Project 2343 | November 2023

SJSU SAN JOSÉ STATE UNIVERSITY

MINETA TRANSPORTATION INSTITUTE

Influence of Pavement Conditions on Commercial Motor Vehicle Crashes

Stephen Arhin, PhD, PE, PTOE, PMP, CRA, F. ITE
Babin Manandhar, EIT
Adam Gatiba, PE

MINETA TRANSPORTATION INSTITUTE

Founded in 1991, the Mineta Transportation Institute (MTI), an organized research and training unit in partnership with the Lucas College and Graduate School of Business at San José State University (SJSU), increases mobility for all by improving the safety, efficiency, accessibility, and convenience of our nation's transportation system. Through research, education, workforce development, and technology transfer, we help create a connected world. MTI leads the Mineta Consortium for Transportation Mobility (MCTM) and the Mineta Consortium for Equitable, Efficient, and Sustainable Transportation (MCEEST) funded by the U.S. Department of Transportation, the California State University Transportation Consortium (CSUTC) funded by the State of California through Senate Bill 1 and the Climate Change and Extreme Events Training and Research (CCEETR) Program funded by the Federal Railroad Administration. MTI focuses on three primary responsibilities:

Research

MTI conducts multi-disciplinary research focused on surface transportation that contributes to effective decision making. Research areas include: active transportation; planning and policy; security and counterterrorism; sustainable transportation and land use; transit and passenger rail; transportation engineering; transportation finance; transportation technology; and workforce and labor. MTI research publications undergo expert peer review to ensure the quality of the research.

Education and Workforce Development

To ensure the efficient movement of people and products, we must prepare a new cohort of transportation professionals who are ready to lead a more diverse, inclusive, and equitable transportation industry. To help achieve this, MTI sponsors a suite of workforce development and education opportunities. The Institute supports educational programs offered by the Lucas Graduate School of Business: a Master of Science in Transportation Management, plus graduate certificates that include High-Speed and Intercity Rail Management and Transportation Security Management. These flexible programs offer live online classes so that working transportation professionals can pursue an advanced degree regardless of their location.

Information and Technology Transfer

MTI utilizes a diverse array of dissemination methods and media to ensure research results reach those responsible for managing change. These methods include publication, seminars, workshops, websites, social media, webinars, and other technology transfer mechanisms. Additionally, MTI promotes the availability of completed research to professional organizations and works to integrate the research findings into the graduate education program. MTI's extensive collection of transportation-related publications is integrated into San José State University's world-class Martin Luther King, Jr. Library.

Report 23-38

Influence of Pavement Conditions on Commercial Motor Vehicle Crashes

Stephen Arhin, PhD, PE, PTOE, PMP, CRA, F. ITE

Babin Manandhar, EIT

Adam Gatiba, PE

November 2023

A publication of the
Mineta Transportation Institute
Created by Congress in 1991
College of Business
San José State University
San José, CA 95192-0219

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
23-38		

4. Title and Subtitle	5. Report Date
Influence of Pavement Conditions on Commercial Motor Vehicle Crashes	November 2023
	6. Performing Organization Code
7. Authors Stephen Arhin, PhD, PE, PTOE, PMP, CRA, F. ITE: 0000-0002-1306-4017 Babin Manandhar, EIT: 0000-0003-3623-8347 Adam Gatiba, PE: 0000-0003-3131-0960	**8. Performing Organization Report** CA-MTI-2343
9. Performing Organization Name and Address Mineta Transportation Institute College of Business San José State University San José, CA 95192-0219	**10. Work Unit No.** **11. Contract or Grant No.** 69A3551747127
12. Sponsoring Agency Name and Address U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology University Transportation Centers Program 1200 New Jersey Avenue, SE Washington, DC 20590	**13. Type of Report and Period Covered** **14. Sponsoring Agency Code**

15. Supplemental Notes

10.31979/mti.2023.2343

16. Abstract

Commercial motor vehicle (CMV) safety is a major concern in the United States, including the District of Columbia (DC), where CMVs make up 15% of traffic. This research uses a comprehensive approach, combining statistical analysis and machine learning techniques, to investigate the impact of road pavement conditions on CMV accidents. The study integrates traffic crash data from the Traffic Accident Reporting and Analysis Systems Version 2.0 (TARAS2) database with pavement condition data provided by the District Department of Transportation (DDOT). Data spanning from 2016 to 2020 was collected and analyzed, focusing on CMV routes in DC. The analysis employs binary logistic regression to explore relationships between injury occurrence after a CMV crash and multiple independent variables. Additionally, Artificial Neural Network (ANN) models were developed to classify CMV crash injury severity. Importantly, the inclusion of pavement condition variables (International Roughness Index and Pavement Condition Index) substantially enhanced the accuracy of the logistic regression model, increasing predictability from 0.8% to 41%. The study also demonstrates the potential of Artificial Neural Network models in predicting CMV crash injury severity, achieving an accuracy of 60% and an F-measure of 0.52. These results highlight the importance of considering road pavement conditions in road safety policies and interventions. The study provides valuable insights for policymakers and stakeholders aiming to enhance road safety for CMVs in the District of Columbia and showcases the potential of machine learning techniques in understanding the complex interplay between road conditions and CMV crash occurrences.

17. Key Words	18. Distribution Statement
Commercial Motor Vehicles, Artificial Neural Network, Confusion Matrix, Road Safety, Pavement Condition	No restrictions. This document is available to the public through The National Technical Information Service, Springfield, VA 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	102	

Form DOT F 1700.7 (8-72)

DOI: 10.31979/mti.2023.2343

Mineta Transportation Institute
College of Business
San José State University
San José, CA 95192-0219
Tel: (408) 924-7560
Fax: (408) 924-7565
Email: mineta-institute@sjsu.edu

transweb.sjsu.edu/research/2343

ACKNOWLEDGMENTS

The authors extend their appreciation to the U.S. Department of Transportation's University Transportation Centers Program and the staff of the Mineta Transportation Institute, including Executive Director Karen Philbrick, PhD and Deputy Executive Director Hilary Nixon, PhD, for funding and contributing to this study. The authors also thank the District Department of Transportation for providing International Roughness Index and Pavement Condition Index data for the DC roads used for the study. Any opinions, findings, and conclusions or recommendations in this material are those of the authors and do not necessarily reflect the views of the U.S. Department of Transportation.

Contents

LIST OF FIGURES

List of Tables

Executive Summary

Commercial Motor Vehicles (CMVs) play a crucial role in the movement of goods, serving as a fundamental driver of the national economy. Nonetheless, the growing incidence of CMV accidents has raised concerns regarding road safety, emphasizing the need for a comprehensive examination of the underlying factors. Many studies identify roadway pavement condition as a potential factor that significantly impacts the frequency of accidents.

This report presents the findings of a study that investigated the influence of pavement conditions, specifically the conditions measured by the International Roughness Index (IRI) and Pavement Condition Index (PCI) that serve as indicators of road quality and deterioration, on CMV crash severity in Washington, DC. Furthermore, the study evaluated the CMV crashes that occurred exclusively on the DC CMV routes within the years 2015–2020 with additional focus on crashes on DC interstate roadways. For this purpose, generalized linear models as well as machine learning techniques (Artificial Neural Network) were employed to model the complex relationships between road pavement conditions and CMV crash severity.

The study utilized traffic crash data from the Traffic Accident Reporting and Analysis Systems Version 2.0 (TARAS2) database and pavement condition data maintained by the District Department of Transportation (DDOT) in the District of Columbia. TARAS2 contains various data fields that include vehicle characteristics, environmental conditions, roadway characteristics, traffic exposure characteristics, as well as crash location, date, time, crash type, crash severity, and information on the individuals involved in the crashes. The project team looked at the recent IRI data for the years 2016, 2017, 2019 and 2021, which were provided by officials at DDOT. DDOT collects pavement condition data using state-of-the-art imaging technology on more than 4,300 lane miles of pavement surface annually on most parts of the roadway network.

To understand the key characteristics of the dataset, descriptive statistics such as the mean, median, and frequency of the variables were computed and are included in the analysis results. The project team computed the descriptive statistics for all crashes and separately for crashes that occurred on interstate routes only. The spatial distribution and density of crashes were analyzed using the ArcGIS Pro software program.

Binary logistic regression was utilized to investigate the link between a binary dependent variable and independent variables (specifically on interstate routes). In this study, the dependent variable signifies injury presence in CMV crashes, while the independent variables encompass several factors. This model yields separate coefficient estimates for each independent variable, explaining their correlation strength and direction with injury probability. Finally, Artificial Neural Network (ANN) models were developed to demonstrate varying degrees of success in classifying injury outcomes following a CMV crash.

The results of logistic regression to find the likelihood of injuries following an interstate CMV crash indicated that the best-performing model was not statistically significant, $\chi2(8) = 13.373$, $p < 0.1$. However, the model explained 41.1% (Nagelkerke R^2) of the variance in injury prediction and correctly classified 78.9% of CMV crash occurrences. This model contained both the IRI and PCI datapoints as independent variables.

The ANN results showed that after the training of the developed models and evaluating them with the test dataset, their accuracy ranged between 46% to 60%. One of the models (Model No. 7) produced the best classification with an accuracy of 60%, F-measure (0.52), and sensitivity (0.47). Another model (Model No. 5) was found to be the most precise model with a precision of 0.67.

These outcomes emphasize the influence of road pavement conditions in shaping CMV crash severity. The study not only sheds light on the correlation between road pavement conditions and CMV crash occurrences but also determines the practical implications for road safety measures.

1. Introduction

Road safety is a critical concern for any country, as traffic crashes can lead to significant human, property, and economic losses. In the United States, Commercial Motor Vehicles (CMVs) play a vital role in the transportation of goods and are an essential part of the nation's economy. However, the increasing number of CMV crashes raises concerns about road safety and necessitates an in-depth analysis of the contributing factors. Roadway pavement conditions have been identified in multiple studies as a potential critical variable that influences the occurrence of crashes.

The International Roughness Index (IRI) and Pavement Condition Index (PCI) are widely recognized and standardized measures used to assess the condition of road pavements. The IRI quantifies the roughness of a road surface based on iterative assessments using computer algorithms, whereas the PCI provides a subjective rating of pavement condition based on distresses such as cracks, potholes, and rutting. Both these indices serve as indicators of road quality and deterioration, directly impacting the safety and performance of CMVs.

On average, CMVs account for approximately 15% of the total traffic volume in the District of Columbia (DC). Trucks and buses make up approximately 5% and 10% of the overall traffic in DC, respectively. Crashes involving CMVs, such as trucks or buses, resulting in fatalities, injuries, property damages, or the need for a tow, are required to be reported to the Federal Motor Carrier Safety Administration. In 2020, truck and bus accidents accounted for nearly 8% of all reported crashes (including fatal, non-fatal, and property damage only incidents), based on vehicle classification in DC.

This report presents the findings of a study investigating the influence of road pavement conditions, as measured by the IRI and PCI, on CMV crashes in the District of Columbia. By analyzing and interpreting comprehensive data sets, this study aims at determining patterns, correlations, and potential causal relationships between pavement conditions and CMV crash occurrences. The report presents various models designed to predict crash severity, based on roadway pavement condition alongside other relevant factors.

1.1 Objective

The primary objective of this report is to analyze the influence of roadway pavement conditions, specifically the conditions measured by IRI and PCI, on CMV crashes in the United States. To achieve this, generalized linear models were employed as well as machine learning techniques as powerful computational tools to model the complex relationships between road pavement conditions and crash severity. The research objectives are:

1. Obtain and analyze relevant data related to CMV crashes, road pavement conditions (IRI and PCI), and other relevant variables from the District Department of Transportation.

2. Identify significant patterns and correlations between road pavement conditions and CMV crashes, shedding light on the extent of their influence on road safety.

3. Develop and train models to predict CMV crash occurrences based on road pavement conditions and other contributing factors.

Ultimately, the research endeavors are to offer insights and recommendations to policymakers and the transportation community to promote safer roads, reduce CMV crash rates, and enhance transportation infrastructure for the benefit of all road users.

1.2 Report Organization

This report is divided into six sections. Excluding the introduction, these are:

Section 2: Literature Review – A review of the existing literature and studies related to road safety, CMV crashes, road pavement conditions, and applications of machine learning techniques in transportation research.

Section 3: IRI, PCI and CMV Crash Data – A narrative of the data sources and the processing techniques applied to prepare the data for analysis and modeling.

Section 4: Methodology – An explanation of the applied generalized linear model and machine learning techniques, model selection, training procedure, and evaluation metrics used to determine the relationship between road pavement conditions and CMV crash occurrences.

Section 5: Results and Analysis – A presentation of the findings and thorough analysis of the influence of road pavement conditions on CMV crash events.

Section 6: Discussion – A discussion of the implications of the results, potential contributing factors, and limitations of the study.

Section 7: Conclusion and Recommendations – A summary of the key findings and their implications for road safety, along with recommendations for policymakers and future research directions.

2. Literature Review

The increasing concern about crashes involving heavy trucks, especially CMVs, has emerged as a critical issue for traffic safety practitioners. In the United States, the number of registered vehicles reached an astounding 276,491,174 in 2019, with large trucks and buses comprising approximately 5% (14,080,676 vehicles) of this total. Regrettably, CMVs are responsible for a significant number of crashes resulting in injuries and fatalities on roadways, and account for about 5% of all police-reported crashes nationally.

Statistics from 2016 to 2020 reveal a worrisome trend. While the number of injury crashes involving large trucks and buses decreased by 4% during this period (112,000 in 2016 to 108,000 in 2020), the number of fatal crashes involving these vehicles increased by 5% (4,396 in 2016 to 4,588 in 2020). In the District of Columbia, on average, CMVs contribute about 15% of the traffic volume. In 2020, truck and bus crashes constituted nearly 8% of all accidents in the DC area, including fatal, non-fatal, and property damage cases, underscoring the ongoing significance of CMV-related crashes as a critical traffic safety concern in the District.[1] This trend of CMV crash occurrence does not comply with the DDOT's Vision Zero Goal, which prioritizes human life and health and having zero fatalities (and serious injuries) on the road every year.

The continuing trend of CMV crashes could potentially have safety implications for the supply chain system, which is already under strain due to the COVID-19 pandemic. With the economy gearing towards post-pandemic recovery, it is imperative to prioritize the efficient and secure management of CMVs. As a result, researchers and practitioners across different domains, including academia, industry, and government, have investigated CMV crashes with the goal of providing measures to mitigate the risks associated with CMVs.

2.1 CMV Crash Severity Statistics

The National Highway Traffic Safety Administration (NHTSA) conducted an in-depth investigation into injury and death rates associated with crashes involving large CMVs, aiming to provide comprehensive insights into the prevailing safety concerns. Per the final report, the leading cause of death for individuals aged 3 to 33 in the United States is motor vehicle crashes.[2] Another NHTSA report published in 2015 revealed that while motor vehicle crashes ranked 13th overall

---

[1] "U.S. Gross Domestic Product (GDP) Attributed to Transportation Functions," Bureau of Transportation Statistics, https://www.bts.gov/content/us-gross-domestic-product-gdp-attributed-transportation-functions-billions-current-dollars.

[2] "Rural/Urban Comparison of Traffic Fatalities," *Traffic Safety Fact Sheets: 2020*, U. S. Department of Transportation, National Highway Traffic Safety Administration's (NHTSA) National Center for Statistics and Analysis, (July 2022).

as a cause of death, when categorized according to unintentional injury deaths, they rank in the more alarming position of being the second leading cause of death.[3]

A net decrease in deaths across several areas was observed by NHTSA in 2018; however, there was an unsettling increase in large-truck occupant fatalities. The 33,654 fatal motor vehicle crashes resulted in 36,560 fatalities. It is particularly noteworthy that urban areas accounted for 18,285 (54%) of the fatal traffic crashes and 19,498 (53%) of the fatalities, signifying a 34% increase in urban fatalities from 2009 to 2018. During the same period, there was a 15% decrease in rural fatalities.[ii] Conversely, crashes involving large trucks and buses experienced a substantial increase of 48% in fatalities within the same period. Furthermore, based on the Large Truck and Bus Crash Fact report, injury-related crashes involving large trucks and buses surged by 70% between 2009 and 2018. Moreover, large truck and bus fatalities per 100 million vehicle-miles traveled increased by approximately 8% from 2016 to 2018.[4] These findings emphasize concerning trends in crash-related fatalities involving large commercial motor vehicles and underscore the need for continued vigilance and targeted safety measures to effectively mitigate such risks on the nation's roads.

2.2 International Roughness Index

The quality of road pavements plays a crucial role in ensuring safe and efficient transportation systems. The IRI is a known accurate and practical method for measuring pavement surface condition and is essential for maintaining and improving roadways. It serves as a uniform, calibrated roughness measurement of paved roads and is used to achieve nationwide consistency and comparability in pavement assessment. In accordance with American Society for Testing and Materials E867, roughness is defined as the deviation of a surface from a true planar surface with dimensions impacting vehicle dynamics and ride quality. The IRI is measured in meters/kilometer (or inches/mile) and is a preferred measure of roughness due to several advantages. Using IRI helps ensure that measurements of road roughness are reliable, meaningful, and compatible with international practices. IRI provides a stable and repeatable way to process road surface data, hence the results are credible over time. It also presents a good overall picture of how roughness affects a vehicle's handling and how drivers feel while driving on the road. IRI can be used effectively regardless of the length of road being measured, and it's easy to calculate an average value. Lastly, the IRI aligns with established international standards, facilitating its correlation with other roughness measures. By adopting the IRI as the standard roughness index, federal agencies ensure

---

[3] "Motor Vehicle Traffic Crashes as a Leading Cause of Death in the United States: 2015," U. S. Department of Transportation, NHTSA, (2018).

[4] "Pocket Guide to Large Truck and Bus Statistics 2018," United States Department of Transportation, Federal Motor Carrier Safety Administration, Office of Analysis, Research, and Technology, August 18, 2018, https://doi.org/10.21949/1502788.

a realistic and practical approach to measuring pavement conditions, with significant implications for roadway quality and safety.[5]

Figure 1. Vehicle Equipped with IRI Measuring Instruments

2.3 Relation of the IRI to Road Crashes

Several studies have investigated the relationship between the IRI and crashes to understand how road roughness affects traffic safety. This section provides an overview of these studies.

Elghriany et al. (2015) conducted a study to investigate the relationship between the IRI and crash rates over time, considering changes in pavement conditions. They compiled and analyzed data from various sources in Ohio, utilizing a statistical analysis approach. The study found that a quadratic relationship was most effective in linking crash rates with pavement roughness. The study suggests that an IRI value of approximately 1.50 m/km indicates a safe roadway, while values above 2.25 m/km suggest a roadway susceptible to much higher crash rates.[6] In a similar study, Mamlouk et al. (2018) investigated the correlation between crash rates and pavement ride quality (roughness) and rut depth on highways in Arizona, North Carolina, and Maryland. The researchers collected two main types of data: crash data and IRI and rut depth data from each state's pavement management system database. Crash rates were measured as the number of crashes per 100 million vehicle-miles of travel which were computed using the Federal Highway

[5] "HPMS Public Release of Geospatial Data in Shapefile Format," U.S. Department of Transportation, Office of Highway Policy Information, 2018, https://www.fhwa.dot.gov/policyinformation/hpms/shapefiles.cfm.

[6] Ahmed Elghriany, Ping Yi, Peng Liu, and Quan Yu, "Investigation of the Effect of Pavement Roughness on Crash Rates for Rigid Pavement," *Journal of Transportation Safety & Security* 8, no. 2 (2016): 164–76, https://doi.org/10.1080/19439962.2015.1025458.

Administration's methodology. To analyze the relationship between crash rate and both the IRI and rut depth, the researchers used sigmoidal function regression analysis. The findings indicated that the crash rate remained relatively stable until reaching an IRI value of 210 inches/mile or a critical rut depth of 0.4 inches. Beyond these thresholds, an increase in the IRI or rut depth corresponded to an increase in the crash rate.[7] Levinson et al. (2019) utilized Geographic Information System (GIS) data on crashes and pavement quality from the Minnesota Department of Transportation to examine the relationship between road quality and crashes over a 12-year duration. The analysis revealed that good pavement quality generally reduces crash rates, and also that the frequency of injury/ property damage crashes increased on hillcrests, sags, and during wet conditions. Poor road quality is linked to more frequent property damage and injury crashes.[8] Buck et al. (2021) analyzed data from multiple sources in California to assess road conditions and their impact on traffic outcomes, including crashes and vehicle speed. The researchers merged data on road quality, traffic conditions, crashes, weather, and geographical features. The results show a positive relationship between road roughness and crash rates, as well as a negative relationship between road roughness and speed. These findings suggest that rougher roads lead to reduced traffic safety and increased delays for motorists. The study highlights the need for site-specific planning for pavement maintenance to optimize safety and road quality.[9]

Some studies, however, determined that there is no association of crashes with pavement condition. For example, Baskara et al. (2019) investigated the impact of pavement condition on the number of crashes on a Malaysian highway. The researchers found that more than 70% of accidents occurred on roads in good condition (low IRI value). It should be noted, though, that most of the subject highways in this study were in good condition.[10] In Ontario, Canada, researchers assessed the impact of maintenance treatments on pavement condition, measured by the IRI, on road safety. They used the empirical Bayes before–after methodology to estimate the effects on crashes for arterial and collector roads separately. The results showed statistically significant reductions ($P < 0.05$) in all crashes and property damage only (PDO) crashes, with approximately a 5% and 7% reduction for arterial roads and approximately an 11% and 13% reduction for collector roads after treating the pavement (leading to better IRI), respectively.

---

[7] Micheal Mamlouk, Mounica Vinayakamurthy, Shane Underwood, and Kamil Kaloush, "Effects of the International Roughness Index and Rut Depth on Crash Rates," *Transportation Research Record: Journal of the Transportation Research Board* 2672 (2018), https://doi.org/10.1177/0361198118781137.

[8] David Levinson, Toshihiro Yokoo, and Mihai Marasteanu, "Pavement Condition and Crashes," *Findings*, February 15, 2019, https://doi.org/10.32866/5771.

[9] S. N. Baskara, H. Yaacob, M. R. Hainin, S. A. Hassan, N. Mashros, N. Z. M. Yunus, N. A. Hassan, et al., "Influence of Pavement Condition towards Accident Number on Malaysian Highway," *IOP Conference Series: Earth and Environmental Science* 220, no. 1 (2019): 012008, https://doi.org/10.1088/1755-1315/220/1/012008.

However, there were no significant changes in fatal plus injury crashes for both road types. The findings highlight the importance of considering site-by-site planning for specific pavement maintenance treatments to improve both IRI levels and safety outcomes (Anarkooli et al.).[11]

Although pavement conditions have been found to significantly influence the severity and frequency of crashes, their impact on CMVs is not as pronounced as it is for passenger vehicles.[13] This could be attributed to the fact that CMVs are generally less sensitive to pavement conditions compared to most passenger vehicles.[14] Moreover, CMV drivers typically perform maneuvers more proactively and avoid sudden actions, unlike passenger vehicle drivers. It is also worth noting that pavement conditions have minimal impact on freeway crashes. This finding is likely due to the well-maintained nature of most freeway pavements, and the dataset used for analyzing freeway crashes does not exhibit significant variation in terms of pavement quality.[xii] Further research is required to gain more insight into how pavement conditions directly translate to CMV safety.

2.4 Other Crash Contributing Factors

The studies mentioned above primarily focused on the impact of IRI values on crashes. However, numerous other research works have identified several factors contributing significantly to CMV crashes. Thus, it is evident that no single set of factors can be solely responsible for these incidents. The factors contributing to CMV crashes encompass human elements, environmental conditions, temporal characteristics, vehicle condition, and more. For instance, one study revealed that temporal characteristics, driver/passenger attributes, and road and environmental factors had the most influence on large truck crashes.[15] Another study also identified human error or driver

[10] Jafari Anarkooli, Iliya Nemtsov Alireza, and Bhagwant Persaud, "Safety Effects of Maintenance Treatments to Improve Pavement Condition on Two-Lane Rural Roads—Insights for Pavement Management," *Canadian Journal of Civil Engineering* 48, no. 10 (2021): 1287–1294.

[11] Margaret Bock, Alexander Cardazzi, and Brad R. Humphreys, “Where the Rubber Meets the Road: Pavement Damage Reduces Traffic Safety and Speed,” *NBER Working Papers* (2021).

[12] Yingfeng Li, Chunxiao Liu, and Liang Ding, “Impact of Pavement Conditions on Crash Severity,” *Accident; Analysis and Prevention* 59 (2013): 399–406, https://doi.org/10.1016/j.aap.2013.06.028.

[13] Sikai Chen, Tariq Usman Saeed, and Samuel Labi, “Impact of Road-Surface Condition on Rural Highway Safety: A Multivariate Random Parameters Negative Binomial Approach,” *Analytic Methods in Accident Research* 16 (2017): 75–89, https://doi.org/10.1016/j.amar.2017.09.001.

[14] Mouyid Islam, and Salvador Hernandez, “Large Truck-Involved Crashes: Exploratory Injury Severity Analysis,” *Journal of Transportation Engineering* 139, no. 6 (2013): 596–604, https://doi.org/10.1061/(ASCE)TE.1943-5436.0000539.

behavior as the foremost contributors of CMV crashes.[16] Deliberate actions, such as speeding and the failure to comply with right-of-way rules, have also been recognized as common contributors to CMV crashes.[17] It is important to acknowledge that each location or jurisdiction has its unique characteristics that impact the occurrence and nature of CMV crashes. The following paragraphs provide summaries of various studies that have been conducted to identify the other significant causes of CMV crashes.

Human factors, including demographic, behavioral factors and physiological factors, contribute to crashes. The American Transportation Research Institute Crash Predictor Report utilized data from the Motor Carrier Management Information System and the Commercial Driver's License Information System to forecast the likelihood of future truck crashes. The updated 2018 report includes an examination of the impact of a driver's age and gender on the probability of violations, crashes, and convictions. The findings reveal that drivers with reckless driving and failure to yield violations face a higher risk of crashes. Additionally, prior convictions for failures to signal and failures to stay in the proper lane are linked to increased crash probability. The report also indicates that drivers with a history of past crashes have a 74% higher likelihood of experiencing a future crash. Moreover, the report highlights that male drivers and those younger than 40 or older than 85 are more prone to being involved in crashes. In another study, the researchers examined the connection between crash rates and the sociodemographic characteristics of the zip codes where at-fault drivers lived, focusing on crashes involving CMVs and automobiles separately. They discovered that various socioeconomic factors, such as household income, education level, poverty level, employment status, driver's age, and rurality, played a significant role in predicting at-fault involvement in CMV-related crashes.[18]

Other studies have linked environmental conditions to CMV crashes. One such study conducted by Nail et al. (2016) investigated the impact of weather-related elements on injury severity in single-vehicle truck crashes. They analyzed 1,721 police-recorded single-vehicle truck crashes in Nebraska. The findings revealed that injury severity in single-vehicle truck crashes was influenced by wind speed, rain, humidity, and air temperature. Specifically, higher wind speeds were associated with increased injury severity. Additionally, rain and warmer air temperatures were

[15] Majbah Uddin, and Nathan Huynh, "Factors Influencing Injury Severity of Crashes Involving HAZMAT Trucks," *International Journal Transportation Science Technology* 7, no. 1 (2018): 1–9, https://doi.org/10.1016/j.ijtst.2017.06.004.

[16] Sunanda Dissanayake, and Siddhartha Kotikalapudi, "Characteristics and Contributory Causes Related to Large Truck Crashes (Phase II) – All Crashes," *Res. Innov. Technol. Adm.* (2012).

[17] Shraddha Sagar, Nikiforos Stamatiadis, Samantha Wright, and Aaron Cambron, "Identifying High-Risk Commercial Vehicle Drivers Using Sociodemographic Characteristics," *Accident Analysis & Prevention* 143 (2020): 105582, https://doi.org/10.1016/j.aap.2020.105582.

linked to more severe crash injuries in single-vehicle truck accidents, while higher humidity levels were associated with less severe injuries.[19]

Two significant factors affecting CMV crashes are vehicle size and weight and vehicle maintenance. Large CMVs, such as tractor-trailers and buses, are challenging to maneuver and take longer to stop, making them prone to rollover accidents. Studies have shown that CMVs involved in crashes are often heavier than non-crash-involved vehicles. Poor vehicle maintenance can also lead to crashes as it increases the likelihood of mechanical failures, such as faulty brakes, tire blowouts, and steering problems.

2.5 Methodologies Used to Model CMV Crashes

Crash modeling practices have been conducted using several methodologies, including spatial analysis, generalized linear modeling, and machine learning techniques. This section presents a summary of these three methodologies.

Spatial Analysis of CMV Crashes: Crash spatial analysis is a process that involves the use of GIS and spatial analysis techniques to examine the spatial patterns and distributions of traffic crashes. It is a method used to study the locations and characteristics of crashes on road networks, highways, or specific regions. For example, Shafabakhsh et al. (2017) investigated accident patterns in Mashhad over a 12-month period (March 2011 to March 2012) using GIS-based geoprocessing and spatial temporal analysis methods. They employed kernel density estimation, neighbor analysis, and K-function to identify crash hotspots. The Spatial Analysis on a Network toolbox in the ArcGIS operational environment was used to analyze fatal, injury, and PDO crashes for the presence of clusters. The researchers geocoded crash incidents and examined the spatial distribution of crash data to determine if it followed a random pattern or exhibited a systematic clustered or regular pattern.[20] Also, according to Zhao et al. (2018), Texas had had the highest number of fatal crashes involving large trucks in the United States since 1994. In response to this finding, the researchers decided to investigate the risk factors associated with Texas roadways to improve their safety and suitability for large trucks. To achieve this, they conducted a risk analysis using historical large-truck crash data from five years (2011 to 2015) obtained through the Texas Department of Transportation's Crash Records Information System. The crash data were analyzed

[18] Bhaven Naik, Li-Wei Tung, Shanshan Zhao, and Aemal J. Khattak, "Weather Impacts on Single-Vehicle Truck Crash Injury Severity," *Journal of Safety Research* 58 (2016): 57–65, https://doi.org/10.1016/j.jsr.2016.06.005.

[19] Gholam Ali Shafabakhsh, Afshin Famili, and Mohammad Sadegh Bahadori. "GIS-Based Spatial Analysis of Urban Traffic Accidents: Case Study in Mashhad, Iran," *Journal of Traffic and Transportation Engineering (English Edition)* 4, no. 3 (2017): 290–299, https://doi.org/10.1016/j.jtte.2017.05.005.

using GIS, and a crash diagram was created to identify hotspots where similar large-truck crashes occurred frequently.[21]

Generalized Linear Modeling (GLM) of CMV Crashes: GLM is a statistical framework used for modeling the relationship between a response variable and one or more predictor variables. It is a generalization of linear regression, allowing for more flexibility in handling data that may not meet the assumptions of traditional linear models. GLM can be applied to a wide range of data types and response distributions. Examples of GLMs include Binary Logistic Regression, Poisson Regression, and Multinomial Logistic Regression. Hao et al. (2016) examined the impact of injury severity in truck crashes on highway-rail crossing accidents. The authors employed an ordered probit model to analyze the factors influencing injury severity of truck drivers in highway-rail grade accidents. Two ordered Probit models were estimated based on vehicle type: one for the truck group and another for the non-truck group. The ordered Probit approach was chosen because the dependent variable had multiple categories (PDO, injury, and fatality) with a natural order. The study utilized ten years of crash data from the Federal Railroad Administration's accident database, starting in 2002. The results indicated that the injury severity models for truck and non-truck drivers differed significantly with a confidence level exceeding 99.5%.[22]

Machine Learning Techniques in Modeling CMV crashes: In addition to conventional modeling approaches for crashes, numerous studies have delved into employing machine learning techniques to build crash prediction models. Traditional regression methods often rely on predefined assumptions, which may hinder their capacity to effectively capture and interpret the relationships between independent and dependent variables. Consequently, researchers have turned to machine learning techniques, specifically unsupervised learning, known for their ability to robustly identify patterns in unlabeled or unseen data and make predictions about future states or conditions. For example, Arhin and Gatiba (2019) employed Artificial Neural Networks (ANNs) to predict injury severity resulting from angle crashes at unsignalized intersections. Their study utilized a dataset of 3,307 crashes that occurred between 2008 and 2015 to develop 25 distinct ANN models. The most accurate model achieved a rate of 85.62% accuracy and was constructed with three hidden layers containing 5, 10, and 5 neurons, respectively.[23] While similar studies have explored the application

---

[20] Qun Zhao, Tyrie Goodman, Mehdi Azimi, and Yi Qi, "Roadway-Related Truck Crash Risk Analysis: Case Studies in Texas," *Transportation Research Record* 2672, no. 34 (2018): 20–28, 2018, https://doi.org/10.1177/0361198118794055.

