CRASHES AND INJURIES ON RURAL ROADS IN ALASKA - TOWARD A BETTER UNDERSTANDING OF RURAL SAFETY ISSUES THROUGH LINKED DATA AND ENVIRONMENTAL FACTORS – TASK B

FINAL PROJECT REPORT

by

Vinod Vasudevan, Ph.D., P.E. (Co-PI)
Professor of Civil Engineering

for

Center for Safety Equity in Transportation (CSET)
USDOT Tier 1 University Transportation Center
University of Alaska Fairbanks
ELIF Suite 240, 1764 Tanana Drive
Fairbanks, AK 99775-5910

In cooperation with U.S. Department of Transportation,
Research and Innovative Technology Administration (RITA)

rural · isolated · tribal · indigenous

DISCLAIMER

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle Crashes and Injuries on Rural Roads in Alaska - Toward a Better Understanding of Rural Safety Issues through Linked Data and Environmental Factors Task B: Environmental and Geometric Safety Factors of Rural Crashes	**5. Report Date** August 21, 2025 **6. Performing Organization Code**
7. Author(s) and Affiliations Vinod Vasudevan, Ph.D., P.E. (PI) Professor of Civil Engineering University of Alaska Anchorage	**8. Performing Organization Report No.** INE/CSET 25.08
9. Performing Organization Name and Address Center for Safety Equity in Transportation ELIF Building Room 240, 1760 Tanana Drive Fairbanks, AK 99775-5910	**10. Work Unit No. (TRAIS)** **11. Contract or Grant No.** Grant # 69A3551747129
12. Sponsoring Organization Name and Address United States Department of Transportation Research and Innovative Technology Administration 1200 New Jersey Avenue, SE Washington, DC 20590	**13. Type of Report and Period Covered** 05/01/2021 – 05/30/2025 **14. Sponsoring Agency Code**

15. Supplementary Notes

Report uploaded to:

16. Abstract

Alaska experiences extreme weather and driving conditions compared to many other locations in the United States. For example, during summer, most traffic activity occurs in daylight, whereas in winter months, the majority of traffic activity occurs in the dark. Similarly, driving conditions change drastically across seasons. During winter, heavy snow and extremely cold temperatures provide challenges for drivers. These conditions lead to safety concerns such as noticeable changes in daylight hours, peak traffic activities in dark hours, and reduced friction values due to snow and ice. Another challenge is the sun's position above the horizon throughout the year, especially during the spring, summer, and fall seasons. The sun's low elevation angle for an extended period can be a primary concern for drivers since the bright sun glare can make it difficult to see one's surroundings. During summer, long day hours and high activity levels may cause fatigue for drivers. Also, a higher number of tourists during summer may change the traffic conditions, posing higher safety risks in some regions. This report summarizes the findings of a study analyzing crash data, and combining external data to develop better understanding safety challenges in Alaska

17. Key Words	18. Distribution Statement
Rural crashes, crash data, risk assessment	

19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
Unclassified.	Unclassified.	30	N/A

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized.

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS				
Symbol	**When You Know**	**Multiply By**	**To Find**	**Symbol**
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
^oF	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	^oC
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS				
Symbol	**When You Know**	**Multiply By**	**To Find**	**Symbol**
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
^oC	Celsius	1.8C+32	Fahrenheit	^oF
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380. (Revised March 2003)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

Alaska experiences extreme weather and driving conditions compared to many other locations in the United States. For example, during summer, most traffic activity occurs in daylight, whereas in winter months, the majority of traffic activity occurs in the dark. Similarly, driving conditions change drastically across seasons. During winter, heavy snow and extremely cold temperatures provide challenges for drivers. These conditions lead to safety concerns such as noticeable changes in daylight hours, peak traffic activities in dark hours, and reduced friction values due to snow and ice.

Another challenge is the sun's position above the horizon throughout the year, especially during the spring, summer, and fall seasons. The sun's low elevation angle for an extended period can be a primary concern for drivers since the bright sun glare can make it difficult to see one's surroundings. During summer, long day hours and high activity levels may cause fatigue for drivers. Also, a higher number of tourists during summer may change the traffic conditions, posing higher safety risks in some regions.

This study showed that rural and urban areas face slightly different challenges. For example, when the sun elevation is critical, types of crashes and where they occur differs from urban crashes. Analysis of crash data with sun elevation showed that sun angle and number of crashes are highly correlated.

Machine learning models showed that First Sequence Event, Road Surface, and Time of Day are the top predictors for injury in Alaskan crashes. The second machine learning model illustrated that Lighting, Driver Age, and Road Surface were the best predictors for 'Collision with an Animal/Pedestrian/Cyclist' and 'Non-Collision' crashes. Conversely, non-collisions were most likely to occur when it was dark with no lighting and the roads were wet, icy, or snowy. This makes sense for most of the run-off the roadway crashes or other single vehicle crashes.

CHAPTER 1. INTRODUCTION

Alaska experiences extreme weather and driving conditions. For example, during summer, most of the driving occurs in daylight, whereas in winter months, most of the traffic activity occurs in the dark. Similarly, the driving conditions change drastically across seasons. During winter, heavy snow and extremely cold temperatures make the road users, both motorized and non-motorized, face extreme challenges. These conditions develop some unique safety concerns. Some of the examples are listed as follows:

- Challenges caused by noticeable changes in daylight hours for the drivers.
- Safety issues associated with peak traffic activities occurring during dark hours.
- Safety issues due to the snow pile-up on roadsides and its adverse impact on reduced sight distances at intersections, especially unsignalized ones.
- Safety issues of snow pile up in obstructing pedestrian and bike paths, posing greater risks.
- Reduced friction values during winter, especially immediately after a snow fall, resulting in increases in the vehicle braking distances.
- Low visibility during a snowstorm.
- Higher risks during spring and fall, due to the slippery roads.

Another challenge is the Sun's position above the horizon throughout the year, especially during the spring, summer, and fall seasons. The Sun's low elevation angle for an extended period can be a primary concern for drivers since the bright sun glare can make it difficult to see the surroundings. During summer, long day hours may cause fatigue for road users. Also, a higher number of tourists during summer may change the traffic conditions, posing higher safety risks in some regions. Due to the heavy snowfall, people prefer SUVs and pickups over sedans. Due to SUVs' high center of gravity, they may not be safer from overturning, a significant risk factor driving in slippery conditions.

The risks presented are generic, and some may only apply to urban areas. However, several of them are applicable even in rural regions. Since the state's rural population is scattered across the state in small villages, delineating traffic risks is challenging. Traditional crash data analysis alone may not provide valuable insights into rural Alaska's issues. Therefore, we need to use advanced computing techniques and alternate data to make sense of the limited available data. Some of the examples of alternate data include weather data by hour/day by region recorded by the National Oceanic and Atmospheric Administration and the Sun's position by time by region (e.g. https://www.suncalc.org). While all these issues cannot be addressed using available data, a detailed analysis of the crash database is expected to help delineate some of the causes of concern in rural road safety. This report summarizes detailed crash analysis

CHAPTER 2. LITERATURE REVIEW

Crash data is the major database that this research used. Therefore, it is important to understand similar studies carried out in other parts of the country and across the world, especially those related to rural areas. This chapter summarizes this activity.

Modeling Pedestrian Injury Severity in Pedestrian-Vehicle Crashes in Rural and Urban Areas explores the severity of pedestrian-vehicle crashes in North Carolina (Chen & Fan, 2019). They wanted to dive into what the significant factors are in these types of crashes, and studied both rural and urban roads. They took crash data from 2005-2012 to model their results using a mixed logit model approach. Early in their research and modeling, they found that the fatality ratio was higher in rural areas than in urban areas. They found that in both areas, the same factors contributed to fatal crashes. These were high speed limits, curvature in roads, condition of driver (age and physical condition), and time of day. The curvature in roads is significant because rural roads are more likely to have more curvature in them than urban areas. This shows a reason for the higher fatality crashes in rural areas than urban areas.

Developing Safety Performance Functions for Severe Distraction-related Crashes along Kentucky's Rural and Urban Two-Lane Roadways studies distraction-related driving in Kentucky (Kumar et al., 2025). The researchers looked at the four-year crash records between 2018 to 2021, noting down roadway and geometric features that may pertain to the cause of the crashes. Because the sample size is small and under-dispersed, they used statistics and modeling to help with their analysis. The results of their study showed rural and urban roads having differing factors that contributed to accidents. Rural roads saw an increase in crash frequencies due to roadside guardrails, wider right-hand shoulders, and both horizontal and vertical curves. Urban roads saw an increase in frequency due to heavy vehicle percentage, and paved shoulders. Both urban and rural roads saw increases due to cell phone usage as well. On rural roads, single vehicle and sideswipe distraction crashes were the most prominent, while urban roads saw more head-on as well as sideswipe crashes.

Comprehensive analysis of single- and multi-vehicle large truck at-fault crashes on rural and urban roadways in Alabama was used to study crash factors involving large trucks in multi vehicle (MV) and single vehicle (SV) crashes (Islam, Jones, & Dye, 2014). The authors used raw crash data from 2010 to 2012 and a mixed logit model to study these factors. Factors that contributed to more rural crashes were found to be wet surfaces, which was not found to be significant in urban roadways. The severity of the crash is also affected by the curvature of the roads. Interestingly, only single vehicle crashes in urban areas were found to be affected significantly by roadway curves in their model. 3 or more single direction lanes in rural areas were found to decrease the likelihood of major injury, and rear end accidents increased the severity in rural MV crashes, but surprisingly decreased them in urban MV accidents.

Continuation from their previous work, researchers in Alabama studied crash factors of crashes involving large trucks on rural and urban roads (Biglari, Kofi, & Jones, 2024). This study analyzed crash data involving large trucks from 2017-2019 in Alabama, and used the same factors as the 2014 study considered. They observed crashes in rural and urban roadways that included single vehicle (SV) and multi-vehicle (MV) crashes. Their tables show that there were higher frequencies of both major and minor injury incidents in the rural SVs and MVs, while there were more crashes that caused no injuries in the urban SVs and MVs.

Champahom, et al. (2021) compared the effects of urban and rural crashes on motorcyclist injury severities using a correlated random parameter ordered probit approach with heterogeneity. The authors collected motorcycle data in Thailand from the years 2016 to 2019, and used a CRPOPHM modeling approach (A correlated random parameters probit model with heterogeneity in means). They found factors such as being male, improper overtaking, and falling asleep were only significant in rural motorcycle accidents. Both urban and rural areas saw significantly increased crashes with motorcycles with a second passenger, which was also the highest factor of all.

Thompson, et. al (2013) examined the environmental, driver and vehicle factors associated with the serious and fatal crashes of older rural drivers. This study found that while drivers of 65 years of age or older are involved in fewer crashes than other ages, they are at an increased risk of death/serious injury. The paper tried to find the factors of these crashes. After reviewing literature, the authors confirmed that regardless of age, rural crashes are more fatal than urban crashes. The authors used crash data from 2004-2008 in South Australia in vehicles that were cars or light passenger vehicles. Using the model and statistics, they concluded that environments with undivided, unsealed, curved and inclined roads lead to more crashes. The authors further explain that this contributes to how there are twice as many fatal crashes in rural roads, because urban roads have more divided, sealed, straight and level roads.

An on-the-spot study of pedestrian crashes on Brazilian Federal District rural highways crossing urban areas used rural highways crossing urban areas crash data in Brazil. The study showed that 82 percent of Brazilians now live in cities, and this caused many urban outskirts to be in the margins of rural highways (Velloso & Jacques, 2012). The authors used data between October 2004 and March 2005 through on-the-spot data and traffic-accident reports, noting down important factors and variables to crashes that occurred. This study found that the attitude of the pedestrian was vital. Alcohol abuse and lack of attention contributed to a lot of the pedestrian crashes.

Shaw et. al (2022) studied urban and rural child deaths from motor vehicle crashes in the US between the years 2015 and 2019. This study considered factors such as region type, restraint use, and the restraint laws. The authors took fatality crash data and analyzed age and type of restraint used. This study found that deaths increased the more rural the area was, with the lowest rates in the most urban counties, and the higher rates were more prominent in rural roadways. This result is consistent with all the other literature they studied that even among adults, fatal crashes were more present in rural areas. Improper use of restraint for the child were one of the major factors playing into increased fatal crashes, especially in rural areas.

Islam and Brown (2017) carried out a comparative injury severity analysis of motorcycle at-fault crashes on rural and urban roadways in Alabama. This study explored the factors that contributed to the severity of crashes in accidents where motorcycles were at fault in both urban and rural areas. Crash data from 2010-2014 in Alabama was studied using statistics and models, and it was found that increased speeds were a more prominent factor in rural environments for fatal injuries. Motorcyclists not using lights was only significant in the rural areas rather than the urban areas, which can be explained by urban roads having more light posts. Female drivers and horizontal curves were surprisingly only significant in urban roads rather than rural roads. However, both environments saw a higher fatality rate with riders not wearing protective headwear.

Truck-involved crashes injury severity analysis for different lighting conditions on rural and urban roadways explores how lighting conditions affect truck-involved crashes on both urban and rural roads (Uddin & Huynh, 2017). The researchers used mixed logit models and the factors and results from 2009 to 2013. Interestingly, the results showed increasing speeds in dark rural roads increased major and minor injury, but increasing speeds decreased minor crashes on urban roads.

Thompson et. al (2013) studied treatments for crashes on rural two-lane highways in Texas. While this study does not compare urban vs rural crashes as much, they did readings on this topic. Their literature review showed that most urban crashes occurred in intersections and driveways, but in rural areas, they occurred away from intersections and driveways.

Analysis of driver injury severity in single-vehicle crashes on rural and urban roadways was used to study the severity of driver injuries of different types of roads (Wu, et. al, 2016). The researchers took single-vehicle crash data from 2010 to 2011 collected in New Mexico, and used a lot of discrete choice models that have been used in previous studies that were similar to this one. Rural crashes saw more severe injury in raining conditions, and no passing zones. Reflective signs and pavement markers could help with these. Urban roads saw more severe crashes in multi-lane roadways and drug impaired drivers.

Characteristics of Intersection Accidents in Rural Municipalities explores factors of accidents in rural areas (Hanna, et. al, n.d.). This study analyzed data from 300 intersections in 42 differing towns and cities in the state of Virginia. They observed and noted factors of crashes between 1969 to 1973 to understand these factors. They found poor sight to be a factor of increased crashes, which is based on factors like speed, and degree of sight obstruction and distance. Other factors include poor roadway conditions, but when the intersections get severely poor, drivers tend to observe more caution, which decreases the frequency of crashes.

