province

In addition to crash distribution, the scatter plot of crash with events plotted on this polygon and shown herein below:

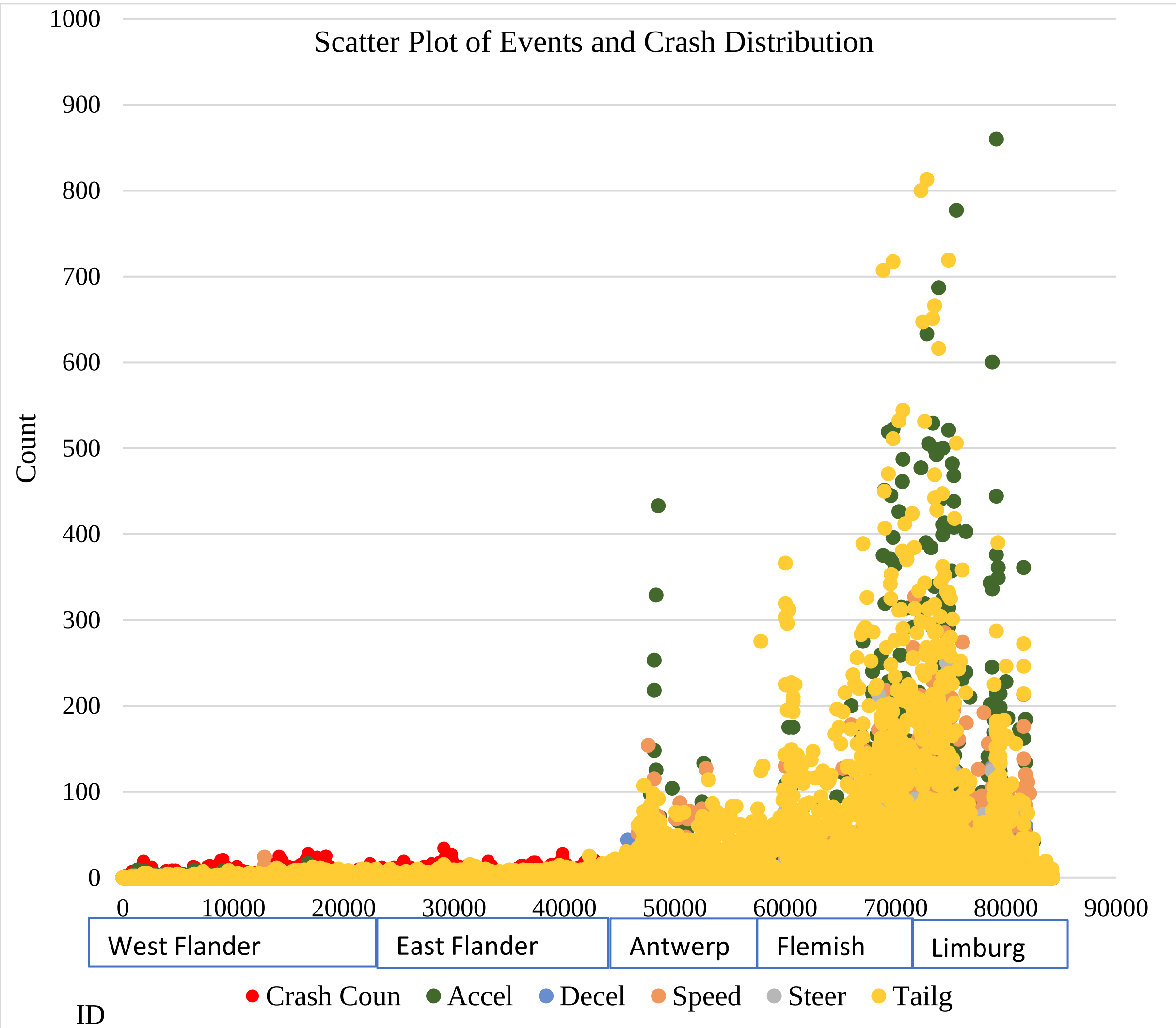

Figure 4.15: Scatter plot of Crash with risky driving events (Medium and High events)

4.2.2 Statistical test of Association between crash and events

Statistical test of association between crash risky driving events performed through chi-squared test of association. This chi-squared test of association answer two questions in this study which are:

Is there a significant association between traffic crashes and risky driving events?

Does the occurrence of risky driving events affect the likelihood of traffic crash?

In order to answer these questions, the following null and alternative hypothesis was developed.

Ho = there is no association between crash and risky driving events within the study area

H1= there is association between crash and risky driving events within the study area

According to the above hypothesis, the test was performed for all events with traffic crash and the following results obtained.

Chi-squared test of crash and Acceleration high event

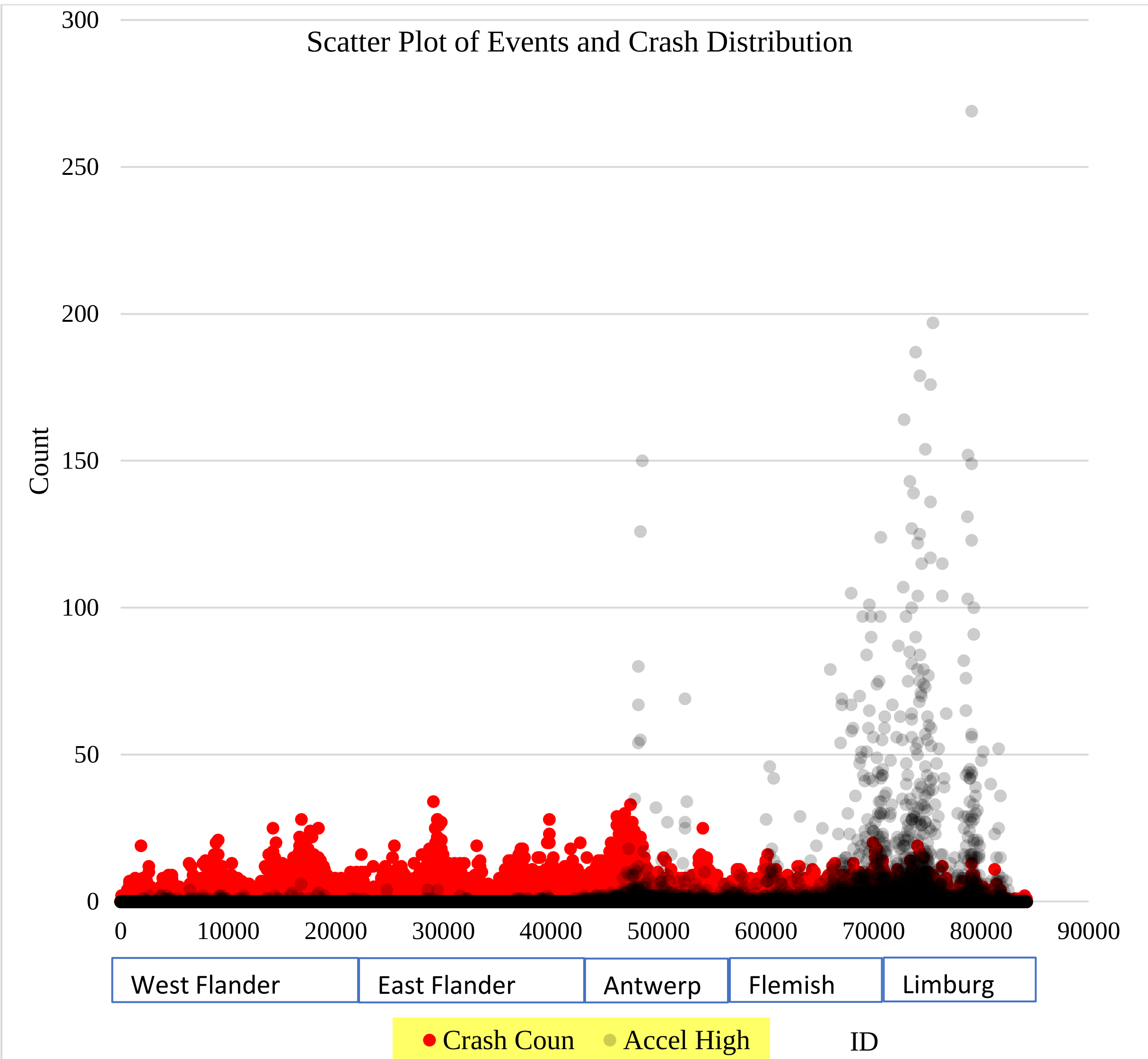

Figure 4.16: Scatter plot of Crash and Acceleration High event

In line with the traffic crash and acceleration high event distribution, a chi-square test was conducted to determine the association between crash and acceleration high event using the following null and alternative hypothesis.

Ho = there is no association between crash and Acceleration high event within the study area

H1= there is association between crash and acceleration high within the study area

Chi- Square test result

Chi-square statistic: 10154.695

P-value: 0.000

Degrees of freedom: 858

The Chi-Square statistics is 10154.695 and Degree of freedom 858 indicates that there is significant difference between observed frequency (actual count of crash and acceleration high event) and expected frequency (the expected counts if there were no association between two variables). The p-value 0.000 means there is strong evidence to reject the null hypothesis. Thus, it suggests there is a significant association between traffic crashes and acceleration high event.

Chi-squared test of crash and Acceleration medium event

Ho = there is no association between crash and acceleration medium event within the study area

H1= there is association between crash and acceleration medium within the study area

Chi-square statistic: 17283.857

P-value: 0.000

Degrees of freedom: 1694

The Chi-Square statistics is 17283.857 and Degree of freedom 1694 indicates that there is significant difference between observed frequency (actual count of crash and acceleration medium event) and expected frequency (the expected counts if there were no association between two variables). The p-value 0.000 means there is strong evidence to reject the null hypothesis. Thus, it suggests there is a significant association between traffic crashes and acceleration medium event.

Chi-Squared test of all variables

The chi- squared test conducted for all variables with traffic crash through the following null and alternative hypothesis.

Ho = there is no association between crash and risky driving events within the study area

H1= there is association between crash and risky driving events within the study area

Table 4.2: Chi-Squared Test of all variables with traffic Crash

Variable	Chi-square statistic	P-value	Degrees of freedom

Acceleration High	10154.695	0.000	858
Acceleration Medium	17283.857	0.000	1694
Deceleration High	570.655	0.000	110
Deceleration Medium	6270.964	0.000	660
Speeding High	3668.340	0.000	1089
Speed Medium	2300.378	0.000	495
Steering High	6485.019	0.000	572
Steering Medium	17429.259	0.000	1925
Tailgating High	5581.095	0.000	792
Tailgating Med	23100.710	0.000	1837

As illustrated in table 4.2, the Chi-Squared analysis for all variables reveals a P-value of 0.000 for all variables, accompanied by elevated chi-squared statistics. Thus, the chi- squared test suggests a significant association between traffic crashes and risky driving events. This evidence supports the rejection of null hypothesis.

In line to the above results, the chi-squared test initiates to assess the correlation between risky driving events and traffic crashes.

4.2.3 Correlation test Crash with risky driving events

As cognizant that, the data set for the study is a count data which is discrete data. The traditional Pearson’s correlation test is not suitable for this data set due to the following reasons: -

- ✓ Assumption of linearity: Pearson's correlation assumes linear relationship of the variables, while count data, the counts discrete and may not exhibit linearity.
- ✓ Normality: Pearson's correlation assumes normal distribution of data. Count data are typically not normally distributed and are mostly skewed distribution.
- ✓ Homoscedasticity: Pearson's correlation assumes the variance of one variable constant across all values of other variables. Count data violates this assumption that the variance may change with the level of count.
- ✓ Discrete Data: Pearson's correlation designed for continuous data whereas count data are discrete and treating them as continuous lead to inaccurate result.

Owing to the above limitations of Pearson's correlation, Spearman's Rank Correlation used for this study due to it is well-suited for data with non-linear, non-normally distributed, and discrete characteristics. It provides a more robust and appropriate measure of association for count data. Spearman's Rank Correlation, also known as Spearman's rho (ρ), is a non-parametric measure of monotonic relationship between variables. Spearman's rho (ρ) used when the data is discrete, ordinal, interval, or non-normally distributed. Accordingly, Spearman's Rank Correlation was conducted for the study data set using SPSS.