[21] Wei Hao, Camille Kamga, Xianfeng Yang, JiaQi Ma, Ellen Thorson, Ming Zhong, and Chaozhong Wu, "Driver Injury Severity Study for Truck Involved Accidents at Highway-Rail Grade Crossings in the United States," *Transportation Research Part F: Traffic Psychology and Behaviour* 43 (2016): 379–386, https://doi.org/ 10.1016/j.trf.2016.09.001.

[22] Stephen A. Arhin, and Adam Gatiba, "Predicting Injury Severity of Angle Crashes Involving Two Vehicles at Unsignalized Intersections Using Artificial Neural Networks," *Engineering,*

of ANN to model crashes, few, if any, have specifically applied it to CMV crashes and the IRI within a highly urbanized setting. Further, Liu et al. (2022) explored the application of machine learning methods to predict the severity of large-truck crashes in the state of Texas using crash records from 2016 to 2019 obtained from the Texas Crash Records Information System. Six machine learning (ML) methods were selected for the prediction task, including four classification tree-based models (Extreme Gradient Boosting tree, Adaptive Boosting tree, Random Forest, and Gradient Boost Decision tree), and two non-tree-based models (Support Vector Machines and K-Nearest Neighbors). The study compared the accuracy levels of all six methods, and the tree-based models demonstrated better performance than the non-tree-based ones. Among the six models, the Gradient Boost Decision tree showed the best prediction performance. While a wide range of ML techniques was employed, the study suggests that exploring other robust models such as tree-based model ANNs could be beneficial.[24]

Summary

As the review of the literature attests, crashes involving CVMs continue to be a significant proportion of crashes in the United States. The IRI serves as a reliable method for measuring pavement surface condition and achieving nationwide consistency in pavement assessment. Several studies have explored the relationship between IRI values and crashes, revealing that higher IRI values are associated with roads susceptible to higher crash rates, while lower values indicate safer roadways. Generally, good pavement quality reduces crash rates. Further research is needed to understand how pavement conditions directly affect CMV safety.

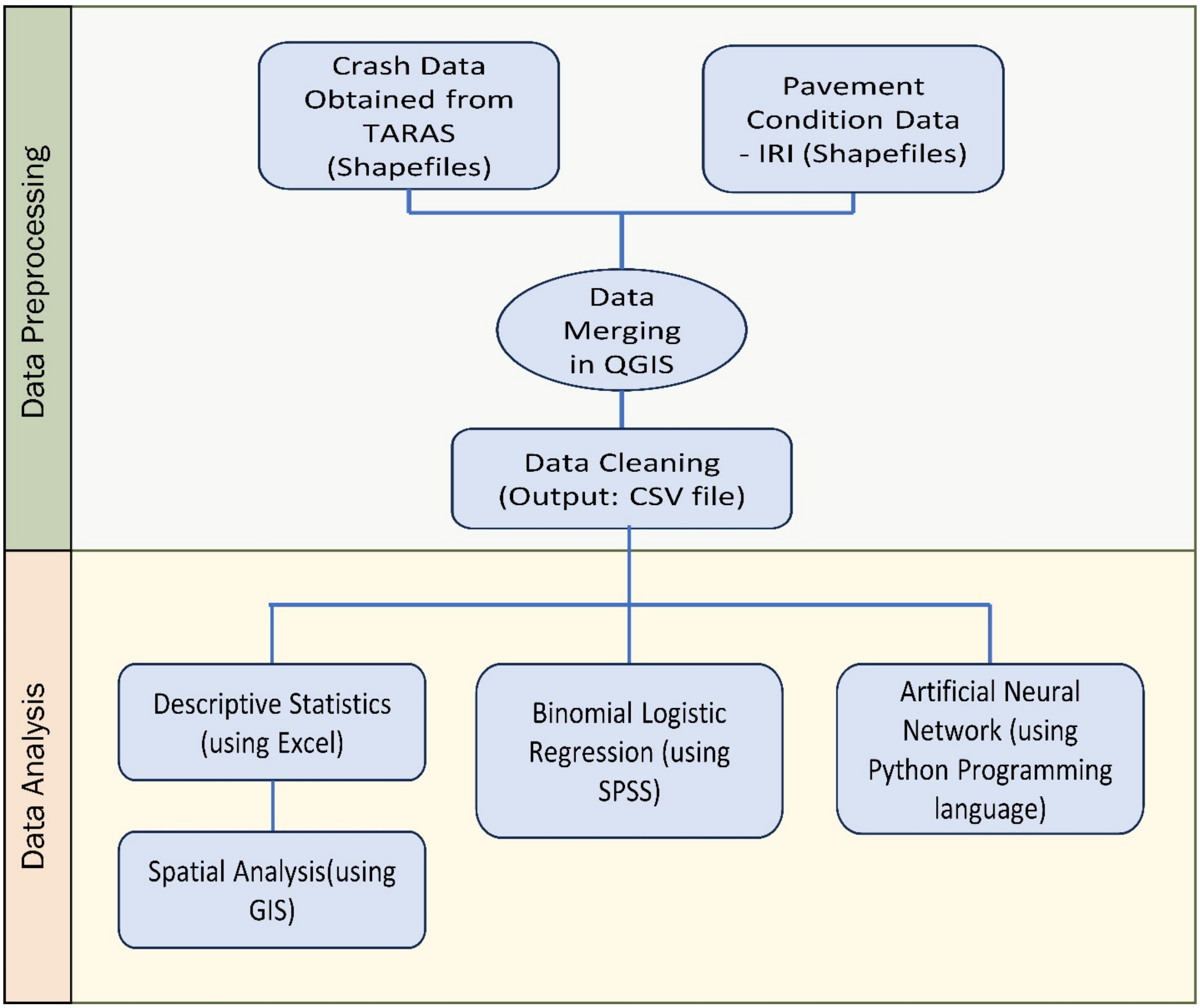
Technology & Applied Science Research 9, no. 2 (2019): 3871–80, https://doi.org/10.48084/etasr.2551.

[23] Jinli Liu, Yi Qi, Jueqiang Tao, and Tao Yueqing, "Analysis of the Performance of Machine Learning Models in Predicting the Severity Level of Large-Truck Crashes," *Future Transportation* 2, no. 4 (2022): 939–55, https://doi.org/10.3390/futuretransp2040052.

3. Research Methodology

This chapter of the report presents the methodology used in analyzing the influence of the IRI on CMV crashes that occur on designated CMV routes in the District of Columbia. Figure 2 shows an overview of the methodology. The details of each step are provided in the following sections of this chapter.

Figure 2. Crash Analysis Methodology



3.1 Description of Study Jurisdiction

This study is based on crash data and pavement condition data (IRI data) collected in DC, which is divided into four unequal quadrants: Northwest (NW), Northeast (NE), Southeast (SE), and Southwest (SW). DC consists of eight wards. As of 2020, the city's population was approximately 689,545, which is a reduction of about 2.9% from the previous year (Census Bureau) (DC Policy).[25] This reduction is attributable to the movement of people out of the city due to the COVID-19 pandemic. Highly urbanized, DC is ranked as the sixth most congested city in the United States, with each driver spending an average of 63 hours stuck in traffic annually. Covering an area of 68.34 square miles, the city has a vast network of 1,503 miles of roads, comprised of local roads, collector roads, minor arterials, principal arterials, freeways, and interstates.[27] It is concerning to note that the American Society of Civil Engineers' 2017 infrastructure report card reveals that around 95% of the roads in DC are in poor condition.[28] Figure 3 presents a map of DC while Figure 4 presents the network of designated CMV routes in DC.

---

[24] "The District of Columbia Gained More Than 87,000 People in 10 Years," American Counts Staff, United States Census Bureau, August 25, 2021, https://www.census.gov/library/stories/state-by-state/district-of-columbia-population-change-between-census-decade.html.

[25] Sunaina Bakshi Kathpalia, "Charts of the Week: A Pandemic-Induced Exodus Has Broken the District's Population Boom" *D.C. Policy Center*, March, 25, 2022, https://www.dcpolicycenter.org/publications/census-shows-pandemic-exodus-has-broken-dc-population-growth/.

[26] "2016 District of Columbia Infrastructure Report Card," ASCE, 2017, https://2017.infrastructurereportcard.org/state-item/district-of-columbia/.

Figure 3. Map of DC

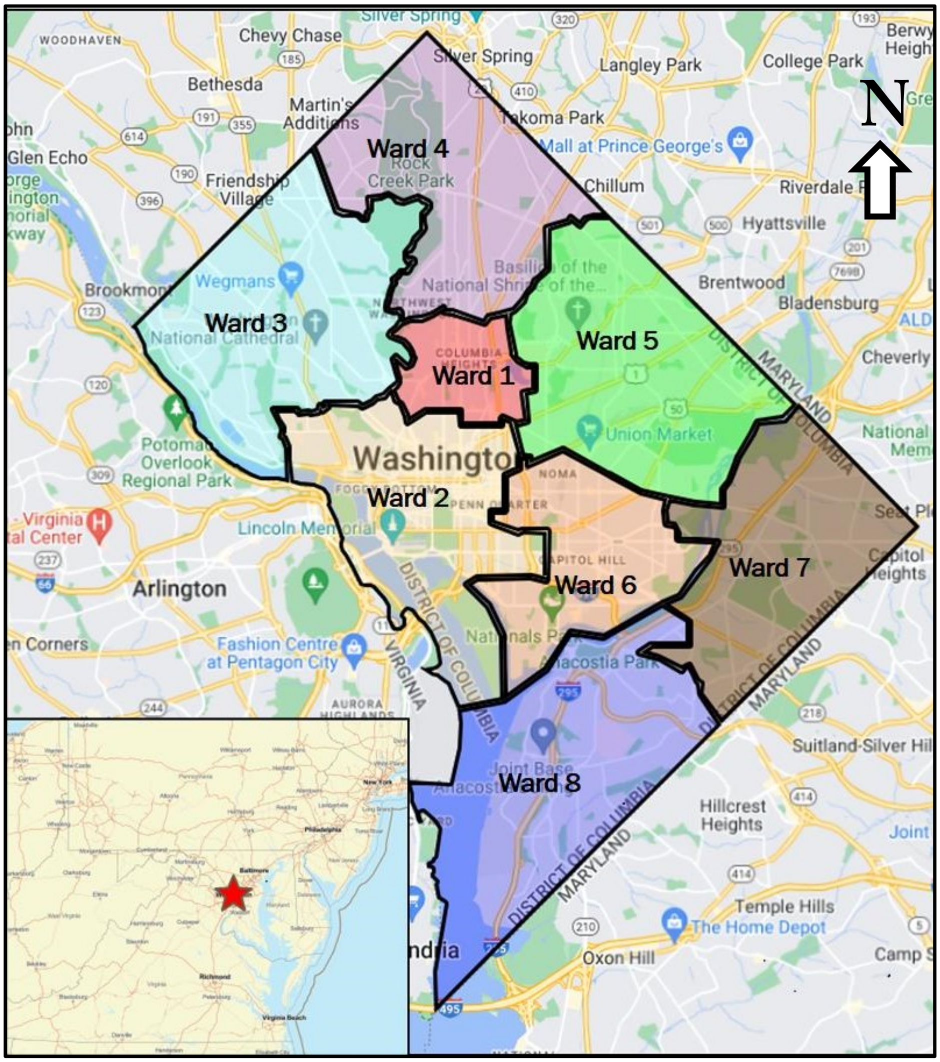
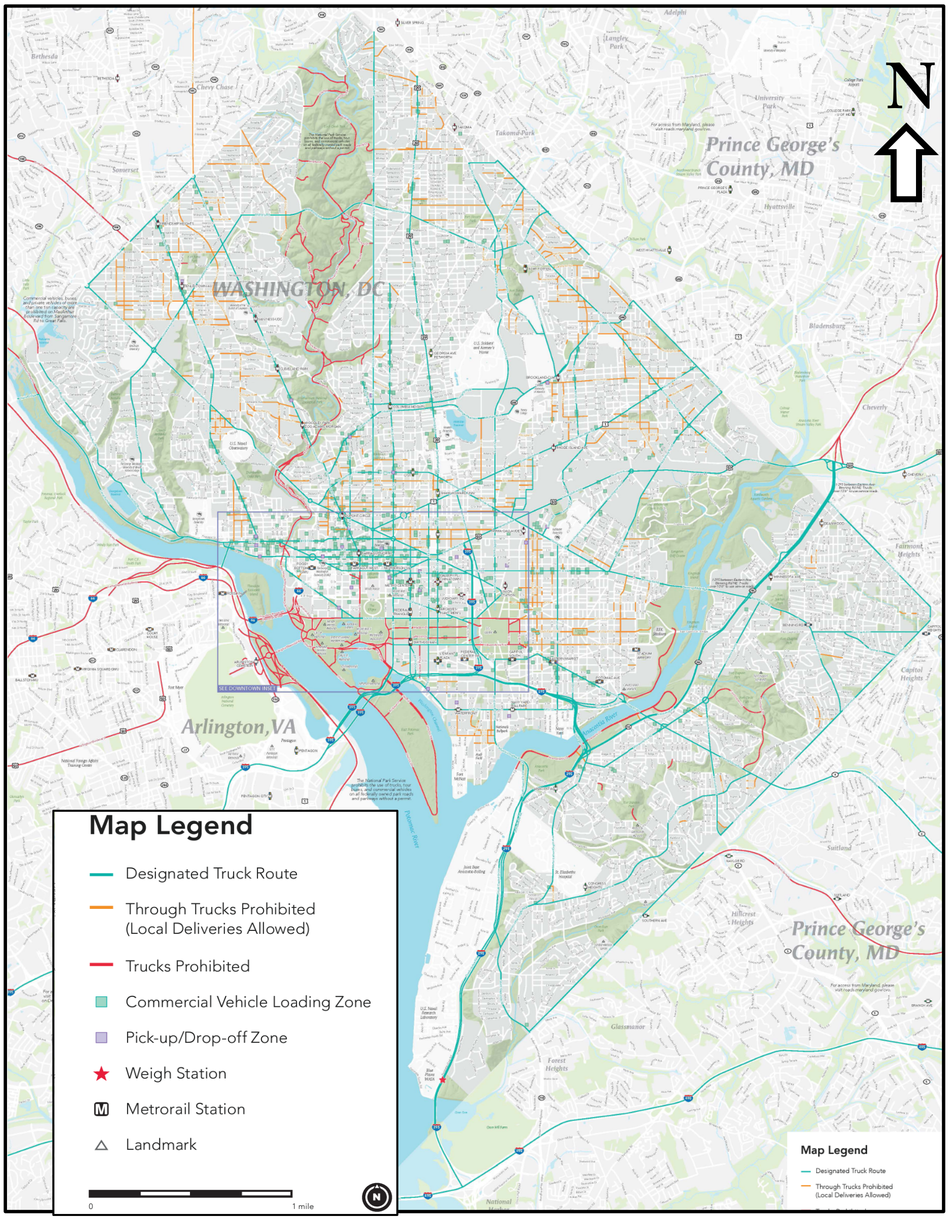


Figure 4. Designated CMV Routes in DC



3.2 Crash and IRI Data Sources and Management

Crash prediction models heavily rely on the data they are built upon, making the accuracy of these models contingent on the quality of available data. To ensure the development of a reliable model, this study utilized traffic crash data from the Traffic Accident Reporting and Analysis Systems Version 2.0 (TARAS2) database and pavement condition data maintained by the District Department of Transportation (DDOT) in DC. These data sources are described below.

3.2.1 Crash Database (TARAS)

The District of Columbia Metropolitan Police Department (DCMPD) electronically records traffic crash information at crash scenes using the Police Department Form number 10 crash reporting form. Subsequently, this crash data is transferred from the DCMPD to the DDOT's database through dedicated servers. In TARAS2, an Oracle-based application, the data is processed and made available for analysis. TARAS2 contains various data fields that can be broadly categorized into vehicle characteristics, environmental conditions, roadway characteristics, and traffic exposure characteristics, as well as crash location, date, time, crash type, crash severity, and information on the individuals involved in the crashes. For the purposes of this research, data was collected for CMV crashes that occurred on designated CMV routes from the years 2016 to 2020. Crash data were obtained in the form of GIS shapefiles.[29]

3.2.2 Pavement Condition Data

Recent IRI data for the years 2016, 2017, 2019, and 2021 were obtained from the DDOT for this study. The DDOT collects pavement condition data using state-of-the-art imaging technology on more than 4,300 lane miles of pavement surface annually on most parts of the roadway network. It should be noted, however, that data was not collected for the year 2018. IRI data were obtained in the form of GIS shapefiles.

3.2.3 Crash and IRI Data Management

To efficiently manage the data integration and analysis, the research team employed QGIS, an open-source geographic information system known for its powerful geospatial capabilities. The system is capable of handling various types of geographical data and performing spatial analysis. The crash data consist of point data while the pavement condition data consist of segment data. These two datasets were matched and aligned. Thus, the crash points/locations were superimposed on their corresponding mile sections of pavement condition data. The output of this process file is a comma-delimited file containing crash data merged with IRI data.

3.2.4 Data Processing

The data obtained from QGIS were further processed by identifying and removing duplicate and incomplete crash records as well as irrelevant data fields. It should be noted that the dataset contained CMV crashes on interstate and non-interstate CVM designated routes. There were two datasets that were obtained from the data processing. The first data set contains all crashes on both interstate and non-interstate CVM designated routes, while the second dataset contains crashes on interstate routes only. Literature has shown that the standard speed at which IRI data are most

---

[27] "Traffic Data," Metropolitan Police, Washington, DC (n.d.), https://mpdc.dc.gov/page/traffic-data.

accurate is about 50 mph. This is because the IRI is sensitive to the same profile wavelengths that cause vehicle vibrations in normal highway conditions.[30] Additionally, visual inspection of the data revealed several inconsistencies between PCI data and IRI data at low speeds (< 50 mph), corroborating the literature. Because of this, while the first dataset was used to understand the overall characteristics of CMV crashes, only the second dataset was used in model development, as explained in the following sections.

3.3 Descriptive Statistics

To understand the key characteristics of the dataset, descriptive statistics such as the mean, median, and frequency of the variables were computed and are included in the results of analysis. These statistics assist in interpreting the data before further analysis is conducted. Descriptive statistics were computed for all crashes and separately for CMV crashes that occurred on DC interstate routes only.

3.4 Spatial Analysis

The spatial distribution and density of crashes were analyzed using the ArcGIS Pro software program. ArcGIS Pro is a geographic information system tool used for creating maps with geographical data and for analyzing mapped information. The coordinates of each crash location are provided in the dataset. These coordinates, together with existing base maps/layers, were used to perform a spatial analysis, including a spatial distribution analysis of crashes based on injury severity and a kernel density analysis for injury crashes.

3.5 Binary Logistic Regression

Binary logistic regression is a statistical method employed to explore the relationship between a binary dependent variable and one or multiple independent variables. In the context of this study, the dependent variable is the occurrence or otherwise of an injury within CMV crashes, while the independent variables include multiple factors, which are described in the following subsection. Through the logistic regression model, distinct coefficient estimates are generated for each independent variable, indicating the strength and direction of the correlation between the predictor (independent variable) and the probability of injury (dependent variable). As explained in Subsection 3.2.4, only crashes that occurred on interstate routes were used for this statistical analysis.

---

[28] Michael W. Sayers, Thomas D. Gillespie, and Cesar A. V. Queiroz, *The International Road Roughness Experiment: Establishing Correlation and a Calibration Standard for Measurements* (University of Michigan, Ann Arbor: Transportation Research Institute, 1986).

3.5.1 Description of Variables

The variables used to develop the models were chosen based on which predictors are identified as significant in previous literature. The crash severity (Injury/No Injury) following a CMV crash is the dependent variable, and the independent variables used in the study include the following:

Month: The crash database contains records of CMV crashes occurring in every month for the 5-year period (2016–2020). The results have detailed the monthly crash frequencies.

Period of the Day: For the purposes of the analysis, the time of reported crashes was categorized into five periods: AM Peak (6 AM–10 AM), Off Peak (10 AM–3 PM), PM Peak (3 PM–7 PM), Night (7 PM–12 AM), and Dawn (12 AM–6 AM).

Day of the Week: This variable contains the day of the week on which a CMV crash occurred.

Lighting Type: The status of the streetlights for every CMV crash record was categorized into four categories: on, off, other, and unknown.

Lighting Condition: This variable describes the lighting conditions at the time of the CMV crash. These conditions are Daylight, Dark-Lighted, Dark-Not Lighted, Dawn, Dusk, Other, and Unknown.

Road Condition: This variable describes the pavement condition at the time of the CMV crash. For the purposes of the study, two main pavement condition categories were identified: wet and dry conditions.

Speed Limit: The CMV crash dataset contains the speed limit of the road at the crash location. These speed limits were segmented into three categories: low (≤ 25 mph), medium (25–55 mph), and high (> 55 mph).

International Roughness Index (IRI): This is a numerical measure of the roughness of the road as measured with an instrumented vehicle.

International Roughness Index (IRI) Category: This variable is a grouping of the PCI values into three categories, namely, rough, marginal, and smooth.

Pavement Condition Index (PCI): This is a numerical index value ranging between 0 and 100 indicating the general condition of the roadway pavement section on which the CMV crash occurred.

Pavement Condition Index (PCI) Category: This variable is a grouping of the PCI values into five categories, namely, poor, fair, satisfactory, good, and excellent.

3.5.2 Model Evaluation for Binomial Regression Analysis

The goodness-of-fit was used to assess the model's predictive performance. Specifically, the Deviance and Pearson chi-square tests and the Hosmer-Lemeshow goodness-of-fit test were used to evaluate the regression model.

Deviance and Pearson chi-square tests:

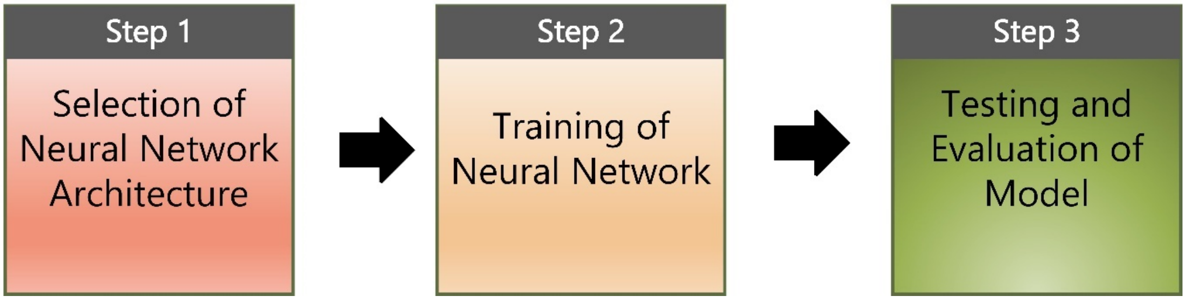
The frequencies of the observed values and the predicted values were compared to determine the overall fit of the regression model. The deviance chi-square test measures the goodness-of-fit based on the difference between the observed and expected frequencies, while the Pearson chi-square test measures the goodness-of-fit based on the difference between the observed and expected frequencies divided by the expected frequencies.

Hosmer-Lemeshow goodness-of-fit test:

The goodness-of-fit indicates how well the data fits the model. In this test, the data sample is divided into multiple groups based on their predicted probabilities, after which the observed and expected frequencies are compared in each group. The chi-square obtained is evaluated to determine if the logistic regression model has a good fit. A large chi-square value indicates a poor fit while a small chi-square value indicates a good fit.

3.6 Artificial Neural Network

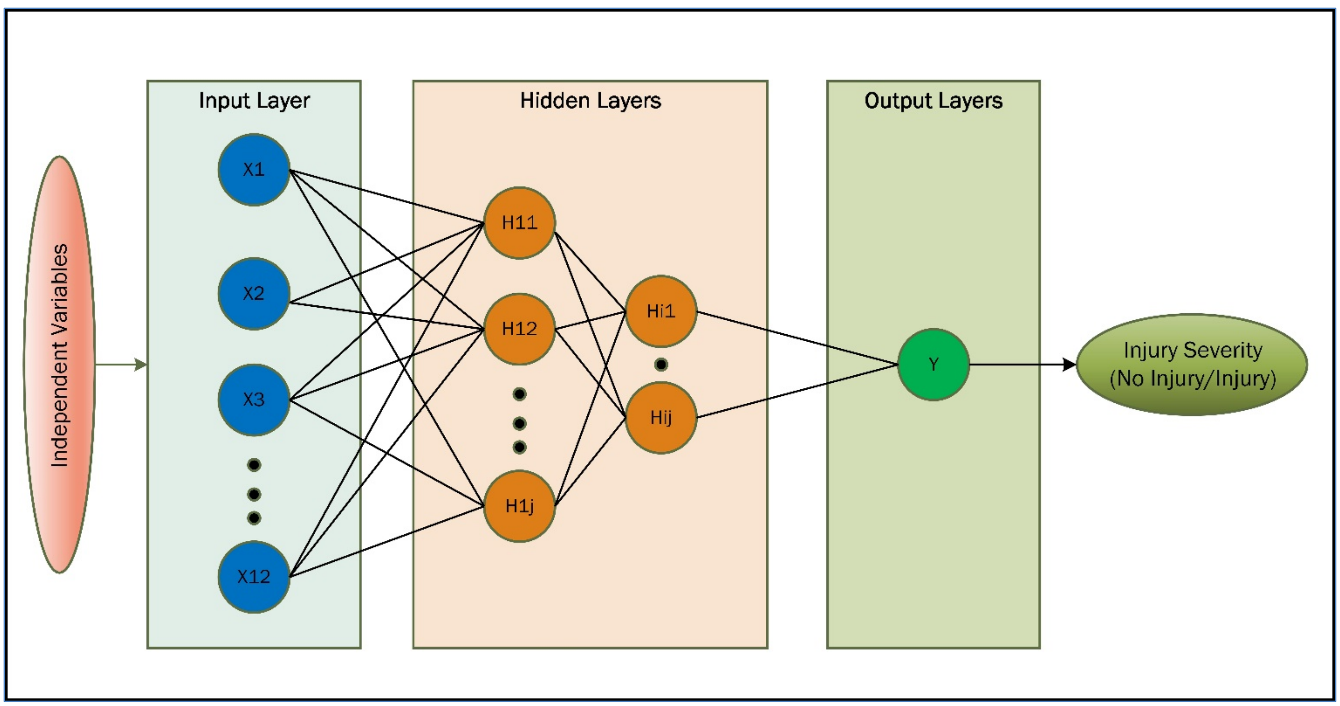
The concept of an Artificial Neural Network (ANN) is fundamental in machine learning. ANNs are inspired by the structure and function of biological neural networks in the human brain and are widely used for various machine learning tasks, including classification, regression, pattern recognition, and more. They consist of interconnected nodes, often referred to as neurons or units, organized into layers. Each neuron processes and transmits information to other neurons based on a set of learned weights and biases. The method by which the weights and bias levels of a network are updated is determined by the learning rule used, which in this study is the multilayer perceptron (MLP). The MLP basically consists of three layers: the input layer, the hidden layer, and the output layer. It is a feedforward learning rule in which information flows from the input layer through the hidden layer to the output layer to produce the outcome or results. The following steps provide a detailed description of how models for classifying CMV crash injury severity were developed using an ANN.



Step 1: Selection of Neural Network Architecture

For the first step of the neural network training, a MLP feedforward ANN was set up with at least an input layer, a hidden layer, and an output layer. Each layer comprises interconnected nodes or neurons, and the hidden and output layer neurons possess nonlinear activation functions. The architecture also includes varying numbers of hidden layers and neurons, which were adjusted until the configuration which yielded the optimal results was obtained. Figure 5 is a depiction of the MLP ANN architecture used in developing the models.

Figure 5. Multilayer Perceptron Artificial Neural Network Model



Step 2: Training of Neural Network

To train the neural network using backward propagation, the following sequence of sub-steps was followed.

Step 2a: Input of Training Dataset into the Network – The training dataset was imported into the network to begin training. The vector of independent variables was fed into each input neuron, which was then connected to the neurons of the first hidden layer. Note that, as explained in Subsection 3.2.4, only crashes that occurred on interstate routes were used for model development. The training process was initiated by randomly selecting weights for all interconnections between the neurons of the input and hidden layers.

Step 2b: Forward Computation – After initializing the weights, forward propagation is executed by multiplying the weights by the input neuron values, and the resulting sum products are saved in the corresponding hidden layer neurons. The weighted sums are then passed through an activation function and, based on the output of the function, the activation of hidden neuron.

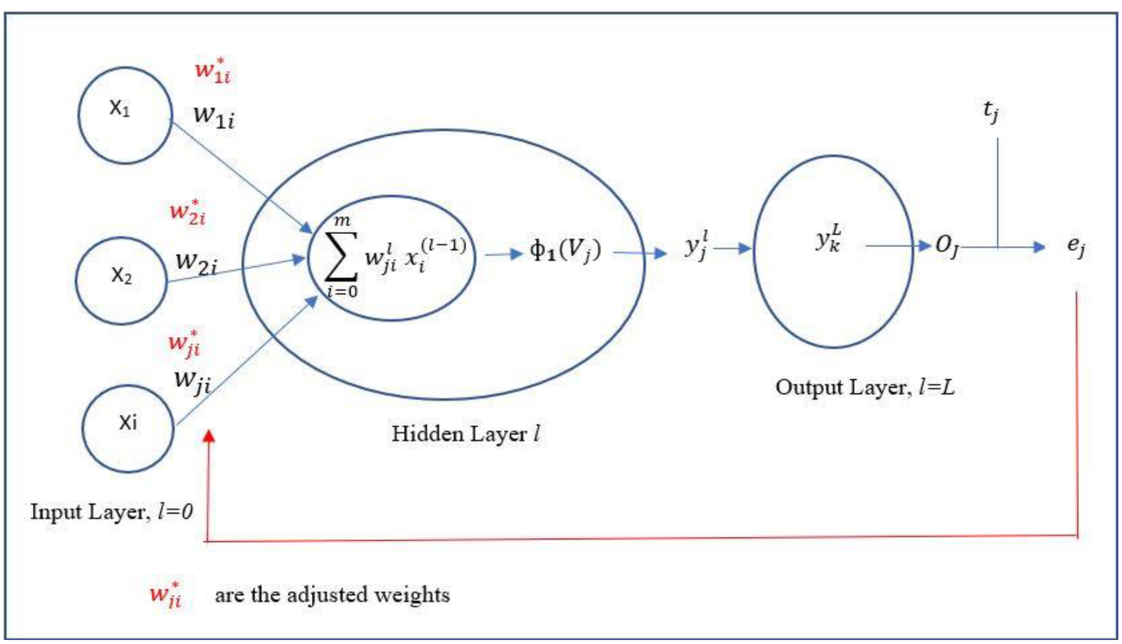
Step 2c: Computation of Error – The error of the *j*-th neuron of the *n*-th iteration is then computed as the difference between the target and the observed output.

Step 2d: Backward Computation – The weights in the network are adjusted based on a local gradient, which is a function of the error computed in step 2c.

Step 2e: Iteration – The procedures in steps b, c, and d are repeated for batches of three observations per iteration until the stopping criterion of 100 epochs is met.

The training process is presented in Figure 6.