Reurings & Janssen (2025) developed accident prediction models for urban and rural carriageways in the Netherlands. This study measured risks and exposures, and researched characteristics within and between road categories. They analyzed crashes on carriageways between the period 2000-2002, which also included crashes that happened on intersections. They found that carriageways on rural roads with only one direction for driving were much safer than ones with two directions. Interestingly, in urban areas, carriageways with 70 km/h speed limit were safer than ones with 50km/h speed limits.

Zhang et al. (2021) explored the effect of changing climates in rural crash injury and fatality patterns in rural, and isolated tribal communities. The researchers looked into the factors that caused crashes in these communities. They took a three-year sample from 2012 to 2014 from Texas, Arkansas, Oklahoma, and Louisiana. This sample included all single-vehicle crashes that were due to rainy conditions. This study found that due to low visibility and low friction on road surfaces, there were increased crashes, especially in rural areas. There were also a lot of single-vehicle crashes in the rain in these areas as well. This study concluded that rural areas were more susceptible due to urban areas having more road maintenance and lighting fixtures.

Champahom et al. (2020) applied hierarchical logistic models to compare urban and rural roadways for severity of rear-end collisions. The researchers used rear-end crash data from 2011 to 2015 on Thai highways and used hierarchical logistic models on the data. The research found that both urban and rural roads saw increased severity at nighttime, lack of seatbelt usage, and size of vehicle. This was more prominent on rural roads than urban roads.

In summary, the studies across the US and the world showed that the crash characteristics of rural areas are different from urban areas due to differences in user behavior, changes in driver environments, and traffic characteristics. Therefore, it is important to study them for Alaska.

CHAPTER 3. CRASH DATA ANALYSIS

Crash data are used as the primary database for this project. Crash data were collected from the Alaska Department of Transportation and Public Facilities (ADOT&PF). Due to changes in policies related to data releases at the ADOT&PF, this project relied on slightly older data. We used crash data from 2009 to 2012 for this project. Although this database is slightly old, its analysis is expected to provide useful input relevant to this study. We used Python programming language for most of these analyses. This section summarizes the major findings from crash data.

As the first step, crashes were divided into rural and urban crashes. Each crash in the database has a "City" field. Any city that the dataset does not mark as a municipality is labeled as "Non-City Area". To define rural and urban areas, any crash under a municipality is defined as "Urban" while any crash labeled as "Non-City Area" is defined as "Rural." Figures 3-1 and 3-2 show the plots of urban and rural crashes respectively. Of the reported crashes between 2009 and 2012, only about 18 percent crashes were from rural areas.

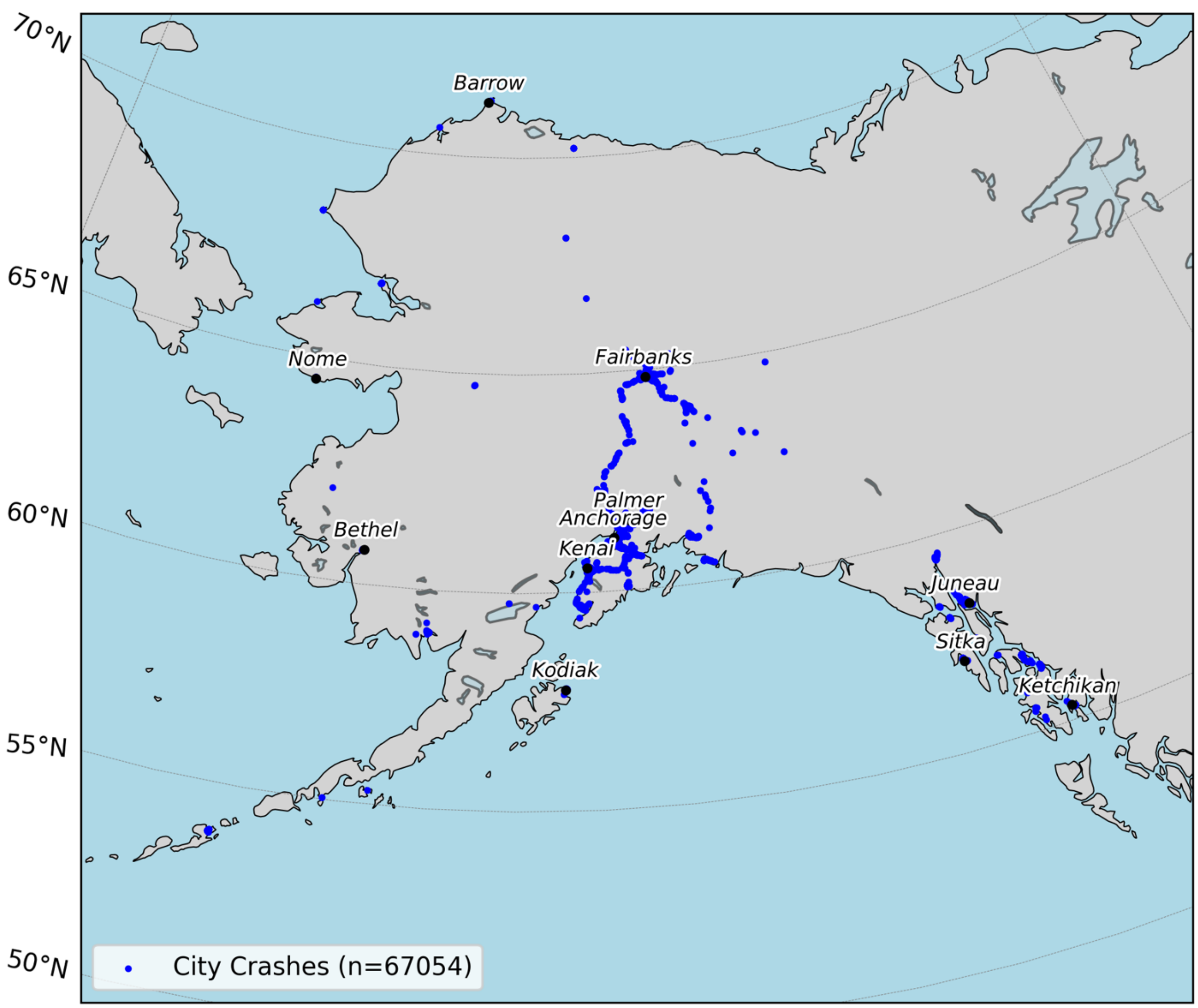

Figure 3-1 Crashes in urban areas (2009 – 2012)

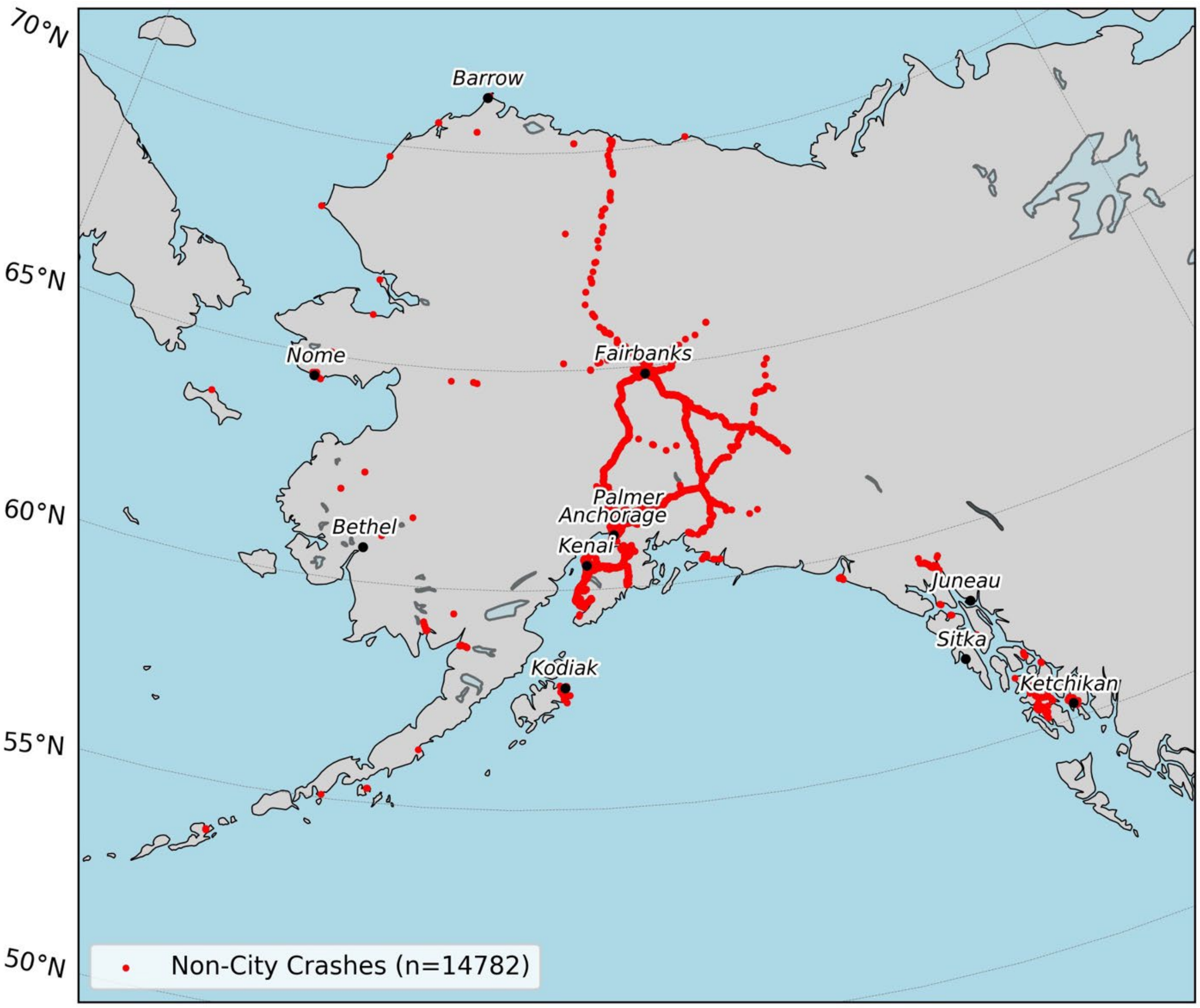

Figure 3-2 Crashes in rural areas (2009 – 2012)

3.1. Rural Vs Urban Comparison

Figures 3-3 and 3-4 show the crash distribution by hour of the day for urban and rural areas. These figures do not show much difference in the crash distribution pattern. In both urban and rural areas, the greatest number of crashes occurred during the peak PM traffic hours. This finding is similar to the findings from the studies from other parts of the country.

Figures 3-5 and 3-6 show the crash distribution by season. In this study, winter is defined as the months November, December, January, February, March, and April. This also corresponds to some of the months with most snow as seen in the weekly snow graph. These figures show that while urban and rural areas maintain similar crash distribution by season, rural areas have slightly higher proportion of crashes during winter season. However, overall, it is worth noting that over 60 percent crashes occur during winter months despite summer months having higher traffic volume due to summer tourism.

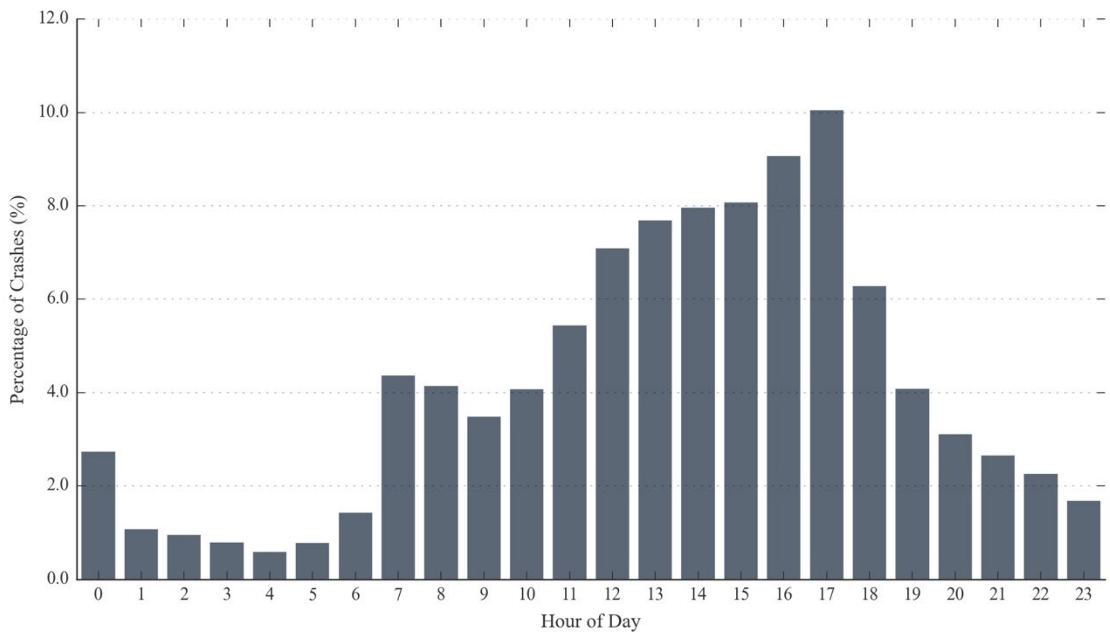

Figure 3-3 Crash distribution by hour in urban areas (2009-2012)