As depicted in table 4.3, Crash has a positive correlation with risky driving events, suggesting that crash increases when the risky driving events increment. Among all variables, tailgating medium exhibit a substantial correlation with crash a value of 0.311, which indicates the risk of crash is high on the frequent tailgating medium locations. In the next section, predicting models of traffic crash performed through the risky driving events as predictor by different statistical models.

Table 4.3: Correlation test

		Crash_ Coun	Accel_ High	Accel_ Med	Decel_ High	Decel_ Med	Speed_ High	Speed_ Med	STEER MED	Steer_ High	Tailg_ High	Tailg_ Med
Crash_Coun	Coeff	1										
	Sig.											
Accel_High	Coeff	.190**	1									
	Sig.	0										
Accel_Med	Coeff	.222**	.724**	1								
	Sig.	0	0									
Decel_High	Coeff	.099**	.351**	.304**	1							
	Sig.	0	0	0								
Decel_Med	Coeff	.171**	.592**	.589**	.485**	1						
	Sig.	0	0	0	0							
Speed_High	Coeff	.205**	.495**	.583**	.284**	.488**	1					
	Sig.	0	0	0	0	0						
Speed_Med	Coeff	.197**	.480**	.534**	.288**	.475**	.641**	1				
	Sig.	0	0	0	0	0	0					
STEERMED	Coeff	.197**	.587**	.674**	.321**	.568**	.547**	.518**	1			
	Sig.	0	0	0	0	0	0	0				
Steer_High	Coeff	.135**	.513**	.478**	.356**	.514**	.399**	.406**	.569**	1		
	Sig.	0	0	0	0	0	0	0	0			
Tailg_High	Coeff	.244**	.395**	.425**	.257**	.406**	.477**	.476**	.398**	.347**	1	
	Sig.	0	0	0	0	0	0	0	0	0		
Tailg_Med	Coeff	.311**	.470**	.565**	.251**	.465**	.580**	.565**	.522**	.369**	.684**	1
	Sig.	0	0	0	0	0	0	0	0	0	0	

4.2.4 Negative Binomial Modeling

Negative Binomial Modeling is used to develop a statistical model for traffic crash count by using the predictors of risky driving events. Accordingly, the Negative Binomial Model is developed to traffic crash count after testing each assumption of Negative Binomial Modeling.

Examining Negative Binomial Modeling Assumptions

To conduct Negative Binomial Modeling, it is necessary to examine the major assumptions before conducting any modeling task which is helpful to avoid misleading results of one or more assumptions.

Assumption 1: Count Data

Negative Binomial Modeling is designed for the count data. Hence, it is necessary that the predictor and dependent variables as a count data. As indicated above, the data used in this study is a count data for all variables.

Assumption 2: Overdispersion

One of the primary focus to use negative binomial modeling is to account overdispersion in the data. Overdispersion occurs when the variance of the data greater than the mean. As shown in table, the variance of the study data greater than the mean which implies there is overdispersion on the data.

Assumption 3: Multicollinearity (the predictor variable should not highly correlate each other).

If the independent variables are highly correlated to each other, it gives misleading result on the model fitting. Spearman's Rank Correlation was conducted for all variables and the magnitude of correlation should be less than 0.8. According to this, all variables have less than 0.8 correlation as indicated in correlation matrix table. Note that the predictors include high and medium level of severity.

Table 4.4: Correlation Matrix of between predictors

		Acceleration	Deceleration	Speed	Steer	Tailgating
Acceleration	Correlation	1.000				
	Sig.					
Deceleration	Correlation	.590**	1.000			
	Sig.	0.000				
Speed	Correlation	.596**	.483**	1.000		
	Sig.	0.000	0.000			
Steer	Correlation	.674**	.568**	.555**	1.000	
	Sig.	0.000	0.000	0.000		
Tailgating	Correlation	.560**	.462**	.619**	.518**	1.000
	Sig.	0.000	0.000	0.000	0.000	

Assumption 4: No significant outlier

According to this assumption, the data points with high leverage points, highly influential points and the data points with significant outlier removed from the analysis and the same data points with unusual observation removed from the analysis.

Assumption 5: Homoscedasticity

The data is Homoscedasticity when the variance along the line of best fit remains the same along the best fit line. If the data shows homoscedasticity, the residuals exhibit an overall fairly random display. The scatter plot of fitted value versus standardized deviance residuals exhibits an overall fairly random display as shown in the figure below. Hence, it can be concluded that there is no heteroscedasticity, and the model shows Homoscedasticity.

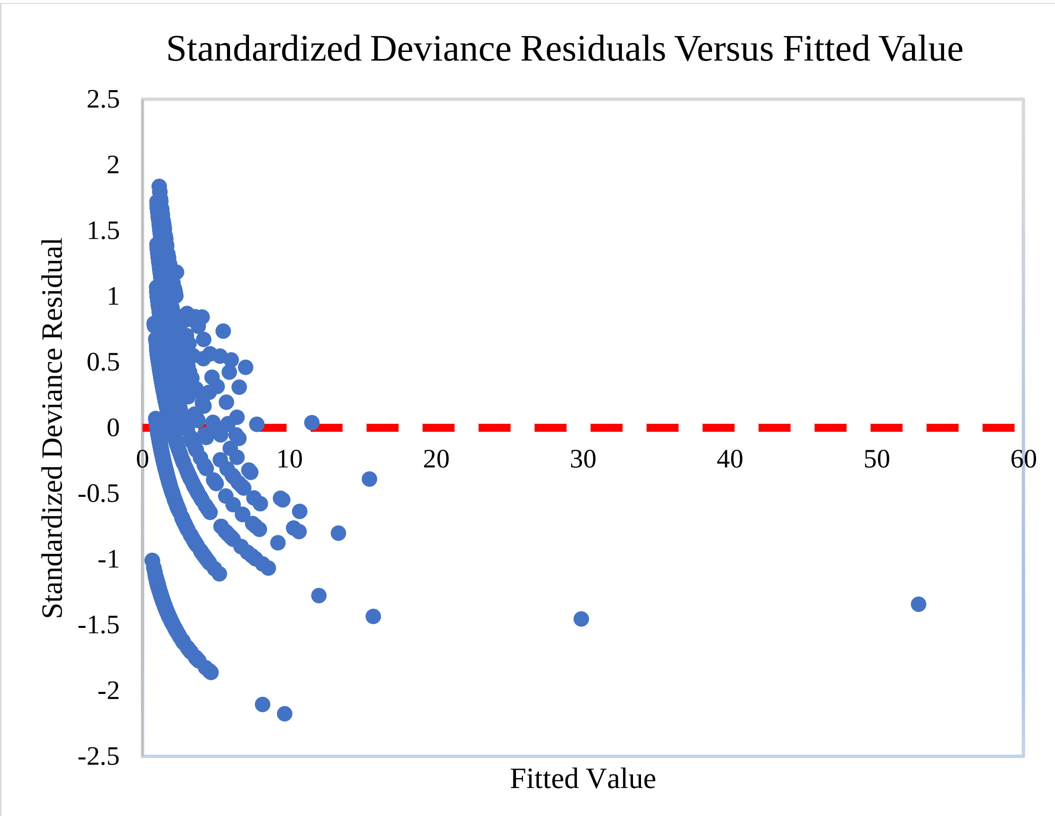

Figure 4.17: Scatter Plot of Standardized Residual Versus Fitted Values

Negative Binomial Regression

A negative binomial regression performed to model traffic crash with the following predictor variables. The predictor variables consist count data of risky driving events which are acceleration (acceleration medium and high event), Deceleration (deceleration medium and high event), Speeding (speed medium and high event), Steering (steering medium and high event), and Tailgating (Tailgating medium and high events). All the predictor and response variable data are counted on 500m strip of roads for the study area. To perform this, 54,007 observations are conducted, a count data of response and predictor variable taken on 54,007 locations from the study area.

$$E[Y \mid X] = \exp(\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k) \text{ ....Eqn-4.2}$$

Whereas: Y is the expected number of traffic crashes, β0 is the intercept, β1, β2, and β3 are the coefficients associated with each risky driving event and X_1, X_2 and X_k are the predictor variables.

The Negative Binomial Regression was conducted by using SPSS statistical software and the following output developed.

Table 4.5: Variables information

Count Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation
Dependent Variable	Crash_Coun	54007	0	12	.52	1.042
Covariate	Accel	54007	0	860	1.14	12.431
	Decel	54007	0	144	.18	2.358
	Speed	54007	0	247	.73	5.899
	Steer	54007	0	1878	1.24	18.766
	Tailg	54007	0	424	1.74	12.129

Table 4.5 shows the variance is greater than the mean for all variables, which is overdispersion on the data, and 54,007 rows of count data are taken to the analysis.

Goodness of Fit

Table 4.6: Goodness of Fit for Negative Binomial

	Value	df	Value/df
Deviance	38611.620	54000	.715
Scaled Deviance	38611.620	54000	
Pearson Chi-Square	50649.643	54000	.938
Scaled Pearson Chi-Square	50649.643	54000	
Log Likelihood[b]	-51176.380		
Akaike's Information Criterion (AIC)	102366.759		
Finite Sample Corrected AIC (AICC)	102366.761		
Bayesian Information Criterion (BIC)	102429.037		
Consistent AIC (CAIC)	102436.037		

Dependent Variable: Crash_Coun
Model: (Intercept), Accel, Decel, Speed, Steer, Tailg

Deviance is a metrics to measures goodness of fit and in this model, deviance is 38,611 which is the difference between the observed data and models prediction. The best model has Value/df close to 1. As shown in the above table, this study has 0.715 value/df which is closer to 1 and means the model fits the data well. Pearson Chi-squared is also another measure of goodness of fit. The value/df closer to 1 is the best model on Pearson Chi-squared test. The model shows a value/df of 0.938 for Pearson Chi-squared test which is closer to 1 and this suggests that the model provide a good fit.

Loglikelihood measures how well the model explains the observed data. A lower (more negative) loglikelihood is desirable for a better fit model. Akaike's Information Criterion (AIC) and Bayesians' Information Criterion (BIC) are comparative measures of model fit and complexity. Lower AIC and BIC values indicate a good model.

Omnibus Test

Omnibus test of the model is conducted to assess if the fitted model performs better than the reduced model (intercept only model). The model with large chi-squared and lower p-value is a better model which indicates fitted model performs better than the intercept only model.

Table 4.7:Omnibus test

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
1368.383	5	.000

Dependent Variable: Crash_Coun
Model: (Intercept), Accel, Decel, Speed, Steer, Tailg

a. Compares the fitted model against the intercept-only model.

According to this study data, the Chi-squared is 1368.383 and the P value of 0.000 which indicates that multiple predictors provide significantly better fit of the data compared to the model with no predictors (intercept only model). In general, inclusion of predictors improves the model’s ability to explain the response variable.

Table 4.8: Significant Variables

Tests of Model Effects

	Type III		
Source	Wald Chi-Square	df	Sig.
(Intercept)	6570.595	1	.000
Accel	4.125	1	.042
Speed	27.307	1	.000
Steer	.067	1	.796
Tailg	361.582	1	.000
Decel	.580	1	.446

Dependent Variable: Crash_Coun
Model: (Intercept), Accel, Speed, Steer, Tailg, Decel

Table 4.8 illustrates the significant effect of each variable on the response variable or traffic crash. According to the negative binomial modeling, in the study area Acceleration, Speed and Tailgating events have statistically significant effect (P<0.05) on the expected count of traffic crash. However, deceleration and steering events do not have a statistically significant effect (P>0.05) on the expected traffic crash count. The statistical significance of speed, acceleration and tailgating events reveals that this variable have a notable impact on the expected traffic crash in the study area.

After conducting a statistically significant test, each predictor's parameter estimate was conducted to know the magnitude of effect for the response variable.