Figure 6. ANN Training Process[xxii]



where,

V_j^l is the weighted sum in *j*-th neuron of the *l*-th hidden layer,

w_{ji}^l is the weight coefficient of the *j*-th neuron of the *l*-th layer that is fed from the *i*-th neuron in layer *l*-1,

$x_i^{(l-1)}$ is the output of the *i*-th neuron in the previous layer *l*-1,

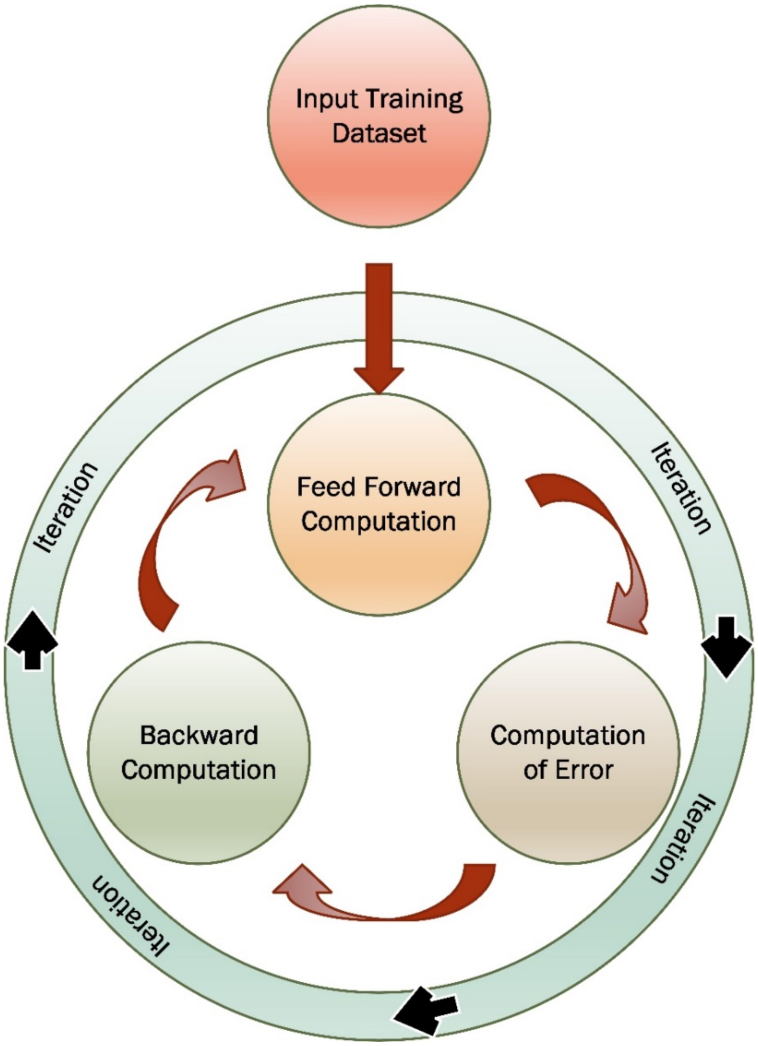
y_j^l is the output of the of the *j*-th neuron in layer *l*-1,

Φ_k is the activation function which is a rectilinear unit function in the hidden layers and a sigmoid function in the output layer. Thus, k=1 in the hidden layer and k=2 in the output layer, and

O_J is the output of the *n*-th iteration.

The ANN training procedure is illustrated in Figure 7.

Figure 7. ANN Training Process



Step 3: Testing and Evaluation of Model

Once the network is trained for the designated number of epochs (100), the model is evaluated using the test dataset. To assess the accuracy of the model, a confusion matrix (CM) was used. A CM contains details on the actual and predicted classifications made by a classification system (in this case, the ANN model). In the CM, each row denotes the instances of an actual class, while each column denotes the instances of a predicted class. Table 1 illustrates the CM for a two-class classifier.

Table 1. Confusion Matrix (CM)

Total No. of Observations		Predicted: Negative	Predicted: Positive
Actual	Negative	True Negative (TN): Instances that are negative and correctly classified as negative	False Positive (FP): Instances that are negative and wrongly classified as positive
	Positive	False Negative (FP): Instances that are positive and wrongly classified as negative	True Positive (TP): Instances that are positive and correctly classified as positive

Based on the CM, the measures shown in Table 2 were computed to evaluate the models developed. Following the evaluation, the number of hidden layers and neurons in the network architecture is modified based on the model's performance, and the training process is repeated. This iterative process is repeated until the best-performing model is achieved based on the choice of parameters such as learning rate, number of neurons/layers, and data quality/quantity.

Table 2. Measures for Evaluation

Measure	Description	Computation
Accuracy (AC)	The accuracy is the proportion of the total number of predictions that were correctly classified.	$\frac{TN+TP}{TN+FP+FN+TP}$
Error Rate (ER)	The error rate is the rate at which predictions were misclassified.	$1-AC$
Precision (P)	This is the proportion of the predicted positive cases that were correct.	$\frac{TP}{FP+TP}$
Sensitivity (S)	This is the proportion of positive cases that were correctly identified.	$\frac{TP}{FN+TP}$
F-measure (F)	This is a measure of the accuracy of the test model computed using Sensitivity and Precision. The value of F ranges from 0 to 1, where 1 shows an excellent model and 0 show a bad model.	$2\,\frac{S.P}{S+P}$

3.6.1 Analysis Software

The ANN algorithm was implemented using Python, a high-level general-purpose programming language. The Anaconda Python distribution, which is an open-source distribution with standard libraries for data processing, analysis, and machine learning, was used. Anaconda also offers several Integrated Development Environments (IDE), including Jupyter, Notebook, and Spyder. In this study, the Spyder IDE (a powerful Python IDE with advanced editing, interactive testing, debugging, and introspection capabilities) was utilized. To simplify data preprocessing, the study imported the following libraries:

> *NumPy*: A package for array processing that can efficiently manipulate large multi-dimensional arrays of data.
>
> *Pandas*: An open-source library designed for high-level data manipulation.

Furthermore, the study imported the following libraries to develop models:

> *TensorFlow*: A system developed by Google Brain that accelerates numerical computations and machine learning.
>
> *Keras*: A high-level application program interface for neural networks that operates on the backend of TensorFlow. It also assists with the development of ANN models.

In addition to Spyder IDE, IBM Statistical Software for Social Scientists (SPSS) was used for binomial regression analysis.

4. Results

4.1 Summary Statistics

This section presents an overview of the CMV crash trends in the designated CMV routes in the District of Columbia for the years 2016 through 2020 and includes a summary of comparative crash statistics.

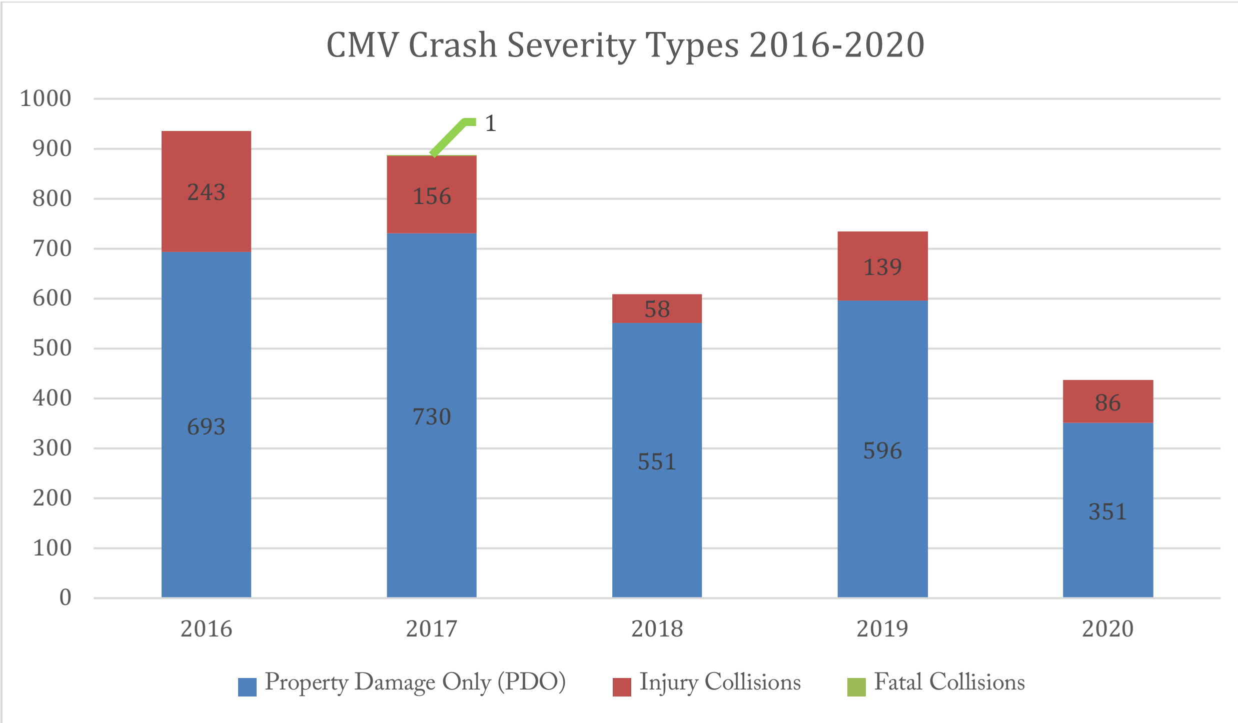
4.1.1 CMV Crash Statistics

A summary of the CMV crashes reported in DC on CMV routes from 2016 through 2020 is presented in Table 3. Figure 8 represents the collision severity distribution of the CMV crashes. It should be noted that these crashes represent only the CMV crashes occurring on designated CMV routes and not the entire DC road network.

Table 3. DC CMV Crashes Quick Facts for 2016–2020

Year	2016	2017	2018	2019	2020
Total Collisions	936	887	609	735	437
Fatal Collisions	0	1	0	0	0
Injury Collisions	243	156	58	139	86
Property Damage Only (PDO) Collisions	693	730	551	596	351

Figure 8. Crash Severity Types for 2016 through 2020



There was a noticeable decline of approximately 40% in the total CMV collisions in 2020 compared to 2019 that occurred specifically on CMV routes, as presented in Table 3. However, the documented injury count to collision ratio in 2020 was higher than that of 2019. The most frequent CMV crash severity type documented in 2020 was PDO, which constitutes nearly 80% (351) of all crashes for that year.

4.1.2 Total Crashes from 2016 through 2020

The trend of all CMV crashes that occurred in DC on CMV routes and corresponding injuries by year from 2016 through 2020 is presented in Figure 9. The figure shows that there was a reduction in injury counts by approximately 38% from 2019 to 2020. Figure 10 presents the overview of injured persons recorded by year from 2016 through 2020.

Figure 9. Traffic Crashes and Injury Crashes from 2016 through 2020

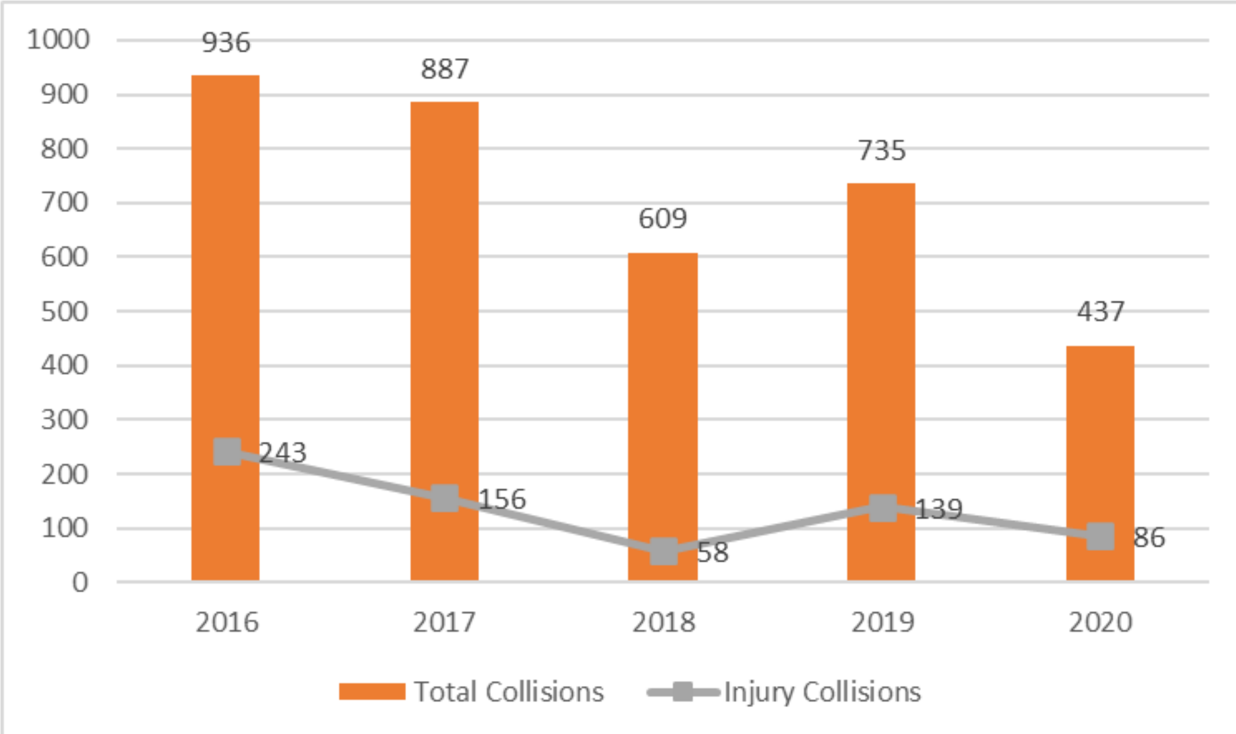
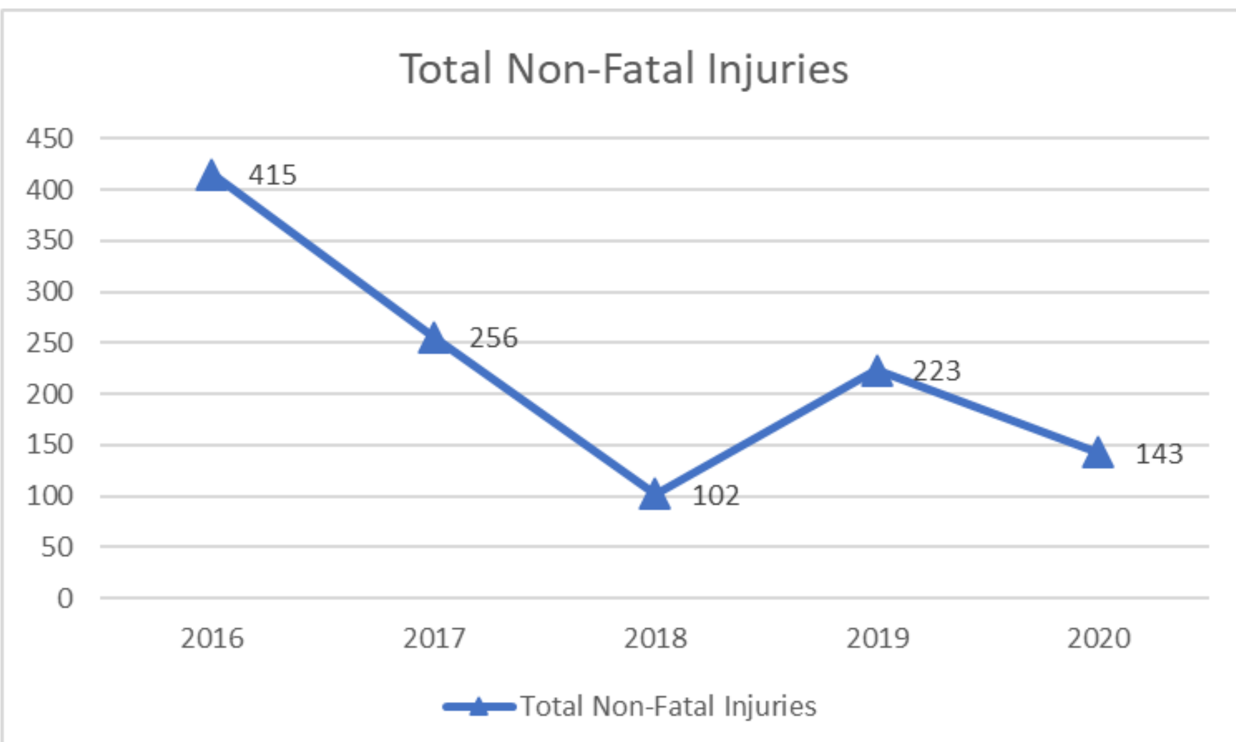


Figure 10. Non-Fatal Collisions from 2016 through 2020



From Figure 10, it can be observed that in 2020, the number of injured people decreased by almost 27% compared to 2019.

4.2 CMV Crash Descriptive

This section presents the descriptive statistics for CMV crashes reported in DC on CMV routes from 2016 to 2020. The analyzed characteristics included crash occurrence time, crash type, roadway user and vehicle contributing factors, road conditions, and geometric characteristics. These characteristics were present for all the CMV crash data points as reported by police officers present at the scene of each crash. It is important to note that these crashes do not represent all the CMV crashes that occurred in DC between 2016 through 2020, but only those that happened on designated CMV routes.

Table 4 presents the frequency distribution of CMV crashes that occurred from 2016–2020 in DC on CMV routes categorized by hour. From the table, a notable number of crashes occurred between 9 AM–10 AM (365 collisions), with the highest number of reported injuries (122) taking place between 11 AM–12 PM. The only CMV-related fatality in the same duration was recorded for hour 23 (11 PM–12 AM). The yearly frequency of CMV collisions and the associated injuries by the hour of the day from 2016 to 2020 is presented in Figures 11 through 15.

Table 4. All Crashes by Hour of the Day in 2016–2020

Hour	Collisions	Fatalities	Injuries
00	32	0	12
01	29	0	18
02	21	0	13
03	23	0	8
04	28	0	16
05	49	0	14
06	115	0	45
07	221	0	81
08	326	0	86
09	365	0	73
10	324	0	116
11	300	0	122
12	298	0	86
13	275	0	107
14	264	0	75
15	213	0	70
16	194	0	56
17	162	0	41
18	114	0	36
19	80	0	23
20	63	0	25
21	50	0	6
22	31	0	7
23	27	1	3
Total	3,604	1	1,139

Figure 11. Total Collisions and Injuries by Hour in 2016

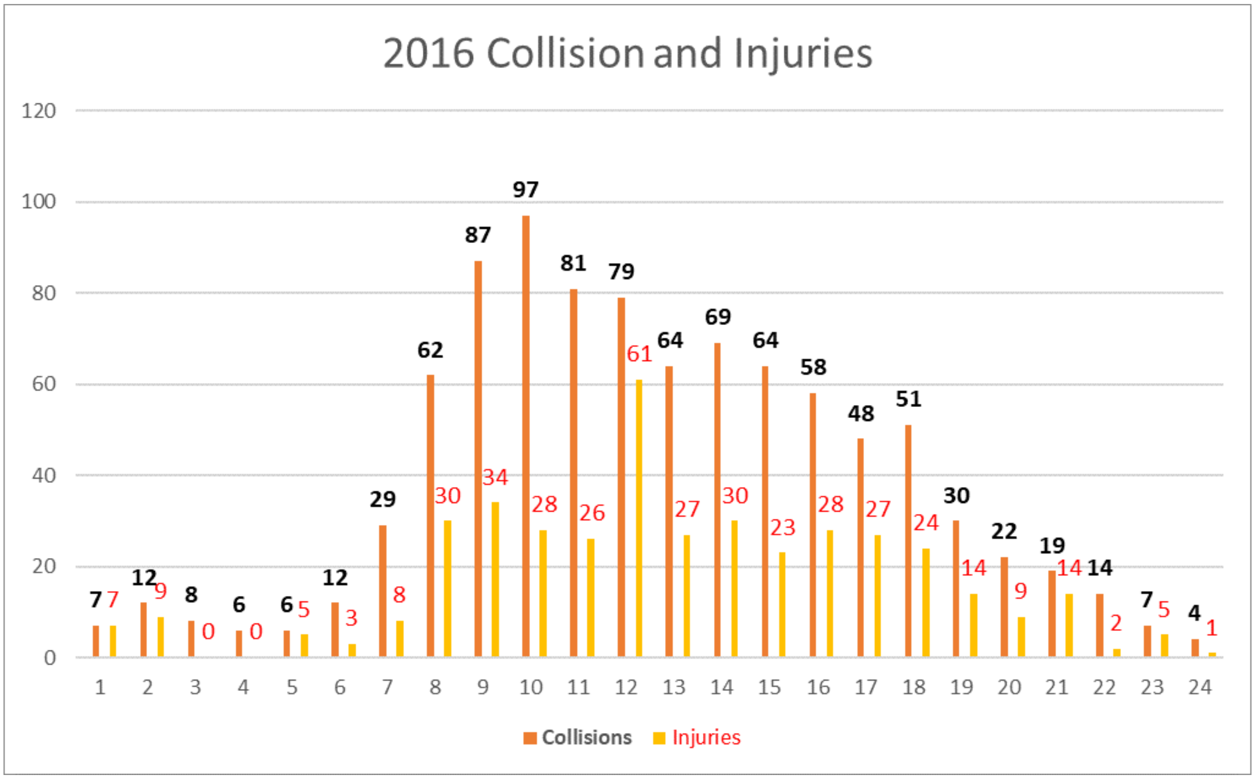


Figure 12. Total Collisions and Injuries by Hour in 2017

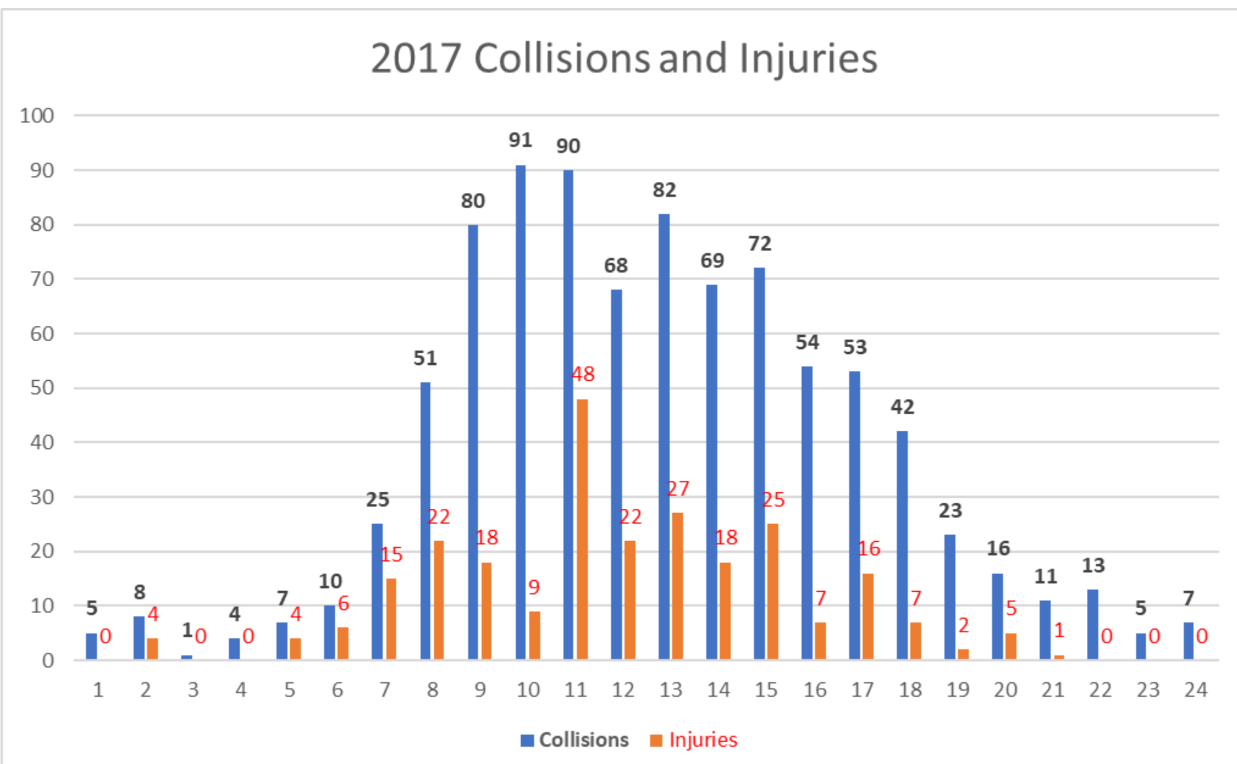


Figure 13. Total Collisions and Injuries by Hour in 2018

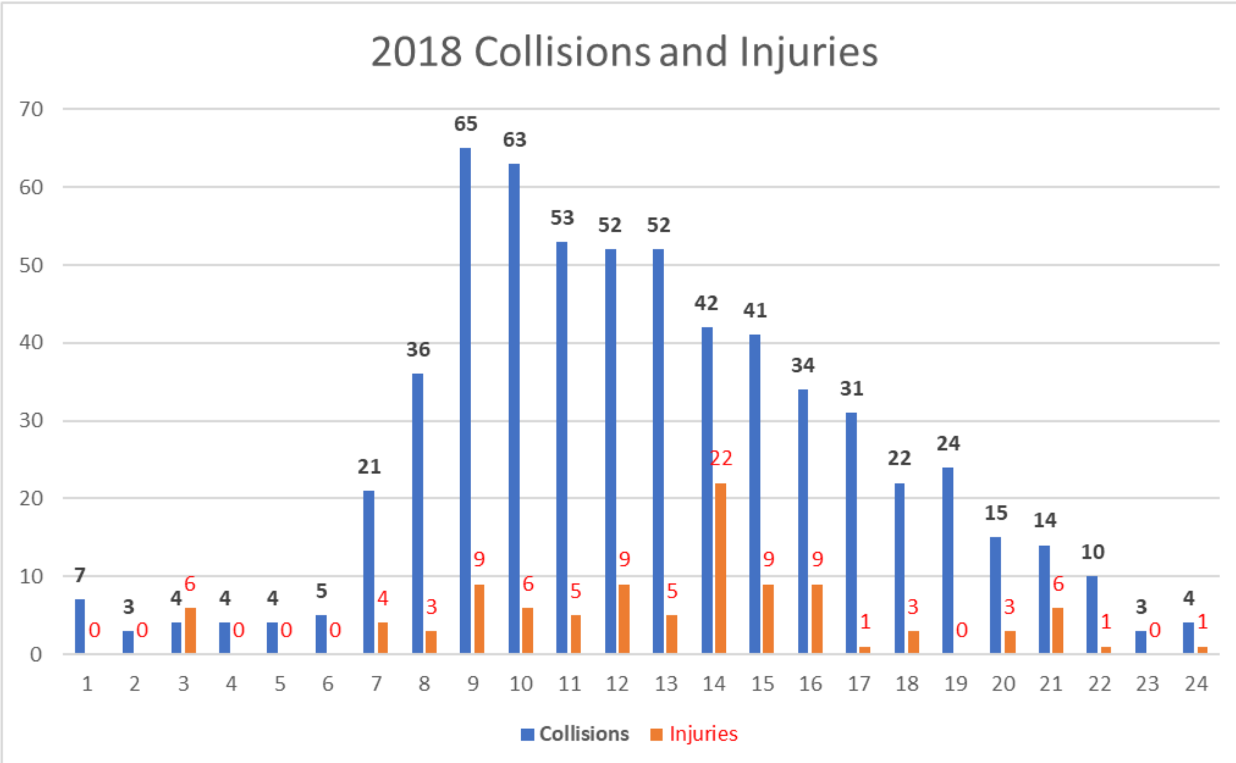


Figure 14. Total Collisions and Injuries by Hour in 2019

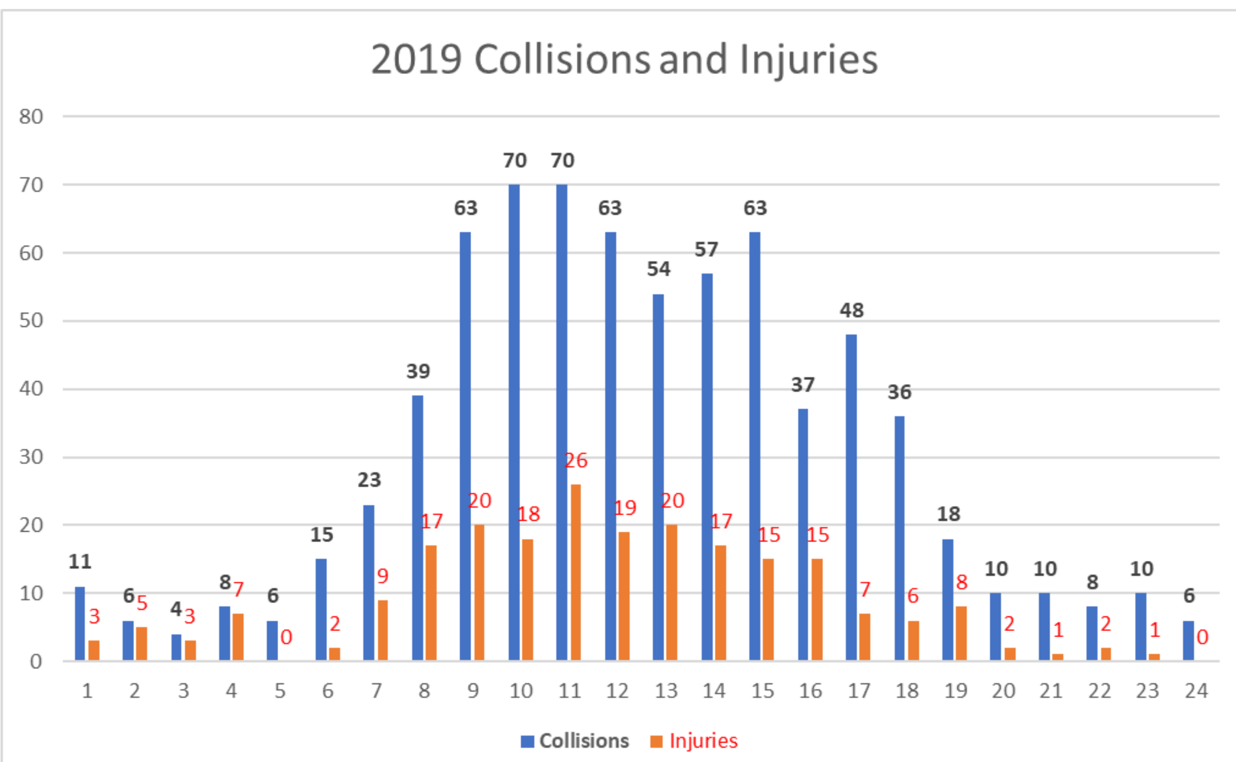
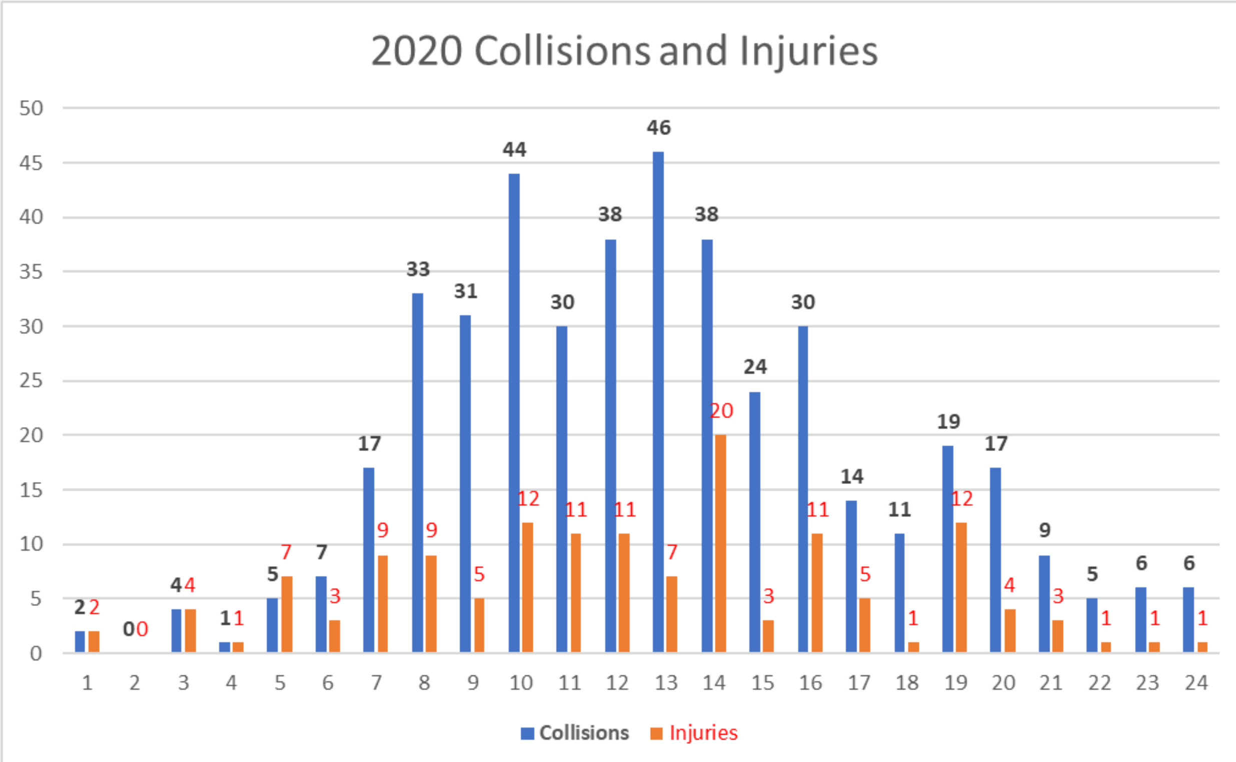


Figure 15. Total Collisions and Injuries by Hour in 2020



4.3 CMV Crashes and Injuries by Day of the Week

Table 5 presents the distribution of CMV crash occurrences on DC CMV routes from 2016–2020, categorized by day. From the table, Sundays had the lowest frequency of both CMV collisions and associated injuries over the 5-year duration. On the other hand, Wednesdays had the highest CMV-related collision counts (693), while most injuries occurred on Tuesdays (242). Figures 16 through 20 present the yearly trends of CMV collisions and corresponding injuries by day from 2016 to 2020. A five-year CMV collision frequency summary is presented in Figure 21.