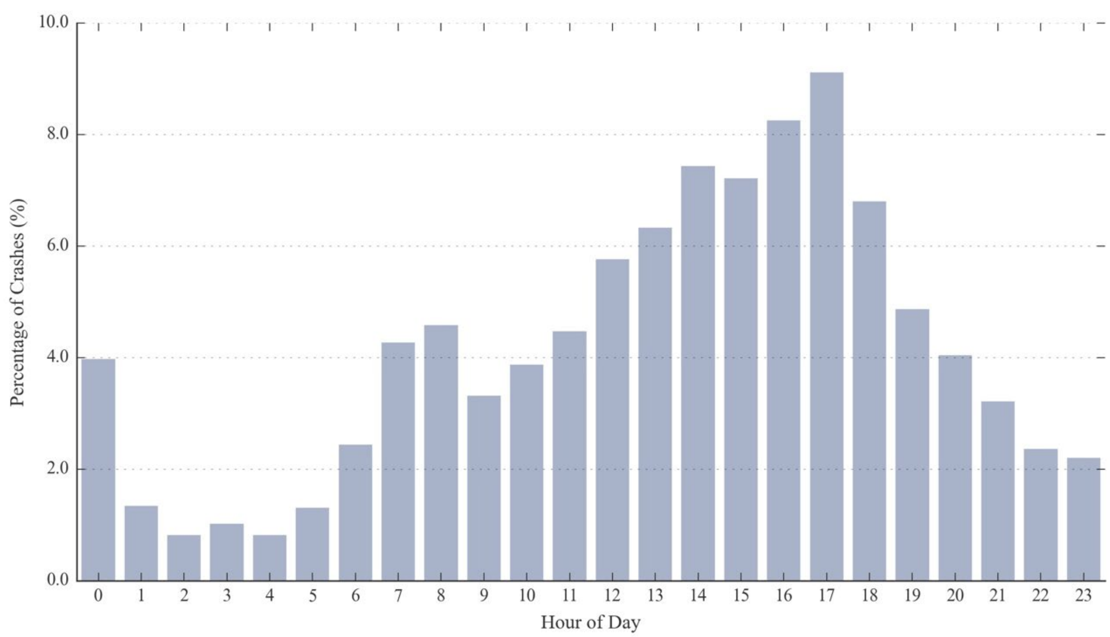

Figure 3-4 Crash distribution by hour in rural areas (2009-2012)

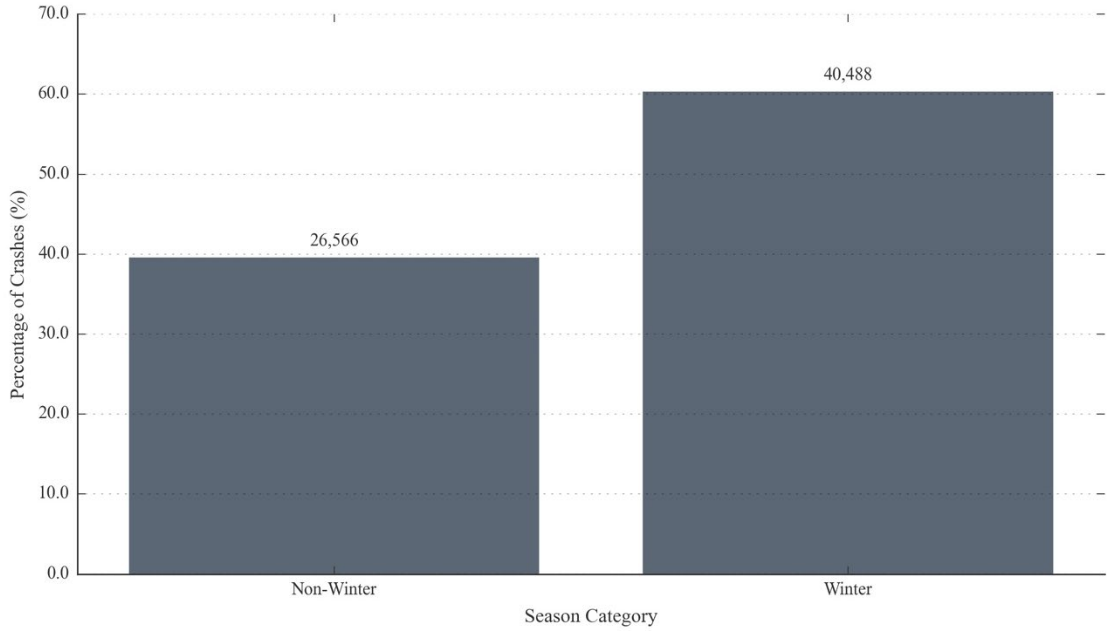

Figure 3-5 Crash distribution by season in urban areas (2009-2012)

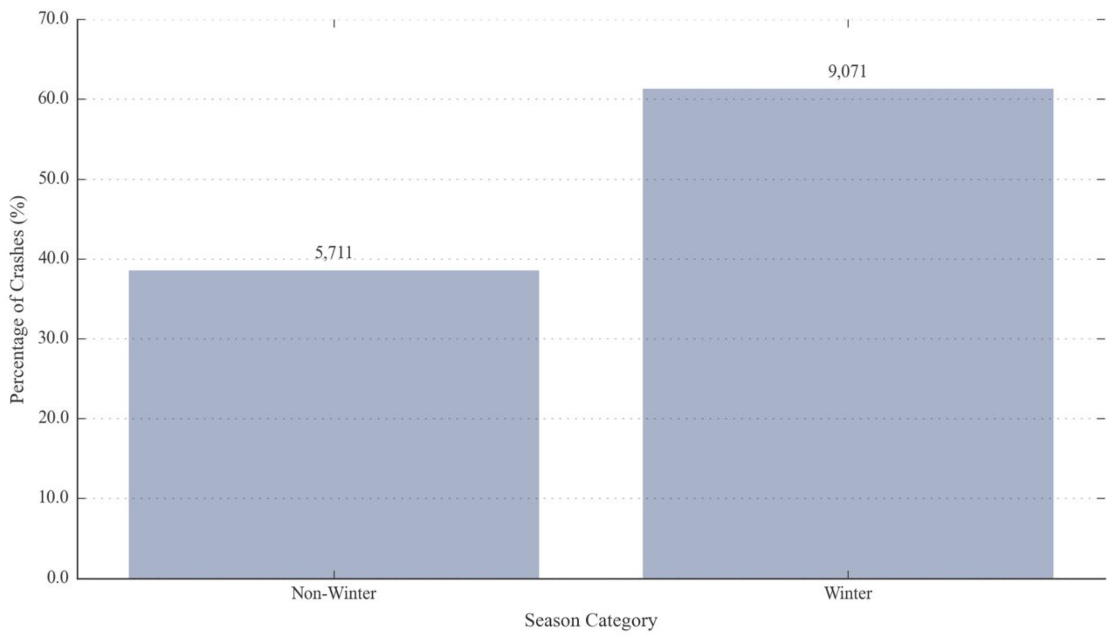

Figure 3-6 Crash distribution by season in rural areas (2009-2012)

Figure 3-7 compares crash distributions by traffic control types between urban and rural areas. This figure shows that the majority of crashes in both urban and rural areas occur at locations without a control. These include mid-block locations. In rural areas over 70 percent of crashes occur at such locations, whereas about 45 percent of the urban crashes occur at similar locations. Proportions of crashes at traffic controlled and stop sign controlled intersections are similar in rural areas (less than 10 percent each), whereas in urban areas, over 30 percent of crashes occur at intersections.

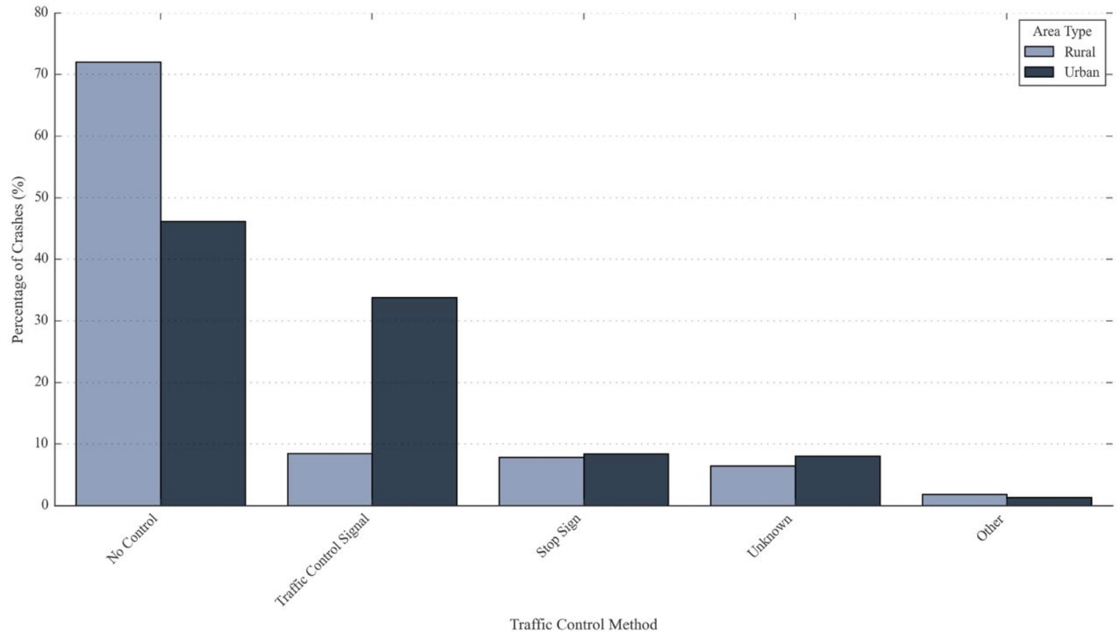

Figure 3-7 Comparison of crash distribution by traffic control type urban Vs. rural

First sequence events of crashes for urban and rural areas are shown in Figure 3-8. This figure shows clear differences between urban and rural crashes. Angle and rear-end collisions are the most occurred crashes in both these regions. However, single vehicle crashes, such as moose-related or ran-off-roadway crashes represent a notable percent of crashes in rural areas. Figure 3-9 compares crash types during winter and non-winter seasons. During winter months, angle collisions were the first sequence of events in about 50 percent of crashes.

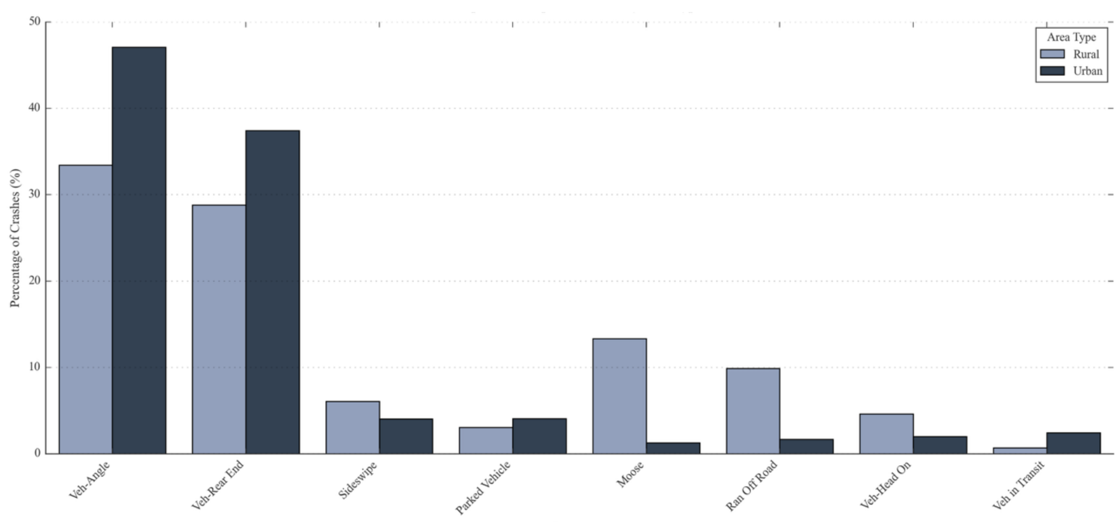

Figure 3-8 Crash type distribution urban Vs. rural

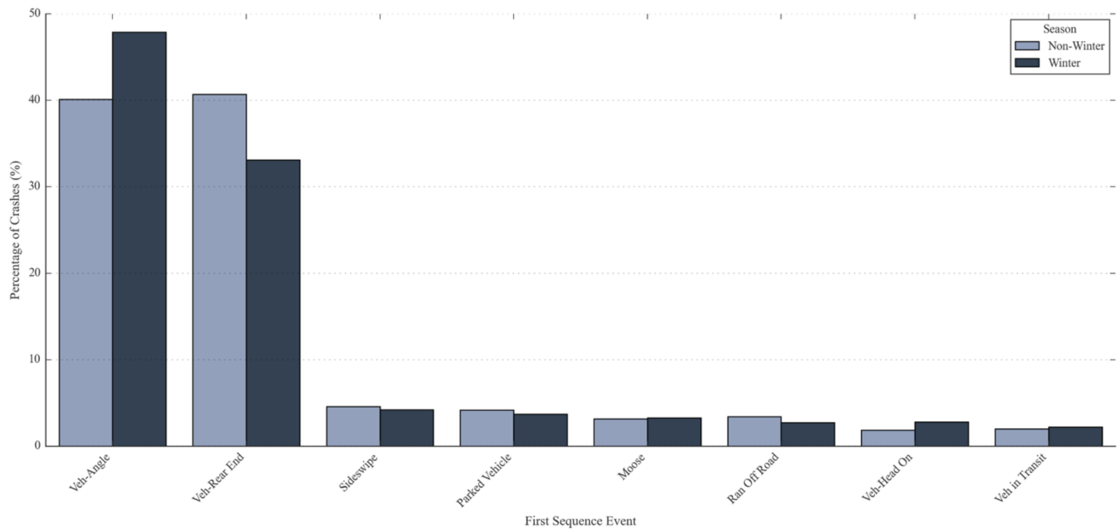

Figure 3-9 Crash type distribution winter Vs. non-winter

Figure 3-10 shows a heat map of week of year and hour or day. This figure shows that in the first part of the year and the last part of the year, i.e., between fall equinox and spring equinox of the following year, a clear demarcation can be observed between office and non-office hours. While majority crashes are between 8:00 am and 7:00 pm, a higher distribution can be observed between 11:30 am and 6:00 pm. Figures 3-11 and 3-12 show the same information for urban and rural regions respectively. These figures also show the similar trend that Figure 3-10 showed.

3.2. Relation Between Crashes and Sun Angle

Figures 3-10 to 3-12 showed a different crash distribution pattern during days with shorter daylight compared with days with longer day light. These differences could be because of sun's position on drivers. At dawn and dusk, lower vertical position of the sun can cause discomfort. Most parts of Alaska experience very short day-hours during winter and even when sun is present, the low vertical angle can create uncomfortable driving condition as the sun will be directly facing driver's eyes. At low angles, even the visors in the vehicles cannot help drivers. This section analyses relationship between crashes and sun angle.