Table 4.9:Parameter Estimates

			95% Wald Confidence Interval		Hypothesis Test				95% Wald Confidence Interval for Exp(B)	
Parameter	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.	Exp(B)	Lower	Upper
(Intercept)	-.732	.0090	-.749	-.714	6570.595	1	.000	.481	.473	.490
Accel	.003	.0014	9.884E-5	.006	4.125	1	.042	1.003	1.000	1.006
Speed	.009	.0018	.006	.013	27.307	1	.000	1.009	1.006	1.013
Steer	.000	.0009	-.002	.002	.067	1	.796	1.000	.998	1.002
Tailg	.019	.0010	.017	.021	361.582	1	.000	1.019	1.017	1.021
Decel	-.004	.0059	-.016	.007	.580	1	.446	.996	.984	1.007
(Scale)	1[a]									
(Negative binomial)	2.158	.0375	2.086	2.233						

Dependent Variable: Crash_Coun

Model: (Intercept), Accel, Speed, Steer, Tailg, Decel

a. Fixed at the displayed value.

b. Note that all predictor variables include high and medium events

As shown in table 4.9, the intercept, accel, speed and tailgating coefficients are significant with P<0.05.

- ✓ **Intercept**

The intercept coefficient is -0.732 and the significance level is very low(P<0.001), indicating that the intercept is statistically significant, Exp(B) is 0.481 which entails that for every one unit decrease in the predictor variables, the expected traffic crash count decrease by a factor of 0.519. The intercept coefficient indicates that the absence of risky driving events contributes to decreased traffic crashes.

- ✓ **Acceleration event**

The coefficient for acceleration is 0.003. A positive coefficient indicates that an increase in the acceleration event is associated with an increase in the expected traffic crash count. The significance level of 0.042 suggests that the acceleration coefficient is statistically significant, with a p-value less than 0.05. The exponentiated value of the acceleration coefficient is 1.003, which means that for every one-unit increase in the acceleration event, the expected traffic crash count increases by a factor of 1.003, or the log count of traffic crash shows a 0.3% increment for a one-unit increase in the acceleration event.

- ✓ **Speeding event**

The coefficient for speed is 0.009 with a significant value of 0.000 which is less than 0.05, indicating that it is statistically significant at a 95% confidence level. The exponentiated value of the speed coefficient is 1.009, meaning that for every one-unit increase in the speeding event, the expected traffic crash count increases by 1.009, or approximately 0.9%. Thus, a one-unit increase in the speeding event is associated with an approximate 0.9% increase in the expected log count of traffic crash, holding all other variables constant.

- ✓ **Tailgating event**

The coefficient for tailgating is 0.019 with a significant value of 0.000 which is less than 0.05. The coefficient is significant at 95% confidence. The exponentiated value of the tailgating coefficient is 1.019, which means that for every one-unit increase in the tailgating event, the expected traffic crash count increases by a factor of 1.019, or approximately 1.9% increase in the expected log count of traffic crash.

- ✓ **Steering and Deceleration event**

As indicated above in the model effects, the deceleration and steering events are not significant and similarly the coefficients in parameter estimate also not significant. Thus, deceleration and steering events are excluded in the model and the following Negative Binomial Modeling is developed for the study area.

$Expected\ Count\ of\ Crash\ = \exp(-0.732 + 0.003 * Accelartion + 0.009 * Speed\ + 0.019 * Tailgating)$…….Eqn-4.3

Based on the result, it is evident that acceleration, speed, and tailgating have a positive and significant effect on the expected count of crashes. This implies that frequent acceleration, speed, and tailgating events are associated with an increased likelihood of traffic crashes.

As cognizant that, the above model is non spatial model which doesn't consider the spatial effects of historical traffic crash. As indicated above, historical traffic crashes show spatial autocorrelation which means the traffic crashes affected by neighborhood effect and during modeling it's necessary to consider this spatial effect through spatial lag or spatial error models. Owing to this, the study considers this spatial effect by incorporating spatial lag model in the above negative binomial model. The spatial lag model considers the lag count of traffic crash by conducting spatial weight matrixes through queen contiguity method. This spatial lag model utilizes the following formulation.

Spatial Lag model $E[Y \mid X] = \exp(\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \rho WY)$ …………..Eqn-4.4

Whereas: Y is the expected number of traffic crashes, β0 is the intercept, β1, β2, and β3 are the coefficients associated with each risky driving event and X_1, X_2 and X_k are the predictor variables, ρ is the spatial autoregressive coefficient, W is the spatial weights matrix and Y is traffic crash count. ρWY is s the spatial lag term, representing the weighted sum of traffic crash in neighboring locations.

Table 4.10: Parmeter Estimate Spatial Lag Model

			95% Wald Confidence Interval		Hypothesis Test				95% Wald Confidence Interval for Exp(B)	
Parameter	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.	Exp(B)	Lower	Upper
(Intercept)	-1.218	0.0090	-1.235	-1.200	18432.569	1	0.000	0.296	0.291	0.301
Accel	0.001	0.0006	4.8E-05	0.003	4.144	1	0.042	1.001	1.000	1.003
Decel	-0.004	0.0032	-0.010	0.003	1.193	1	0.275	0.996	0.990	1.003
Speed	0.005	0.0009	0.003	0.007	33.180	1	0.000	1.005	1.003	1.007
Steer	0.000	0.0003	-0.001	0.000	2.414	1	0.120	1.000	0.999	1.000
Tailg	0.011	0.0004	0.010	0.012	738.068	1	0.000	1.011	1.010	1.012
lg_Cr_C	0.540	0.0054	0.530	0.551	10079.666	1	0.000	1.717	1.699	1.735
(Scale)	1[a]									
(Negative binomial)	.331[b]	0.0222	0.291	0.378						

Dependent Variable: Crash_Coun
Model: (Intercept), Accel, Decel, Speed, Steer, Tailg, lg_Cr_C (lag of crash count)
a. Fixed at the displayed value.

b. Hessian matrix singularity is caused by the scale or negative binomial parameter.

Likelihood ratio -48796, AIC 97609 and BIC 97680. The likelihood ratio, AIC and BIC values of the spatial lag model are lower than the non-spatial model which entails the spatial lag model better fit and considered the spatial effects in the model. The parameter estimates of spatial lag model shows that the lag count of traffic crashes has a positive significant effect on the expected traffic crash in addition to the Acceleration, Speed and Tailgating event.

Table 4.11: Summary of Goodness of fit and performance evaluation

	Non-Spatial NB	**Spatial Lag NB**
AIC	102366.759	97609
BIC	102366.761	97680
Pseudo R-squared	0.0353	0.2031
MPB	0.145	0.2291
MAD	0.867	0.8466

Based on AIC and BIC values, the spatial lag Negative Binomial (NB) model performed better than the non-spatial model. Pseudo R-squared value also favors the spatial lag NB model. Mean Prediction Bias (MPB) measures the magnitude and direction of bias in the estimates of developed model. Mean Absolute

Deviation (MAD) also measures the average magnitude of errors between predicted and actual values. The Non spatial NB model has lower MPB and higher MAD value compared to Spatial lag Negative Binomial Model.

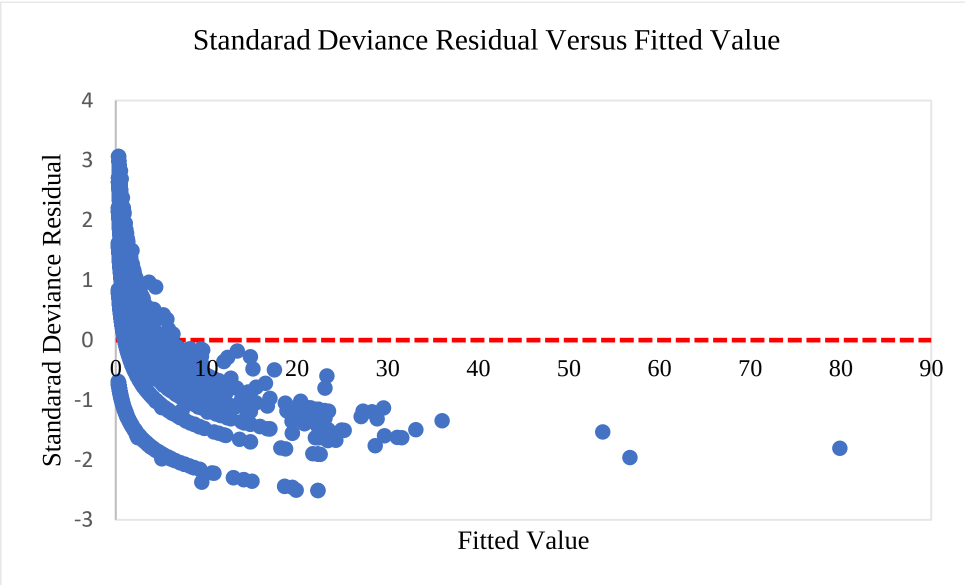

Figure 4.18: Fitted value and deviance residual plot of spatial lag NB model

The findings suggest that drivers who exceed the speed limit and engage in aggressive acceleration behaviors are more likely to be involved in a crash. These behaviors make it more difficult for drivers to control their vehicles, increasing the chances of an accident. Additionally, the high frequency of tailgating indicates that drivers in the study area tend to closely follow other vehicles, leaving less time to react to sudden events and reducing the safe following distance. In addition, the spillover effect of traffic crash contributes significant positive effect on the expected count of traffic crashes.

It is important for policymakers and road safety institutions in the study area to consider interventions targeting these aggressive fast-driving and close-following behaviors. Implementing measures to address these risky driving behaviors, such as increased enforcement, public awareness campaigns, and educational programs, can reduce the number of crashes and improve overall road safety in the region.

Furthermore, this study's results can inform the development of predictive models for expected crash rates, which can be utilized by future autonomous vehicles operating in the region. By incorporating this model into autonomous vehicle systems, proactive measures can be taken to identify high-risk locations and potentially prevent crashes by adjusting driving behaviors and routes when necessary.

Overall, this study gives valuable insights into the common driving behavior in the study area and highlights the aggressive driving practices in the region. The findings contribute to the body of research on driving behavior and can inform evidence-based interventions and policies to reduce crashes and promote safer road environments.

4.3 Risk Map of Expected Crashes

On top of the above predictive analysis, this study's main goal is to produce a usable map that shows the risk map of expected crashes from the above models. Accordingly, the risk map of the study area developed through ARC GIS software after generating the expected number of traffic crashes based on the above predictive models. The risk map developed by predicting the number of traffic crashes from each strip and based on the strip ID'S then the spatial data plotted on the study area and the following maps are developed. The risk map was developed for both non spatial and spatial models.