Table 5. Crashes by Day of the Week for 2016–2020

Day	Collisions	Fatalities	Injuries
Sunday	86	0	23
Monday	586	0	191
Tuesday	673	1	242
Wednesday	693	0	201
Thursday	645	0	196
Friday	656	0	204
Saturday	265	0	82
Total	3,604	1	1,139

Figure 16. Total Collisions and Injuries by Day in 2016

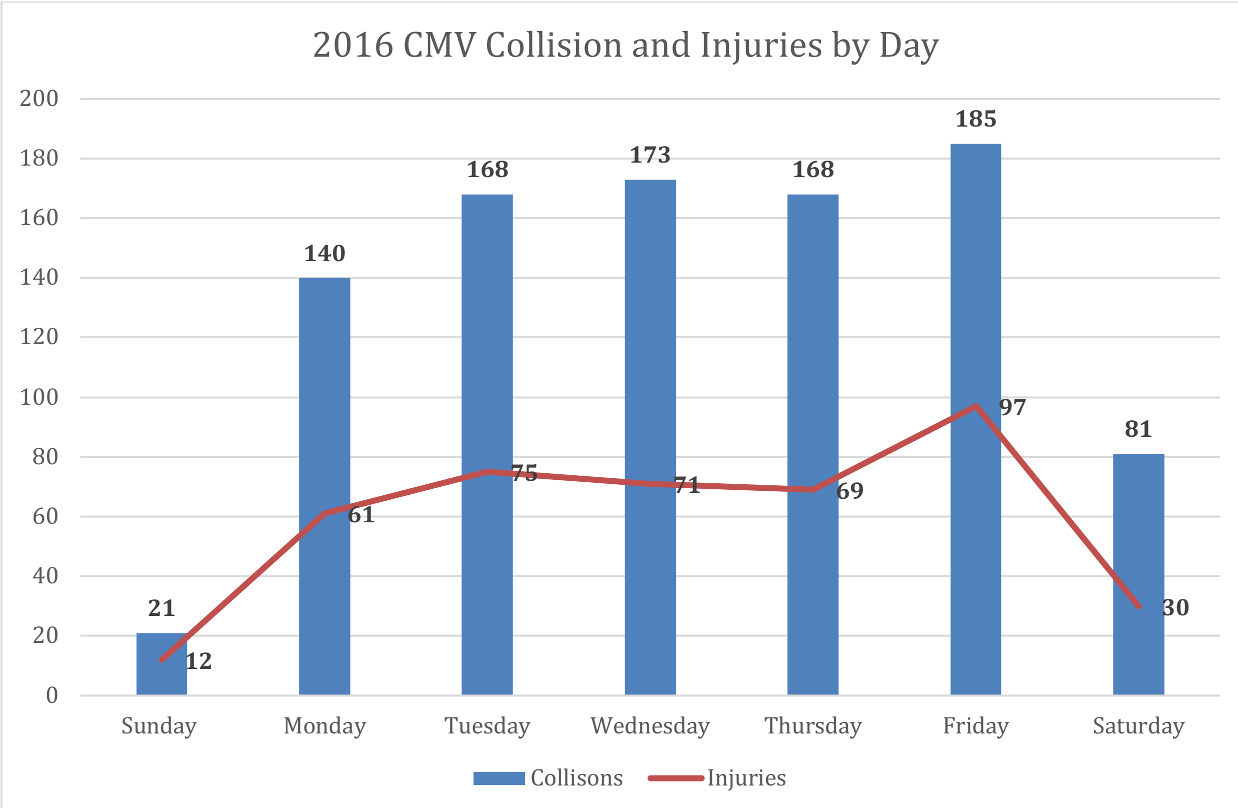


Figure 17. Total Collisions and Injuries by Day in 2017

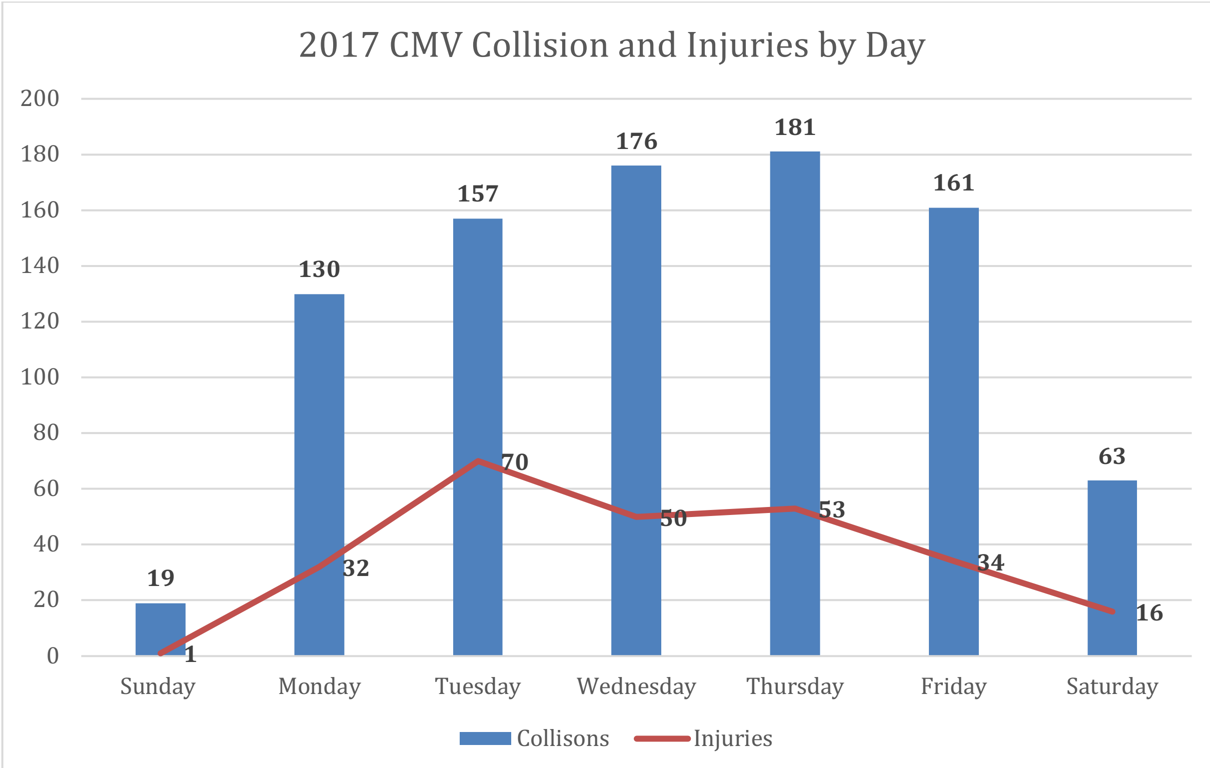


Figure 18. Total Collisions and Injuries by Day in 2018

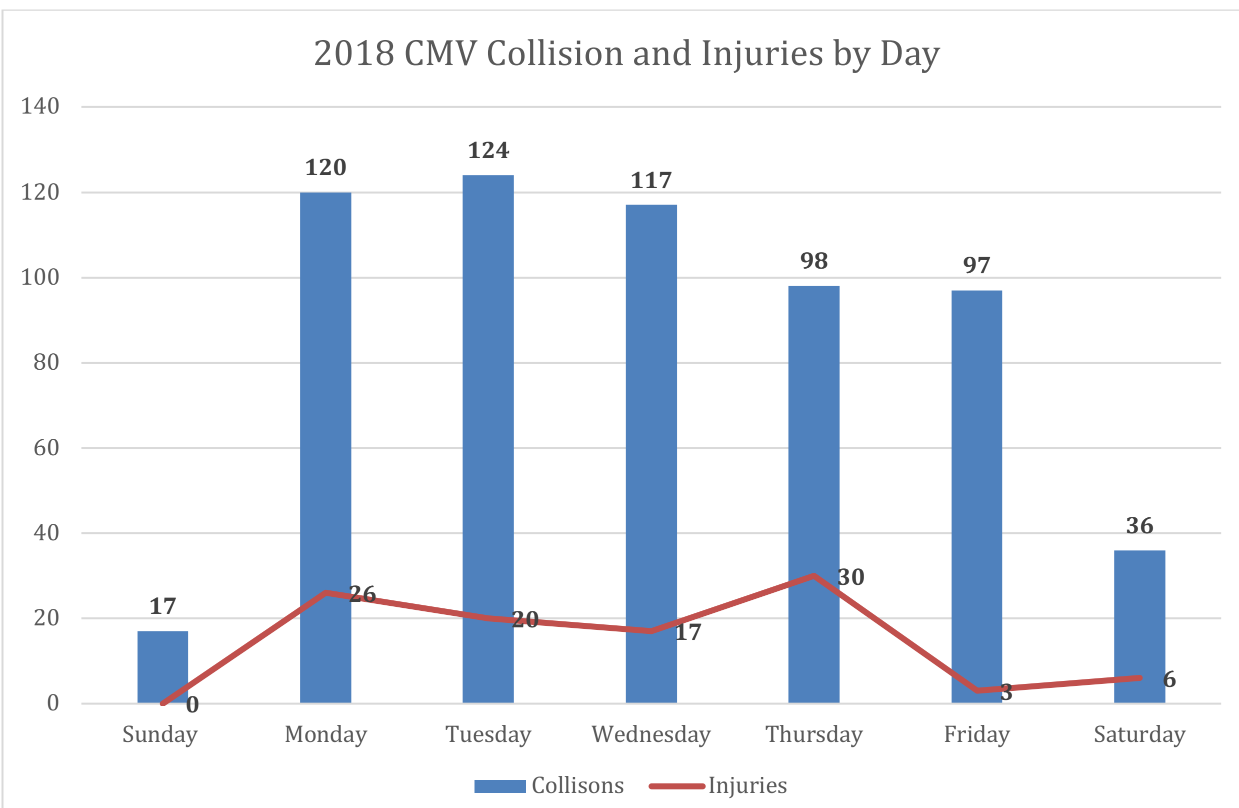


Figure 19. Total Collisions and Injuries by Day in 2019

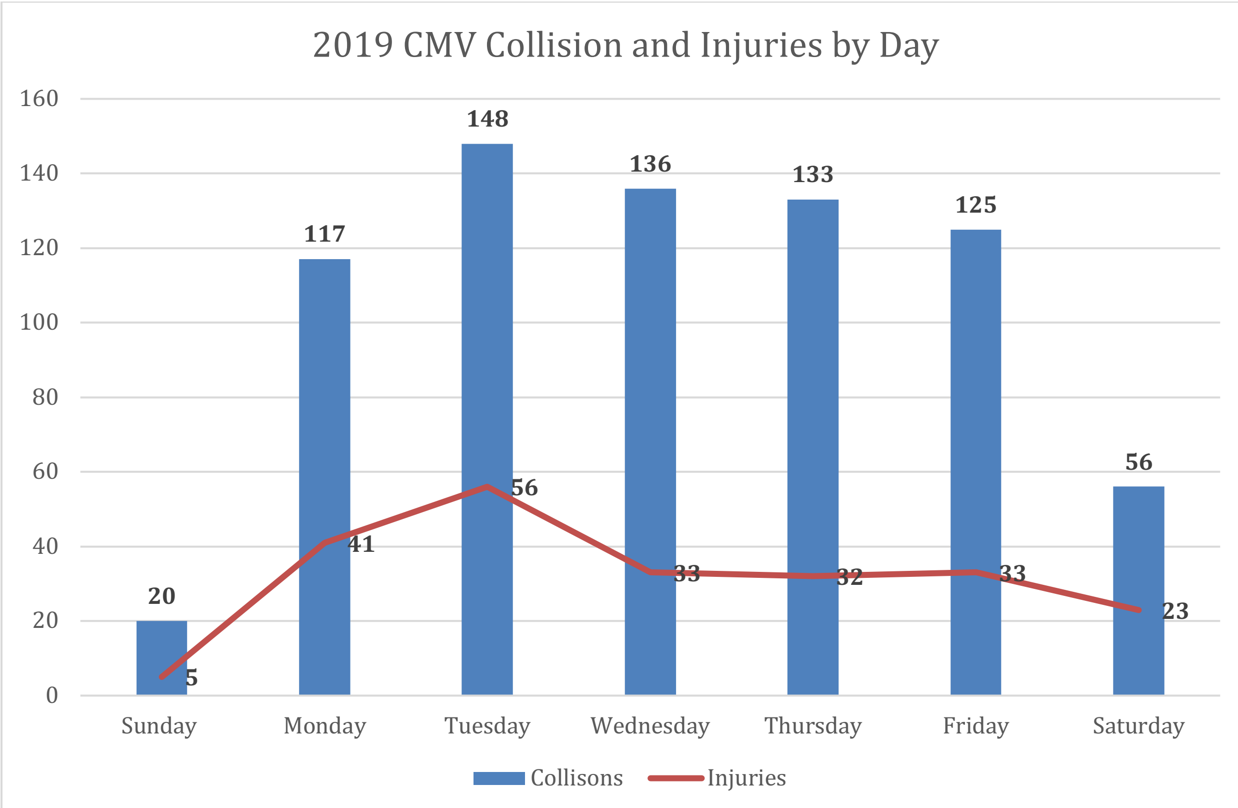


Figure 20. Total Collisions and Injuries by Day in 2020

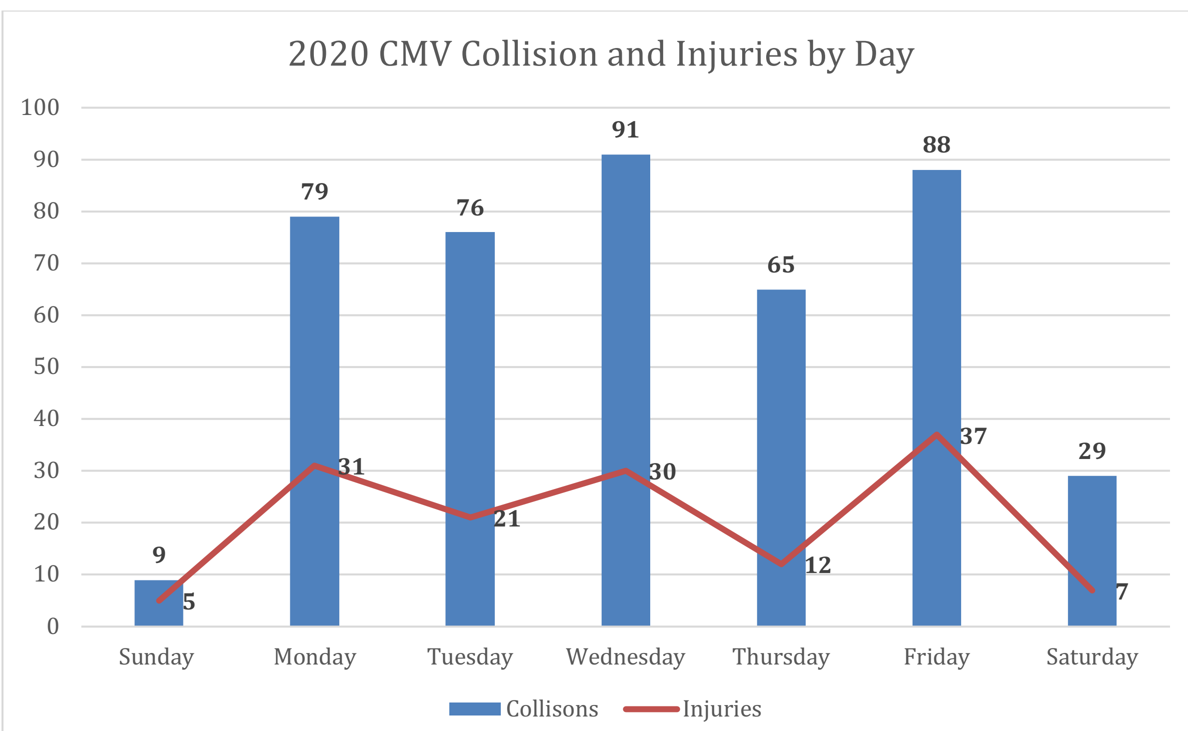
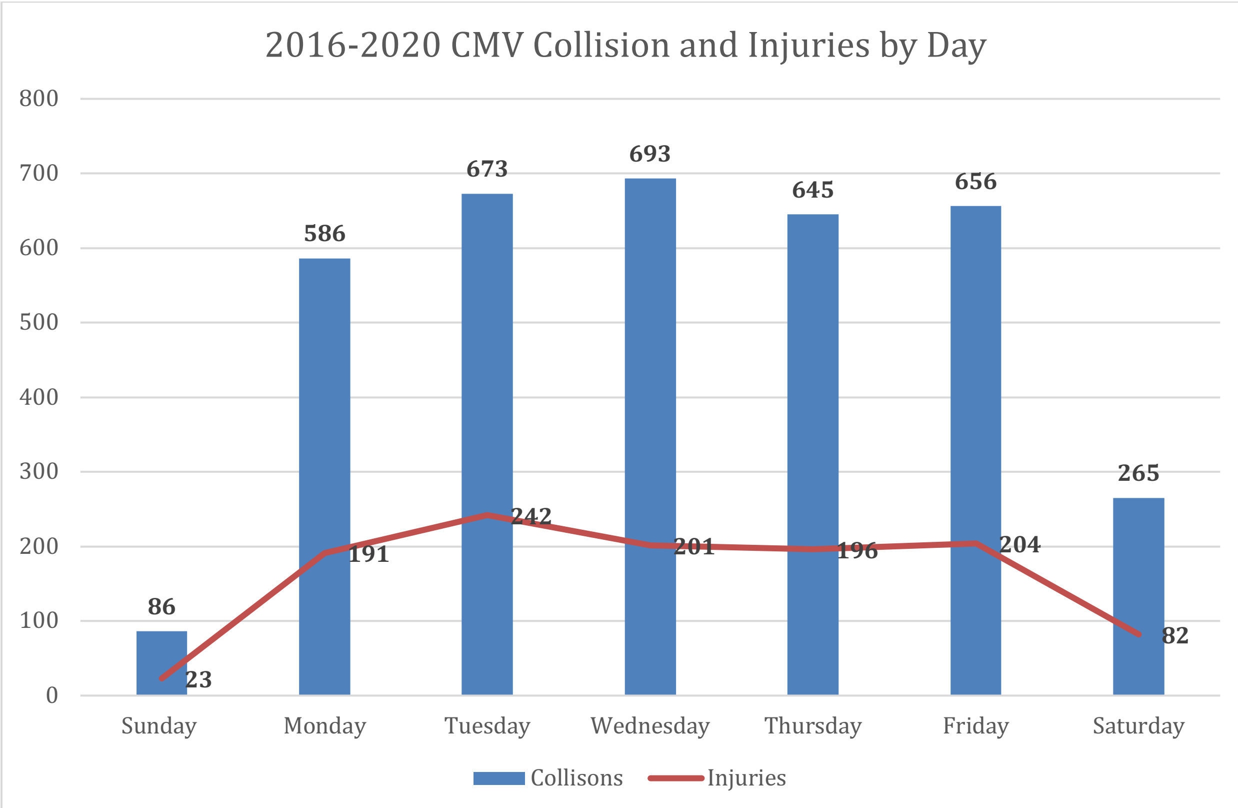


Figure 21. Total Collisions and Injuries by Day 2016–2020



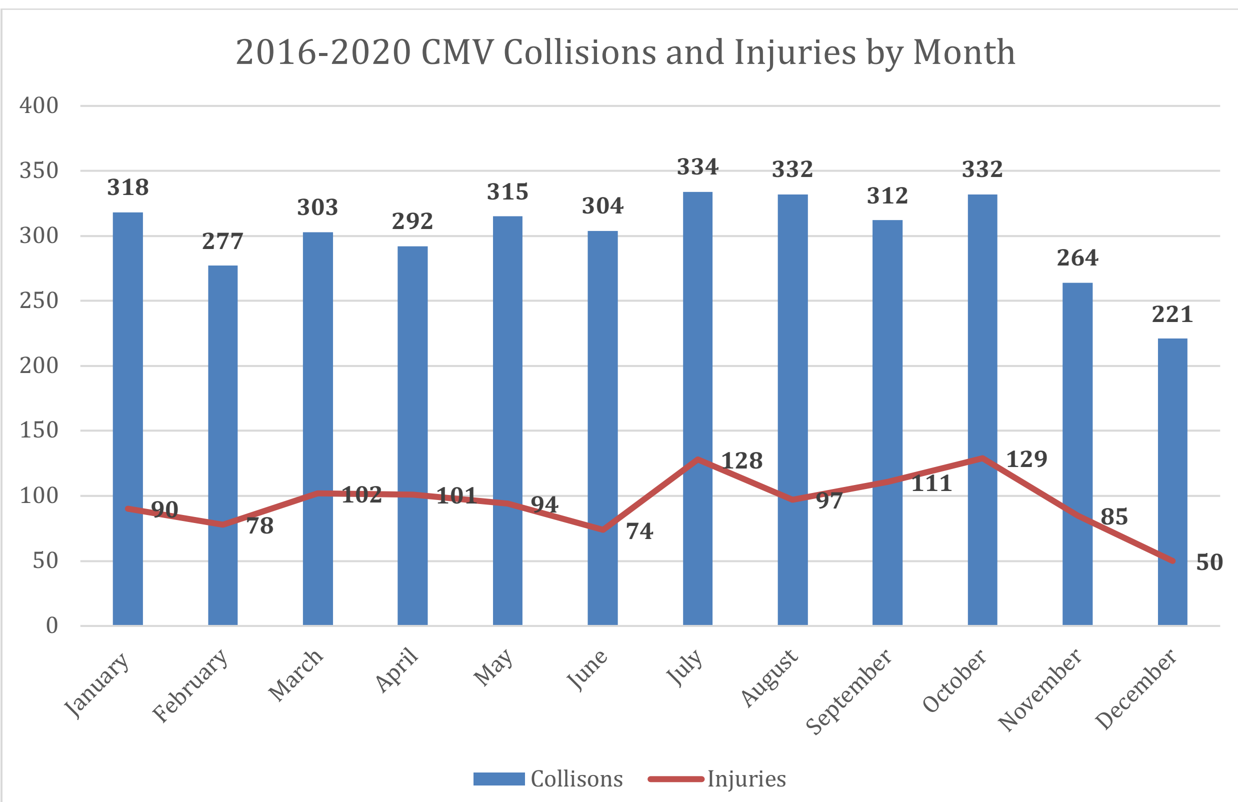
4.4 CMV Crashes and Injuries by Month of the Year

Table 6 presents the frequency of the 2016–2020 DC CMV crashes by month. From the table, it can be observed that May 2017 had the highest number of CMV collisions on the designated CMV routes (102), whereas the highest number of CMV-related injuries took place in October 2019. Figure 22 presents a 5-year compilation of all the monthly CMV crashes and the related injuries.

Table 6. CMV Crashes by Month for 2016–2020

Year	2016		2017		2018		2019		2020	
Month	Collisions	Injuries	Collisions	Injuries	Collisions	Injuries	Collisions	Injuries	Collisions	Injuries
January	77	20	75	35	48	11	65	17	53	7
February	64	18	70	16	54	9	49	15	40	20
March	64	31	90	28	63	4	52	24	34	15
April	75	63	69	3	49	19	78	11	21	5
May	70	28	102	25	48	0	62	23	33	18
June	93	40	83	18	47	0	51	11	30	5
July	98	58	93	28	43	3	62	19	38	20
August	87	34	82	26	56	7	69	26	38	4
September	69	40	82	25	65	14	64	25	32	7
October	97	48	42	15	72	20	83	26	38	20
November	66	14	69	33	39	14	50	11	40	13
December	76	21	30	4	25	1	50	15	40	9
Total	936	415	887	256	609	102	735	223	437	143

Figure 22. Total Collisions and Injuries by Month 2016–2020



4.5 CMV Crashes and Injuries by Quadrant

Table 7 presents the number of collisions and injuries that occurred in the CMV routes in all four DC quadrants. From the table, the Northwest quadrant had the highest number of collisions (49%) as well as injuries (39%), while the Southwest quadrant had the lowest number of collisions and injuries (7% and 13%, respectively). Figures 23 and 24 present the yearly CMV crashes and the total number of injuries caused by CMV collisions that occurred in all DC quadrants.

Table 7. CMV Crashes by Quadrant for 2016–2020

Quadrant	Collisions	Collision Percentage	Fatalities	Injuries	Injury Percentage
NW	1,767	49%	0	444	39%
NE	1,095	30%	1	339	30%
SW	261	7%	0	152	13%
SE	481	13%	0	204	18%
Total	3,604	100%	1	1,139	100%

Figure 23. Total Collisions by Quadrant 2016–2020

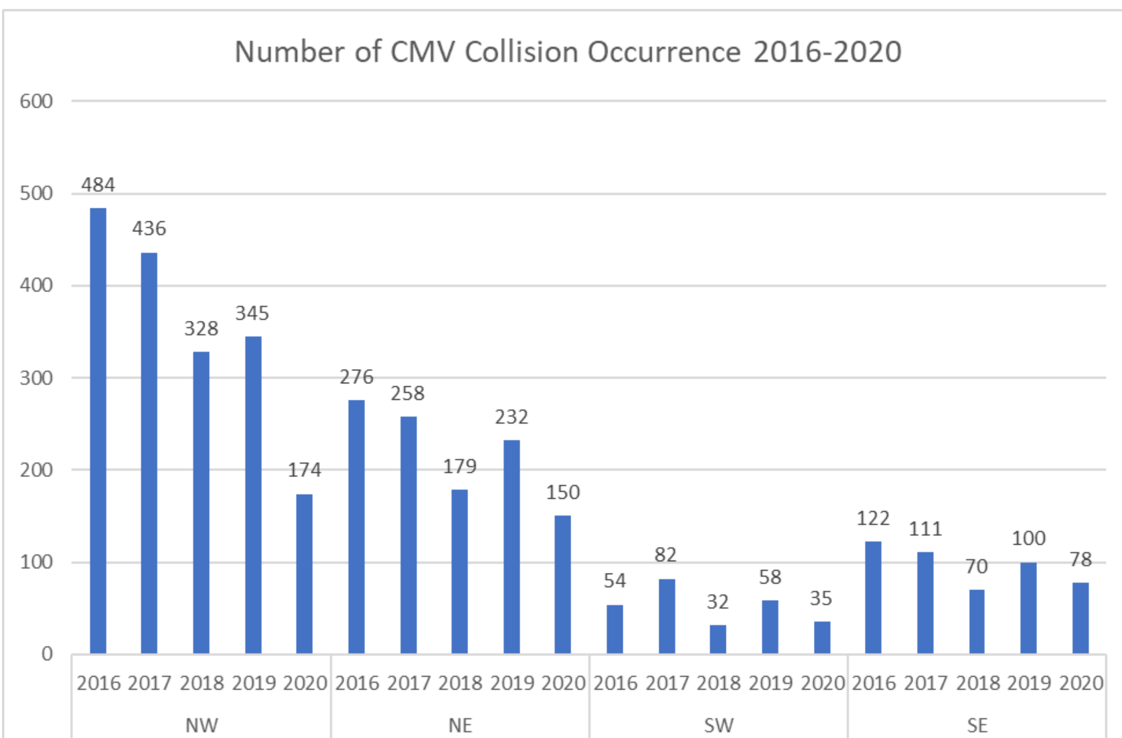
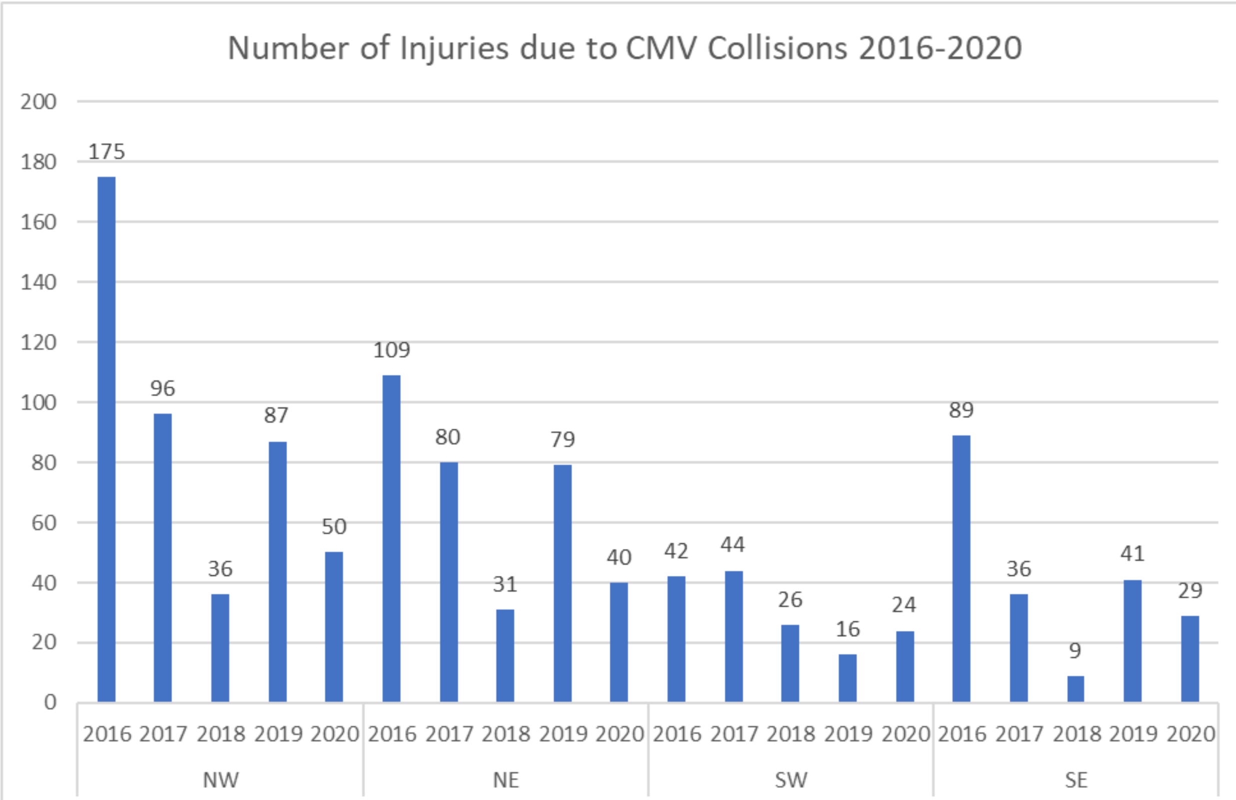


Figure 24. Total Injuries by Quadrant 2016–2020



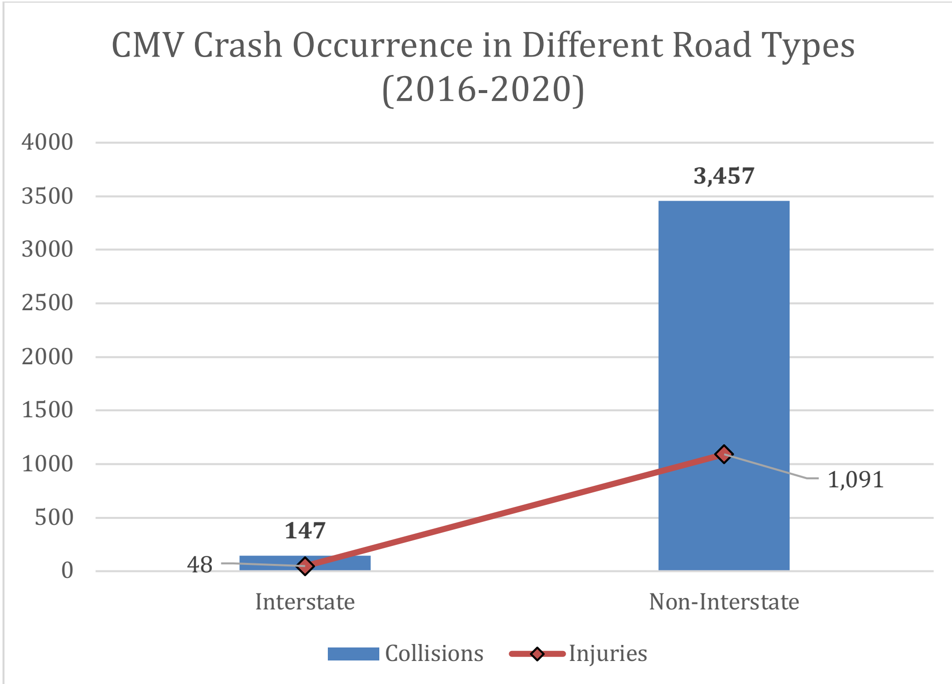
4.6 CMV Crashes and Injuries by Road Type

Table 8 presents the frequency distribution of all CMV collisions that occurred in the two types of roadways (interstate and non-interstate) from 2016–2020. Only 4.08% of the collisions occurred on the interstates. The injuries associated with these CMV crashes on different road types are presented in Figure 25. It must be noted that the District of Columbia has only about 16 miles of roadways classified as interstate or freeway.