Using standard libraries in Python, sun's position was determined for each crash using the latitude and longitude values. The sun's position has two parts, elevation and (vertical angle) and direction (horizontal angle). Since the driving direction information is inaccurate in most of the reported crashes, that information is not used for this study. Therefore, the first step was to determine the range of vertical angle that is overrepresented in crashes. Figures 3-13 and 3-14 show the plot between number of crashes and vertical angle of sun's position at the time of crash for urban and rural areas respectively. Both these figures clearly show that number of crashes are the highest around five degrees, and that the crashes are very low when the elevation is above 20 degrees. Figures 3-15 and 3-16 show similar information as percent crashes.

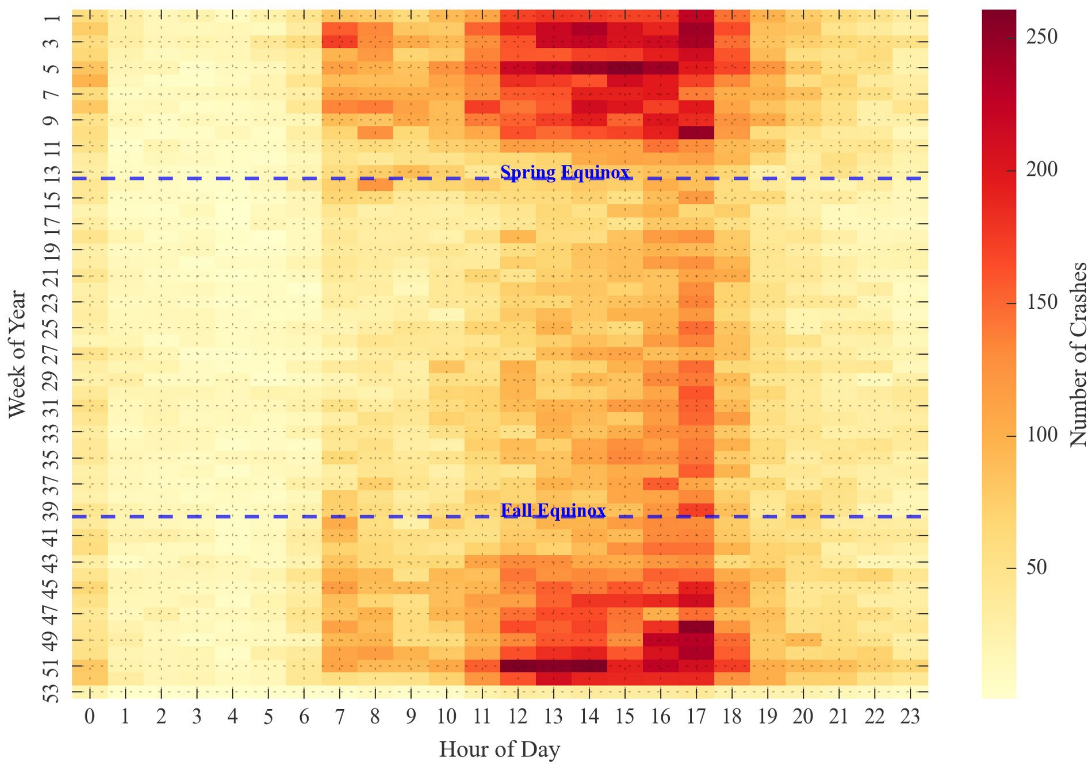

Figure 3-10 Week of the year Vs Hour of day heat map (overall)

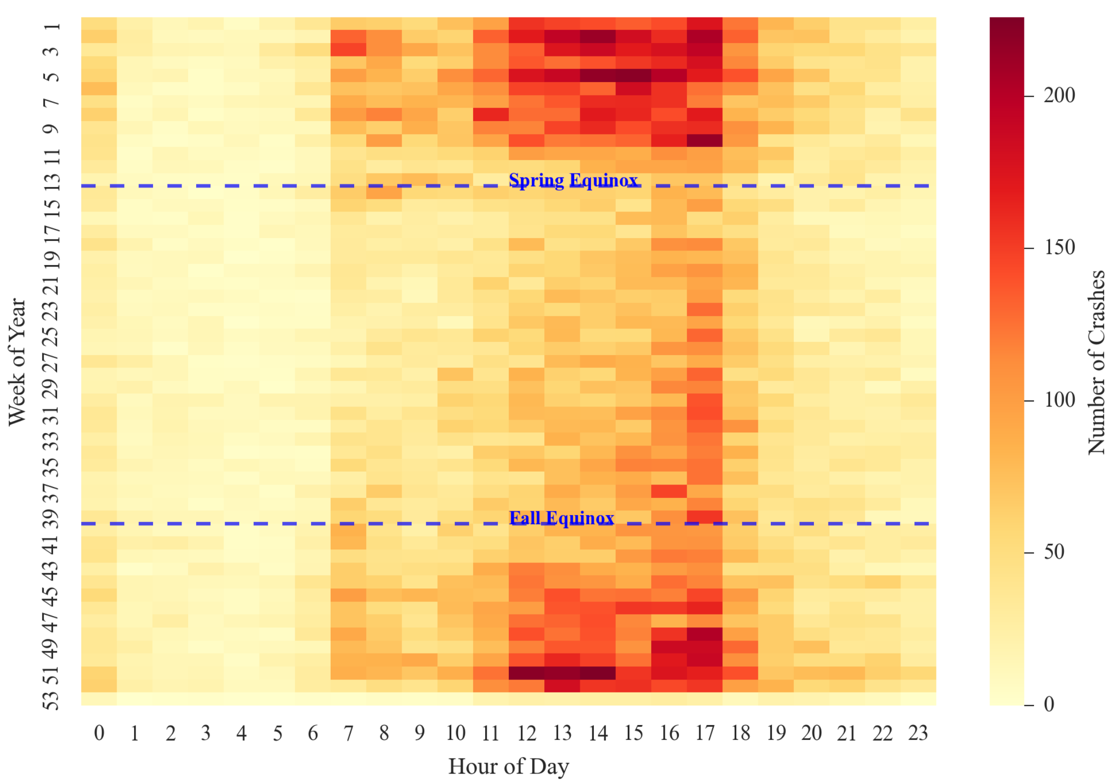

Figure 3-11 Week of the year Vs Hour of day heat map (urban)

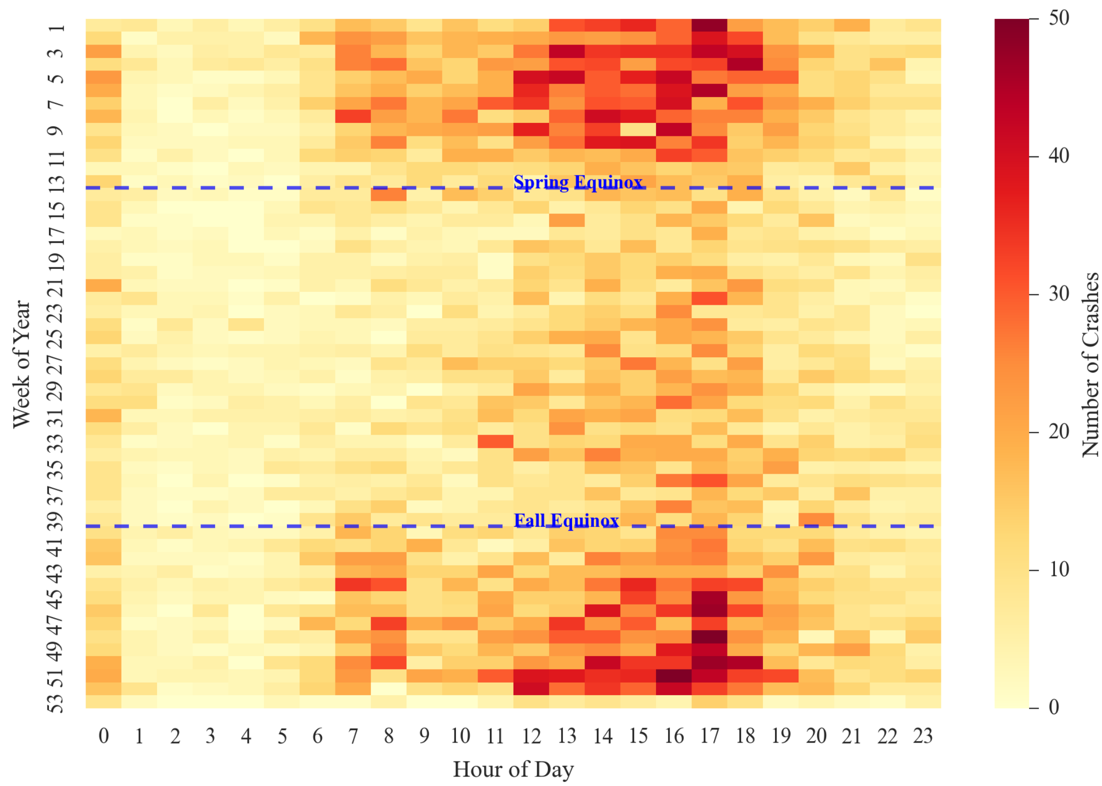

Figure 3-12 Week of the year Vs Hour of day heat map (rural)

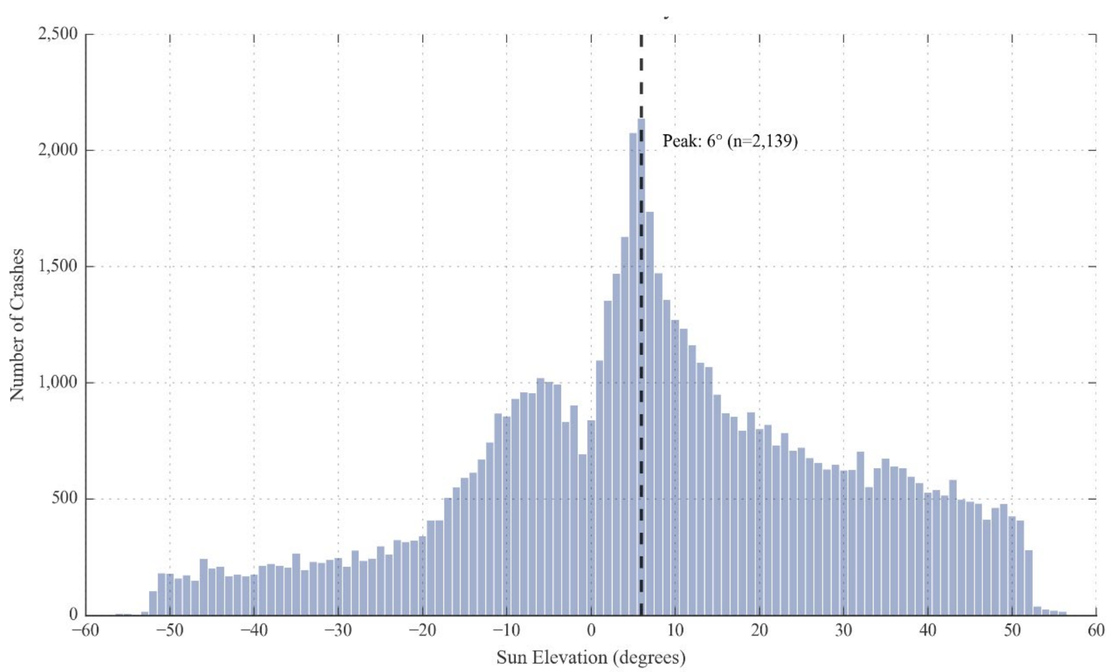

Figure 3-13 Number of crashes Vs Sun elevation (urban)

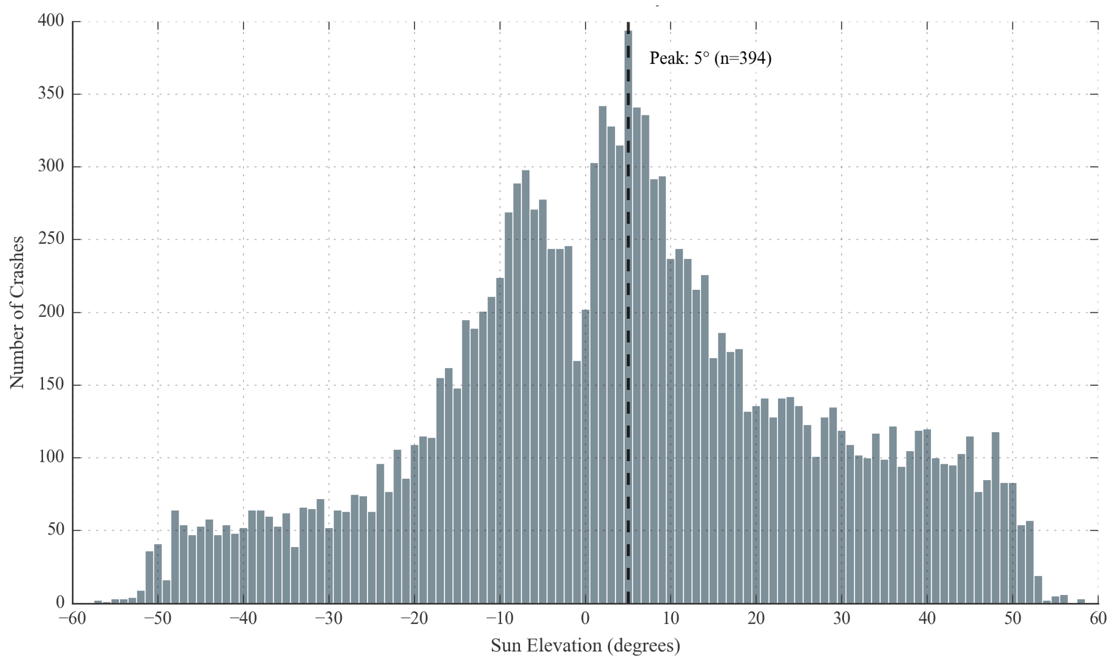

Figure 3-14 Number of crashes Vs Sun elevation (rural)