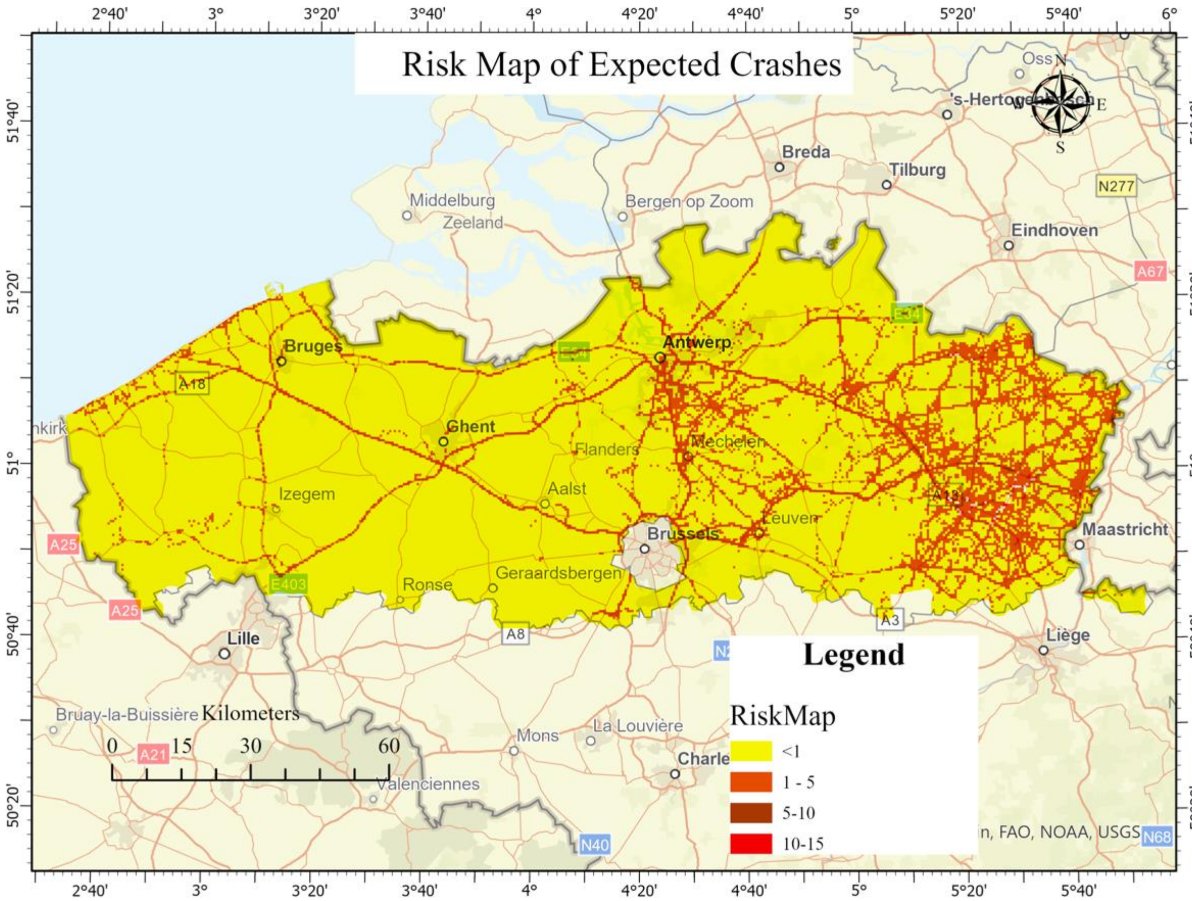

Figure 4.19: Risk Map of Expected Traffic Crash from Non-Spatial Model

The risk map in figure 4.19 serve as a comprehensive visualization of the anticipated future traffic crash derived from the predictive negative binomial model. This map delineates the areas of heightened risk of traffic crash.

Notably, the map reveals substantial concentration of expected traffic crashes in the vicinity of Limburg Province. This concentration is attributed to the occurrence of frequent risky events in this province of the study region. The elevated risk in Limburg Province underscores the urgency of targeted intervention to address and mitigate the factors contributed to the increased likelihood of traffic crash to this area.

Further, the risk map identifies motorways and surrounding of major cites as risky section. This designation is as a result of frequent occurrence of risky events (acceleration, speed and tailgating events) in these areas.

The proximity of these high risky sections to major urban centers and motorways underscores the implementation of strategic traffic safety measures and intervention in these areas.

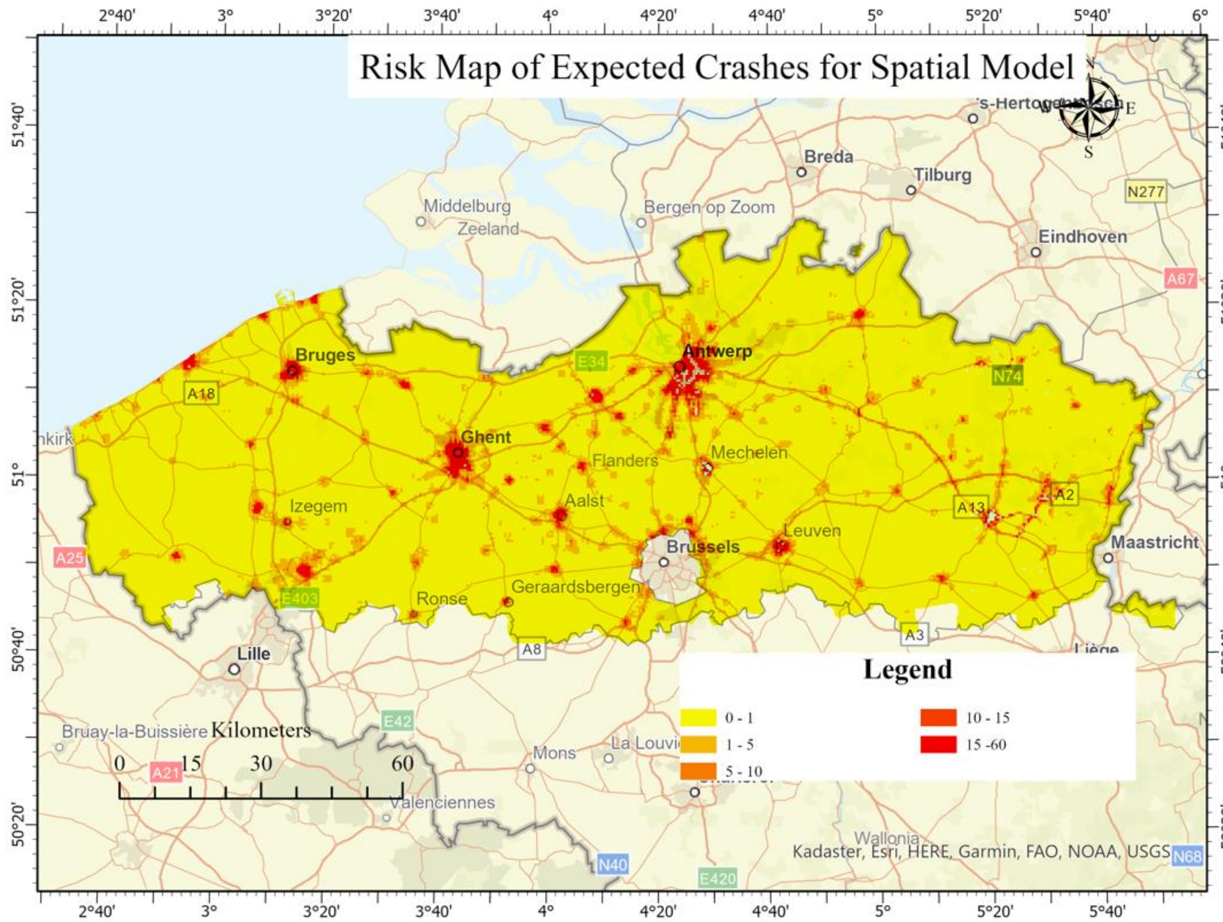

Figure 4.20: Risk Map of expected crashes for spatial model

The risk map of spatial model considers the spatial effect of historical traffic crash in addition to the risky event. Consideration of spatial effect in the spatial model is essential to reduce the spatial errors. The risk map of spatial model corroborates findings from historical traffic crashes. These historically identified hot spot sections align with flagged as risky section in the current risk map from predictive model. These risk map provide a consistent and reinforcing validation of areas requiring immediate intervention and attention. Recognizing these historical hot spots as current risk areas in the above predictive model emphasizes the persistent nature of traffic safety challenge in these areas.

In general, the risk map not only pinpoint areas of elevated future traffic crashes but also reveals the spatial distribution of this risks, enabling targeted and proactive intervention. The integration of spatial effect of historical traffic crash further strengthens the reliability of identified risk areas and guiding effective strategies for enhancing overall road safety within the study region.

The developed risk map is an indicative to policy makers to take action on aggressive fast driving and close following drivers on the indicated locations of the study area.

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

The study mainly aims to assess the risky driving events at crash proxy locations through spatial and statistical analysis. The study confirmed the significant risky driving events at crash spot through spatial and predictive analysis in the study area. Finally, the study conquered that Negative Binomial Modeling can model traffic crash count with significant predictors and the following concluding remarks are investigated in this study;

- ✓ The bivariate and spatial distribution mapping are an indicative of spatial relationship of traffic crashes and risky driving events in the study area. High frequency of crashes and risky driving events occurred around Hasselt, Genk, Mechelen, Antwerp Cities and along Motorways. The local bivariate relationship shows a positive linear relationship between traffic crashes and vehicle control events in many locations within the study area.

- ✓ The Multivariate Clustering was developed to figure out the locations with high incident rate of crashes and risky driving events. The clustering reveals that highly frequent risky driving events and crashes are in the Limburg province of Belgium. This occurred due to the fact that most of the naturalistic driving data was collected in Limburg province. This clustering is crucial for the Limburg province to implement interventions in areas where crashes and risky driving events are frequent.

- ✓ Chi-Squared test of association was established to analyze the relationship between traffic crashes and risky driving events. The test aimed to evaluate both the null and alternative hypothesis. Consequently, a Chi-Squared test of association was carried out for each risky driving event with traffic crash, revealing a noteworthy association between all risky driving events and traffic crashes, as indicated a significance level of $P < 0.00$ and higher Chi-squared value.

- ✓ A correlation test was performed between traffic crashes and risky driving events using Spearman's Rank Correlation test to explore the strength and direction of the relationship. The results of correlation test indicate a significant positive correlation between traffic crashes with the risky driving events. This suggests that an increase in risky events is associated with heightened risk of traffic crashes.

- ✓ Negative Binomial Modeling was conducted for the study to develop a model that predicts the expected traffic crash count from risky driving events. The Negative Binomial Model was performed after examining all the assumptions of Negative Binomial Regression i.e discrete data, overdispersion, multicollinearity test, no significant outlier and Homoscedasticity test. Negative Binomial model developed by using acceleration, deceleration, speeding, steering and tailgating event as predictor and traffic crash count as response variable. Accordingly, the model specifies that acceleration, speeding, and tailgating events have a significant and positive relation with the expected traffic crash. However, steering and deceleration events are insignificant to the model to predict the expected traffic crash. The model suggests that a frequent occurrence of medium to high acceleration, speed and tailgating event is associated with an increased likelihood of traffic crashes. The model reveals that Drivers who exceed speed limits and engage in excessively fast driving are at a heightened risk of being involved in traffic accidents. Into consideration of spatial effect of traffic crash, the spatial lag negative binomial model developed by using lag count traffic crash as predictor in addition to vehicle control events. The spatial lag model also specifies acceleration, speeding, tailgating events and lag count of crashes have a significant and positive relation with the expected traffic crash. The spatial lag model reveals that traffic crash spillover effect shows significant effect on the expected traffic crash, in addition to the impact of vehicle control events. Moreover, the spatial lag model demonstrates lower AIC and BIC value compared to the non-spatial model.

- ✓ Additionally, the models reveal that when drivers operate their vehicles at high speeds and exhibit aggressive acceleration behavior, they may experience reduced control over their vehicles, consequently elevating the risk of accidents. Moreover, a high frequency of tailgating events suggests that drivers tend to closely follow other vehicles, leading to diminished safe following distances and less reaction time for unforeseen events.

- ✓ In practical terms, the prevalent driving behavior in the study area leans towards aggressive, high-speed driving with a tendency to follow other vehicles closely. This driving behavior is the most frequently observed pattern in the region. Therefore, it is imperative for policy makers and road safety institutions in the study area to consider implementing interventions aimed at mitigating aggressive driving, excessive speeding, and close-following behaviors.

- ✓ The study aligns with existing research conducted in the region, which consistently highlights aggressive driving as a common behavior among commuters in the Flanders region. This research reinforces the prevalence of such behavior through quantitative and spatial analysis. Furthermore, the findings can be applied in the context of future autonomous vehicles operating in the region, enabling them to predict the expected crash rates and proactively take remedial measures in locations exhibiting frequent risky driving events.