Table 8. CMV Crashes by Road Type 2016–2020

Road Type	Collisions	Percentage
Interstate	147	4.07
Non-Interstate	3,457	95.93
Total	3,604	100.00

Figure 25. Total Collisions and Injuries by Hour 2016–2020



4.7 CMV Crash Occurrence Type

Table 9 and Figures 26 and 27 provide the summary of the cumulative CMV crash count, categorized by crash type from 2016 through 2020. The tally of associated injuries, as well as the PDO CMV crashes, are included in the table. It can be observed that sideswipe crashes in the same direction were the prevailing type of CMV collision (1,769), overall accounting for 49% of all CMV crashes. Additionally, the same crash type resulted in the highest number of PDO crashes (1,516) within the five-year duration. Rear to rear crashes were the least frequent type of CMV collisions (37), and they led to the fewest injuries (10). Most injuries resulted from front-to-rear collisions resulting in 438 injuries in 2016–2020. A front-to-rear crash was also responsible for the one fatality that occurred during the study's time span.

Table 9. CMV Crash Occurrence Types 2016–2020

Type of Crash	Total Crashes	Fatal Crashes	Injuries	PDO Crashes
Angle	256	0	108	200
Front to Front	126	0	54	96
Front to Rear	740	1	438	514
Other	212	0	71	169
Rear to Front	181	0	37	154
Rear to Rear	37	0	10	28
Rear to Side	85	0	17	67
Sideswipe, Opposite Direction	107	0	26	89
Sideswipe, Same Direction	1,769	0	361	1,516
Unknown	91	0	17	80
Total	3,604	1	1,139	2,913

Figure 26. Types of CMV Collisions 2016–2020

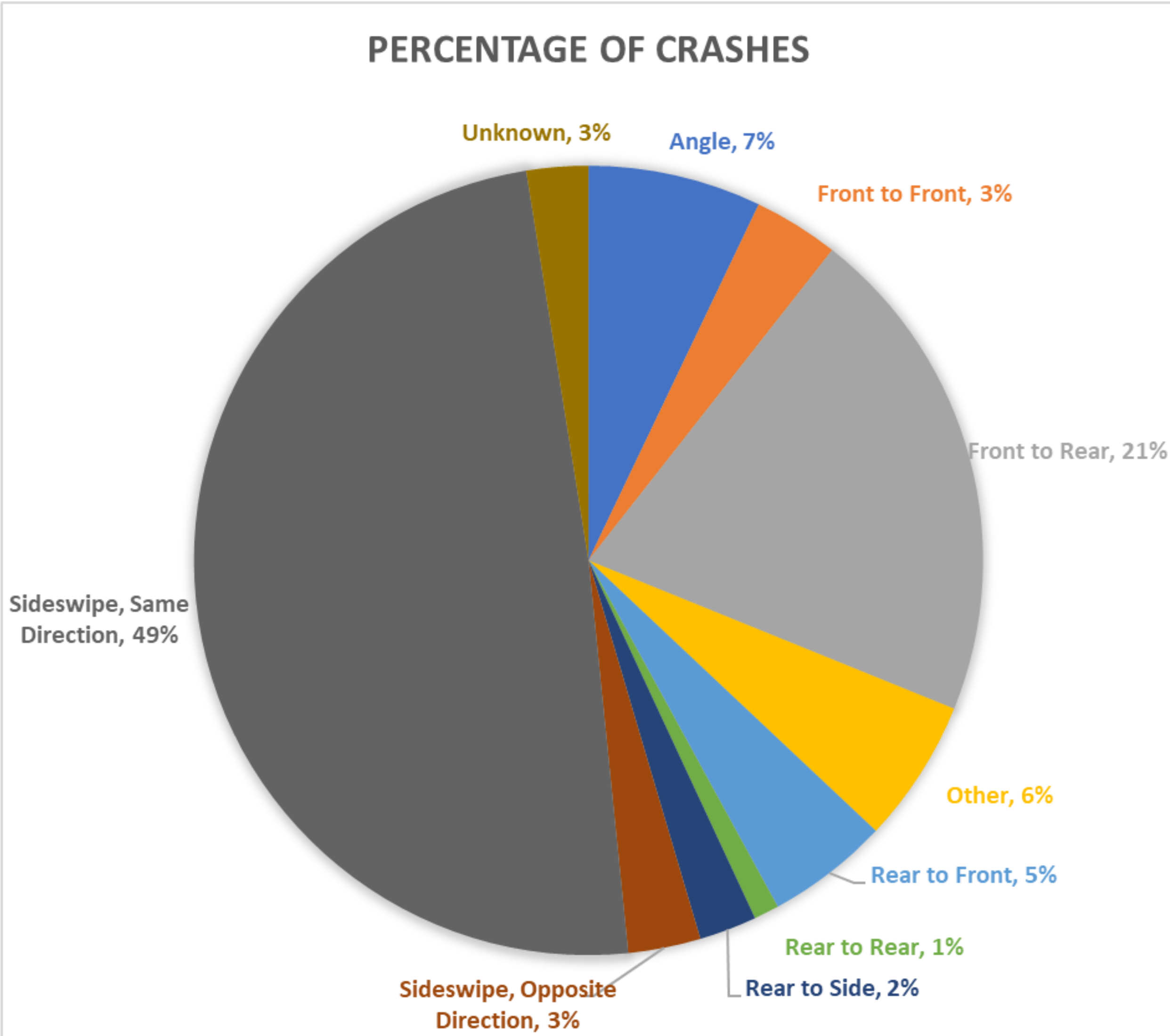
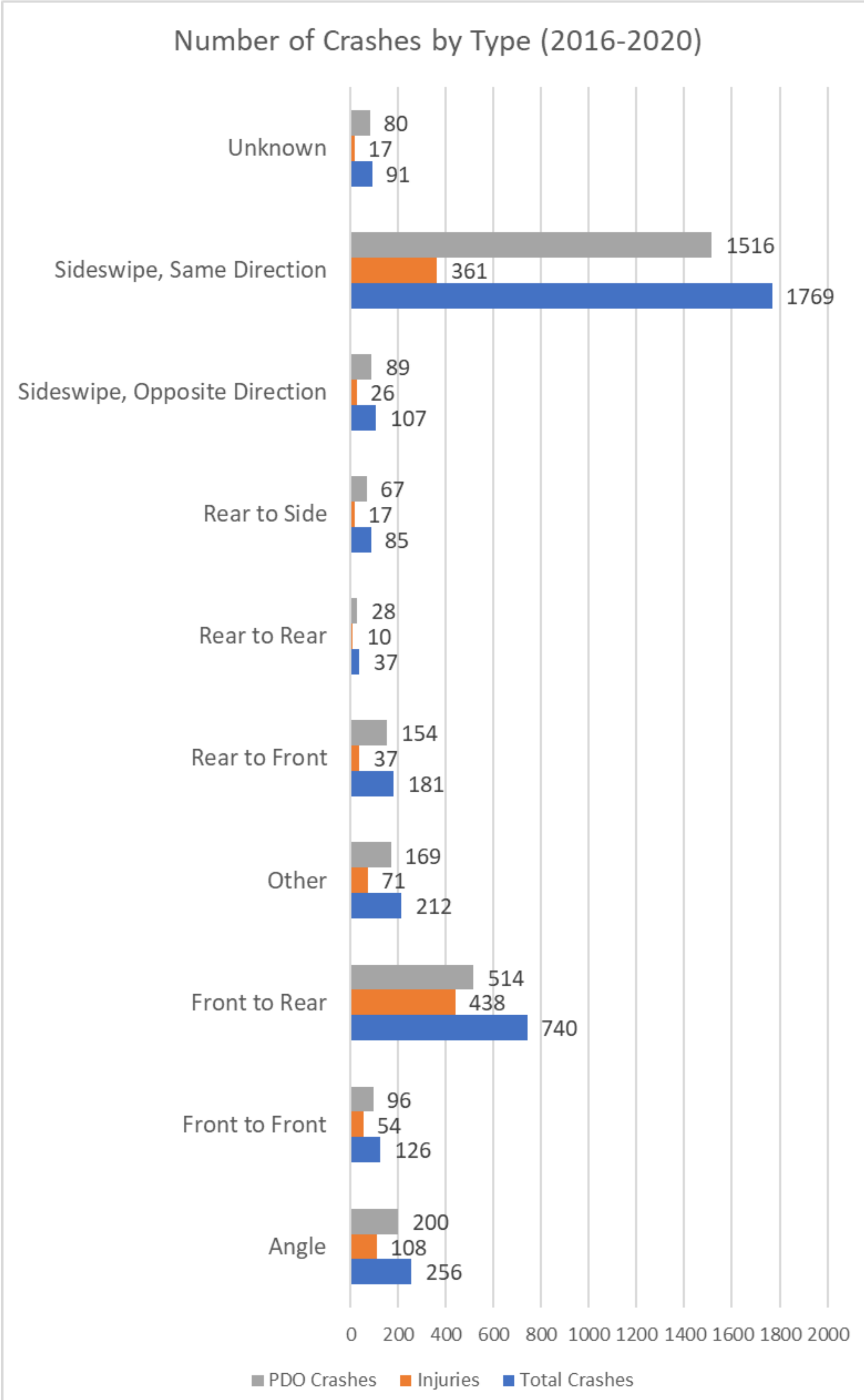


Figure 27. CMV Crash Type Frequency, Injuries, and PDO Collisions 2016–2020



4.8 CMV Crashes by Vehicle Classification

The frequency of the different types of CMVs that were involved in the CMV collisions from 2016–2020 is presented in Table 10. Figures 28 and 29 display the percentages of the involved CMVs in collisions and associated injuries, respectively, in pie charts.

Table 10. Summary of CMV Crashes by Vehicle Type 2016–2020

Year		Construction/ Industrial Equipment	Large/ Heavy Truck	Trailer	Total
2016	Collisions	6	882	48	936
	Fatalities	0	0	0	0
	Injuries	2	386	27	415
2017	Collisions	9	834	44	887
	Fatalities	0	1	0	1
	Injuries	0	242	14	256
2018	Collisions	6	572	31	609
	Fatalities	0	0	0	0
	Injuries	0	94	8	102
2019	Collisions	7	683	45	735
	Fatalities	0	0	0	0
	Injuries	1	206	16	223
2020	Collisions	3	409	25	437
	Fatalities	0	0	0	0
	Injuries	0	142	1	143
Total		34	4,451	259	4,744

Figure 28. Collisions due to CMV Type 2016–2020

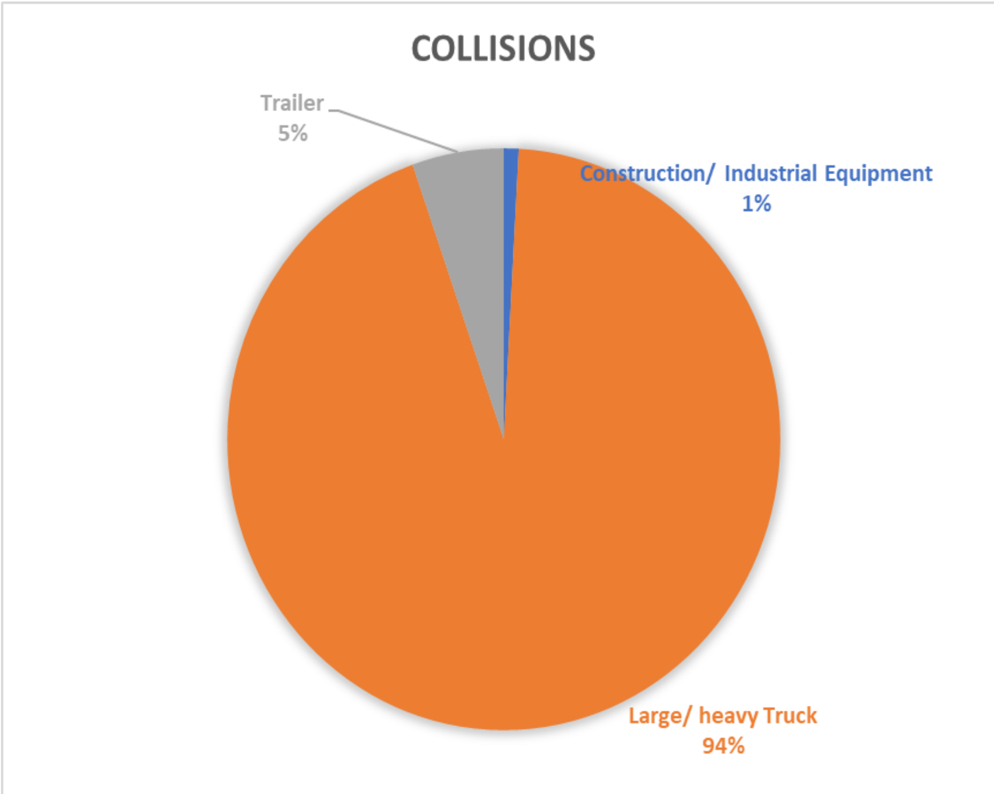
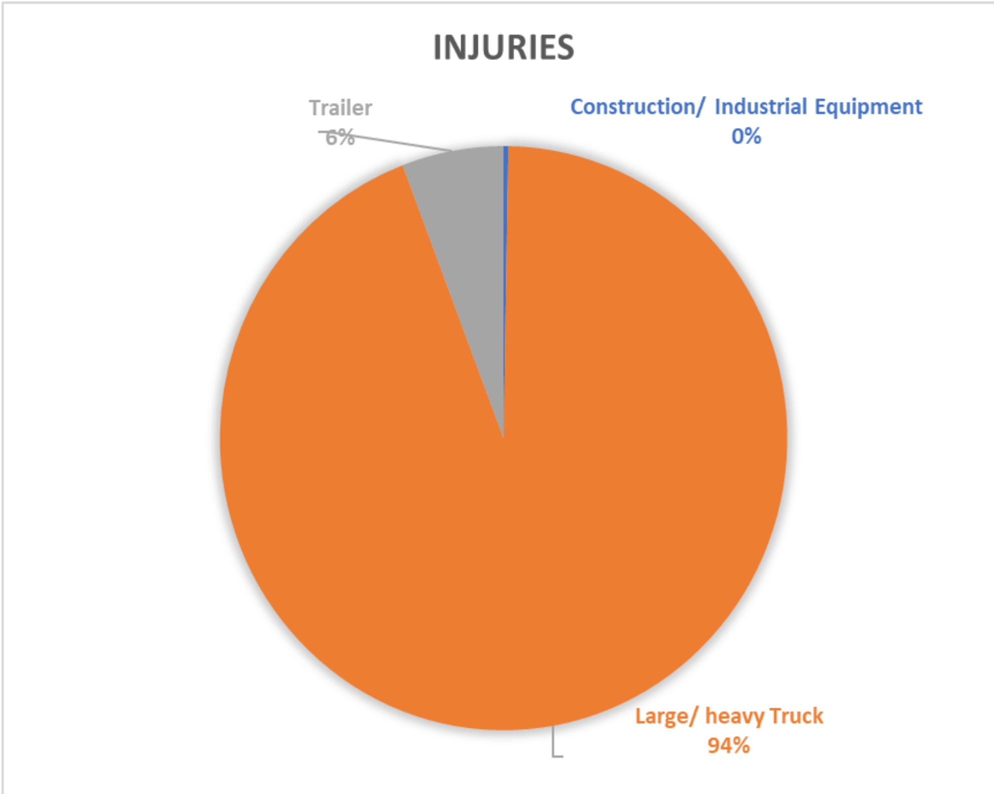


Figure 29. Injuries Resulting from Different CMV Types 2016–2020



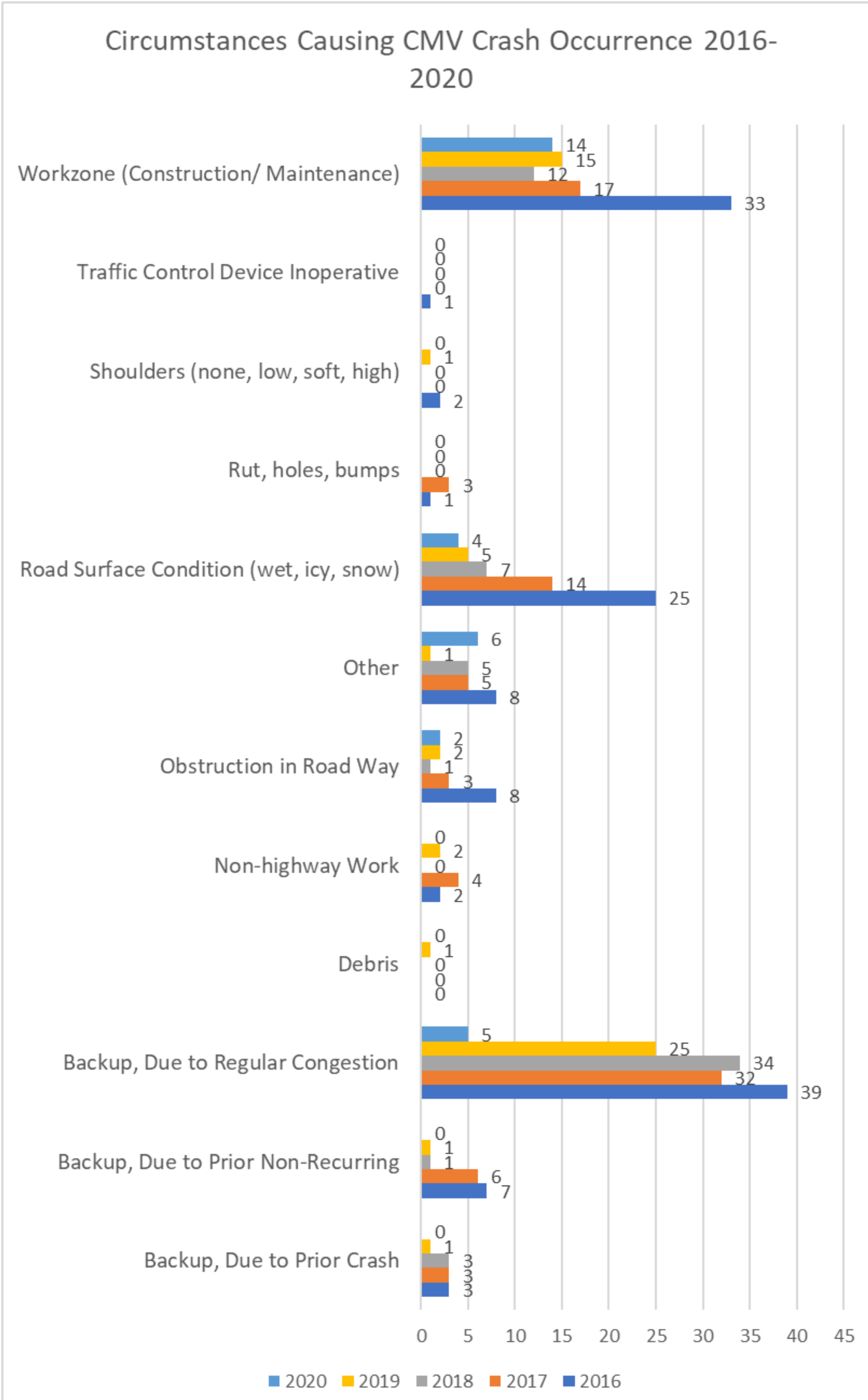
4.9 CMV Crash Contributing Factors

A comprehensive overview of the contributing factors reported for CMV crashes in DC from 2016 to 2020 is presented in Table 11 and Figure 30. Excluding the categories "None" and "Unknown;" the most prominently reported CMV crash-contributing factors during the five-year interval were "backup, due to regular congestion" and "workzone (construction/maintenance)."

Table 11. CMV Crash Contributing Factors 2016–2020

Contributing Factors	2016	2017	2018	2019	2020	TOTAL	Overall Percentage
Backup, Due to Prior Crash	3	3	3	1	0	10	0%
Backup, Due to Prior Non-Recurring	7	6	1	1	0	15	0%
Backup, Due to Regular Congestion	39	32	34	25	5	135	4%
Debris	0	0	0	1	0	1	0%
Non-Highway Work	2	4	0	2	0	8	0%
Obstruction in Roadway	8	3	1	2	2	16	0%
Other	8	5	5	1	6	25	1%
Road Surface Condition (wet, icy, snow)	25	14	7	5	4	55	2%
Rut, Holes, Bumps	1	3	0	0	0	4	0%
Shoulders (low, soft, high)	2	0	0	1	0	3	0%
Traffic Control Device Inoperative	1	0	0	0	0	1	0%
Workzone (construction/maintenance)	33	17	12	15	14	91	3%
Unknown	220	255	185	241	163	1,064	30%
None	587	545	361	440	243	2,176	60%
Total	936	887	609	735	437	3,604	100%

Figure 30. CMV Crash Contributing Factors 2016–2020



4.10 Binary Logistic Regression

Of the 3,604 CMV crash data points that were recorded on DC CMV routes, the project team filtered the crashes that could be associated with the IRI and PCI dataset. A total of 2,390 data points were obtained. Since the project also involves assessing the quality of road surfaces as a factor for evaluation, the analysis focused exclusively on the IRI data. It should be noted that IRI values are obtained when the measuring-device-equipped vehicle operates at speeds greater than 50 mph. As a result, only crashes that occurred on interstates were filtered for binomial logistic regression. There were 147 CMV crashes that occurred on DC CMV routes that were also interstates. The following subsections provide a description of those crash data points.

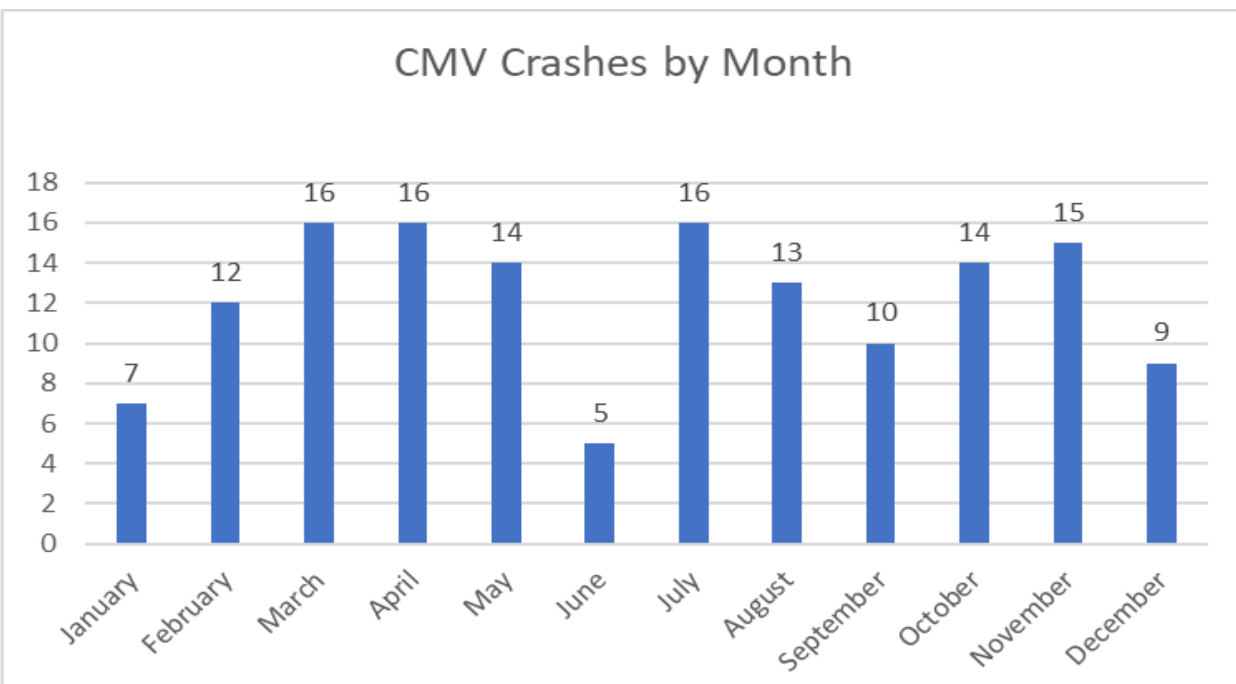
4.10.1 CMV Crashes in DC Interstates by Month

Table 12 presents the distribution of injury and non-injury CMV crash occurrences on interstates from 2016–2020, categorized by month. Figures 31 presents a bar graph of these CMV collisions. From the bar graph, it can be seen that most of the CMV crashes occurred in the months of March, April, and July, collectively constituting almost one-third of the total number of crashes. June had the lowest instances of CMV crashes in the five-year duration.

Table 12. CMV Interstate Crashes 2016–2020 by Month

	Non-Injury	Injury
January	7	0
February	8	4
March	11	5
April	10	6
May	10	4
June	3	2
July	10	6
August	11	2
September	5	5
October	10	4
November	10	5
December	4	5
Total	99	48

Figure 31. Interstate CMV Crashes 2016–2020 by Month



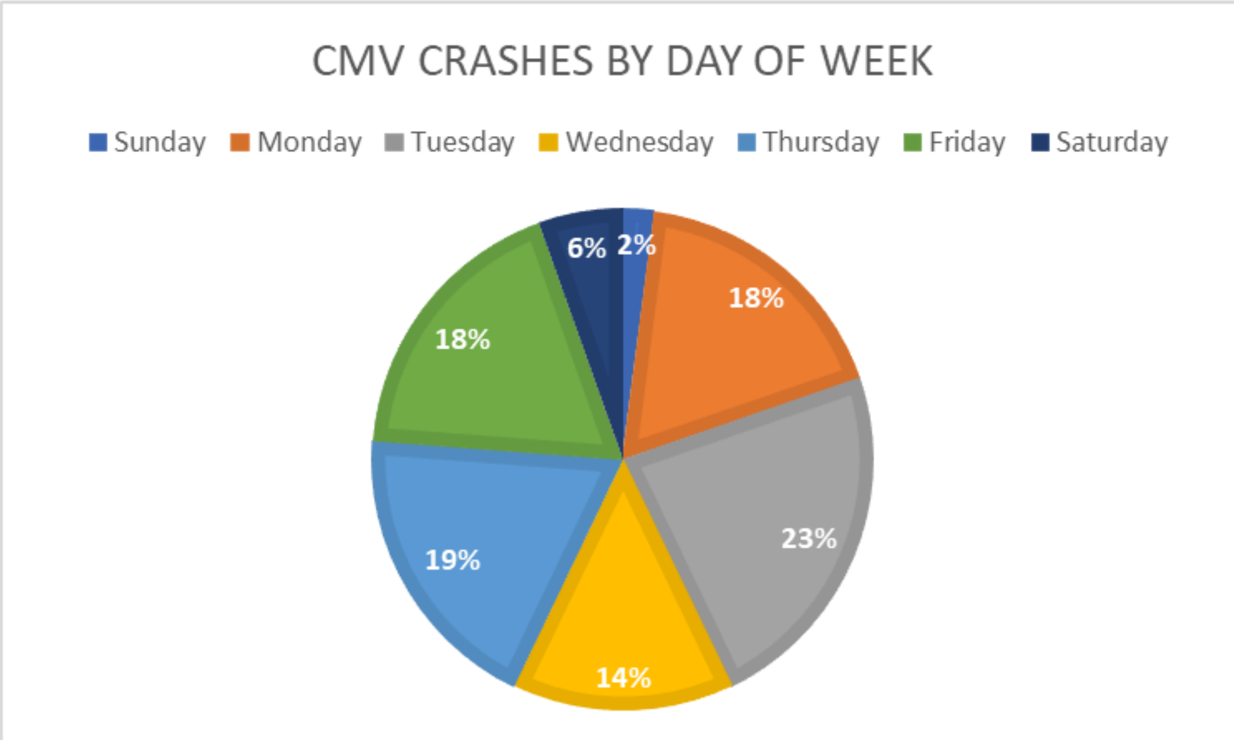
4.10.2 CMV Crashes in DC Interstates by Day of the Week

Table 13 presents the distribution of injury and non-injury CMV crash occurrences on interstates from 2016–2020, categorized by day. From the table, it can be seen that Sundays had the lowest frequency of CMV collisions, and these resulted in the lowest non-injury (2) as well as injury (1) counts. On the other hand, Tuesdays had the highest CMV collision injury (11) and non-injury counts (23). Figures 32 presents a pie chart of these CMV collisions by day from 2016 to 2020.

Table 13. CMV Interstate Crashes 2016–2020 by Day

Row Labels	Non-Injury	Injury
Sunday	2	1
Monday	18	8
Tuesday	23	11
Wednesday	12	9
Thursday	20	8
Friday	17	10
Saturday	7	1
Total	99	48

Figure 32. Interstate CMV Crashes 2016–2020 by Day



4.10.3 CMV Crashes in DC Interstates by Period of the Day

Figures 33 and 34 present the CMV injury and non-injury crashes and total percentage distribution of CMV crashes, respectively, by period of the day. The AM Peak (7 AM–10 AM), Off-Peak (10 AM–4 PM), PM Peak (4 PM–7 PM), and Night (7 PM–7 AM) details are provided. It should be noted that the Off-Peak is twice as long as the AM and PM Peaks, and Night is twice as long as Off-Peak.