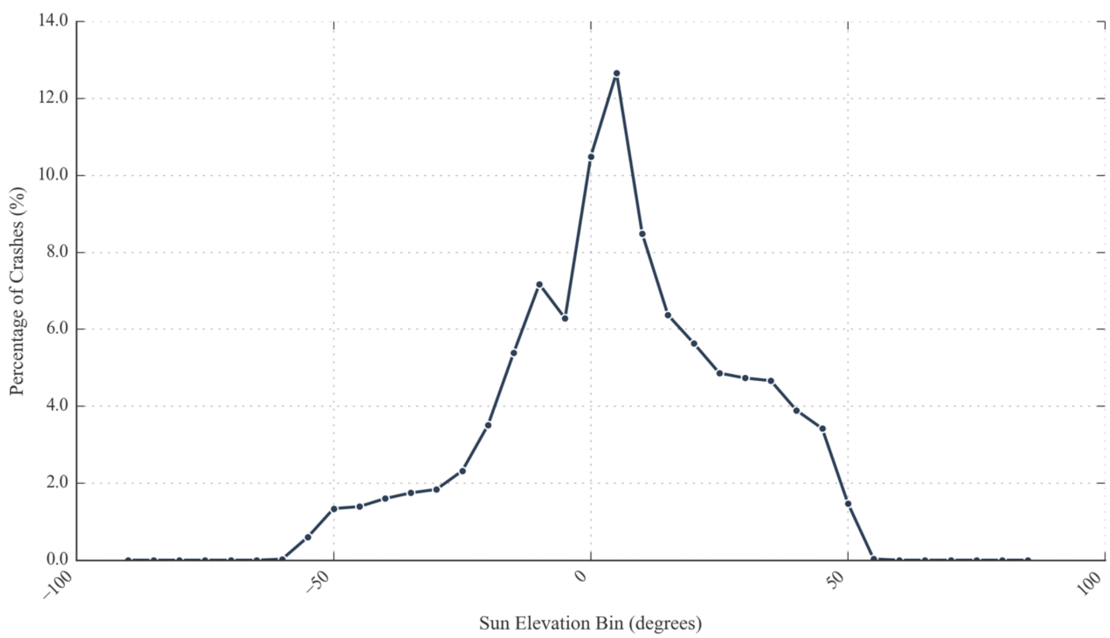

Figure 3-15 Percent of crashes Vs Sun elevation (urban)

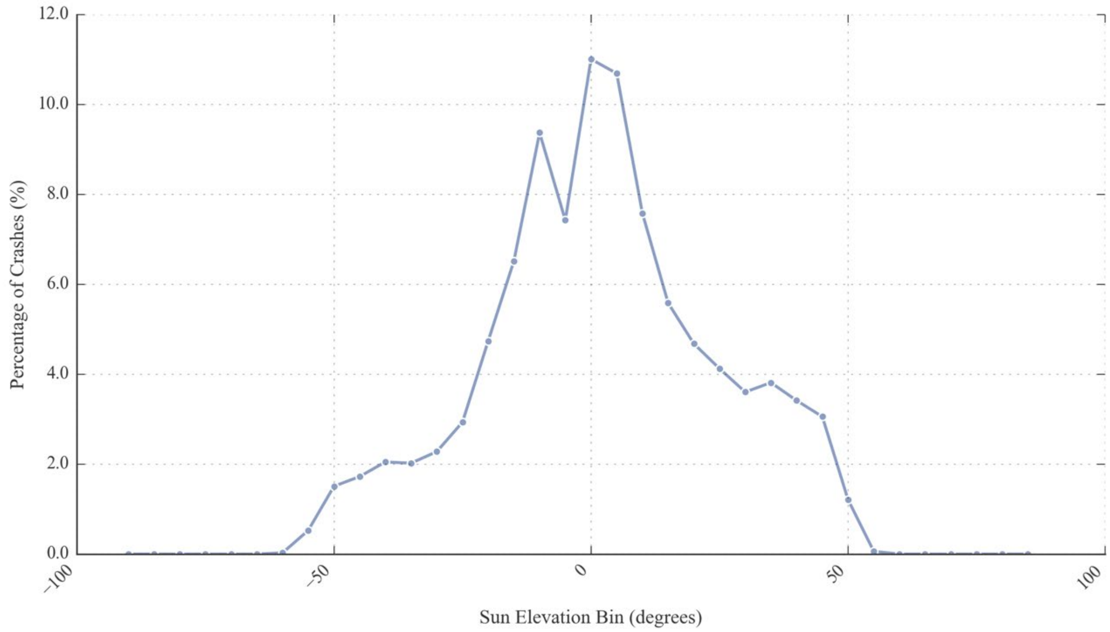

Figure 3-16 Percent of crashes Vs Sun elevation (rural)

Figures 3-13 to 3-16 showed that a higher proportion of crashes occur when the sun elevation is between 0 and 15 degrees. This pattern is more visible in the heat map presented in Figure 3-17 to 3-19. Figure 3-17 shows the heat map of crashes by the week of the year and sun elevation for the overall data. Figures 3-18 and 3-19 show the same information for urban and rural areas respectively. In all these figures, higher proportion of crashes occurring between 0 and 15 degrees of sun elevation during periods of shorter day lights are visible. Between the spring equinox and fall equinox, these changes are less significant. During summer months, most of the days, sun angles are above the critical values of 0 to 15 degrees. In days when the sun is between these values, this occurs only for a short duration. Comparison of Figures 3-18 and 3-19 show that the changes are drastic in urban areas than rural areas. This could be due to the higher proportion of vehicles on the road. Figure 3-8 showed that the majority of crashes (over 80 percent for urban and over 70 percent for rural) involve more than one car.

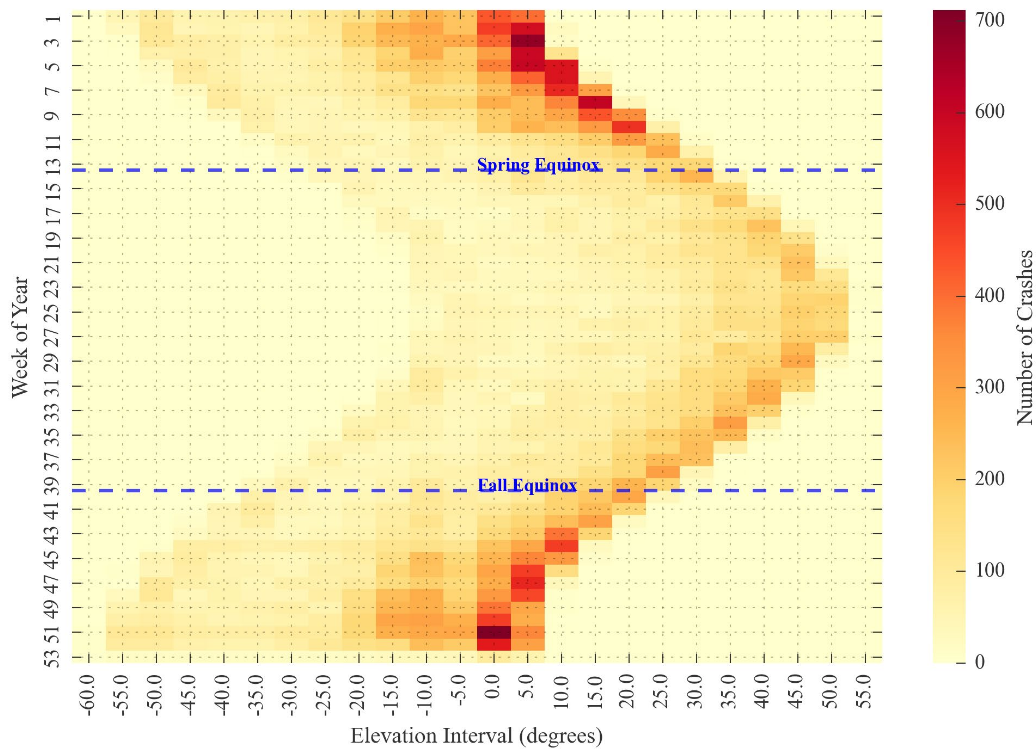

Figure 3-17 Crash heat map: Week Vs. Elevation (overall)

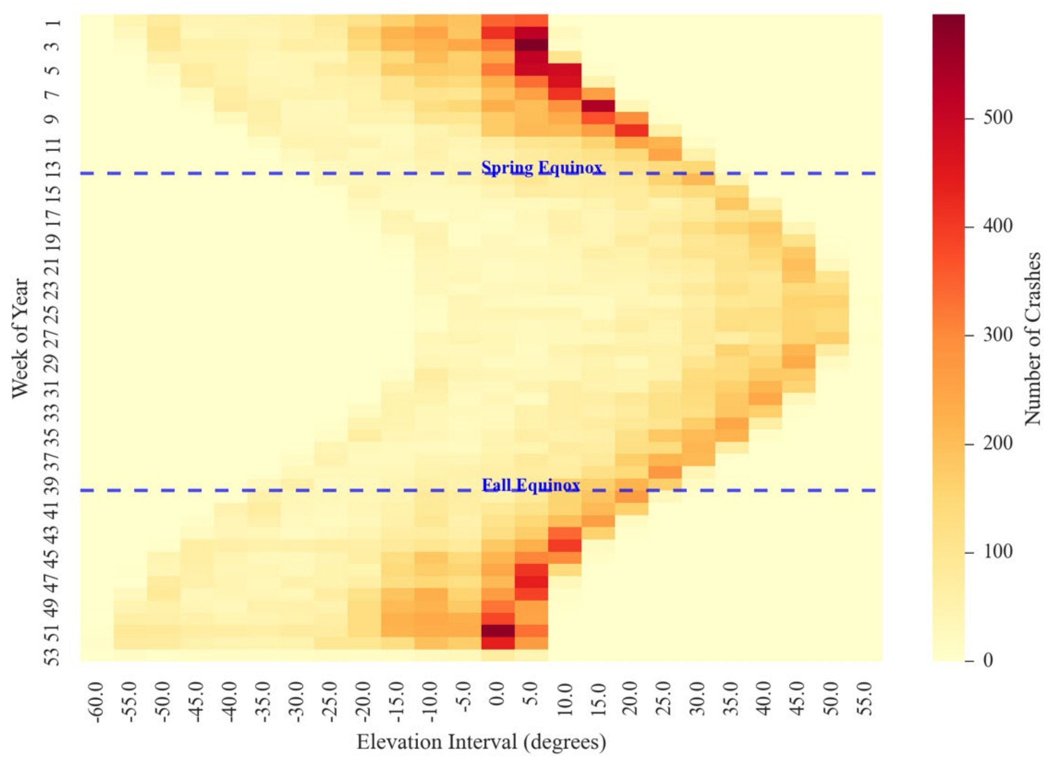

Figure 3-18 Crash heat map: Week Vs. Elevation (urban)

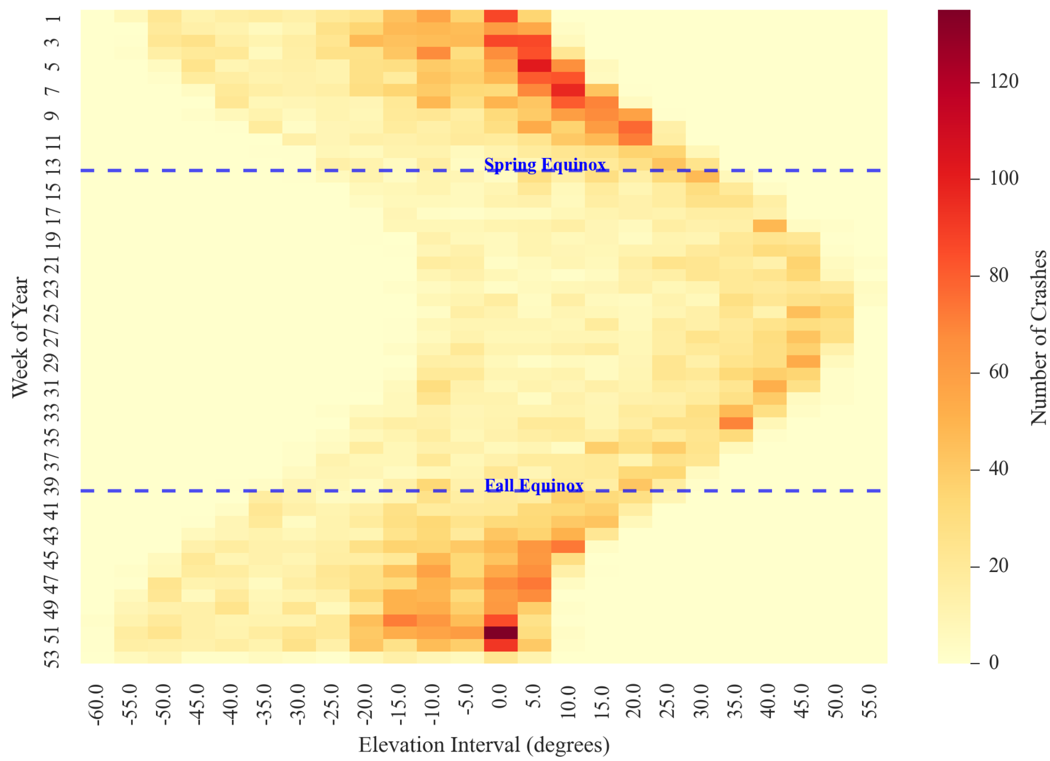

Figure 3-19 Crash heat map: Week Vs. Elevation (rural)

Further analysis of crashes shows that of the 81,836 overall crashes in the analysis period, about 37.5 percent (or 30,657) occur when the sun is positioned at critical elevation range of 0 to 15 degrees. The similar values for urban and rural areas are 38.0 and 34.8 percent respectively. This means that crashes in urban areas slightly more affected by the critical sun elevation than rural areas. Looking at the types of crashes that occur during this period, Figures 3-20 to 3-22 shows that angle and rear-end collisions are the two major types of crashes occurring during the critical elevation period. This result is similar to the overall crashes.

Figures 3-23, 3-24, and 3-25 show the crash distribution by traffic control type at critical sun elevation for overall state, urban areas, and for rural areas respectively. These figures show that compared to the whole data, when sun elevation is within the critical range, a higher proportion of crashes occur at areas without any control. These locations include intersections without any control. At such locations, drivers' inability to see oncoming vehicles on the major road can result in vehicles entering the major road without proper gap, causing angle or rear-end collisions.