5.2 Recommendations

On the top of this research findings, assessing risky driving events at historical crash spot locations is a unique type of research which fills the gap of lack of crash data in naturalistic driving study. Hence, this type of study is essential to recognize the possible expected traffic crash from risky driving events and to figure out the significant risky driving events in the study area through quantitative and spatial analysis. Since traffic crashes have many complex factors, the researcher of this study recommends the practical use of the study result for policy makers and to make required intervention in the study area. Further, the study only considered the risky driving events occurred at crash spot locations in addition to this the road condition as well as the environment need to be considered. In line to this study, the following practical recommendation are drawn to the study area: -

- ➢ **Targeted Interventions in High-Risk Areas**: Implement targeted interventions, such as increased law enforcement presence, road signage improvements, and public awareness campaigns, in the identified high-risk areas, particularly around major cities and along motorways.

- **Aggressive Driving Education**: Develop and implement programs to mitigate aggressive driving behaviors. This could include awareness campaigns, defensive driving courses, and outreach initiatives to promote safer driving practices. Promote driver education programs that focus on the risks associated with aggressive acceleration.

- **Speed Limit Enforcement**: Strengthen law enforcement efforts to monitor and enforce speed limits on roadways, especially in areas where the study identifies a significant positive relationship between speeding events and the expected traffic crash. This may involve increased use of speed cameras, implement automated speed enforcement systems to deter speeding, and launch public awareness campaigns to educate drivers about the dangers of excessive speeding.

- **Tailgating Awareness Campaigns**: Promote safe following distances and emphasize the importance of maintaining a proper distance between vehicles to reduce the risk of rear-end collisions.

- **Infrastructure Improvements**: Evaluate and enhance road infrastructure, especially in high-risk areas, to address factors contributing to risky driving events. This could include improvements in road design, traffic flow management, and the installation of safety features.

- **Technology Integration**: Technology can be leveraged to detect risky driving behaviors and provide immediate feedback to drivers in real time. Promote use of advanced driver assistance systems (ADAS) that can provide warnings about following vehicles too closely. Explore the integration of autonomous vehicles that can leverage the developed model to predict crash rates.

- **Collaboration with Local Authorities**: Collaborate with local authorities and traffic management agencies to implement and monitor the effectiveness of interventions. Regular communication and coordination with local stakeholders are essential for successful road safety initiatives.

- **Continued Research and Monitoring**: Support ongoing research and monitoring efforts to continually assess the effectiveness of implemented interventions. Regular updates and adjustments to strategies based on new data and findings will contribute to sustained road safety improvements.

- **Policy Development and Enforcement**: Develop and enforce policies that specifically target behaviors identified in the study, such as aggressive acceleration, speeding and tailgating. Regularly update policies to address emerging driving behaviors and technology trends.

These recommendations are intended to address the identified risky driving behaviors and improve road safety in the Flanders region. By taking a multifaceted approach, involving policy development, public awareness, and technological advancements, the region can work toward reducing traffic crashes, ultimately making its roadways safer for all travelers.

5.3 Limitations and Future Research

To make the research manageable, this study is limited in scope, time and coverage areas and has the following limitations.

- ✓ Data collection: the study acknowledged limited data collection coverage area, and the collected data covers only the Flanders region of Belgium. This limited geographical coverage affect the generalizability of the findings. In addition, any errors, missing data, or biases in the data collection may influence the study results.

- ✓ External Factors: the study primarily focused only the risky driving events. However, traffic crashes have many contributing factors related to weather conditions, road, vehicle, and environmental factors. This limitation affects the study holistic understanding of traffic crashes in the study area.
- ✓ Temporal Scope: The study lacks a temporal dimension, which could be crucial for understanding how driving behavior and crash patterns evolve. A more dynamic analysis could provide insights into the temporal dimension of risky driving behavior and historical crash patterns.
- ✓ Model Limitations: To improve the model's predictive power, it's necessary to consider and assess other machine learning models.
- ✓ Seasonal Variation: Seasonal variations include weather conditions and changes in traffic patterns, which can influence driving behaviors and crash rates. The study limited to explore the impact of seasonal variation on risky driving events and their association with traffic crash.
- ✓ Socio-Demographic Factors: the study doesn't consider socio-demographic factors like age, experience and gender differences among different drivers. Factors like age, experience, gender and personal traits could influence how drivers respond to certain events.

The study's findings need to be interpreted based on the above limitations, and the following future research areas are recommended.

- ✓ Comprehensive Geographic Coverage: Further studies should aim comprehensive geographic coverage to enhance the generalizability of the findings across different regions of Europe and Belgium. This allows researchers to assess the universality of the relationship between risky driving and traffic crashes.
- ✓ Longitudinal Analysis: Incorporating a temporal dimension in the analysis would enable researchers to track driving behavior changes and crash patterns over time.
- ✓ Behavioral Studies: conducting in-depth behavioral studies to understand the underlying reasons for aggressive driving, speeding, and tailgating could inform targeted intervention. Exploring psychological and sociological factors contributing to these behaviors may enhance the effectiveness of the proposed intervention.
- ✓ Comparative Analysis: Comparing the findings with similar studies conducted in other regions or countries offer valuable insight into cultural and environmental influencing driving behavior. Cross-comparative study with other regions also helps to assess the universality of the relationship between risky driving and crash.
- ✓ Influence of seasonal variation: To address the effect of weather condition and traffic pattern variations, future studies need to consider seasonal variation on risky driving behavior and traffic crashes.
- ✓ Influence of Road Infrastructure and Traffic Flow dynamics: Road infrastructure and traffic flow dynamics like traffic congestion and road geometry parameters affect traffic crashes and driving behavior. To create a holistic understanding of traffic crash patterns and risky driving behavior, it is necessary to consider the impact of road infrastructure and traffic flow.
- ✓ Integration with Autonomous Vehicles: As mentioned in the conclusion, exploring how autonomous vehicles can utilize the findings for proactive measures is a fascinating avenue. Future research could delve into developing practical applications and strategies for incorporating these findings into autonomous vehicle systems for enhanced safety.

REFERENCES

Abd Rahman, N. I., Dawal, S. Z. M., & Yusoff, N. (2020). Driving mental workload and performance of ageing drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, *69*, 265–285.

Adminaité-Fodor, D., Stipdonk, H., Thomas, P., Yannis, G., Avenoso, A., Jost, G., Carson, J., & ETSC. (2021). Ranking EU progress on road safety. June. www.etsc.eu/pin

Ahmed, M. M., & Abdel-Aty, M. (2015). Evaluation and spatial analysis of automated red-light running enforcement cameras. *Transportation Research Part C: Emerging Technologies*, *50*, 130–140. https://doi.org/10.1016/j.trc.2014.07.012

Alam, M. S., & Tabassum, N. J. (2023). Spatial pattern identification and crash severity analysis of road traffic crash hot spots in Ohio. *Heliyon*, *9*(5), e16303. https://doi.org/10.1016/j.heliyon.2023.e16303

Alkaabi, K. (2023). Identification of hotspot areas for traffic accidents and analyzing drivers' behaviors and road accidents. *Transportation Research Interdisciplinary Perspectives*, *22*(August), 100929. https://doi.org/10.1016/j.trip.2023.100929

Alrassy, P., Smyth, A. W., & Jang, J. (2023). Driver behavior indices from large-scale fleet telematics data as surrogate safety measures. *Accident Analysis & Prevention*, *179*, 106879.

Arvin, R., Kamrani, M., & Khattak, A. J. (2019). The role of pre-crash driving instability in contributing to crash intensity using naturalistic driving data. *Accident Analysis and Prevention*, *132*(April), 105226. https://doi.org/10.1016/j.aap.2019.07.002

Bagdadi, O. (2013). Assessing safety critical braking events in naturalistic driving studies. *Transportation Research Part F: Traffic Psychology and Behaviour*, *16*, 117–126. https://doi.org/10.1016/j.trf.2012.08.006

Becker, N., Rust, H. W., & Ulbrich, U. (2022). Weather impacts on various types of road crashes: a quantitative analysis using generalized additive models. European transport research review, 14(1), 1-18.

Cai, M., Yazdi, M. A. A., Mehdizadeh, A., Hu, Q., Vinel, A., Davis, K., Xian, H., Megahed, F. M., & Rigdon, S. E. (2021). The association between crashes and safety-critical events: Synthesized evidence from crash reports and naturalistic driving data among commercial truck drivers. *Transportation Research Part C: Emerging Technologies*, *126*(October 2019), 103016. https://doi.org/10.1016/j.trc.2021.103016

Carter, C. J, May, A, Smith,F, & Fairclough,S. (1995). An evaluation of an in-vehicle headway feedback system. *Contemporary Ergonomics*, 287-287.

Charly, A., & Mathew, T. V. (2023). Identifying risky driving behavior: a field study using instrumented vehicles. *Transportation Letters*, *00*(00), 1–15. https://doi.org/10.1080/19427867.2023.2233782

Chen, S., Kuhn, M., Prettner, K., & Bloom, D. E. (2019). The global macroeconomic burden of road injuries: estimates and projections for 166 countries. *The Lancet Planetary Health*, *3*(9), e390–e398. https://doi.org/10.1016/S2542-5196(19)30170-6

Cheng, Z., Zu, Z., & Lu, J. (2019). Traffic crash evolution characteristic analysis and spatiotemporal hotspot identification of urban road intersections. *Sustainability (Switzerland)*, *11*(1). https://doi.org/10.3390/su11010160

Daniels, S., Martensen, H., Schoeters, A., Van den Berghe, W., Papadimitriou, E., Ziakopoulos, A., Kaiser, S., Aigner-Breuss, E., Soteropoulos, A., Wijnen, W., Weijermars, W., Carnis, L., Elvik, R., & Perez, O. M. (2019). A systematic cost-benefit analysis of 29 road safety measures. *Accident Analysis and Prevention*, *133*(October), 105292. https://doi.org/10.1016/j.aap.2019.105292

Dozza, M. (2020). What is the relation between crashes from crash databases and near crashes from naturalistic data? Journal of Transportation Safety and Security, 12(1), 37–51. https://doi.org/10.1080/19439962.2019.1591553

Elvik, R. (2008). Dimensions of road safety problems and their measurement. *Accident Analysis and Prevention*, *40*(3), 1200–1210. https://doi.org/10.1016/j.aap.2008.01.004

Erdogan, S., Ilçi, V., Soysal, O. M., & Kormaz, A. (2015). a Model Suggestion for the Determination of the Traffic Accident Hotspots on the Turkish Highway Road Network: a Pilot Study. *Boletim de Ciências Geodésicas*, *21*(1), 169–188. https://doi.org/10.1590/s1982-21702015000100011

European Commission. (2020). Next steps towards Vision Zero.

ETSC. (2022). *Road Death in European Union - Latest Data*. Retrieved from https://etsc.eu/euroadsafetydata/

Flahaut, B., Mouchart, M., San Martin, E., & Thomas, I. (2003). The local spatial autocorrelation and the kernel method for identifying black zones: A comparative approach. *Accident Analysis & Prevention*, *35*(6), 991–1004.

Fridulv Sagberg, Rob Eenink (SWOV), M. H. (TNO), & Ruth Welsh (LBORO), M. W. (KfV T. L. (Or Y. N. van N. (SWOV) R. S. (TNO). (2011). Recommendations for a large-scale European naturalistic driving.

Galovski, T. E. (2006). *Road rage: Assessment and treatment of the angry, aggressive driver*. American Psychological Association.

Gitelman, V., Bekhor, S., Doveh, E., Pesahov, F., Carmel, R., & Morik, S. (2018). Exploring relationships between driving events identified by in-vehicle data recorders, infrastructure characteristics and road crashes. *Transportation Research Part C: Emerging Technologies*, *91*(April), 156–175. https://doi.org/10.1016/j.trc.2018.04.003

Gordon, T. J., Kostyniuk, L. P., Green, P. E., Barnes, M. A., Blower, D., Blankespoor, A. D., & Bogard, S. E. (2011). Analysis of crash rates and surrogate events: Unified approach. *Transportation Research Record*, *2237*, 1–9. https://doi.org/10.3141/2237-01

Guo, F., & Fang, Y. (2013). Individual driver risk assessment using naturalistic driving data. *Accident Analysis and Prevention*, *61*, 3–9. https://doi.org/10.1016/j.aap.2012.06.014

Guo, F., Klauer, S. G., Hankey, J. M., & Dingus, T. A. (2010). Near crashes as crash surrogate for naturalistic Driving Studies. *Transportation Research Record*, *2147*, 66–74. https://doi.org/10.3141/2147-09

Guo, F., Wang, X., & Abdel-Aty, M. A. (2010). Modeling signalized intersection safety with corridor-level spatial correlations. *Accident Analysis and Prevention*, *42*(1), 84–92. https://doi.org/10.1016/j.aap.2009.07.005

Haddon, W. (1968). The changing approach to the epidemiology. Prevention, and amelioration of trauma 58(8), 1431–1438.

Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge University Press.

J., O. P., Ernst, L., & Belinda, N. (2010). Detecting Unsafe Roadways with Spatial Statistics: Point Patterns and Geostatistical Models. *Journal of Transportation Engineering*, *136*(5), 457–464. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000048

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, *31*(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

Joint, M. (1997). *Road Rage: AAA Foundation for Traffic Safety*.

Jones, B., Janssen, L., & Mannering, F. (1991). Analysis of the frequency and duration of freeway accidents in Seattle. *Accident Analysis & Prevention*, *23*(4), 239–255.

Joshua, S. C., & Garber, N. J. (1990). Estimating truck accident rate and involvements using linear and Poisson regression models. *Transportation Planning and Technology*, *15*(1), 41–58. https://doi.org/10.1080/03081069008717439

Jovanis, P. P., & Chang, H. L. (1986). Modeling the Relationship of Accidents To Miles Traveled. *Transportation Research Record*, 42–51.

Khattak, M. W., Pirdavani, A., De Winne, P., Brijs, T., & De Backer, H. (2021). Estimation of safety performance functions for urban intersections using various functional forms of the negative binomial regression model and a generalized Poisson regression model. *Accident Analysis & Prevention*, *151*, 105964.

King, A. P., & Eckersley, R. J. (2019). *Chapter 2 - Descriptive Statistics II: Bivariate and Multivariate Statistics* (A. P. King & R. J. B. T.-S. for B. E. and S. Eckersley (eds.); pp. 23–56). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-08-102939-8.00011-6

Kunnawee, K., Ketphat, M., & Jiwattanakulpaisarn, P. (2012). Application of the theory of planned behaviour to predict young drivers' speeding behaviour. *Injury Prevention*, *18*(Suppl 1), A200.1-A200. https://doi.org/10.1136/injuryprev-2012-040590t.5

Lee, J. D., McGehee, D. V., Brown, T. L., & Reyes, M. L. (2002). Collision warning timing, driver distraction, and driver response to imminent rear-end collisions in a high-fidelity driving simulator. *Human Factors*, *44*(2), 314–334. https://doi.org/10.1518/0018720024497844

Lertworawanich, P. (2006). Safe-Following Distances Based on the Car-Following Model. *International Seminar on Intelligent Transport Systems (ITS) in Road Network Operations, 2006, Kuala Lumpur, Malaysia*, 1–19.

Ljung, M. (2002). DREAM: Driving Reliability and Error Analysis Method. Department of Computer and Information Science (IDA).

Lomax, R. G., & Hahs-Vaughn, D. L. (2013). *An introduction to statistical concepts*. Routledge.

Lord, D. (2006). Modeling motor vehicle crashes using Poisson-gamma models: Examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. *Accident Analysis & Prevention, 38*(4), 751–766.

Lord, D., & Bonneson, J. A. (2007). Development of accident modification factors for rural frontage road segments in Texas. *Transportation Research Record*, *2023*(1), 20–27.

Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*, *44*(5), 291–305.

Lord, D., & Park, P. Y.-J. (2008). Investigating the effects of the fixed and varying dispersion parameters of Poisson-gamma models on empirical Bayes estimates. *Accident Analysis & Prevention, 40*(4), 1441–1457.

Lord, D., Washington, S. P., & Ivan, J. N. (2005). Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis & Prevention*, *37*(1), 35–46.

Mchugh, M. L. (2013). The Chi-square test of independence Lessons in biostatistics. *Biochemia Medica*, *23*(2), 143–149. http://dx.doi.org/10.11613/BM.2013.018

Miaou, S.-P. (1994). The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. *Accident Analysis & Prevention*, *26*(4), 471–482.

Miaou, S.-P., & Lord, D. (2003). Modeling traffic crash-flow relationships for intersections: dispersion parameter, functional form, and Bayes versus empirical Bayes methods. *Transportation Research Record*, *1840*(1), 31–40.

Mitra, S. (2009). Spatial autocorrelation and Bayesian spatial statistical method for analyzing intersections Prone to injury crashes. *Transportation Research Record*, *2136*, 92–100. https://doi.org/10.3141/2136-11

Mohaymany, A. S., Shahri, M., & Mirbagheri, B. (2013). GIS-based method for detecting high-crash-risk road segments using network kernel density estimation. *Geo-Spatial Information Science*, *16*(2), 113–119. https://doi.org/10.1080/10095020.2013.766396

Moons, E., Brijs, T., & Wets, G. (2009). Identifying hazardous road locations: Hot spots versus hot zones. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5730 LNCS*, 288–300. https://doi.org/10.1007/978-3-642-10649-1_17

Moran, P. A. P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*, *37*(1/2), 17–23. https://doi.org/10.2307/2332142

National Highway Traffic Safety Administration. (2015). Traffic safety facts, 2012 data: pedestrians. *Annals of Emergency Medicine, 65(4), 452*.

NHTSA. (2015). *Traffic safety facts research note: overview of motor vehicle crashes in 2015.* Retrieved from National Highway Traffic Safety Administration: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812348

NHTSA. (2020). *Examination of the Traffic Safety Environment.* : United States. Department of Transportation. National Highway Traffic Safety Administration. Office of Behavioral Safety Research

Pande, A., Chand, S., Saxena, N., Dixit, V., Loy, J., Wolshon, B., & Kent, J. D. (2017). A preliminary investigation of the relationships between historical crash and naturalistic driving. *Accident Analysis and Prevention*, *101*, 107–116. https://doi.org/10.1016/j.aap.2017.01.023

POLIRURAL, D. I. (2022). *Flanders Belgium*. Retrieved from https://hub.polirural.eu/flanders-belgium

Rahman, M. M., Teo, S. H., & Gan, L. M. (2016). Speeding driving behaviour: Age and gender experimental analysis. *MATEC Web of Conferences*, *74*, 5–8. https://doi.org/10.1051/matecconf/20167400030

Reason J. (1997). Managing the Risks of Organizational Accidents. Aldershot, Hants, Ashgate.

Schagen, V., Agberg, I., & Fridulv. (2012). The Potential Benefits of Naturalistic Driving for Road Safety Research: Theoretical and Empirical Considerations and Challenges for the Future. Procedia - Social and Behavioral Sciences, 48, 692–701. https://doi.org/10.1016/j.sbspro.2012.06.1047

Schagen, V., Ingrid, W., Backer, R., Grondhal, Agatha Hoedemaeker, M. L. T., Andrew, M., Fridulv, S., & Martin, W. (2011). Towards a large-scale European Naturalistic Driving study: final report of PROLOGUE (Issue July). https://dspace.lboro.ac.uk/dspace-jspui/handle/2134/9341

Selvi, H. Z., & Caglar, B. (2018). Using cluster analysis methods for multivariate mapping of traffic accidents. *Open Geosciences*, *10*(1), 772–781. https://doi.org/10.1515/geo-2018-0060

Shaon, M. R. R. (2019). Identifying and incorporating driver behavior variables into crash prediction models. Dissertation Abstracts International: Section B: The Sciences and Engineering, 80(11-).

Soltani, A., & Askari, S. (2017). Exploring spatial autocorrelation of traffic crashes based on severity. *Injury*, *48*(3), 637–647. https://doi.org/10.1016/j.injury.2017.01.032

Talbot, R., Fagerlind, H., & Morris, A. (2013). Exploring inattention and distraction in the SafetyNet Accident Causation Database. Accident Analysis and Prevention, 60, 445–455. https://doi.org/10.1016/j.aap.2012.03.031

Talbot, R., Hancox, G., Pilkington, F. C., Filtness, A., Brijs, K., Brijs, T., Ross, V., Katrakazas, C., Haddad, C. al, Taveria, R., Kaiser, S., & Mateus, T. (2020). Framework for operational design of experimental work in i -DREAMS Project identification Document summary. 814761, 1–90.

Tarko, A. P., Saunier, N., & Sayed, T. (2009). Surrogate measures of safety ANB20 (3) Subcommittee on Surrogate Measures of Safety ANB20 Committee on Safety Data Evaluation and Analysis

Van den Berghe, W. (2020). *European Road Safety Observatory, European Commission.* Brussels: European Commission, Directorate General for Transport.

VanVoorhis, C. R. W., & Morgan, B. L. (2007). Understanding power and rules of thumb for determining sample sizes. *Tutorials in Quantitative Methods for Psychology*, *3*(2), 43–50.

VIAS institute. (2022). *Road Safety BAROMETER*. Retrieved from https://www.vias.be/storage/main/barometer-2022-q4-en.html#V4_Differences_between_the_road_users

vrt NWS. (2022, Feburary). *The number of road deaths increased in Flanders in 2021*. Retrieved from https://www.vrt.be/vrtnws/en/2022/02/21/the-number-of-road-deaths-increased-in-flanders-in-2021/

Wagner, E., Atkins, R., Berning, A., Robbins, A., Watson, C., Anderle, J., & Administration, N. H. T. S. (2020). *Examination of the Traffic Safety Environment During the Second Quarter of 2020: Special Report*. *October*, 30p. https://rosap.ntl.bts.gov/view/dot/50940%0Ahttps://trid.trb.org/view/1744054

Wang, J. H. (2011). Assessing drivers' tailgating behavior and the effect of advisory signs in mitigating tailgating. *In Driving Assessment Conference.* University of Iowa.

WHO. (2018). *Global status report on road safety 2018*. Retrieved from https://www.who.int/publications/i/item/9789241565684

WHO. (2021). *Road traffic injuries.* Retrieved from https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

Wu, K. F., Aguero-Valverde, J., & Jovanis, P. P. (2014). Using naturalistic driving data to explore the association between traffic safety-related events and crash risk at driver level. *Accident Analysis and Prevention*, *72*, 210–218. https://doi.org/10.1016/j.aap.2014.07.005

Xu, Y., Bao, S., & Pradhan, A. K. (2021). Modeling drivers' reaction when being tailgated: A Random Forests Method. *Journal of Safety Research*, *78*, 28–35. https://doi.org/10.1016/j.jsr.2021.05.004