Figure 33. Interstate CMV Crashes with and without Injuries (2016–2020)

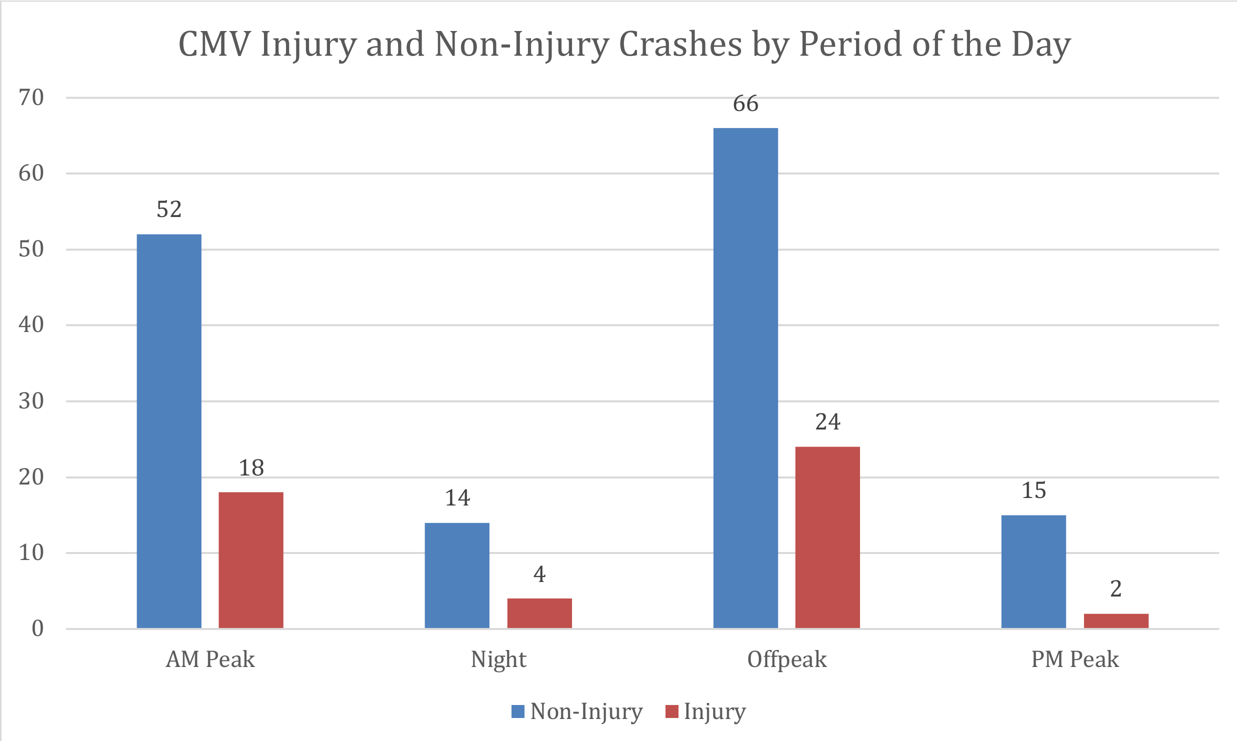
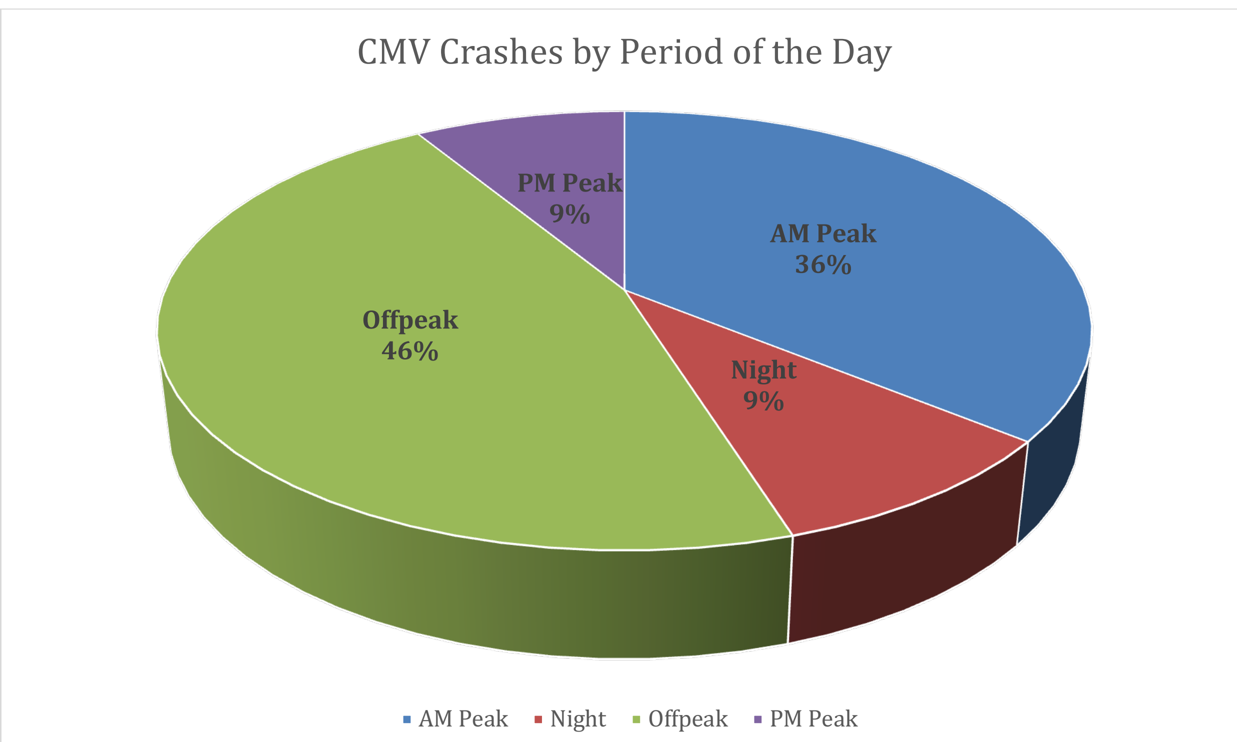


Figure 34. Interstate CMV Crashes by Time of Day



From the figures, it can be seen that almost half of all CMV crashes that occurred on interstates happened during the Off-Peak (10 AM–4 PM). The lowest incidences of injury crashes occurred during the PM Peak (4 PM–7 PM) and Night (7 PM–7 AM) as observed in Figure 34.

4.10.4 Street Lighting Type and Lighting Conditions

Figure 35 presents the distribution of street lighting types in injury and non-injury interstate CMV crashes. The lighting condition during CMV crashes is shown in Figure 36.

Figure 35. Street Lighting Type during Interstate CMV Crashes

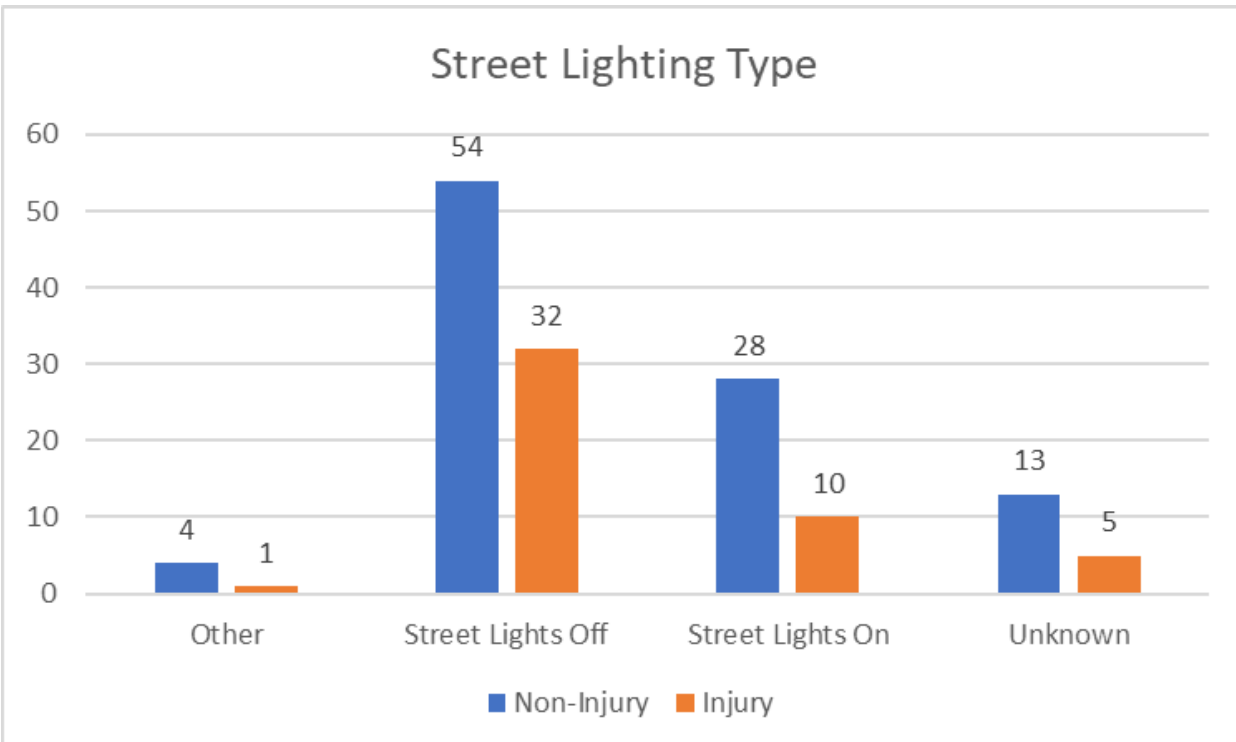
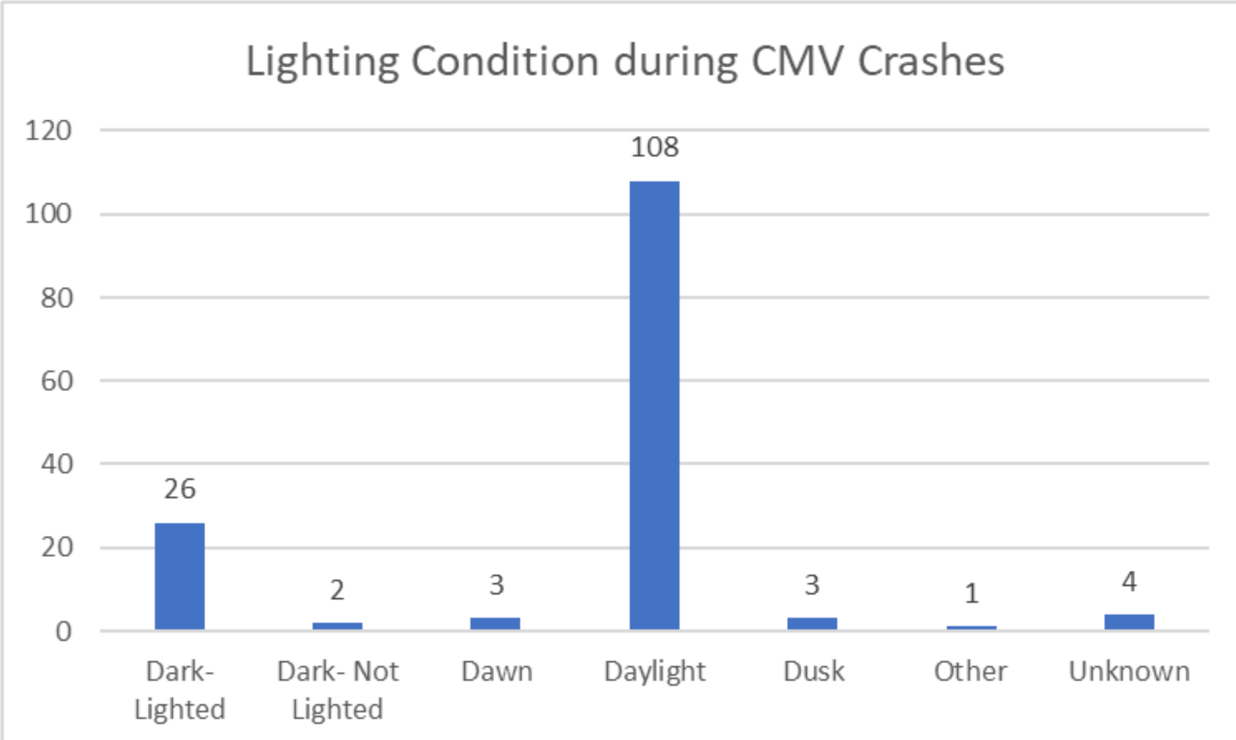


Figure 36. Street Lighting Conditions during Interstate CMV Crashes

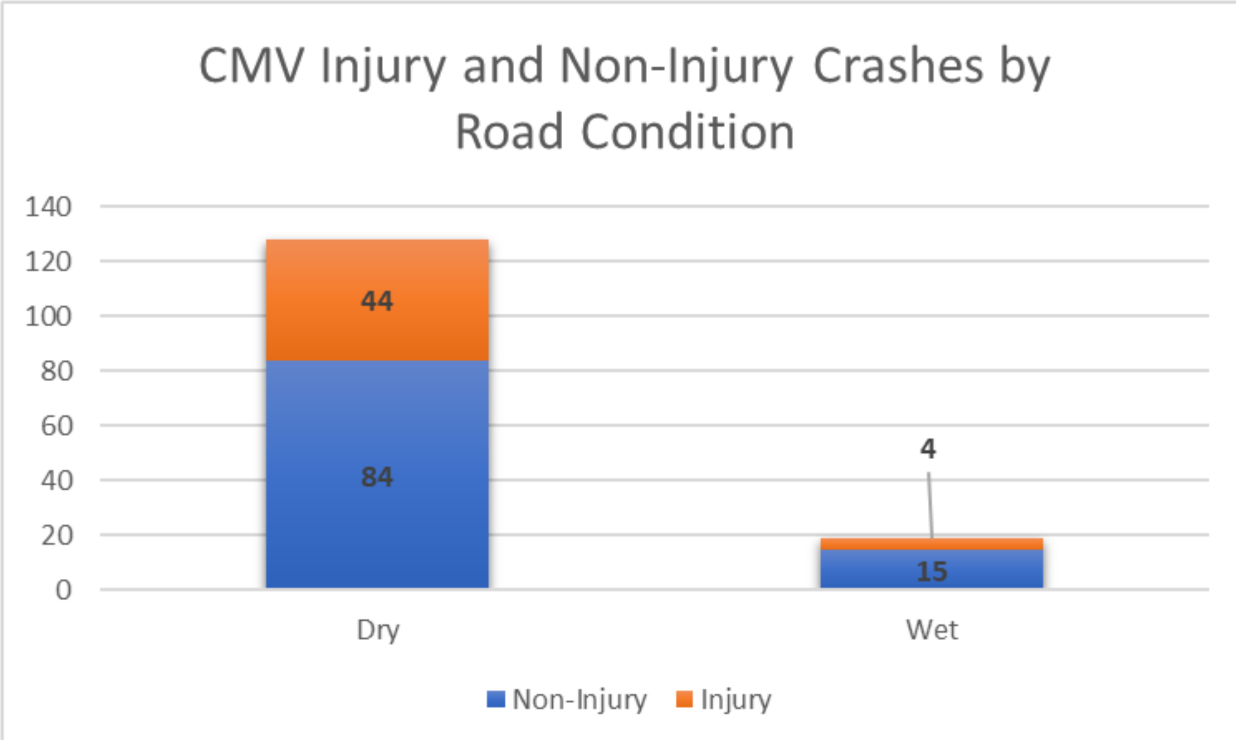


As shown in the figures, most CMV crashes occurred during the daytime when the streetlights were off. From Figures 34 through 36, typically most crashes were common during the day (Off-Peak), which is counterintuitive to expectations of crashes occurring in poorly lit conditions/during peak periods with heavy vehicular volume.

4.10.5 Road Condition for CMV Crash Data Points

The frequency of crashes on different road types, divided into the crashes that involved injuries and those that did not, is shown in Figure 37. Most crashes (87%) occurred on dry roadways.

Figure 37. Injury Counts in Interstate CMV Crashes (by Road Condition)



From the Figure, it can be seen that the observed injury to non-injury ratio for CMV crashes occurring on dry roads was approximately 1:2, while the ratio for crashes recorded on wet roads was about 1:4.

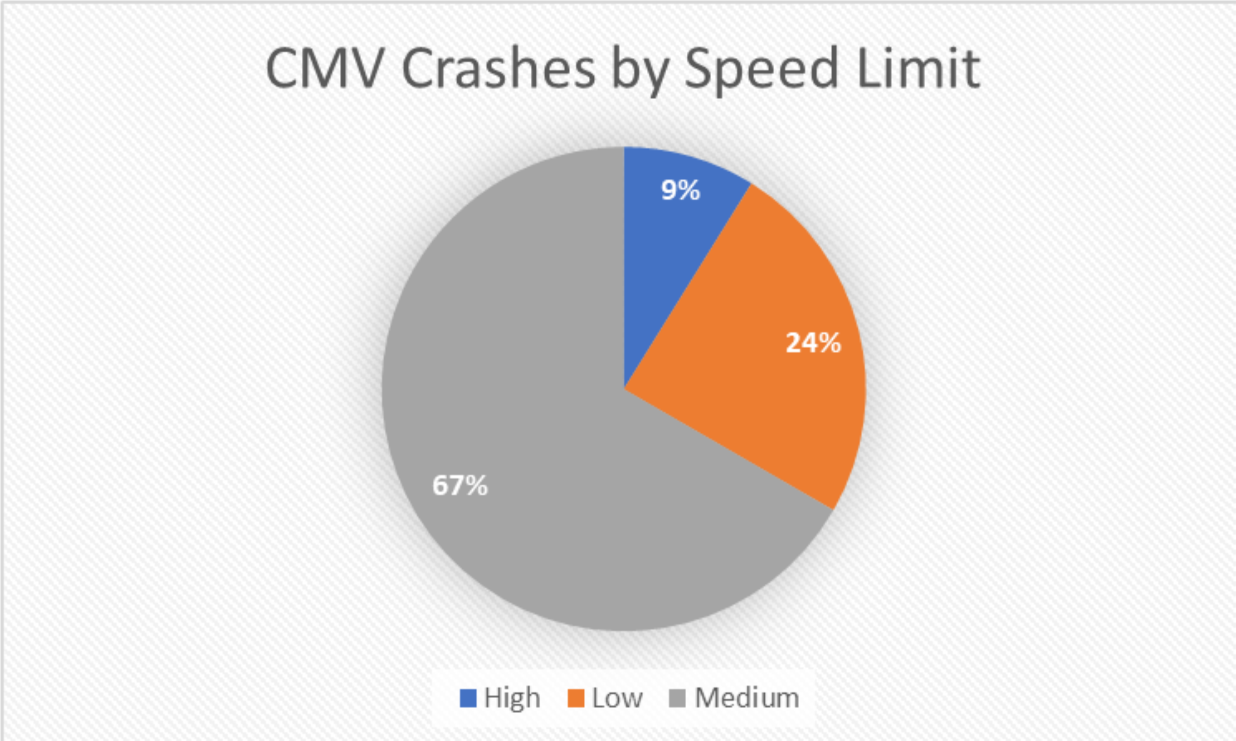
4.10.6 Speed Limit during CMV Crashes

The 147 CMV crashes that were recorded on DC interstates were further categorized by speed limit. Figure 38 presents a pie chart of the CMV crash frequency distribution by speed limit. The speed limit categories are: low (≤ 25 mph), medium (25–55 mph), and high (> 55 mph). The associated injury and non-injury counts are presented in Table 14.

Table 14. Speed Limit Details in Interstate CMV Crashes

Row Labels	Non-Injury	Injury
High	9	4
Low	28	8
Medium	62	36
Total	99	48

Figure 38. Interstate CMV Distribution by Road Speed Limit



4.10.7 IRI Data for Interstate CMV Crashes

The IRI data was obtained for the DC interstates that were geospatially associated with individual CMV crash events. IRI scores can be broadly categorized into Rough ($\leq$ 2.5), Marginal (2.6–3.4), and Smooth ($\geq$ 3.5) based on the measured values for the road segment. The IRI categories associated with all interstate CMV crashes are presented in Figure 39. Figure 40 presents the corresponding injury and non-injury numbers in relation to the different IRI categories.

Figure 39. Interstate CMV Crash Distribution by Different Road IRI Categories

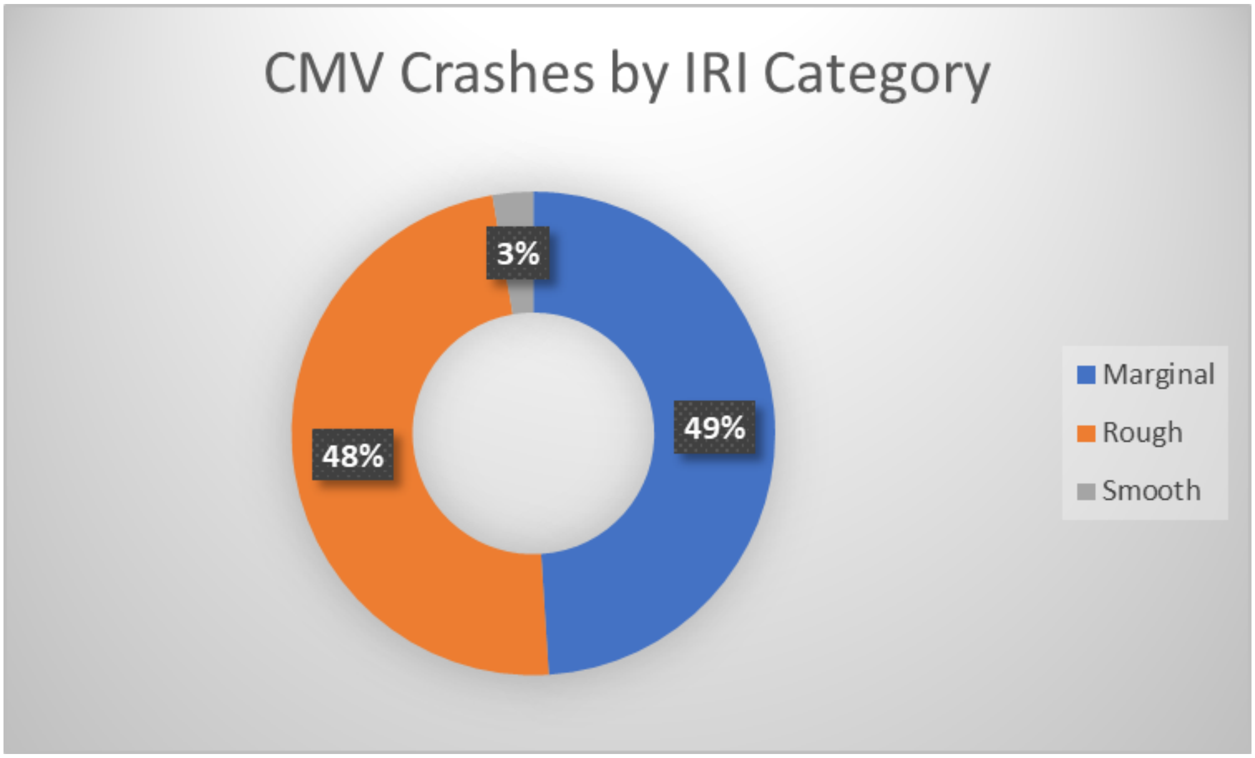
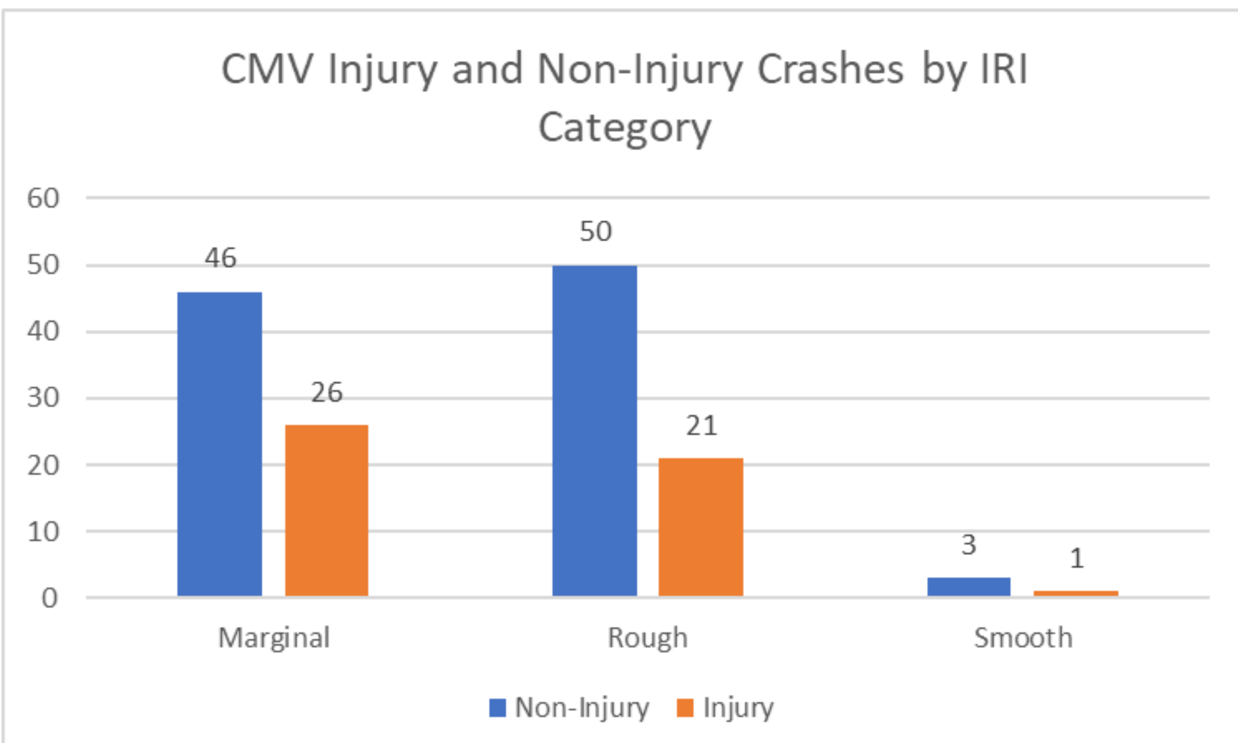


Figure 40. Injury Details in Interstate CMV Crash Distribution by Different Road IRI Categories



4.10.8 PCI Data for Interstate CMV Crashes

PCI data was also obtained for DC interstates and were geospatially associated with all the CMV crash events. There are five categories for PCI scores, which range from Poor to Excellent. The PCI categories associated with all interstate CMV crashes are presented in Figure 41. Figure 42 presents the injury and non-injury numbers corresponding to the different PCI categories.

Figure 41. Interstate CMV Crash Distribution by Different Road PCI Categories

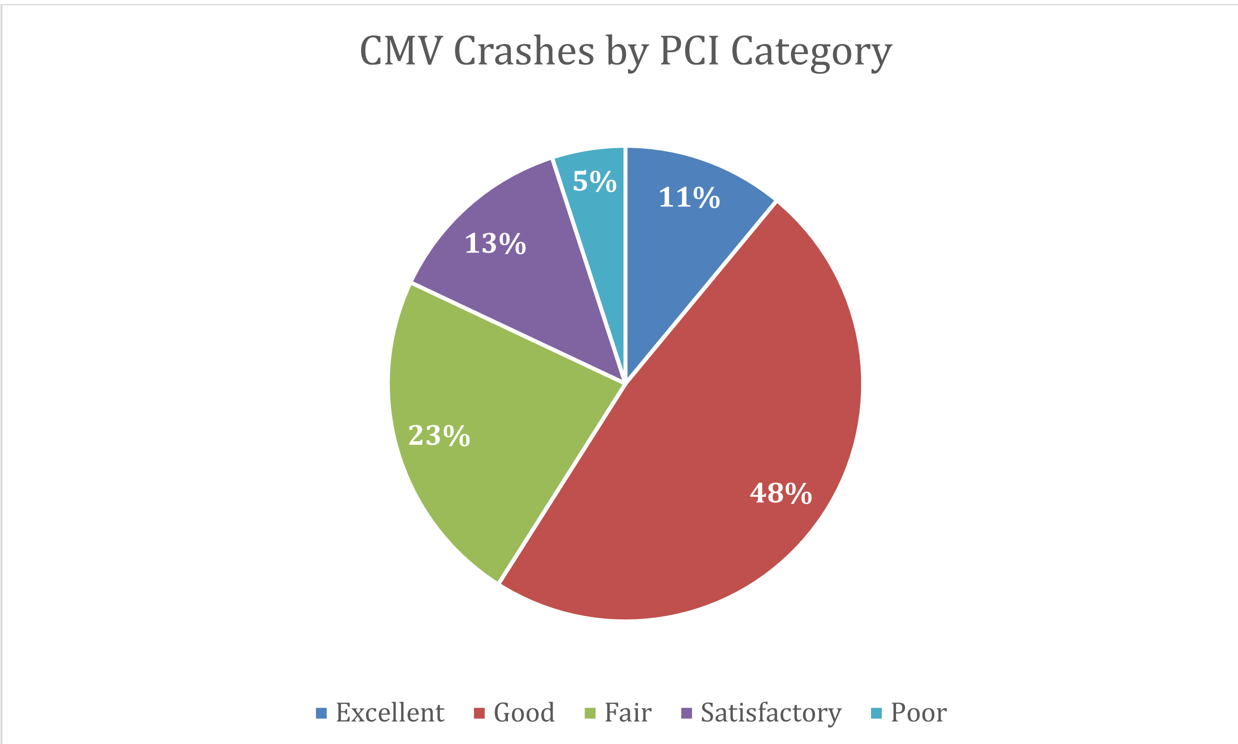
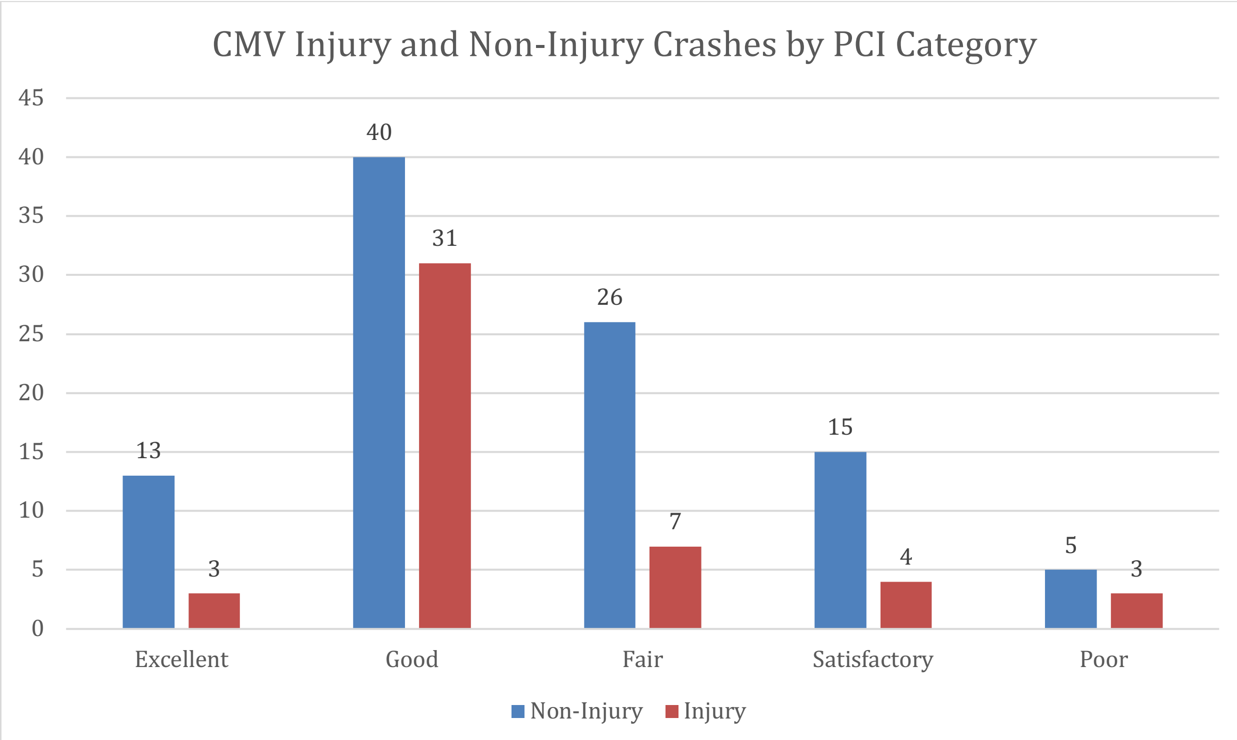


Figure 42. Injury Details in Interstate CMV Crash Distribution by Different Road PCI Categories



4.11 Binary Logistic Regression Output

This section presents the outcomes of the regression analysis conducted in SPSS to predict the incidence of injuries from CMV crashes that occurred on interstates. The analysis is based on the different IRI categories (from one dataset) that were associated with the individual CMV crash incidence from another dataset by matching the coordinates. The "Enter Variable Selection Method" was used to determine the influence of different predictor variables in the regression equation. For instance, Block 1 contains variables that were specifically associated with each individual CMV crash (month/time of day/road condition, etc.), while Block 2 contains the IRI condition of the road during the year in which the CMV crash in question occurred. An additional block with PCI data (Block 3) was also evaluated. Table 15 presents the model summary table for all three blocks.

Table 15. Logistic Regression Model Summary

Block	-2 Log likelihood	Cox & Snell R-Square	Nagelkerke R-Square
1	184.878	0.006	0.008
2	144.167	0.246	0.343
3	134.341	0.295	0.411

Cox & Snell R-square and Nagelkerke R-square metrics can be used for comparing different logistic regression models or assessing the goodness-of-fit of a single model. Both values can range from 0 to 1, with higher values indicating a better fit between the model and the observed data. From Table 15, for Block 1, the explained variation (the model's ability to predict outcomes) in the dependent variable, after the logistic regression, range from 0.6% (0.006 for Cox & Snell R-square) to 0.8% (0.008 for Nagelkerke R-square). Hence, the results indicate a low level of goodness-of-fit. This range of values suggests that the model is not a good fit for the data, and it accounts for only up to a 0.8% proportion of the variability in the response variable. However, by adding variables, such as IRI values (continuous values) and IRI categories (nominal values), to Block 2, the outcome prediction rose to 34% (Nagelkerke R-square value of 0.343). Moreover, the addition of PCI values and PCI Categories in Block 3 indicated the higher predictability of the model (41%). Hence, the addition of road pavement condition variables indicates a stronger relationship or a better predictability than Blocks 1 and 2. This interpretation is also supplemented by conducting the Hosmer-Lemeshow test to assess whether there are significant discrepancies between the observed and predicted outcomes.