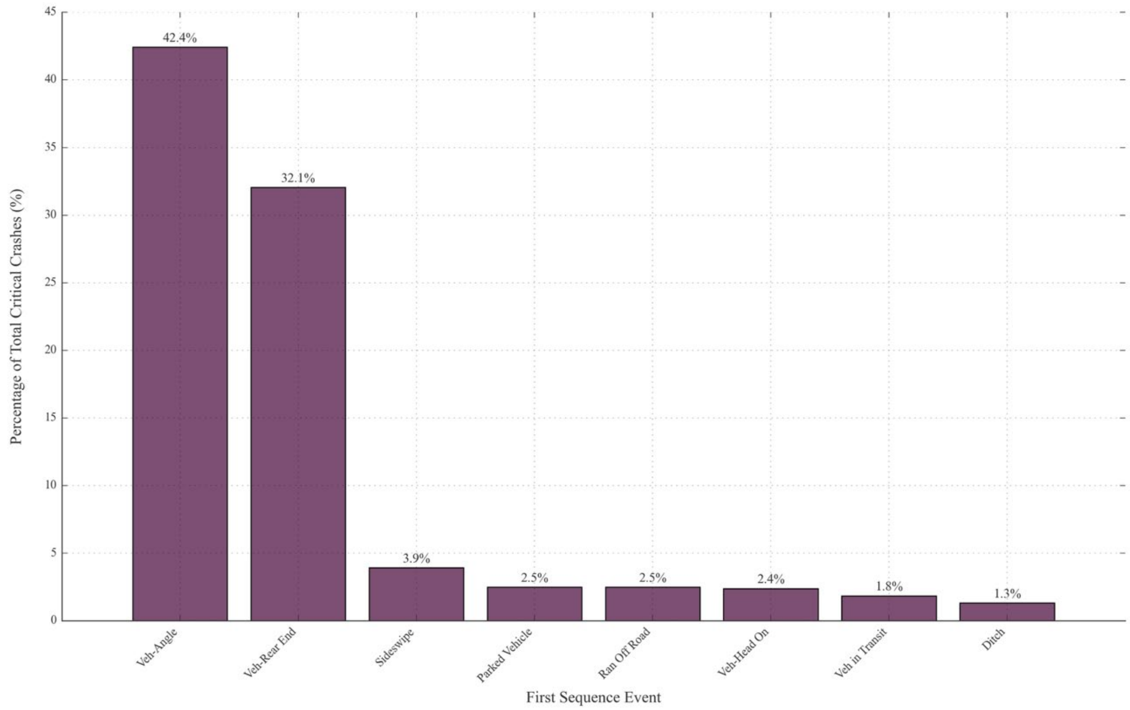

Figure 3-20 Crash type distribution at critical sun elevation (overall)

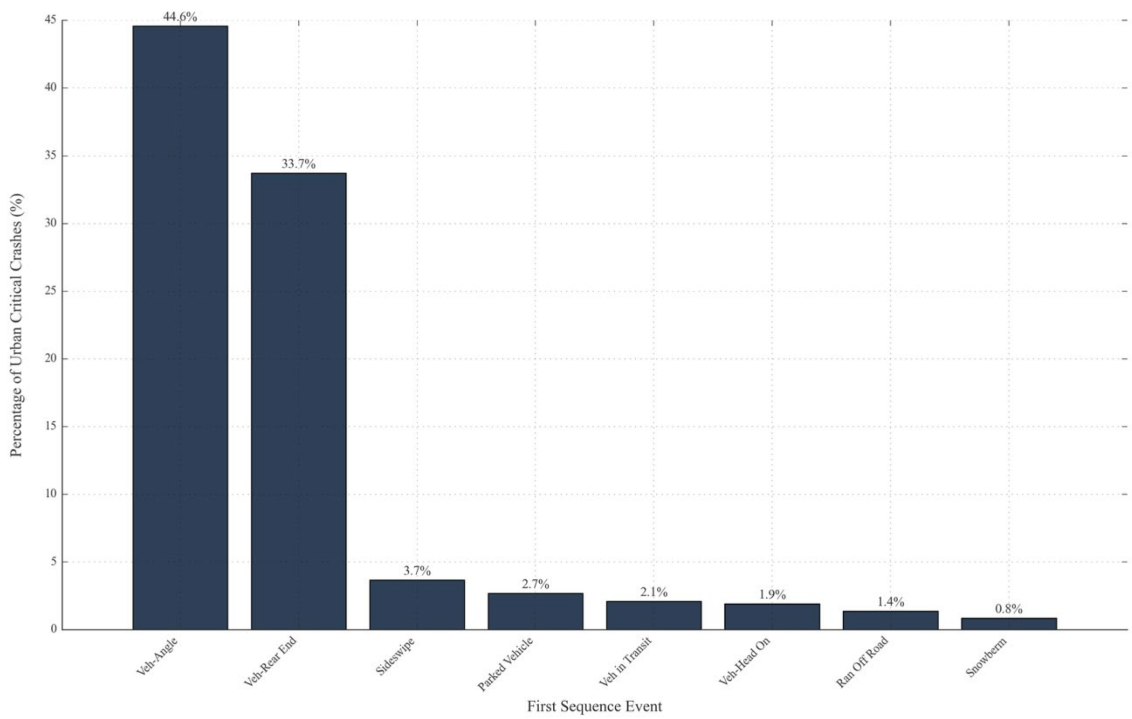

Figure 3-21 Crash type distribution at critical sun elevation (urban)

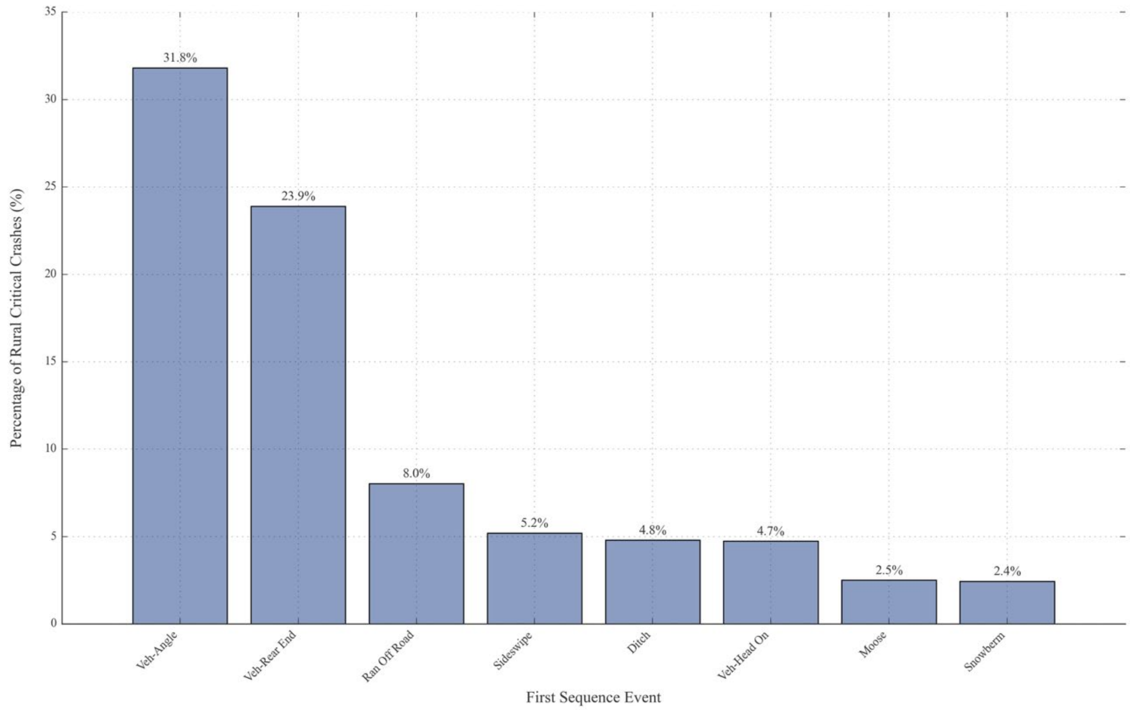

Figure 3-22 Crash type distribution at critical sun elevation (rural)

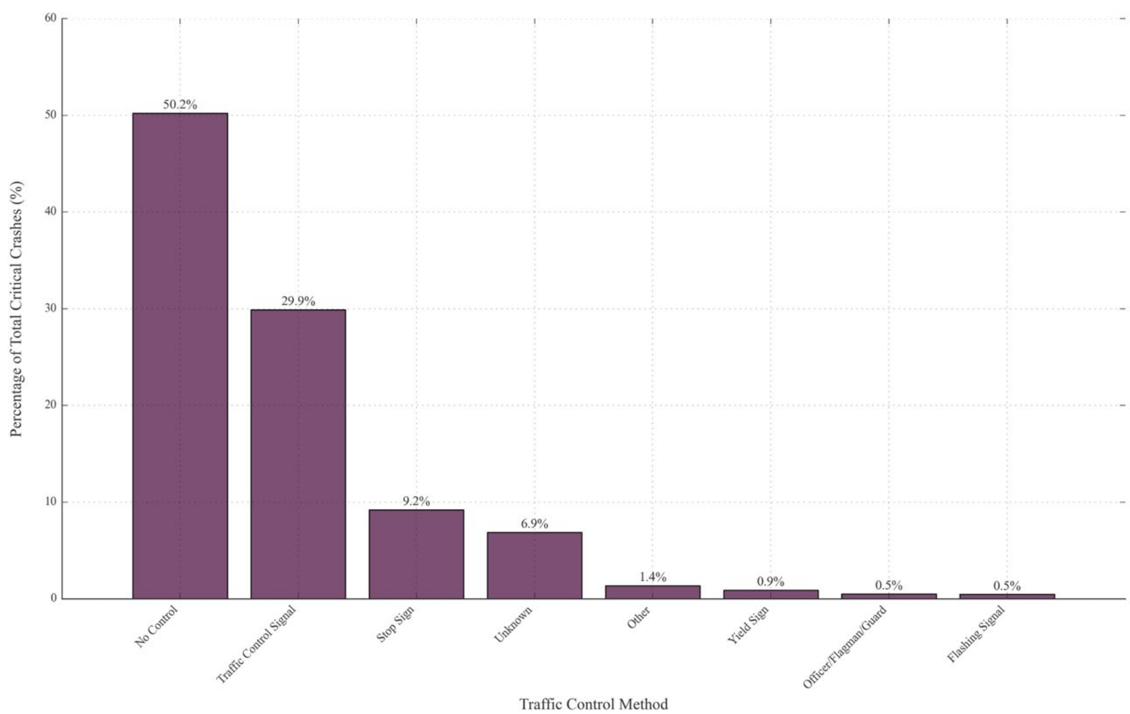

Figure 3-23 Crash distribution by traffic control type at critical sun elevation (overall)

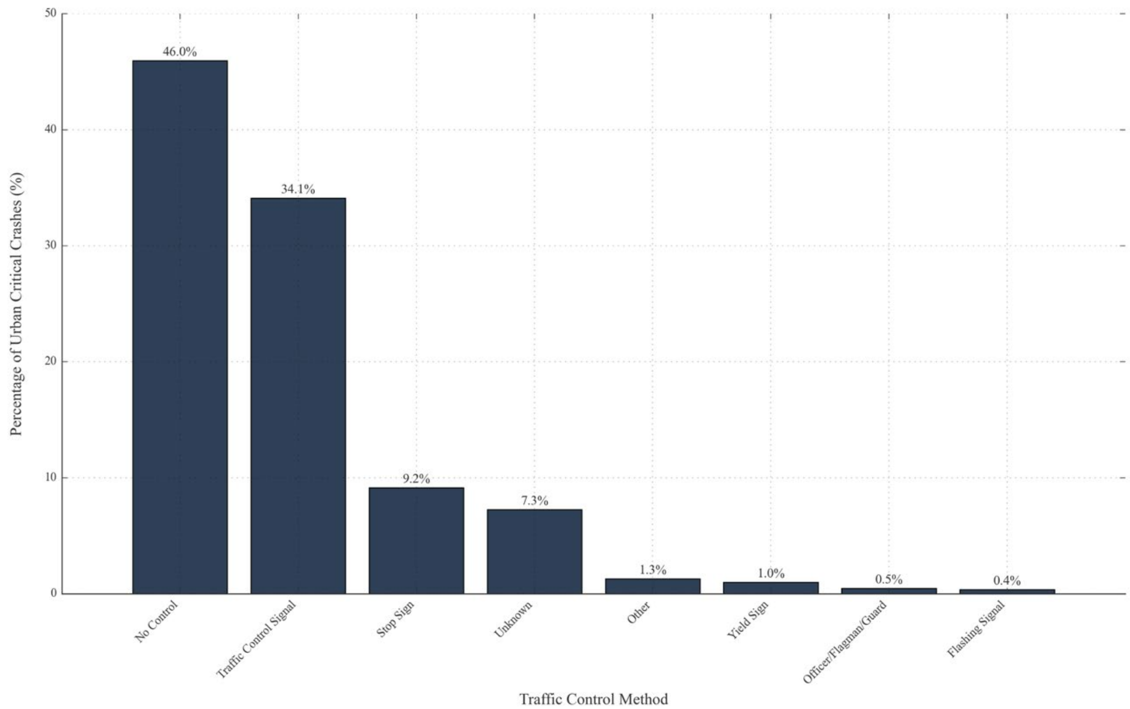

Figure 3-24 Crash distribution by traffic control type at critical sun elevation (urban)

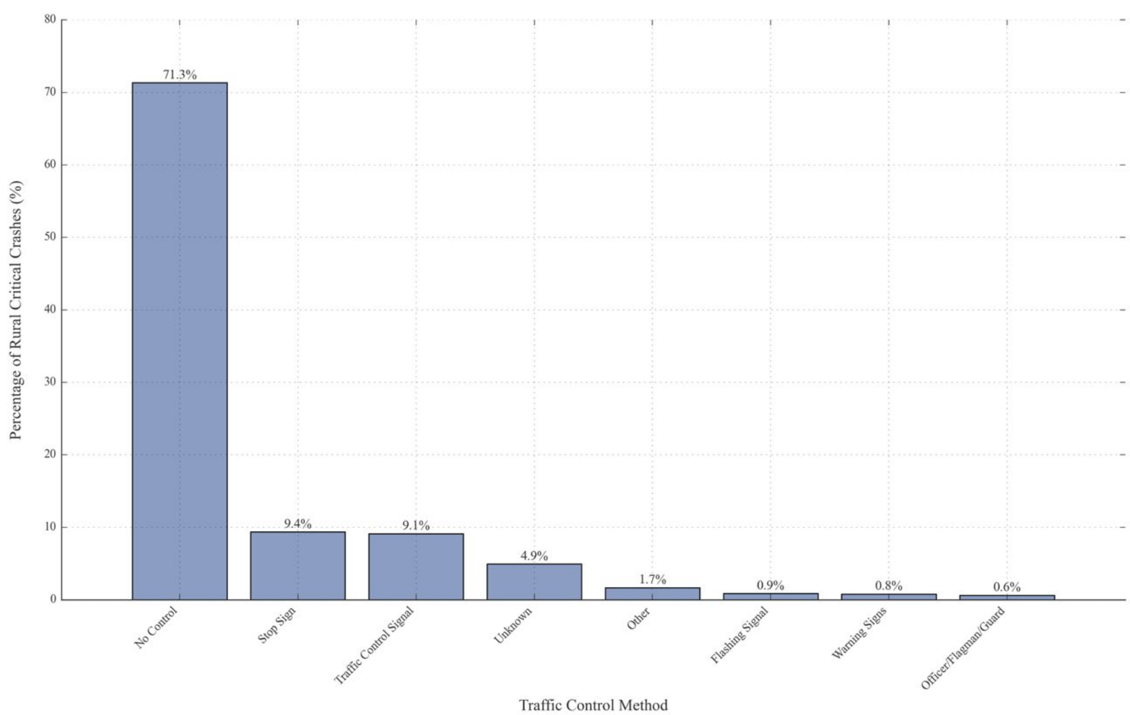

Figure 3-25 Crash distribution by traffic control type at critical sun elevation (rural)

CHAPTER 4. ADVANCED DATA ANALYSIS

The previous chapter presented a basic analysis of crash data. To better understand the relationship among different variables within the database, they need to be analyzed in detail. Traditional crash data analysis alone may not provide valuable insights into road safety issues. Therefore, in this project, advanced computing analysis techniques, such as machine learning, are employed, incorporating additional solar altitude data to enhance our understanding of the factors contributing to road crashes in Alaska. One of the project's goals was to identify risk factors for crashes in Alaska, with a particular focus on the state's unique environmental factors, including extended and limited daylight hours, as well as icy conditions that persist for much of the year. To conduct our analysis, we employed a combination of machine learning and statistical techniques on the state crash database.