APPENDIX A

Local Moran's Test for Traffic Crash

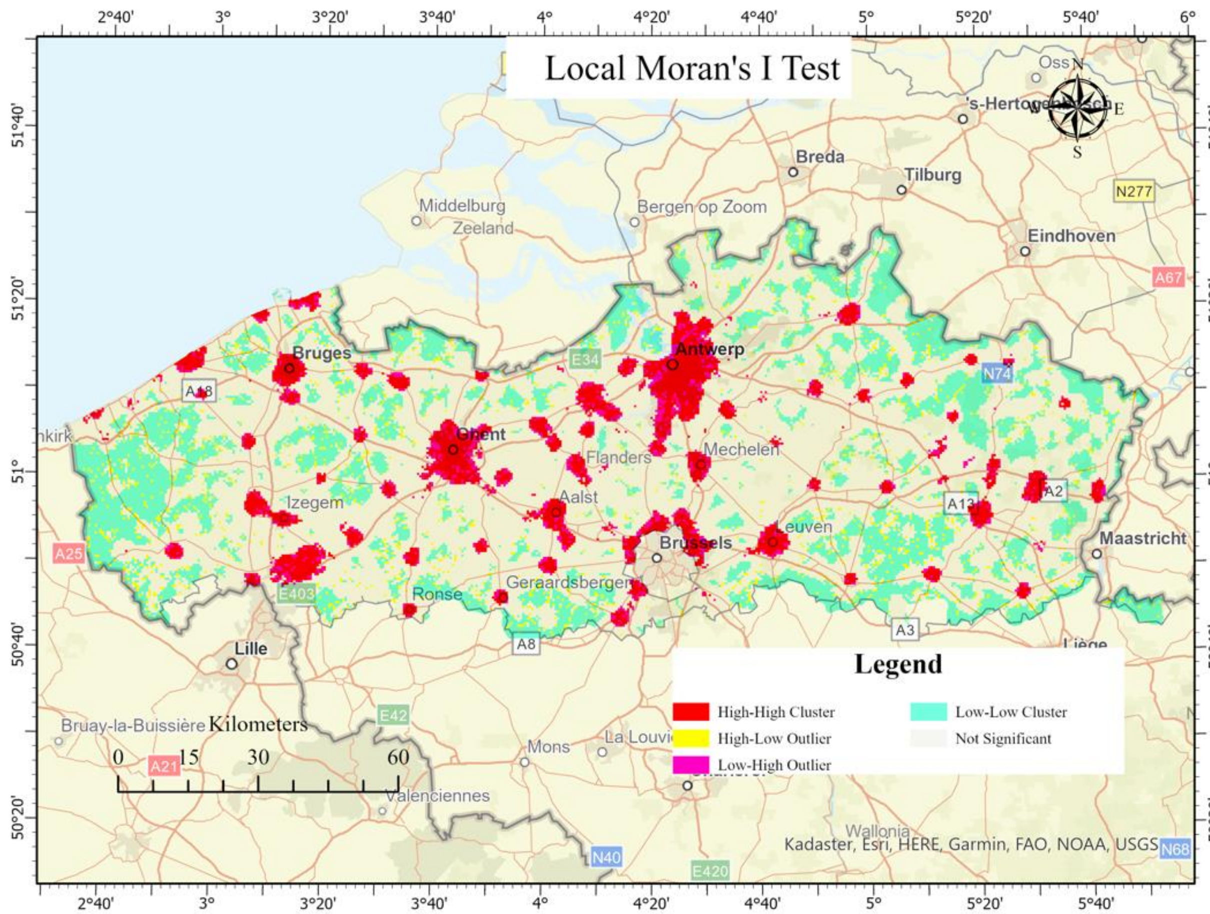

Figure A1: Local Moran's test for Traffic Crash

The local Moran's test for traffic crash indicates High- High clusters in Antwerp, Ghenet, Mechelen, Hassel, Genk, Brugges, Leuven and Aalst and along motorways in the study area. This indicates that in many sections historical traffic crashes show spatial autocorrelation or the traffic crash one location affected by the neighboring historical traffic crash. This spatial autocorrelation phenomena is common on traffic crash due to the fact that traffic crash affected by different spatial effects like the geometric characteristics of the road section (curvature, terrain, the pavement condition, topography, drainage condition, traffic volume, traffic safety measures of the road) and environmental condition of the road i.e weather, snow, rain, and flooding. The geometric parameters and environmental conditions has a spatial association to the nearby locations due to this traffic crash shows spatial autocorrelation.

APPENDIX B

Bivariate Mapping of Traffic crash and Risky Event

The Bivariate mapping developed after conducting the counts of events and crashes through ARC GIS PRO. To perform this mapping, the cutting count of events and crash was decided through consideration of different research and it is to be decided as the location with greater than five crash and five risky driving event considered as high spot locations.