The results of Hosmer-Lemeshow Test are presented in Table 16.

Table 16. Hosmer-Lemeshow Test

Block	Chi-square	df	Sig. (p-values)
1	7.02	8	0.534
2	2.735	8	0.95
3	3.373	8	0.1

The Hosmer-Lemeshow test is a commonly used diagnostic tool in logistic regression analysis, which is used to assess the fit of the model and the observed data set. From Table 16, none of the blocks were found to be statistically significant (since the obtained p-values > 0.05, hence not significant). Hence, the results of the Chi-square test suggest that there is no statistically significant difference between the expected and predicted frequencies.

Following the Hosmer-Lemeshow test, the results of the probability estimation of an event for binomial logistic regression are presented in Table 17. The prediction pertains to the occurrence of injury after a CMV crash.

Table 17. Classification Table

Block 1	INJY_CRSH_	0	89	10	89.9
		1	26	22	45.8
	Overall Percentage				75.5
Block 2	INJY_CRSH_	0	88	11	88.9
		1	24	24	50
	Overall Percentage				76.2
Block 3	INJY_CRSH_	0	90	9	90.9
		1	22	26	54.2
	Overall Percentage				78.9

From Table 17, it can be observed that the model was able to correctly predict the likelihood of the injury following a CMV crash approximately 75.5% of the time when no pavement condition variables were present (Block 1). By adding variables IRI and IRI category, the prediction of the model rose to almost 76.2% (Block 2). The inclusion of PCI and PCI category increased the overall correct prediction percentage to nearly 78.9% (Block 3).

The contribution of each independent variable, or predictor, to the regression model along with the statistical significance of Block 3 (highest overall prediction percentage) has been presented in Table 18.

Table 18. Variables in the Equation

		B	S.E.	Wald	df	Sig. (p-value)	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1a	April			10.216	11	.511			
	August	-3.448	1.405	6.021	1	.014	.032	.002	.500
	December	.543	1.146	.225	1	.636	1.722	.182	16.282
	February	-1.577	1.172	1.812	1	.178	.207	.021	2.053
	January	-22.477	13194.781	.000	1	.999	.000	.000	.
	July	-1.107	1.009	1.204	1	.273	.331	.046	2.389
	June	.943	1.398	.456	1	.500	2.568	.166	39.747
	March	-.803	.987	.662	1	.416	.448	.065	3.098
	May	-.807	1.056	.585	1	.444	.446	.056	3.532
	November	-.577	1.110	.270	1	.603	.561	.064	4.949
	October	-.461	1.030	.200	1	.655	.631	.084	4.751
	September	.478	1.103	.188	1	.665	1.613	.186	14.029
	AM Peak			8.565	4	.073			
	Dawn	3.238	1.633	3.930	1	.047	25.483	1.037	626.013
	Night	1.515	1.356	1.249	1	.264	4.552	.319	64.950
	Off Peak	-.416	.602	.476	1	.490	.660	.203	2.148
	PM Peak	-2.220	1.096	4.103	1	.043	.109	.013	.931
	Friday			7.302	6	.294			
	Monday	-.747	.814	.843	1	.359	.474	.096	2.335
	Saturday	-3.554	1.835	3.752	1	.053	.029	.001	1.043
	Sunday	-.477	1.589	.090	1	.764	.620	.028	13.973

	B	S.E.	Wald	df	Sig. (p-value)	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
Thursday	-.867	.790	1.205	1	.272	.420	.089	1.976
Tuesday	-.449	.794	.320	1	.571	.638	.135	3.024
Wednesday	.767	.951	.650	1	.420	2.152	.334	13.870
Other			1.153	2	.562			
Street L	1.375	1.371	1.006	1	.316	3.955	.269	58.108
Unknown	1.593	1.529	1.086	1	.297	4.920	.246	98.427
Dark-Li			1.308	6	.971			
Dark-No	-1.933	2.363	.669	1	.413	.145	.001	14.866
Dawn	-22.876	18238.576	.000	1	.999	.000	.000	.
Daylight	.722	1.085	.443	1	.505	2.059	.246	17.258
Dusk	-20.505	19211.640	.000	1	.999	.000	.000	.
Other	19.781	40192.969	.000	1	1.000	389867878.327	.000	.
Unknown	.250	1.843	.018	1	.892	1.284	.035	47.571
Road Condition (1)	-1.360	.820	2.752	1	.097	.257	.051	1.280
High (Speed Limit)			7.582	2	**.023**			
Low (Speed Limit)	-.874	.923	.896	1	.344	.417	.068	2.549
Medium (Speed Limit)	.924	.833	1.232	1	.267	2.520	.493	12.892
Marginal			1.124	2	.570			
Rough	.548	.765	.513	1	.474	1.729	.386	7.741
Smooth	1.285	1.753	.538	1	.463	3.616	.116	112.272
IRI	-.002	.004	.365	1	.546	.998	.990	1.005
PCI	-.018	.035	.262	1	.609	.983	.918	1.051

	B	S.E.	Wald	df	Sig. (p-value)	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
Excellent			7.789	4	.100			
Fair	-1.385	1.401	.977	1	.323	.250	.016	3.901
Good	1.010	.919	1.208	1	.272	2.746	.453	16.636
Poor	-.742	2.389	.096	1	.756	.476	.004	51.472
Satisfactory	.426	1.311	.106	1	.745	1.532	.117	20.011
Constant	-.240	4.064	.003	1	.953	.787		

From Table 18, only the independent variable Speed Limit had a significance value (or p-value) of less than 0.05 (0.023), indicating that it was a statistically significant predictor of the occurrence of injury following a CMV crash in the DC interstate road network.

Hence, the results of the logistic regression performed to find the likelihood of injuries following an interstate CMV crash indicate that the best-performing model (Block 3) is not statistically significant, $\chi2(8) = 13.373$, $p < 0.1$. However, the model can explain 41.1% (Nagelkerke R-square) of the variance in injury prediction and correctly classified 78.9% of CMV crash occurrences. The details of the individual variable contribution have been provided in the Appendix.

4.12 Geospatial Visualization

This section of the report presents the visual depiction of all CMV crashes that occurred on DC CMV routes throughout the study period (2016–2020), as well as the CMV crashes that occurred only on interstates. A Geographic Information System (GIS) software, ArcGIS, was used to superimpose all the CMV crash locations onto a map of DC. The software was used to filter injury and non-injury crashes by year, time of day, and road condition to better understand crash patterns and identify high-frequency crash zones in DC. All CMV crashes occurring on DC CMV routes from 2016–2020 have been presented in Figure 43. A heatmap was also generated and is displayed in Figure 44. By visually examining the CMV crash data locations, it is possible to gain insights into how the patterns of activity have changed over time. Moreover, it helps identify problematic areas in need of intervention.

Figure 43. ArcGIS Map of All CMV Crashes on CMV Routes 2016–2020

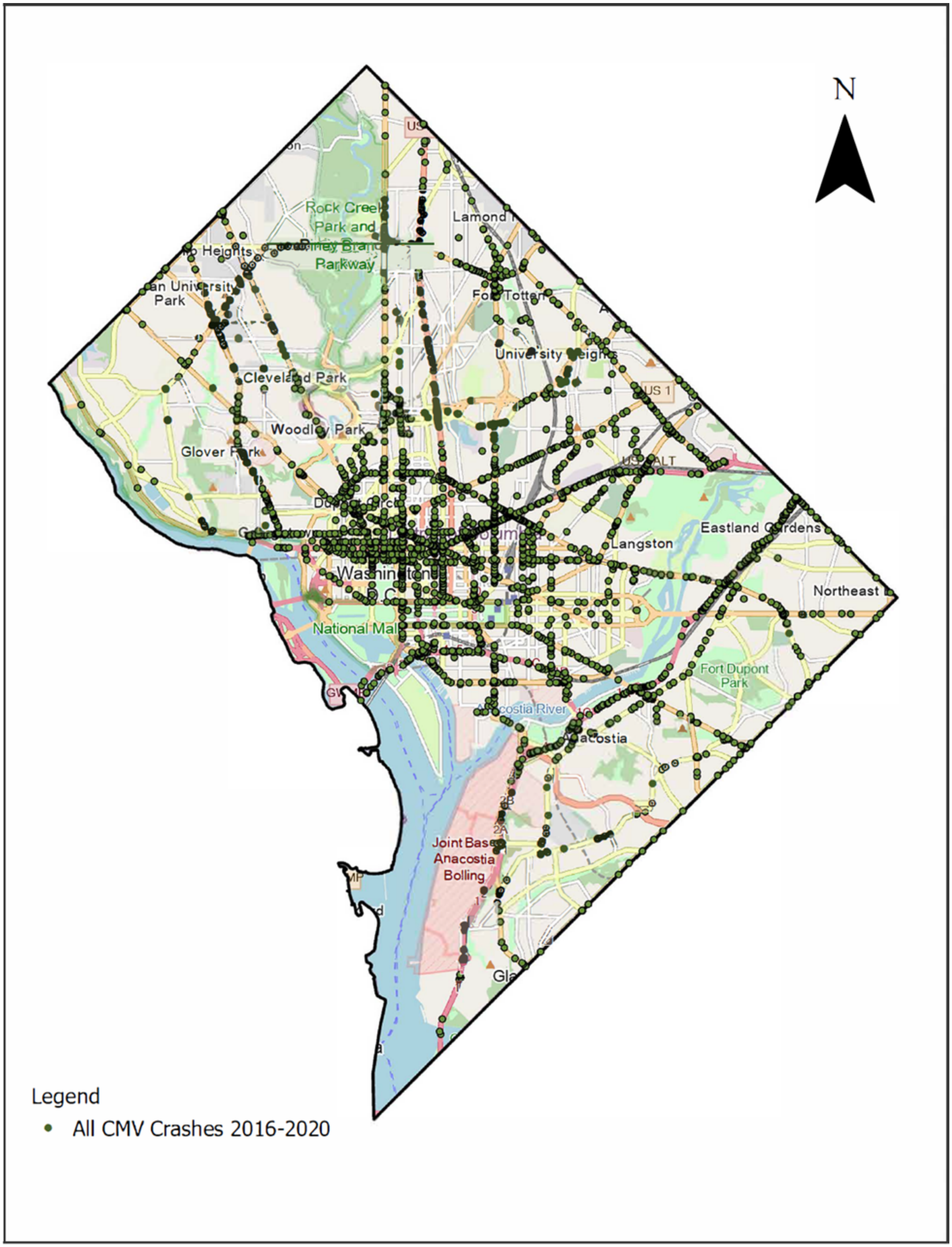
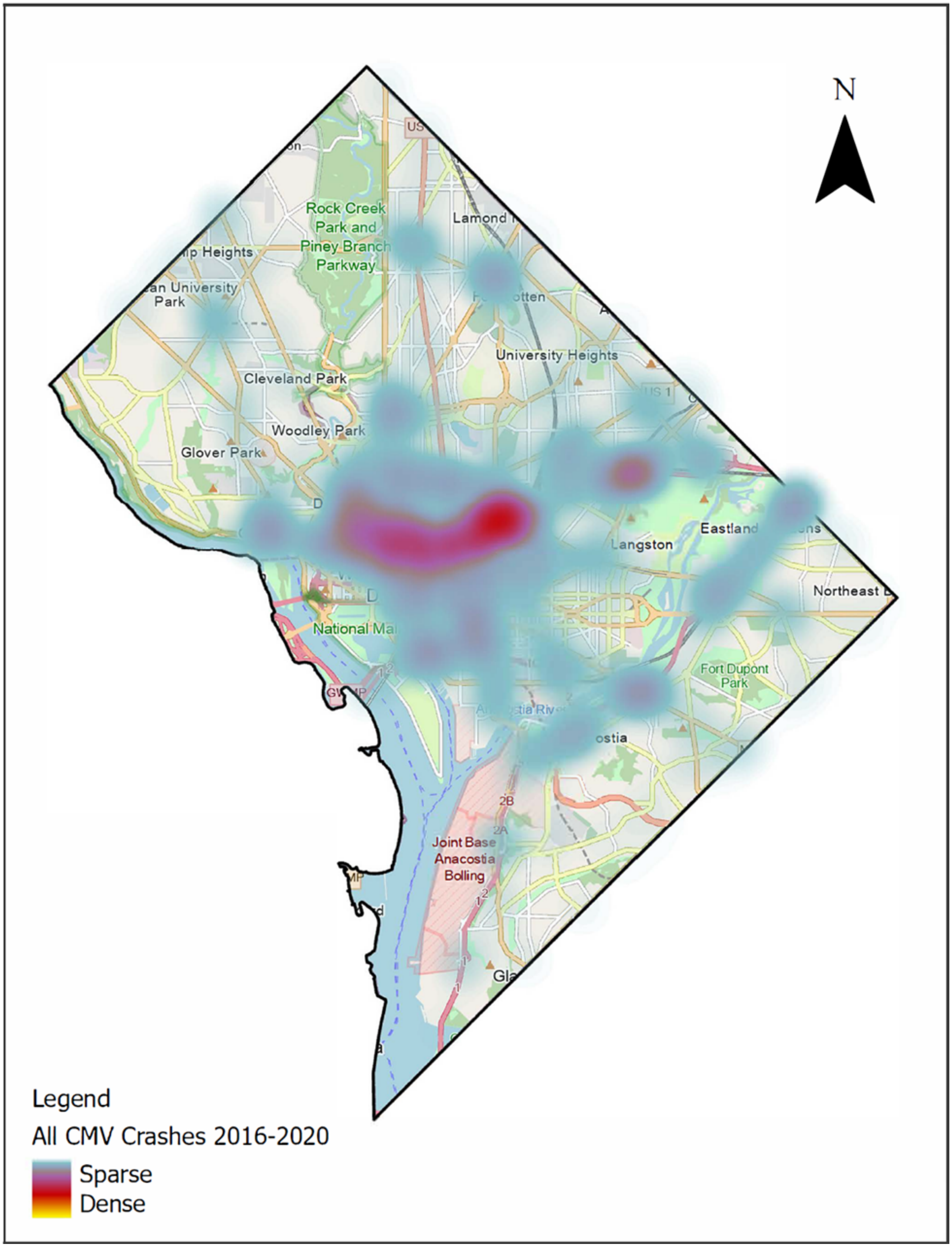


Figure 44. ArcGIS Heatmap of All CMV Crashes on CMV Routes 2016–2020



Figures 45 and 46 show that the frequency of CMV crashes resulting in injuries over the study duration is concentrated near the lower part of the Northwest Quadrant, which happens to be around the downtown area of DC. ArcGIS was also used to plot and generate a heat map of all CMV crashes that resulted in one or more injuries.

Figure 45. ArcGIS Map of All CMV Injury Crashes on CMV Routes 2016–2020

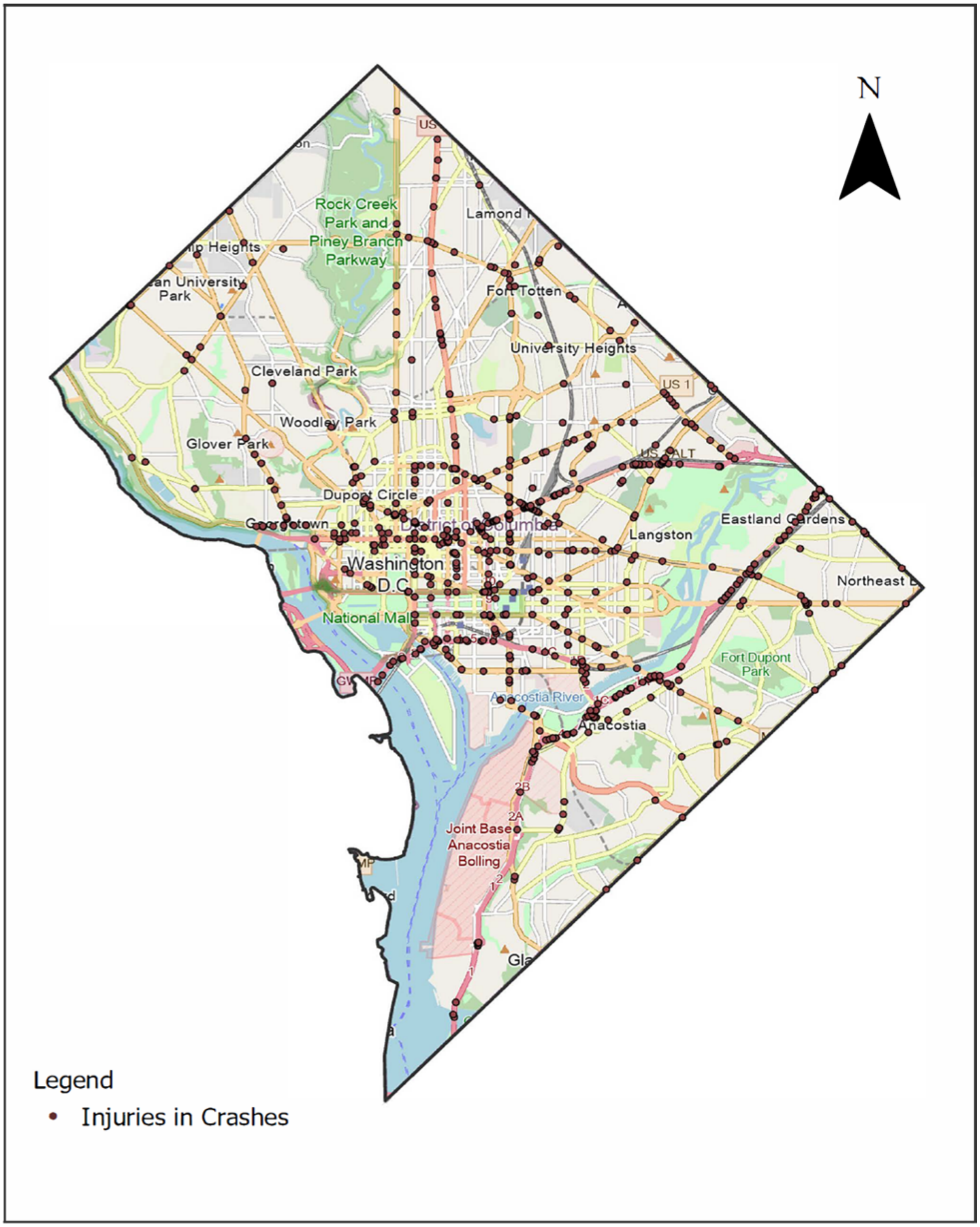
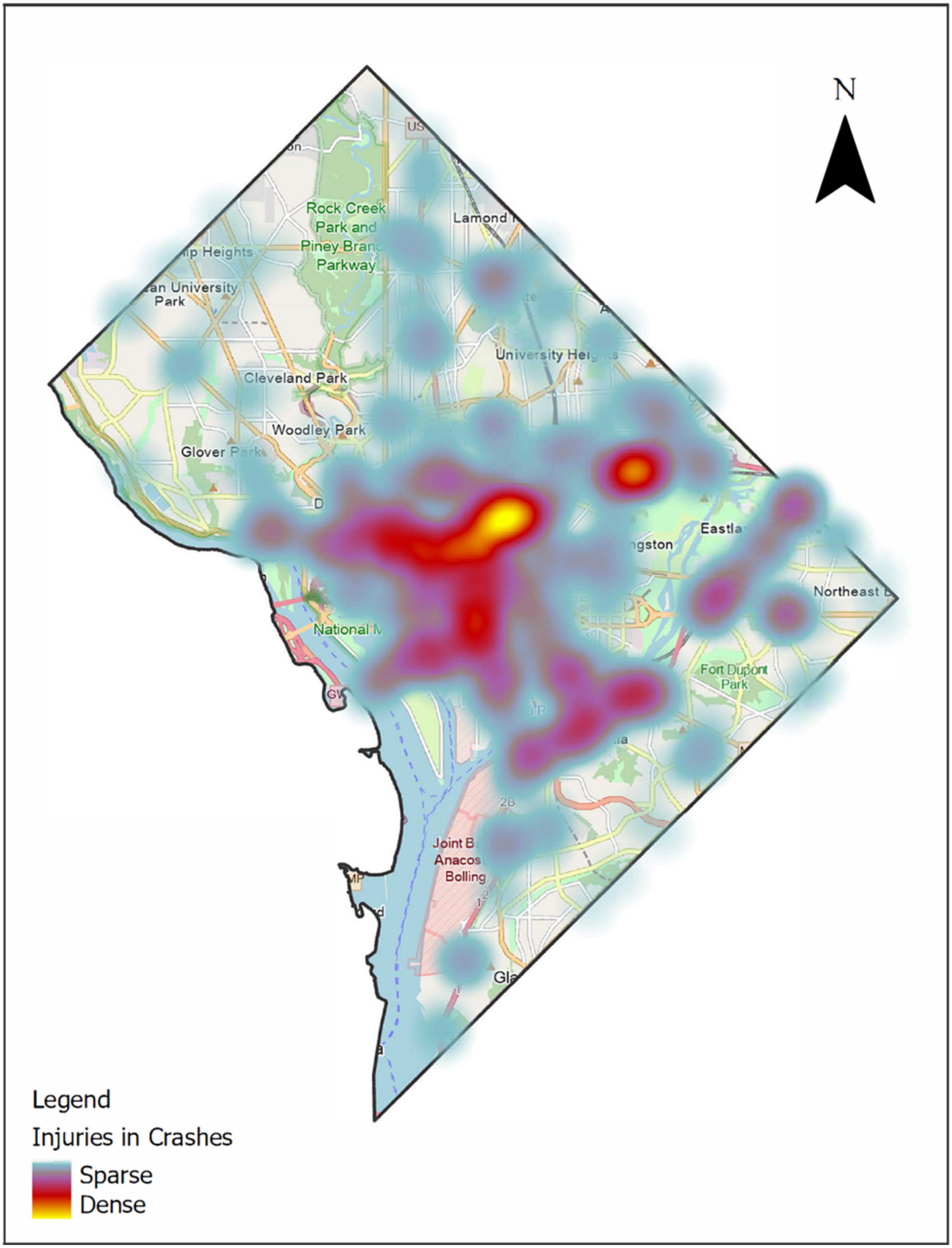


Figure 46. ArcGIS Heatmap of All CMV Injury Crashes on CMV Routes 2016–2020



The heatmaps displaying 5-year CMV crashes during various times of the day are depicted in Figures 47 through 50. Respectively, the heatmaps for the AM Peak (7 AM–10 AM), Mid-Day Peak (10 AM–4 PM), PM Peak (4 PM–7 PM), and Night Time (7 PM–7 AM) CMV crashes have been presented.

Figure 47. ArcGIS Heatmap of 2016–2020 CMV Crashes (AM Peak)

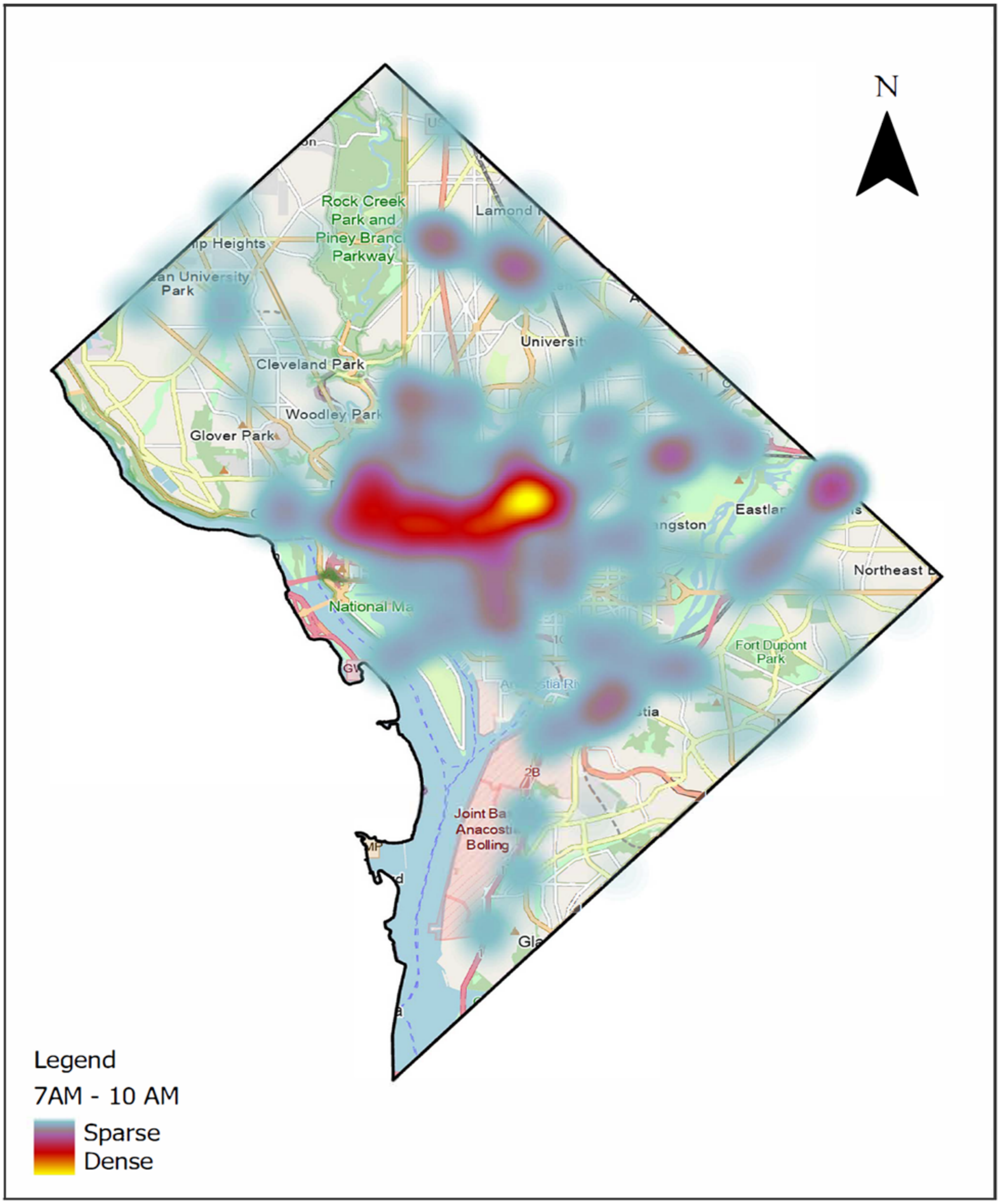


Figure 48. ArcGIS Heatmap of 2016–2020 CMV Crashes (Mid-Day Peak)

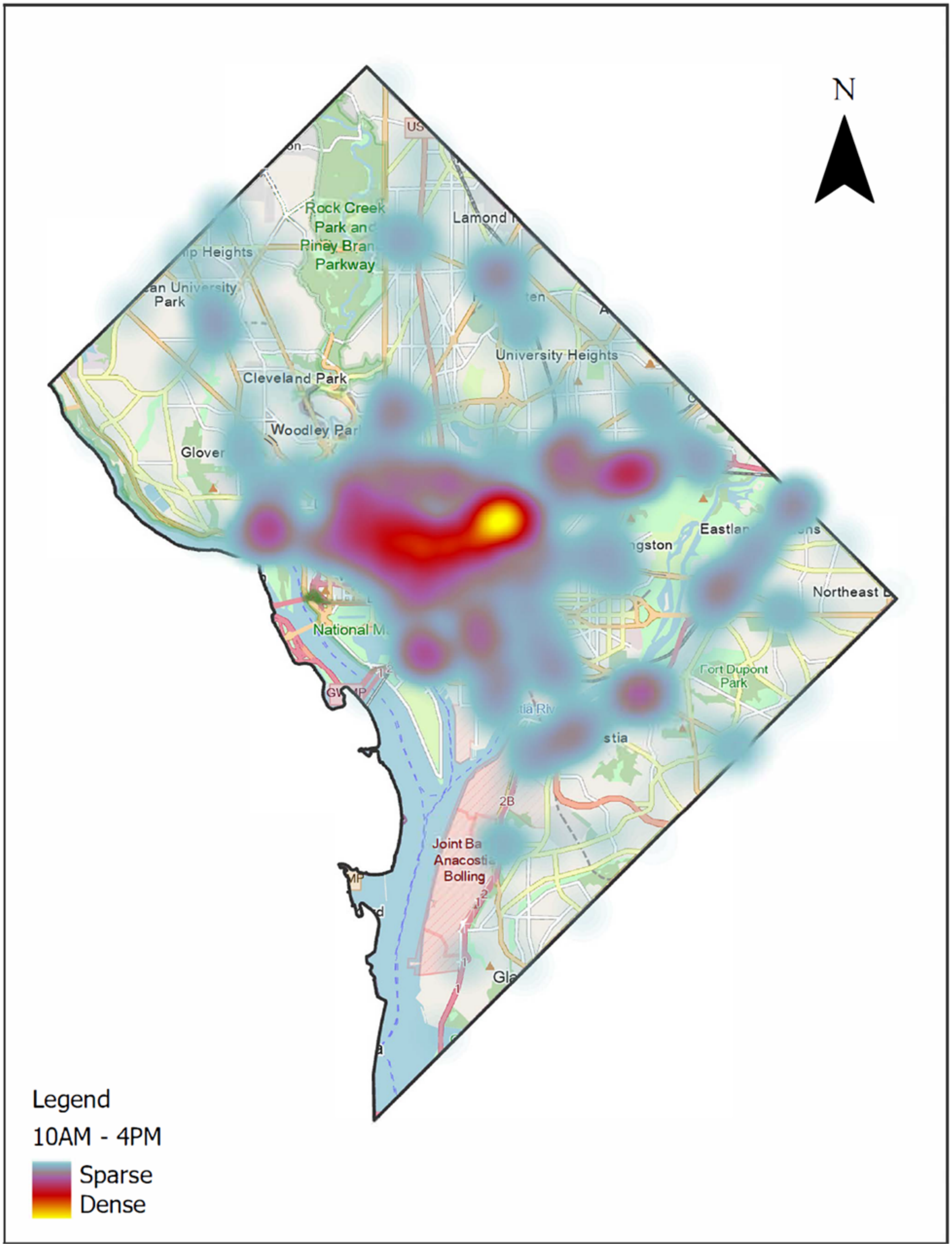


Figure 49. ArcGIS Heatmap of 2016–2020 CMV Crashes (PM Peak)