This chapter presents the application of advanced tools, including machine learning, to examine various safety aspects of road safety in Alaska. In this section, we address three research questions relevant to the project. They are discussed in detail in each of the following subsections.

4.1. Research Question 1

The first research question was: what are potential risk factors for injury in Alaskan crashes? To determine these factors, we trained four machine learning models to predict Crash Severity, which is the most severe injury that occurred in each crash. Crash Severity had six possible values, listed below in Table 4.1, although we dropped unknown values (U) from the dataset. These injury classifications were made by police officers at the scene of the crash.

Table 4- 1 Crash severity definition

Letter	Definition	Percentage of Crashes
O	No Apparent Injury	64.1%
C	Suspected Possible Injury	13.0%
B	Suspected Minor Injury	14.3%
A	Suspected Serious Injury	2.5%
K	Fatal	0.5%
U	Injury Unknown	5.6%

The first model we trained was a CART decision tree. Due to the severe imbalance in Crash Severity, our first model only predicted the dominant class, O. To combat this, we tried grouping C, B, and A into their own class, but the resulting decision tree still predicted only O. However, when we grouped all classes with injury/fatalities (C, B, A, and K) into one Injury class, and grouped O into a Non-Injury class, the model began to predict both Injury and Non-Injury.

Next, we looked at the feature importance coefficients of the decision tree, and omitted features that had little or no importance for prediction. We narrowed down the best features for injury prediction to Road Surface, Time of Day, and First Sequence Event. We also found that Solar Altitude could be interchanged with Time of Day for identical results, since both are related to daylight. Including both features reduced model accuracy, since they were highly correlated. After hyper tuning the depth and minimum samples per split, the decision tree's best accuracy was 69.5% on both the train and test sets.

The final decision tree model is shown below, visualized using the dtreeviz library. Note that Snow, Ice, and Dry are Road Surface values, and Collision with Animal/Pedestrian/Cyclist and Non-Collision are First Sequence Event values.

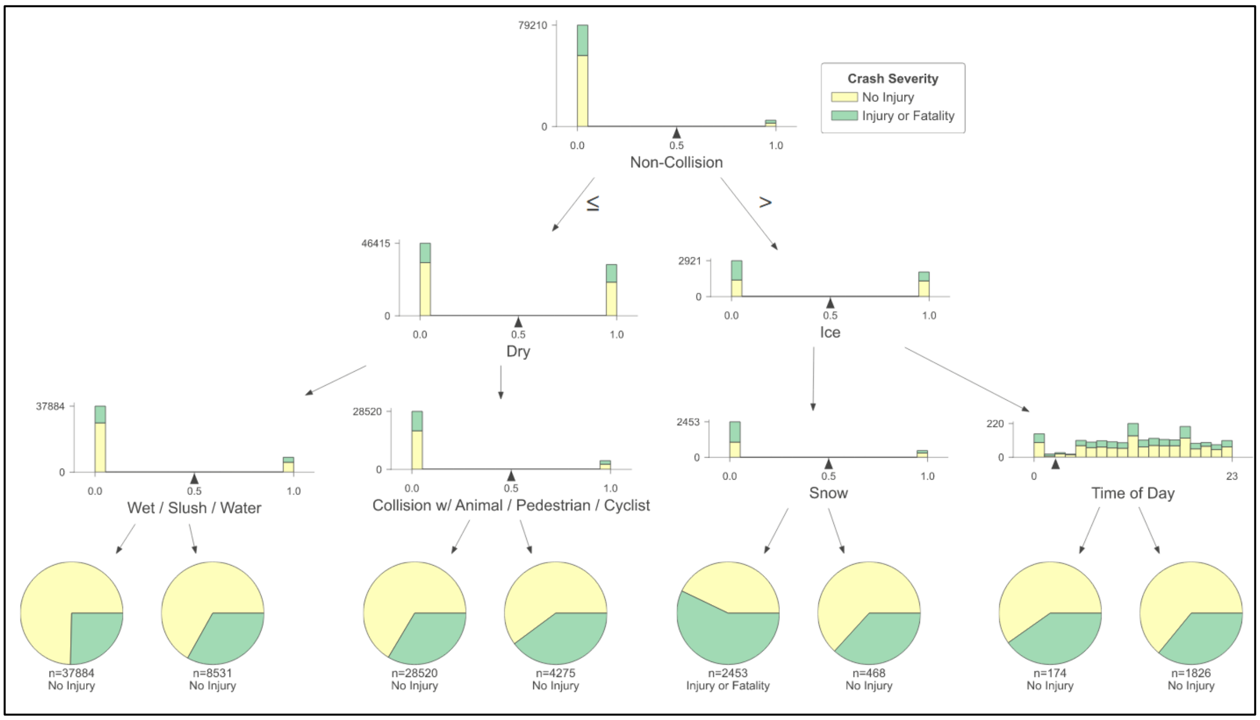

Figure 4-1 Optimal decision tree model for Injury and no-injury crashes

We also trained an XGBoost, Logistical Regression, and Support Vector Machine model. Each model's accuracies and top predictors are listed in the table below. Again, we used feature importance and hyperparameter tuning to get the optimal accuracy for each model below. To find the model's top predictors, we used a combination of visualizing the model (XGBoost) and getting the importance score of each model's features (Logistic Regression). For the SVM model, we inferred the best predictors by finding which ones resulted in the highest accuracy. Table 4.2 summarizes the results.

Table 4- 2 Summary of Decision Tree Analyses

Model	Accuracy	Top Predictors (high to low)
Decision Tree	Train: 69.5% Test: 69.9%	First Sequence Event Road Surface Time of Day
XGBoost	Train: 70.0% Test: 69.9%	First Sequence Event Road Surface Time of Day
Logistic Regression	Train: 55.1% Test: 54.9 %	First Sequence Event Road Surface
Linear Support Vector Machine	Train: 69.5% Test: 69.5%	First Sequence Event Road Surface

4.2. Research Question 2

The second research question was: when is each type of crash likely to occur? We defined crash types by First Sequence Event, which indicated the causal event of the crash. First Sequence Event had five different crash types, listed below, along with definitions and percentage of each class. Again, we dropped the unknown entries from our dataset before analysis. Table 4.3 shows the classes and its definitions.

Similar to the first research question, we trained two machine learning models and determined their top predictors using accuracy and feature importance scores. We initially trained a Decision Tree and XGBoost model to predict all four types of crashes, but ran into the issue of imbalanced data again. The minority class, 'Non-Collision,' was not predicted at all by either model. We tried a combination of oversampling the 'Non-Collision' minority class and undersampling the majority class 'Collision with a Motor-Vehicle In-Transport.' This led to all classes being predicted, but both models had an accuracy loss of 20 to 30 percent, which was not ideal. We also used iterative imputing to create simulated 'Non-Collision' entries, but this also resulted in poor model accuracy.

Next, we tried dividing 'First Sequence Event' into two separate data groups. We grouped 'Collision with a Motor Vehicle In-Transport' and 'Collision with Fixed Object' into one dataset, since these were the two largest classes. 'Collision with an Animal/Pedestrian/Cyclist' and 'Non-Collision' were grouped into another dataset, since they were the minority classes. This resulted in much better model accuracy, although the downside was that we were not comparing all crash types against each other.

Table 4- 3 Classes and definitions for crash types

Class Name	Definition	Percentage
Collision with a Motor Vehicle In-Transport	Causal event of crash involves two motor vehicles colliding	66.9%
Collision with a Fixed Object	Casual event of crash involves vehicle colliding with a fixed object (a sign, guardrail, etc.)	16.6%
Collision with an Animal/Pedestrian/Cyclist	Causal crash event involves vehicle striking an animal, pedestrian, or cyclist.	9.0%
Non-Collision	Causal crash event does not involve a collision of any type (loss of control on road, rollovers, etc.)	5.7%
FSE Other/Unknown	Causal event unknown, or does not fall into the above four categories	1.8%

Again, we used a combination of feature importance scores and hyperparameter tuning to find the optimal accuracy and predictors for each model. The results are listed in Tables 4-4 and 4-5.

Table 4- 4 Summary of models predicting crashes involving multi vehicles and fixed object

Predicting 'Collision with a Motor Vehicle In-Transport' and 'Collision with Fixed Object'		
Model	**Accuracy**	**Top Predictors**
XGBoost	Train: 81.2% Test: 81.0 %	Lighting Time of Day
Decision Tree	Train: 81.4% Test: 81.3%	Lighting Time of Day

Table 4- 5 Summary of models predicting crashes involving animals or non-motorist and non-collision

Predicting 'Collision with an Animal/Pedestrian/Cyclist' and 'Non-Collision'		
Model	**Accuracy**	**Top Predictors**
XGBoost	Train: 72.7% Test: 72.1 %	Lighting Driver Age Road Surface
Decision Tree	Train: 70. 1% Test: 70.0%	Lighting Driver Age Road Surface

Figure 4-2 shows the optimal Decision Tree model for predicting 'Collision with a Motor Vehicle In-Transport' and 'Collision with Fixed Object.'

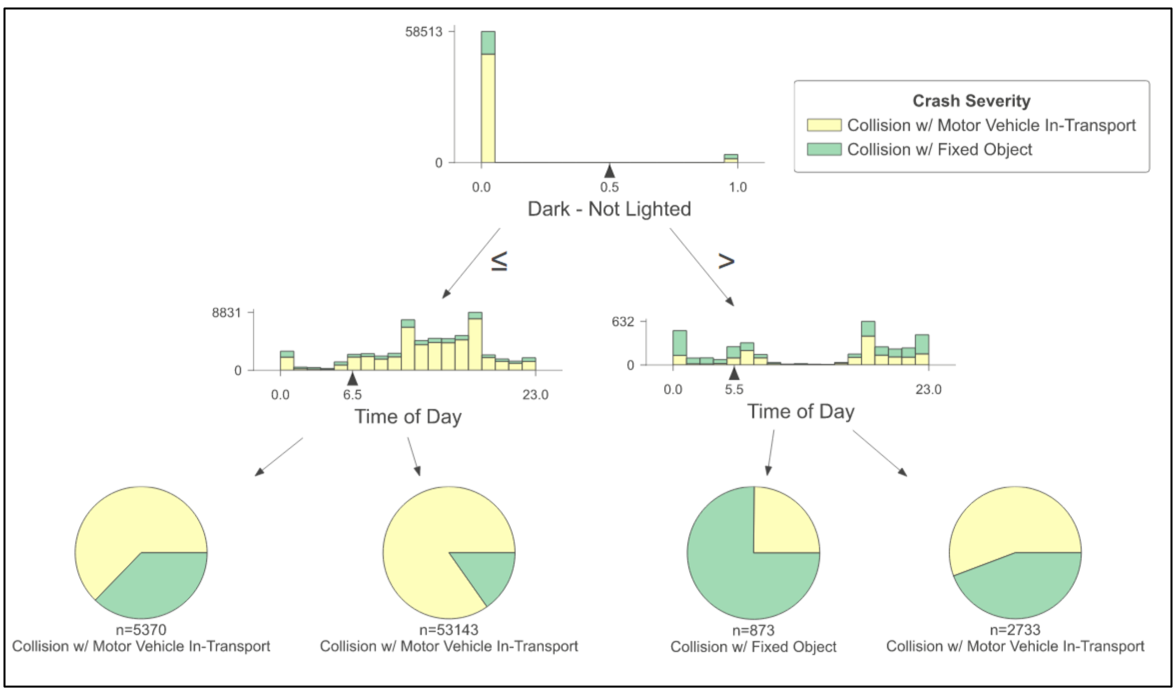

Figure 4-2 Optimal decision tree model for predicting different types of collisions

4.3. Research Question 3

The third research question was: are there factors outside of our dataset that impact crash rates? For this question, we were interested specifically in solar altitude, which was not included in our original data. When the sun has a low solar altitude, it can result in glare and poor visibility, which we hypothesized might have an impact on crash rates.

We first calculated solar altitude for each crash using latitude, longitude, and time of day, and grouped it into bins of ten degrees. Graphing these solar altitude bins showed that low solar altitude seemed to correlate with crash rates. However, solar altitude is not fixed – each month does experience the same solar altitude. Because of this, it was possible that the increased crashes observed at low solar altitudes were merely happening because Alaska experiences more low solar altitude.