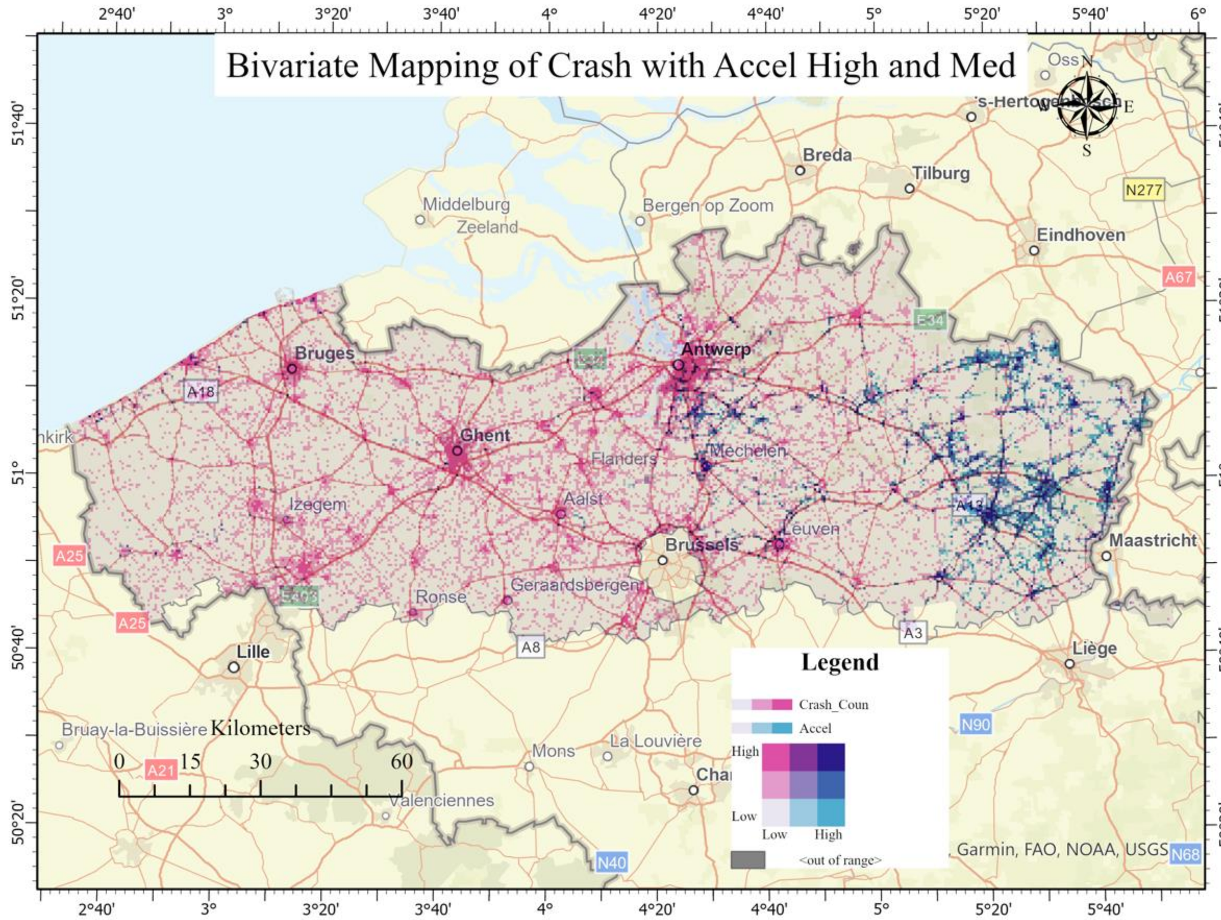

Figure B-1: Bivariate Mapping of Crash with Acceleration High and Medium events

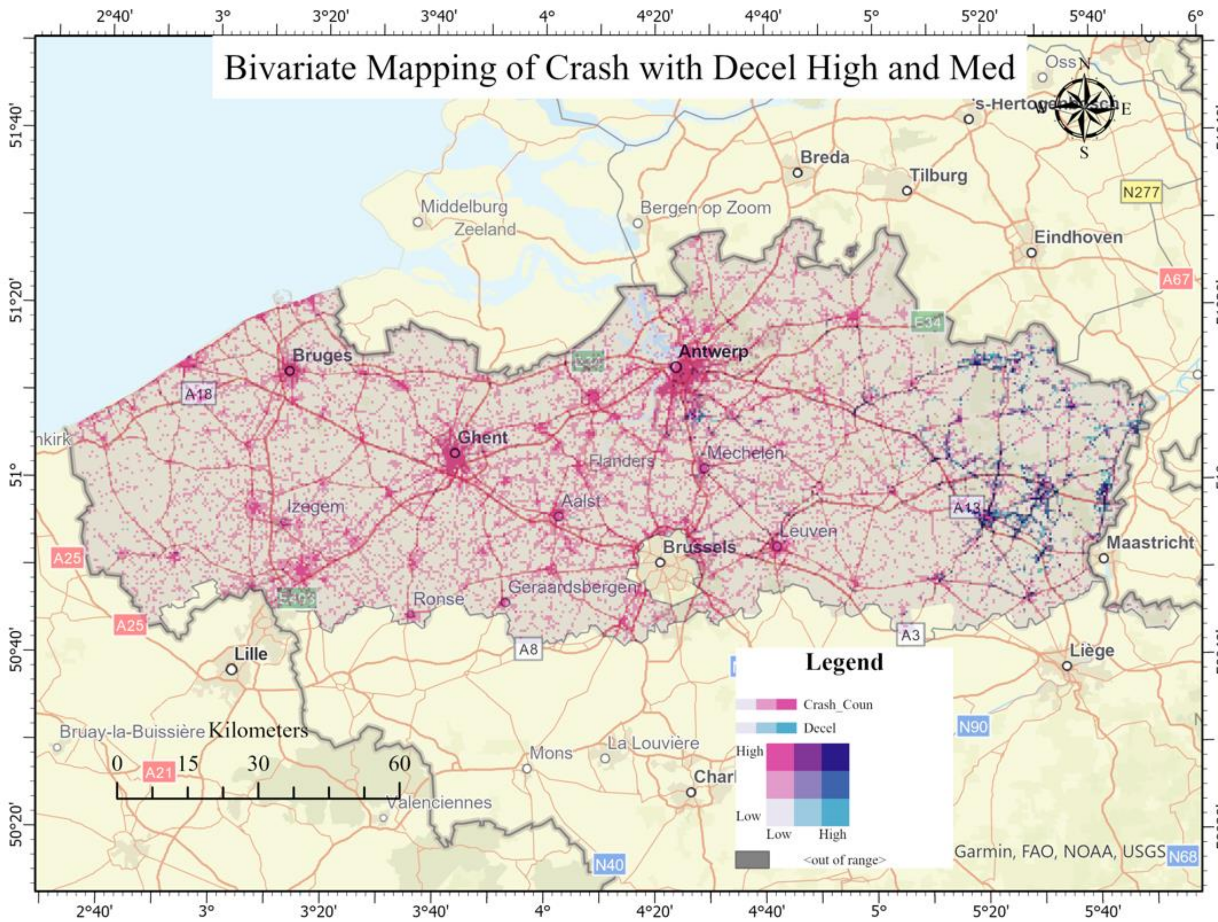

Figure B-2: Bivariate Mapping of Crash with Deceleration High and Medium events

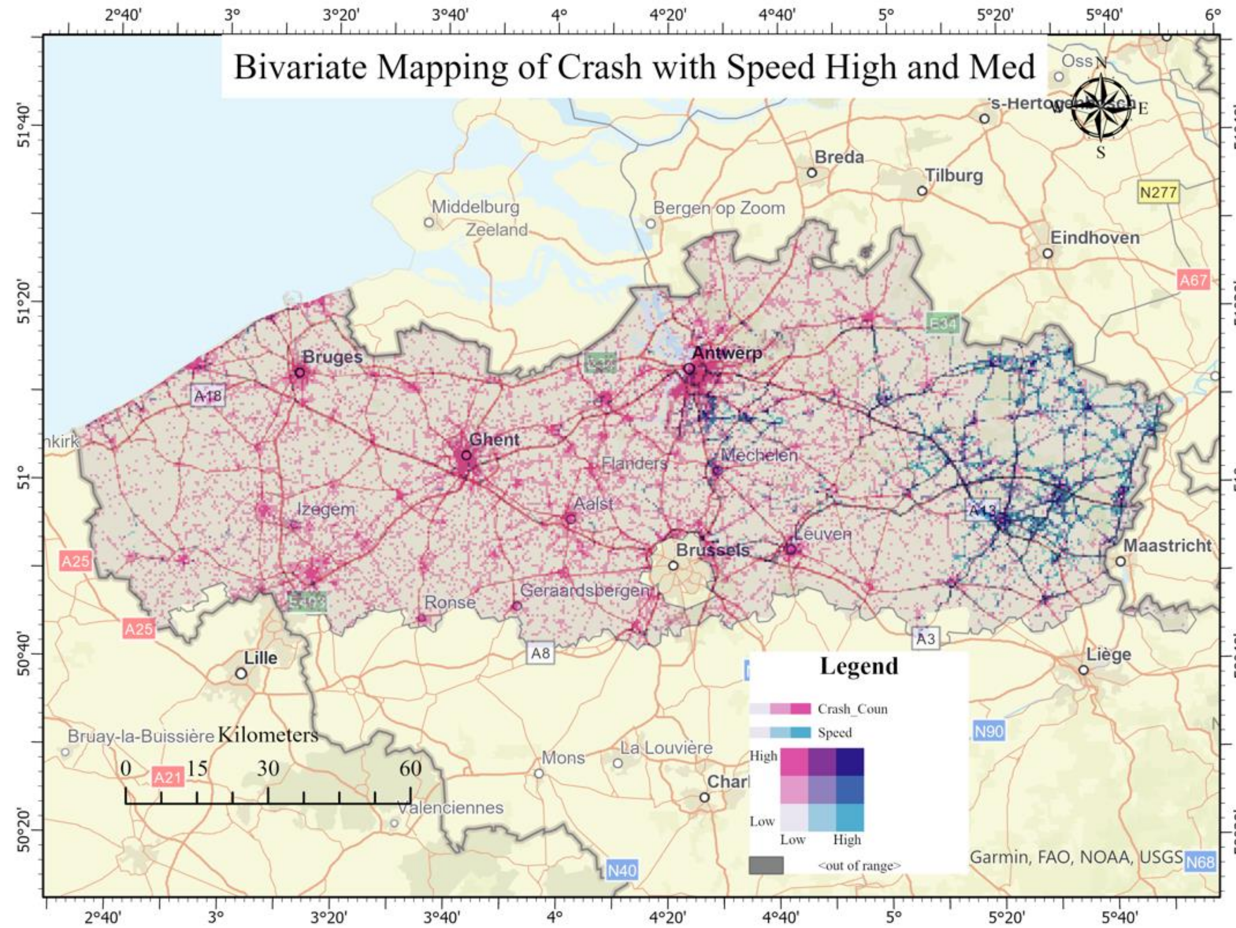

Figure B-3: Bivariate Mapping of Crash with Speed High and Medium events

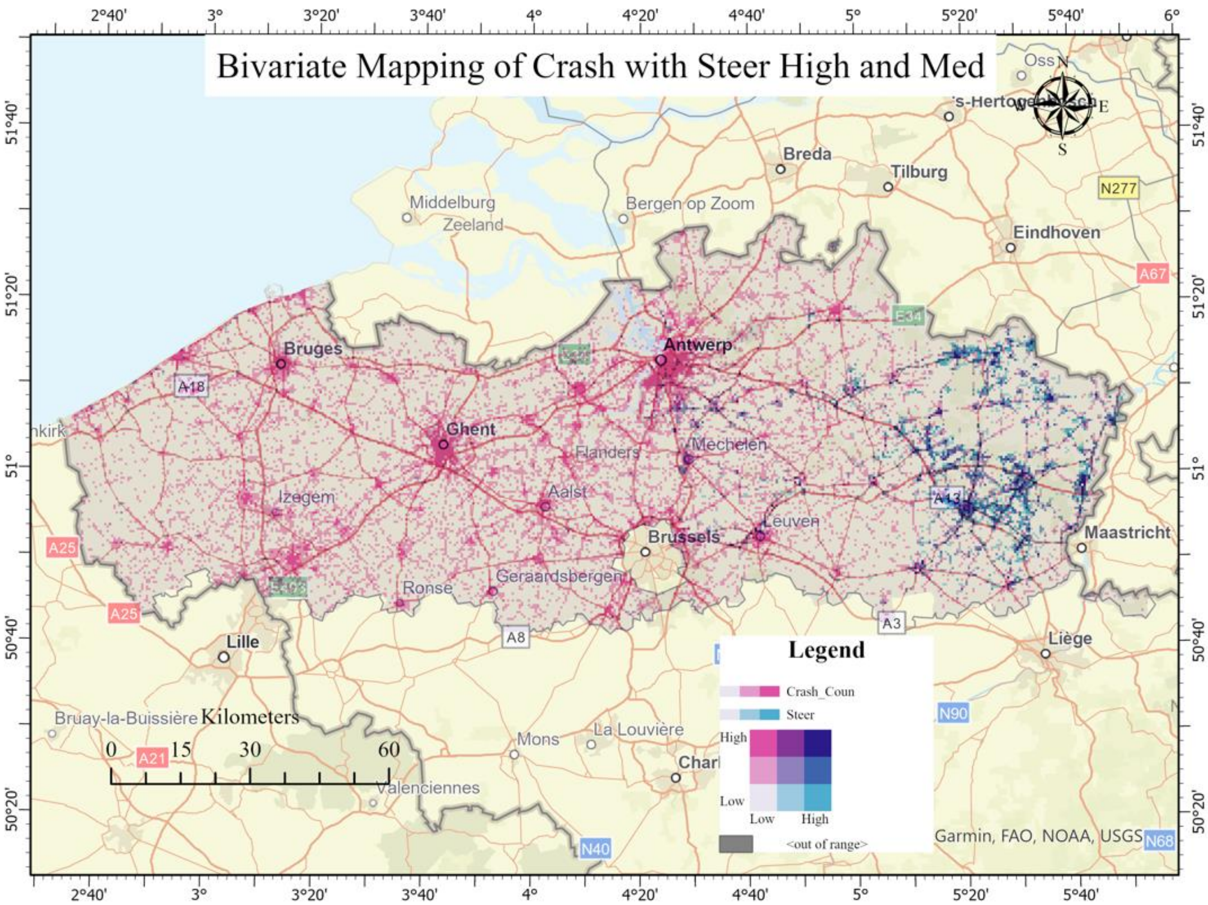

Figure B-4: Bivariate Mapping of Crash with Steer High and Medium events

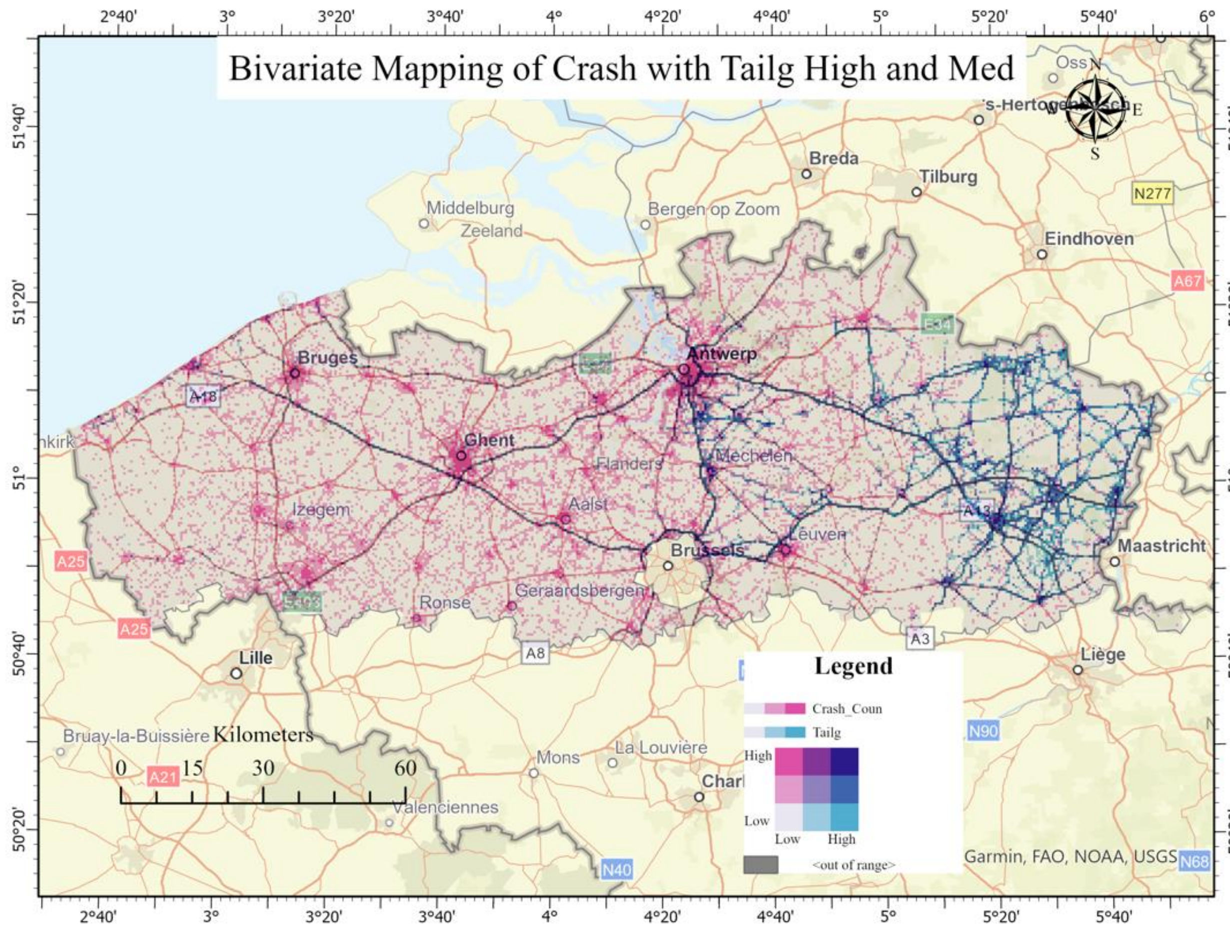

Figure B-5: Bivariate Mapping of Crash with Tailgating High and Medium event

The above choropleth mapping indicates that traffic crash and risky driving event has high frequency on the cities around Antwerp, Hasselt, Mechelen and Genk. The locations with the highest frequency of crash and risky events are indicated with purple color whereas the locations with low frequency of crash and low risky events are shown with light grey color. The locations with high frequency of traffic crash are indicated by cyan color and also the sections with only high number of risky events shown with blue colors. The sections with high number of crashes and risky event are an indicative of the relationship between traffic crash and risky events within the study area.

APPENDIX C

Spatial Lag Model

Goodness of Fit[a]

	Value	df	Value/df
Deviance	53385.461	53999	.989
Scaled Deviance	53385.461	53999	
Pearson Chi-Square	69280.700	53999	1.283
Scaled Pearson Chi-Square	69280.700	53999	
Log Likelihood[b]	-48796.672		
Akaike's Information Criterion (AIC)	97609.343		
Finite Sample Corrected AIC (AICC)	97609.346		
Bayesian Information Criterion (BIC)	97680.518		
Consistent AIC (CAIC)	97688.518		

Dependent Variable: Crash_Coun
Model: (Intercept), Accel, Decel, Speed, Steer, Tailg, lg_Cr_C
a. Information criteria are in smaller-is-better form.
b. The full log likelihood function is displayed and used in computing information criteria.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
6146.372	6	.000

Dependent Variable: Crash_Coun
Model: (Intercept), Accel, Decel, Speed, Steer, Tailg, lg_Cr_C
a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

Source	Type III Wald Chi-Square	df	Sig.
(Intercept)	18432.569	1	.000
Accel	4.144	1	.042
Decel	1.193	1	.275
Speed	33.180	1	.000
Steer	2.414	1	.120
Tailg	738.068	1	.000
lg_Cr_C	10079.666	1	.000

Dependent Variable: Crash_Coun
Model: (Intercept), Accel, Decel, Speed, Steer, Tailg, lg_Cr_C

Note that: Lag count of traffic crash developed by queen contiguity method of spatial weight matrix. The queen contiguity method considers all neighbors traces to the polygon.