To determine if this was the case, the subset of crashes that occured in Anchorage (which was approximately 60% of our dataset) was further analyzed. We then calculated baseline solar altitude tables for each month in Anchorage. These tables showed us the percentages each month spent at different solar altitudes. We also weighted the baseline to account for increased rush hour traffic between 7am – 9am and 4pm – 6pm. This meant that if a specific solar altitude range happened frequently during rush hour, its baseline value was increased, since more cars were driving during that time.

Finally, we compared the number of crashes that occurred at different solar altitudes to the baseline and performed a significance test on our results. We found that solar altitude is indeed correlated with crash rates ($p = 0.02$). The graph of actual crashes by solar altitude compared to the baseline is shown in Figure 4-3.

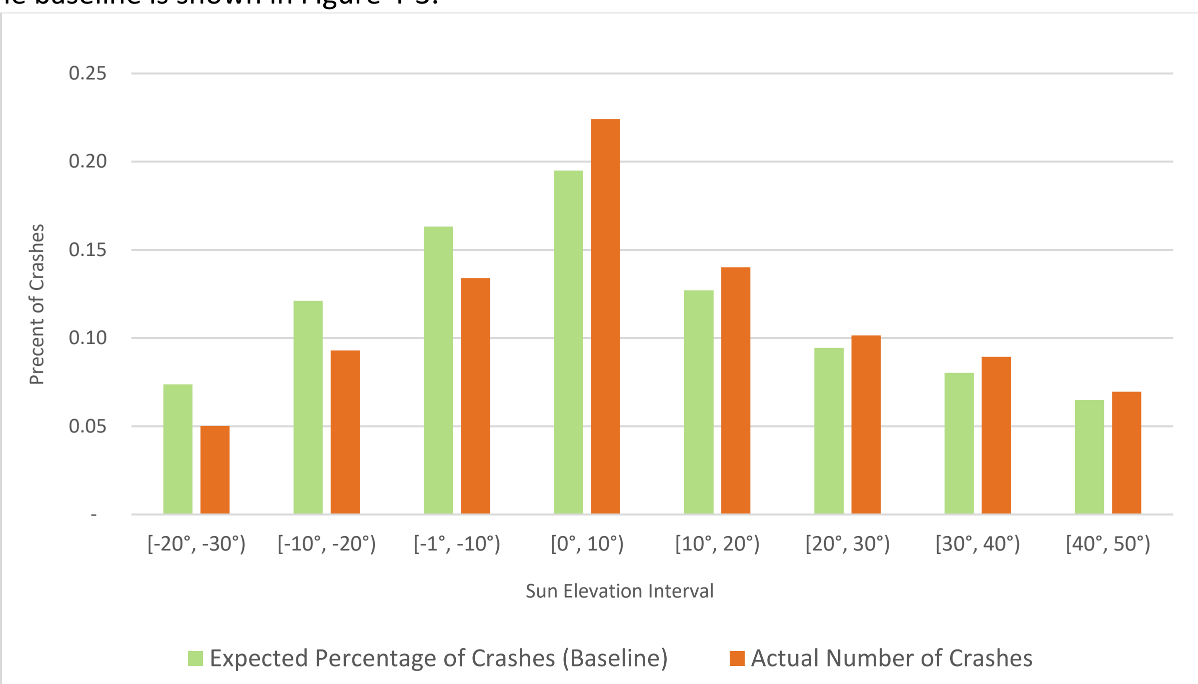

Figure 4-3 Expected Vs actual number of crashes for different sun elevation bins

4.4. Discussions on Results

For the first research question, the XGBoost model had the highest accuracy predicting injury. We found that First Sequence Event, Road Surface, and Time of Day are the top predictors for injury in Alaskan crashes. Specifically, when the First Sequence Event of the crash is 'Non-Collison' and the crash occurs on roads that are not icy or snowy, injury is most likely to occur. We hypothesize that this increased chance of injury could be due to drivers exercising less caution and driving faster, since the road conditions are better.

Conversely, when the First Sequence Event is not a 'Non-Collision' and the roads are not dry, injury is least likely to occur. This is likely due to drivers exercising more caution and driving at slower speeds because the roads conditions are poor. The key limitation of this analysis was that we used certain types of data in our analysis: specifically environmental, road, temporal, and crash-related data. It is possible that there were other, better predictors in the data that we did not include in our analysis.

The second research question tried to predict when each type of crash likely to occur. The analysis found that Lighting and Time of Day were the best predictors for 'Collision with a Motor Vehicle In-Transport' and 'Collision with Fixed Object' crashes. The Decision Tree model was the most accurate at predicting these types of crashes. A collision with a fixed object was most likely to occur between midnight and 5am, on roads that had no lighting. Conversely, a collision with another vehicle was most likely to happen between 6am and 11pm in broad daylight.

Lighting, Driver Age, and Road Surface were the best predictors for 'Collision with an Animal/Pedestrian/Cyclist' and 'Non-Collision' crashes. The XGBoost model had the highest accuracy predicting these crash types. Collisions with animals/pedestrians/cyclists were most likely to occur when it was not dark with no lighting (i.e., during the day or on lit roads) and the driver was over the age of 31. Conversely, non-collisions were most likely to occur when it was dark with no lighting and the roads were wet, icy, or snowy.

The biggest limitation of this analysis was that we had to break crash types (delineated by First Sequence Event) into two datasets due to imbalanced data. This meant that we were not comparing all the crash types to each other. A second limitation is the one mentioned above: since we only used certain types of data in our analysis, there might be other predictors that we missed.

The third question tried to identify if data not listed in the database can be used to analyze the data. This study found that solar altitude is significantly correlated with crash rates ($p = 0.02$), and that the most crashes occur when the sun is between 0 and 9 degrees in the sky, when glare is very high. This study also found that there are more crashes in general during daylight (positive solar altitude) compared to the baseline, and less crashes at night (negative solar altitude) than expected from the baseline.

The biggest limitation to this analysis was not knowing the hourly traffic rates for Anchorage. Although we calculated solar altitude baselines and weighted them to account for rush hour, knowing traffic volume would allow us to control for their impact on crash rates, and know with more certainty that solar altitude was correlated with crash rates.

CHAPTER 5. SUMMARY AND FUTURE WORK

Alaska experiences extreme weather and driving conditions compared to many other locations in the United States. For example, during summer, most traffic activity occurs in daylight, whereas in winter months, the majority of traffic activity occurs in the dark. Similarly, driving conditions change drastically across seasons. During winter, heavy snow and extremely cold temperatures provide challenges for drivers. These conditions lead to safety concerns such as noticeable changes in daylight hours, peak traffic activities in dark hours, and reduced friction values due to snow and ice.

Another challenge is the sun's position above the horizon throughout the year, especially during the spring, summer, and fall seasons. The sun's low elevation angle for an extended period can be a primary concern for drivers since the bright sun glare can make it difficult to see one's surroundings. During summer, long day hours and high activity levels may cause fatigue for drivers. Also, a higher number of tourists during summer may change the traffic conditions, posing higher safety risks in some regions.

This study showed that rural and urban areas face slightly different challenges. For example, when the sun elevation is critical, types of crashes and where they occur differs from urban crashes. Analysis of crash data with sun elevation showed that sun angle and number of crashes are highly correlated.

Machine learning models showed that First Sequence Event, Road Surface, and Time of Day are the top predictors for injury in Alaskan crashes. The second machine learning model illustrated that Lighting, Driver Age, and Road Surface were the best predictors for 'Collision with an Animal/Pedestrian/Cyclist' and 'Non-Collision' crashes. Conversely, non-collisions were most likely to occur when it was dark with no lighting and the roads were wet, icy, or snowy. This makes sense for most of the run-off the roadway crashes or other single vehicle crashes.

One of the major limitations to this study was unavailability of traffic flow for various segments within the state. Although we calculated solar altitude baselines and weighted them to account for rush hour, knowing traffic volume would allow us to control for their impact on crash rates, and know with more certainty that solar altitude was correlated with crash rates.

This study explored some of the key parameters for analysis. In future, it is recommended to include more data in our analysis. It was not possible to include every data field to the high cost of data cleaning and limitations of some of the machine learning models. We only explored sun elevation as an external variable. Even in case of sun angle, horizontal angle, which would have played a crucial role was ignored due to concerns on the accuracy of the crash data. It would be worth exploring the relation of horizontal angle to the crashes. Since Alaska weather is highly variable across state, it would be worth exploring relationship between weather and crashes. Although the weather is one of the fields in the crash data, verifying that and gathering more information from nearest weather stations would be worth exploring.

REFERENCES

Armstrong, K. A., Watling, H., Watson, A., & Davey, J. (2017). Profile of urban vs rural drivers detected drink driving via Roadside Breath Testing (RBT) in Queensland, Australia, between 2000 and 2011. *Transportation Research Part F: Traffic Psychology and Behaviour*, *47*, 114–121. https://doi.org/10.1016/j.trf.2017.03.003

Banerjee, A., Kumar, P. B., Haleem, K., Justice, D., Brittenham, E., & Oliver, J. (2025). Developing Safety Performance Functions for Severe Distraction-Related Crashes along Kentucky's Rural and Urban Two-Lane Roadways. *Transportation Research Record: Journal of the Transportation Research Board*, *2679*(2). https://trid.trb.org/View/2410529

Biglari, S., Kofi Adanu, E., & Jones, S. (2024). A sequel to "Comprehensive analysis of single- and multi-vehicle large truck at-fault crashes on rural and urban roadways in Alabama": Accounting for temporal instability in crash factors. *Accident Analysis & Prevention*, *206*, 107723. https://doi.org/10.1016/j.aap.2024.107723

Champahom, T., Jomnonkwao, S., Watthanaklang, D., Karoonsoontawong, A., Chatpattananan, V., & Ratanavaraha, V. (2020). Applying Hierarchical Logistic Models to Compare Urban and Rural Roadway Modeling of Severity of Rear-End Vehicular Crashes. *Accident Analysis & Prevention*, *141*(0). https://trid.trb.org/View/1707605

Chen, Z., & Fan, W. (David). (2019). Modeling Pedestrian Injury Severity in Pedestrian-Vehicle Crashes in Rural and Urban Areas: Mixed Logit Model Approach. *Transportation Research Record: Journal of the Transportation Research Board*, *2673*(4), 1023–1034. https://doi.org/10.1177/0361198119842825

Factors Contributing to Pedestrian and Bicycle Crashes on Rural Highways. (n.d.). https://www.fhwa.dot.gov/publications/research/safety/10052/10052.pdf

Freightech. (2019). Freightech.us. https://www.freightech.us/post/differences-between-driving-in-urban-and-rural-areas

Hanna, J., Flynn, T., Smith, W., & Tyler, W. (n.d.). *Characteristics of Intersection Accidents in Rural Municipalities*. Retrieved July 10, 2025, from https://onlinepubs.trb.org/Onlinepubs/trr/1976/601/601-016.pdf

Islam, S., & Brown, J. (2017). A comparative injury severity analysis of motorcycle at-fault crashes on rural and urban roadways in Alabama. *Accident Analysis & Prevention*, *108*, 163–171. https://doi.org/10.1016/j.aap.2017.08.016

Islam, S., Jones, S. L., & Dye, D. (2014). Comprehensive analysis of single- and multi-vehicle large truck at-fault crashes on rural and urban roadways in Alabama. *Accident Analysis & Prevention*, *67*, 148–158. https://doi.org/10.1016/j.aap.2014.02.014

Reurings, M., & Janssen, T. (2025). *Accident Prediction Models for Urban and Rural Carriageways Based on Data from The Hague Region Haaglanden*. Trb.org. https://trid.trb.org/View/830457

Se, C., Champahom, T., Jomnonkwao, S., Chaimuang, P., & Ratanavaraha, V. (2021). Empirical comparison of the effects of urban and rural crashes on motorcyclist injury severities: A correlated random parameters ordered probit approach with heterogeneity in means. *Accident Analysis & Prevention*, *161*, 106352. https://doi.org/10.1016/j.aap.2021.106352

Shaw, K. M., West, B., Kendi, S., Zonfrillo, M. R., & Sauber-Schatz, E. (2022). Urban and Rural Child Deaths from Motor Vehicle Crashes: United States, 2015-2019. *The Journal of Pediatrics*, *250*, 93–99. https://doi.org/10.1016/j.jpeds.2022.07.001

Sidhu, C. (n.d.). *Preliminary Assessment of the Increased Speed Limit on Rural Interstate Highways in Illinois*. Retrieved July 10, 2025, from https://onlinepubs.trb.org/Onlinepubs/trr/1990/1281/1281-009.pdf

Thompson, J. P., Baldock, M. R. J., Mathias, J. L., & Wundersitz, L. N. (2013). An examination of the environmental, driver and vehicle factors associated with the serious and fatal crashes of older rural drivers. *Accident Analysis & Prevention*, *50*, 768–775. https://doi.org/10.1016/j.aap.2012.06.028

Uddin, M., & Huynh, N. (2017). Truck-involved crashes injury severity analysis for different lighting conditions on rural and urban roadways. *Accident Analysis & Prevention*, *108*, 44–55. https://doi.org/10.1016/j.aap.2017.08.009

Velloso, M. S., & Jacques, M. A. P. (2012). On-the-spot study of pedestrian crashes on Brazilian Federal District rural highways crossing urban areas. *Transportation Research Part F: Traffic Psychology and Behaviour*, *15*(5), 588–599. https://doi.org/10.1016/j.trf.2012.06.001

Wu, Q., Zhang, G., Zhu, X., Liu, X. C., & Tarefder, R. (2016). Analysis of driver injury severity in single-vehicle crashes on rural and urban roadways. *Accident Analysis & Prevention*, *94*, 35–45. https://doi.org/10.1016/j.aap.2016.03.026

Yook, D., Lee, J., & Sunnie Haam. (2021). Identifying The Differences In The Causal Factors Of Truck -involved crashes in Rural AND Urban areas. *International Journal of Transport Development and Integration*, *5*(3), 231–241. https://doi.org/10.2495/tdi-v5-n3-231-241

Zhang, G., Yang, H., Yu, H., Li, Z., Zou, R., Yuan, R., Prevedouros, P., & Ma, T. (2021.). Extracting Rural Crash Injury and Fatality Patterns due to Changing Clkimates in RITI Communities Based on Enhanced Data Analkysis and Visualizations Tools (Phase II) Final Project Report.. Retrieved July 12, 2025, from https://scholarworks.alaska.edu/bitstream/handle/11122/13024/CSET_1903_Final_Report_Zhang.pdf?sequence=1&isAllowed=y