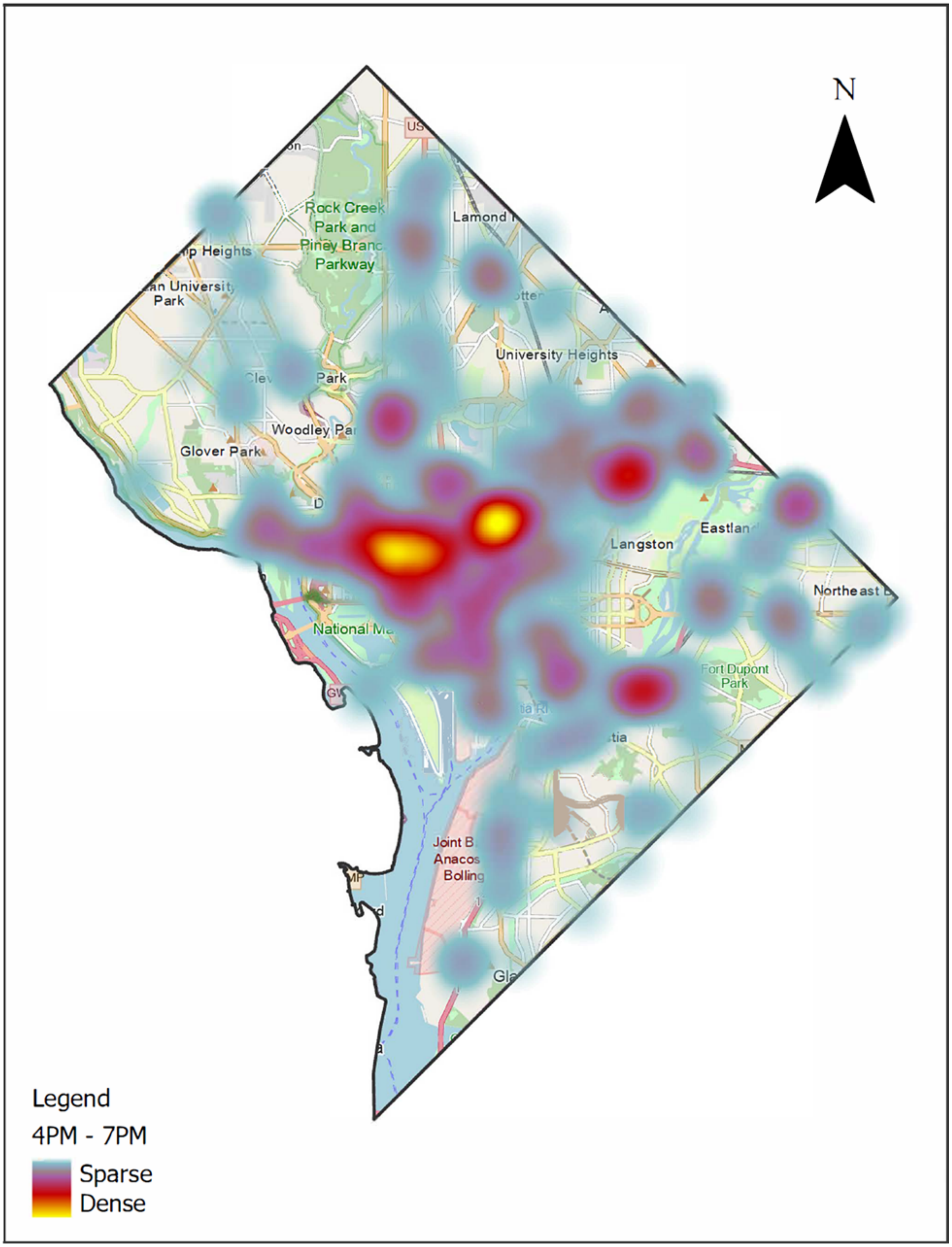
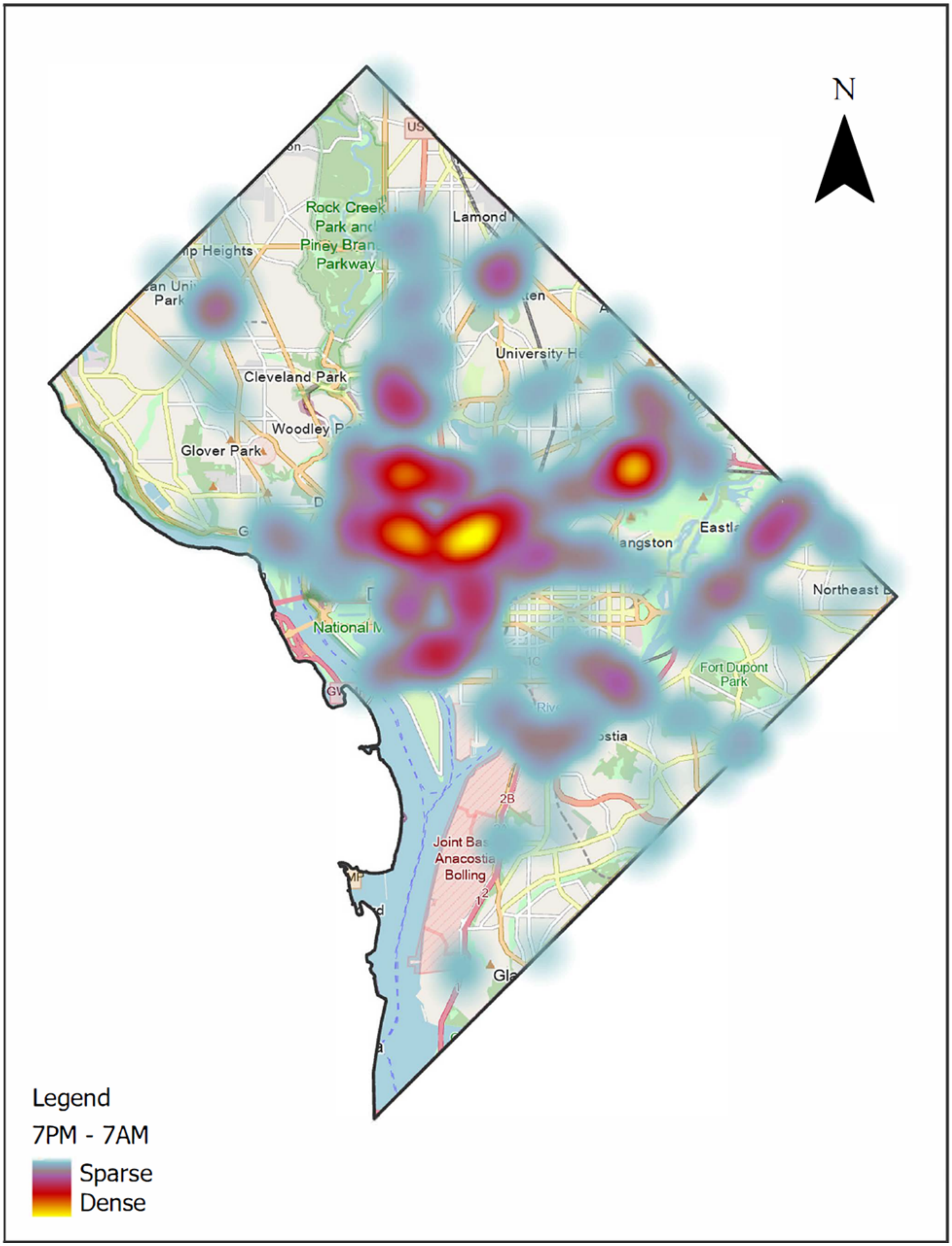


Figure 50. ArcGIS Heatmap of 2016–2020 CMV Crashes (Night Time)



The CMV crashes occurring exclusively on interstates were also filtered for geospatial analysis. Figures 51 and 52 present the GIS scatterplot and heat map of these crashes, respectively.

Figure 51. ArcGIS Scatterplot of 2016–2020 CMV Crashes Occurring on DC Interstates

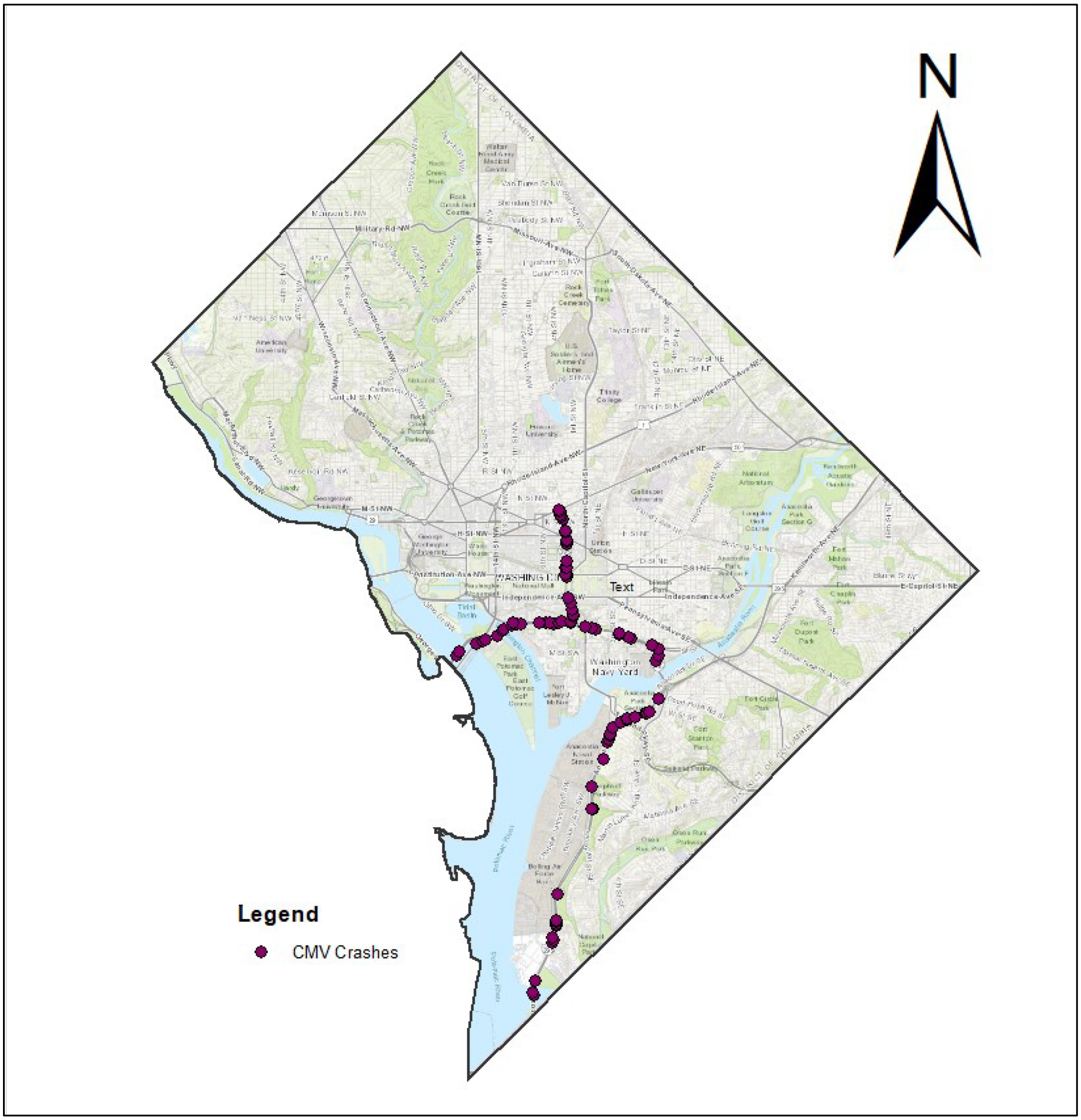
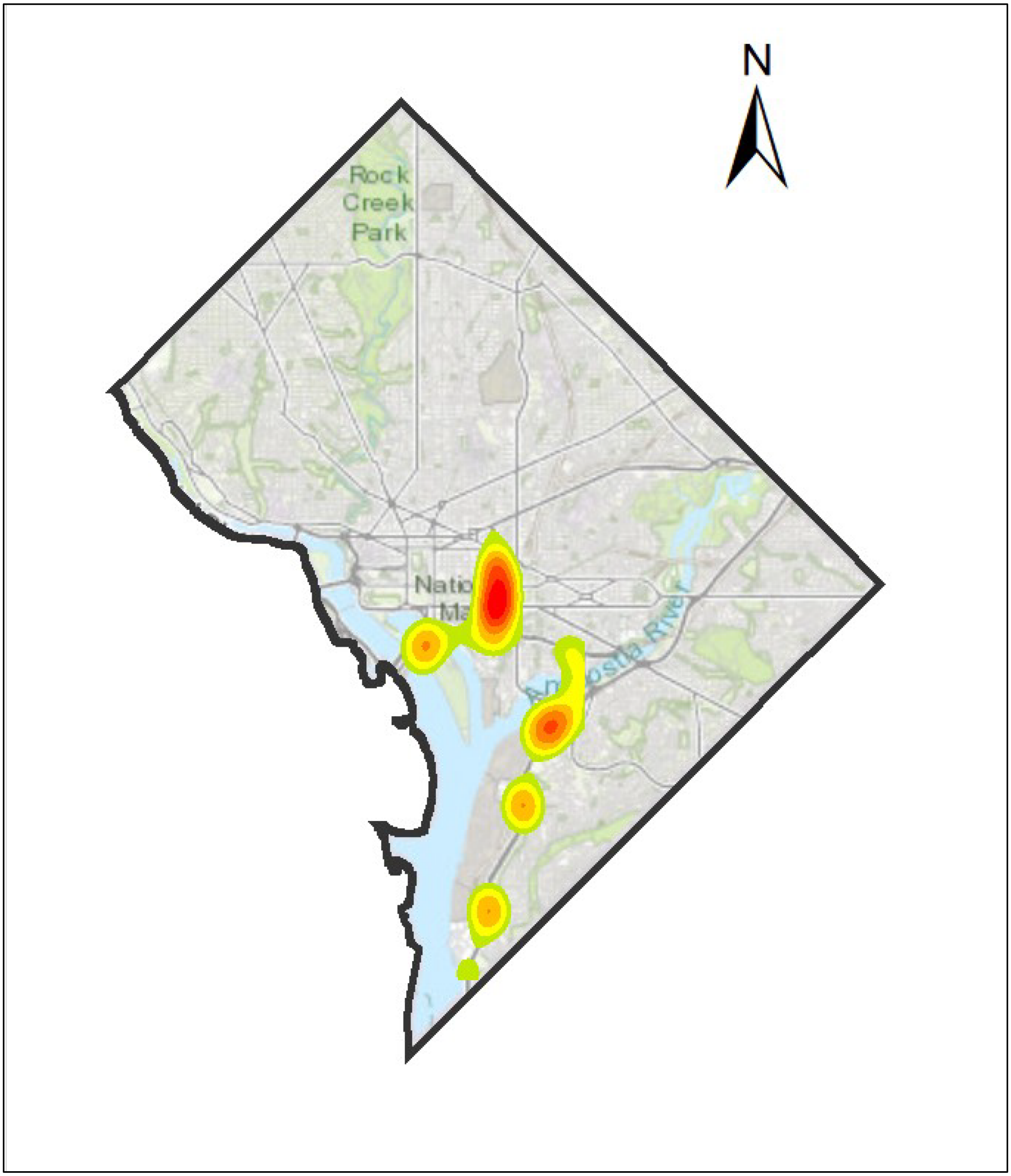


Figure 52. ArcGIS Heatmap of 2016–2020 CMV Crashes Occurring on DC Interstates



4.13 ANN Results

The results of the classification of crashes using an ANN are presented in this section. Ten distinct ANN models were developed using the training dataset. Each model was trained with batches of three observations per iteration until the stopping criterion of 100 epochs was met. The performance of each model was then evaluated using the test dataset (which constitutes 25% of the total dataset). The performance of the models after training and testing are presented in Figure 53, which shows the number of models developed, the structure of the neural network architecture, and the confusion matrix of each model. The performance measures (accuracy, error rate, sensitivity, precision, and F-measure) of each model were computed using the equations presented in Table 2.

The architecture of Model numbers 1 and 2 consists of three layers: 1 input layer, 1 hidden layer, and 1 output layer. The number of neurons in the hidden layer ranges from 5 to 200. The architecture of Model numbers 3 and 4 consists of four layers: 1 input layer, 2 hidden layers, and 1 output layer. The number of neurons in the hidden layers ranges from 5 to 30. The architecture of Model numbers 5 and 6 consists of 5 layers: 1 input layer, 3 hidden layers, and 1 output layer. The number of neurons in the hidden layers ranges from 5 to 10. The architecture of Model numbers 7 and 8 consists of 6 layers: 1 input layer, 4 hidden layers, and 1 output layer. The number of neurons in the hidden layers ranges from 5 to 100. The architecture of Models number 9 and 10 consists of 7 layers: 1 input layer, 5 hidden layers, and 1 output layer. The number of neurons in the hidden layers ranges from 5 to 200.

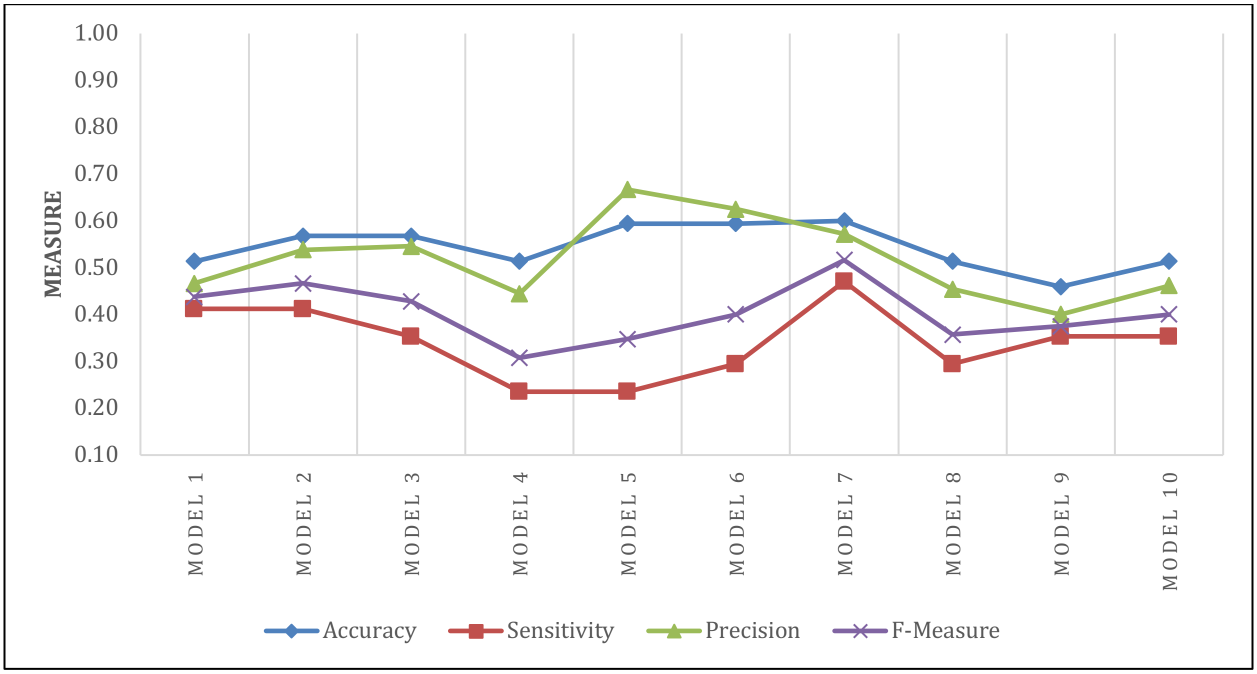
The indices in the confusion matrices of each model in Figure 53 are arranged in accordance with the convention shown in Table 1. The top row values are the number of true negative (TN) and false positive (FP) predictions, respectively, while the values in the bottom row are the false negative (FN) and true positive (TP) predictions, respectively. The accuracy, sensitivity, precision, and F-measure performance measures range from 0 to 1, with a value closer to 1 showing models with better performance and conversely a value closer to 0 showing worse performance.

The results of the analysis tabulated in Figure 53 show that, after the models' training and evaluating with the test dataset, the models' accuracy ranges from 46% to 60%. Model number 7 produced the best classification: accuracy (60%), F-measure (0.52), and sensitivity (0.47). With regards to the precision measure, Model number 5 was the most precise model with a precision of 0.67. The variation of performance measures with varying models is shown in Figure 54.

Figure 53. ANN Performance Measure

Model No.	No. of Hidden Layers	No. of Neurons	Confusion Matrix		Accuracy	Error Rate	Sensitivity	Precision	F-Measure
1	1	55	12	8	0.51	0.49	0.41	0.47	0.44
			10	7					
2	1	30	14	6	0.57	0.43	0.41	0.54	0.47
			10	7					
3	1	20	15	5	0.57	0.43	0.35	0.55	0.43
	2	30	11	6					
4	1	5	15	5	0.51	0.49	0.24	0.44	0.31
	2	15	13	4					
5	1	5	18	2	0.59	0.41	0.24	0.67	0.35
	2	10	13	4					
	3	5							
6	1	2	17	3	0.59	0.41	0.29	0.63	0.40
	2	5	12	5					
	3	2							
7	1	20	14	6	0.60	0.40	0.47	0.57	0.52
	2	60	9	8					
	3	80							
	4	20							
8	1	5	14	6	0.51	0.49	0.29	0.45	0.36
	2	20	12	5					
	3	80							
	4	100							
	5	20							
9	1	10	11	9	0.46	0.54	0.35	0.40	0.38
	2	30	11	6					
	3	60							
	4	200							
	5	60							
	6	20							
10	1	5	13	7	0.51	0.49	0.35	0.46	0.40
	2	20	11	6					
	3	60							
	4	120							
	5	60							
	6	5							

Figure 54. Variation of ANN Performance Measure



5. Discussion

The primary objectives of this study are to understand the relationships between road pavement conditions and CMV crash occurrences, develop predictive models for crash severity, and provide recommendations for improving road safety measures. This section presents a discussion on the findings and insights derived from the analysis conducted.

The analysis revealed a significant decline of approximately 31.6% in the total CMV collisions in 2020 compared to 2019. However, a contrasting trend was observed in the number of fatalities, which increased by almost 36% in 2020. Most crashes resulted in property damage only, accounting for nearly 74%, while injury and fatality collisions constituted about 26% and 0.2%, respectively. Furthermore, it was observed that the occurrence of CMV crashes varied significantly between different types of roadways. Specifically, only 4.08% of crashes occurred on interstates, while the majority were on non-interstate roadways.

Based on the literature, accurate measurements of CMV crashes are obtained when measuring-device-equipped vehicles operate at speeds greater than 50 mph. Due to this, to investigate the influence of roadway pavement condition on CMV crashes, the analysis in this study was based on crashes that only occurred on interstate routes. Firstly, a binary logistic regression was conducted to develop a generalized linear model to predict the severity of crashes. Using the "Block Entry" technique, variables were included in the model in batches. This allowed for the observation of the change in variance explained by the model with the variable's addition.

It was observed that the inclusion of road pavement condition variables (IRI and PCI) significantly improved the model's goodness-of-fit. The Cox & Snell R-square and Nagelkerke R-square metrics indicated that these variables contributed to explaining a notable proportion of the variance in CMV crash injury severity prediction. Without the inclusion of pavement condition variables, the explained variation (the model's ability to predict outcomes) was only 0.8%, while with the inclusion of IRI and PCI, the explained variance increased significantly to 41% (as presented in Table 13). Additionally, the results indicate that the model was able to correctly predict the likelihood of an injury following a CMV crash approximately 79% of the time. However, it is important to note that while IRI and PCI strengthened the predictive power of the model, the overall model fit was not statistically significant, suggesting that other contributing factors also play a role in CMV crash severity. These results corroborate the findings of other literature that suggest that pavement conditions have minimal impact on freeway crashes due to the well-maintained nature of most freeway pavements, and the dataset used for analyzing freeway crashes does not exhibit significant variation in terms of pavement quality.[xii] The results indicated that an ANN model consisting of 4 hidden layers with 20, 60, 80, and 20 neurons, respectively, exhibited the best classification performance with an accuracy of 60% and an F-measure of 0.52.

These findings demonstrate the potential of machine learning techniques to contribute valuable insights into the complex relationships between road pavement conditions and CMV crash occurrences.

6. Conclusions and Recommendations

This study examines the influence of roadway pavement conditions, particularly measured by the International Roughness Index (IRI) and Pavement Condition Index (PCI), on commercial motor vehicle (CMV) crashes in the District of Columbia. More specifically, the study focuses on the CMV crashes that occurred in DC on CMV routes. Findings from the study underscore the significance of road pavement conditions in determining the likelihood of injury following a CMV crash. Two analyses were considered: binary logistic regression and artificial neural network. The logistic regression models indicate that incorporating road pavement conditions significantly improve the predictive accuracy of injuries in the occurrence of CMV crashes. The inclusion of IRI and PCI variables increases prediction accuracy by a considerable margin, highlighting the notable relationship between road pavement conditions and CMV crash severity. Furthermore, the ANN models predicted the severity of CMV crashes with an accuracy of up to 60%. This suggests that artificial neural networks have the potential to predict the relationship between road pavement conditions and CMV crash severity. The study not only sheds light on the impacts of road pavement conditions on CMV crashes but also informs the practical implications for road safety measures.

Based on the finding of this research, the following recommendation are offered:

- Policymakers and the transportation community must prioritize the maintenance and improvement of the road pavement conditions of CMV routes to reduce the severity of CMV crashes when they occur. By fostering quality and safer road infrastructure, the collective losses due to CMV crashes can be reduced.

- Continue supporting research initiatives aimed at understanding the complex relationship between road pavement conditions and CMV crashes. Encouraging innovation in road construction materials and maintenance techniques can lead to long-term improvements in road safety.

- Greater investments in robust data collection and analysis systems to monitor road quality can guide evidence-based policy decisions.

- Conduct road safety awareness campaigns to educate drivers on how road quality affects crash likelihood and severity to encourage them to adapt their driving behavior accordingly.

Bibliography

Anarkooli, Jafari, Alireza, Iliya Nemtsov, and Bhagwant Persaud. "Safety Effects of Maintenance Treatments to Improve Pavement Condition on Two-Lane Rural Roads—Insights for Pavement Management." *Canadian Journal of Civil Engineering* 48, no. 10 (2021): 1287-1294.

Arhin, Stephen A. and Adam Gatiba. "Predicting Injury Severity of Angle Crashes Involving Two Vehicles at Unsignalized Intersections Using Artificial Neural Networks." *Engineering, Technology & Applied Science Research* 9, no. 2 (2019): 3871–80. https://doi.org/10.48084/etasr.2551.

ASCE. "2017 Infrastructure Report Card, 'District of Columbia'." 2017. https://2017.infrastructurereportcard.org/state-item/district-of-columbia/.

Baskara, S. N., H. Yaacob, M. R. Hainin, S. A. Hassan, N. Mashros, N. Z. M. Yunus, N. A. Hassan, et al. "Influence of Pavement Condition towards Accident Number on Malaysian Highway." *IOP Conference Series: Earth and Environmental Science* 220, no. 1 (2019): 012008. https://doi.org/10.1088/1755-1315/220/1/012008.

Bock, Margaret, Alexander Cardazzi, and Brad R. Humphreys. "Where the Rubber Meets the Road: Pavement Damage Reduces Traffic Safety and Speed." *NBER Working Papers* (2021).

Chen, Sikai, Tariq Usman Saeed, and Samuel Labi. "Impact of Road-Surface Condition on Rural Highway Safety: A Multivariate Random Parameters Negative Binomial Approach." *Analytic Methods in Accident Research* 16 (2017): 75–89. https://doi.org/ 10.1016/j.amar.2017.09.001.

Dissanayake, Sunanda and Siddhartha Kotikalapudi. "Characteristics and Contributory Causes Related to Large Truck Crashes (Phase II) – All Crashes," *Res. Innov. Technol. Adm.* (2012).

Elghriany, Ahmed, Ping Yi, Peng Liu, and Quan Yu. "Investigation of the Effect of Pavement Roughness on Crash Rates for Rigid Pavement." *Journal of Transportation Safety & Security* 8, no. 2 (2016): 164–76. https://doi.org/10.1080/19439962.2015.1025458.

Hao, Wei, Camille Kamga, Xianfeng Yang, JiaQi Ma, Ellen Thorson, Ming Zhong, and Chaozhong Wu. "Driver Injury Severity Study for Truck Involved Accidents at Highway-Rail Grade Crossings in the United States." *Transportation Research Part F: Traffic Psychology and Behaviour* 43 (2016): 379–386. https://doi.org/ 10.1016/j.trf.2016.09.001.

Islam, Mouyid, and Salvador Hernandez. “Large Truck-Involved Crashes: Exploratory Injury Severity Analysis.” *Journal of Transportation Engineering* 139, no. 6 (2013): 596–604. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000539.

Kathpalia, Sunaina Bakshi. "Charts of the Week: A Pandemic-Induced Exodus Has Broken the District’s Population Boom." D. C. Policty Center. March 25, 2022. https://www.dcpolicycenter.org/publications/census-shows-pandemic-exodus-has-broken-dc-population-growth/.

Levinson, David, Toshihiro Yokoo, and Mihai Marasteanu. “Pavement Condition and Crashes,” *Findings* (2019). https://doi.org/10.32866/5771.

Li, Yingfeng, Chunxiao Liu, and Liang Ding. “Impact of Pavement Conditions on Crash Severity.” *Accident; Analysis and Prevention* 59 (2013): 399–406. https://doi.org/10.1016/j.aap.2013.06.028.

Liu, Jinli, Yi Qi, Jueqiang Tao, and Tao Tao. “Analysis of the Performance of Machine Learning Models in Predicting the Severity Level of Large-Truck Crashes.” Future Transportation 2, no. 4 (2022): 939–55. https://doi.org/10.3390/futuretransp2040052.

Mamlouk, Micheal, Mounica Vinayakamurthy, Shane Underwood, and Kamil Kaloush. “Effects of the International Roughness Index and Rut Depth on Crash Rates.” *Transportation Research Record: Journal of the Transportation Research Board* 2672, no. 40 (2018). https://doi.org/10.1177/0361198118781137.

Metropolitan Police Washington, DC. “Traffic Data.” (n.d.). https://mpdc.dc.gov/page/traffic-data.

Naik, Bhaven, Li-Wei Tung, Shanshan Zhao, and Aemal J. Khattak. “Weather Impacts on Single-Vehicle Truck Crash Injury Severity.” *Journal of Safety Research* 58 (2016): 57–65. https://doi.org/10.1016/j.jsr.2016.06.005.

Sagar, Shraddha, Nikiforos Stamatiadis, Samantha Wright, and Aaron Cambron. “Identifying High-Risk Commercial Vehicle Drivers Using Sociodemographic Characteristics.” *Accident Analysis & Prevention* 143 (2020): 105582. https://doi.org/10.1016/j.aap.2020.105582.

Sayers, Michael W., Thomas D. Gillespie, and Cesar A. V. Queiroz. *The International Road Roughness Experiment: Establishing Correlation and a Calibration Standard for Measurements* Ann Arbor: University of Michigan, Transportation Research Institute, 1986.

Shafabakhsh, Gholam Ali, Afshin Famili, and Mohammad Sadegh Bahadori. "GIS-Based Spatial Analysis of Urban Traffic Accidents: Case Study in Mashhad, Iran." *Journal of Traffic and Transportation Engineering (English Edition)* 4, no. 3 (2017): 290–299. https://doi.org/10.1016/j.jtte.2017.05.005.

Uddin, Majbah and Nathan Huynh. "Factors Influencing Injury Severity of Crashes Involving HAZMAT Trucks." *International Journal Transportation Science Technology* 7, no. 1 (2018): 1–9. https://doi.org/10.1016/j.ijtst.2017.06.004.

American Counts Staff. "The District of Columbia Gained More Than 87,000 People in 10 years." United States Census Bureau. August 25, 2021. https://www.census.gov/library/stories/state-by-state/district-of-columbia-population-change-between-census-decade.html.

United States Department of Transportation. Bureau of Transportation Statistics. "U.S. Gross Domestic Product (GDP) Attributed to Transportation Functions." https://www.bts.gov/content/us-gross-domestic-product-gdp-attributed-transportation-functions-billions-current-dollars.

United States Department of Transportation, Federal Motor Carrier Safety Administration, Office of Analysis, Research, and Technology. "Pocket Guide to Large Truck and Bus Statistics 2018." (2018). https://doi.org/10.21949/1502788.

United States Department of Transportation, National Highway Traffic Safety Administration's (NHTSA) National Center for Statistics and Analysis. "Rural/Urban Comparison of Traffic Fatalities." *Traffic Safety Fact Sheets: 2020* (July 2022).

United States Department of Transportation, NHTSA. "Motor Vehicle Traffic Crashes as a Leading Cause of Death in the United States: 2015." (2018).

United States Department of Transportation, Office of Highway Policy Information. "HPMS Public Release of Geospatial Data in Shapefile Format." (2018). https://www.fhwa.dot.gov/policyinformation/hpms/shapefiles.cfm.

Zhao, Qun, Tyrie Goodman, Mehdi Azimi, and Yi Qi. "Roadway-Related Truck Crash Risk Analysis: Case Studies in Texas." *Transportation Research Record* 2672, no. 34 (2018): 20–28. https://doi.org/10.1177/0361198118794055.

About the Authors

Stephen Arhin, PhD, PE, PTOE, PMP, CRA, Fellow ITE

Dr. Stephen Arhin is a Professor and the Chair of the Civil and Environmental Engineering Department at Howard University. He has more than 27 years of experience in all facets of traffic and transportation engineering. He is a registered Professional Engineer (P.E.) in DC, DE, MD, and VA, and a registered Professional Traffic Operations Engineer (PTOE). He is also a Certified Research Administrator (CRA), a certified Project Management Professional (PMP), and a Fellow of the Institute of Transportation Engineers (ITE). He has extensive experience in working with state and local transportation agencies on a wide variety of safety, operations, and design projects in addition to private industry. Dr. Arhin has authored and co-authored several project reports, published articles in peer-reviewed journals, and presented at conferences on such topics as pedestrian and bicycle safety, countdown pedestrian traffic signals, intelligent transportation technologies, pavement condition monitoring, crash data analysis, traffic volume trends, mitigation of reflective cracking in composite highway pavements, truck weight enforcement, and red-light violation. He is a member of ITE, TRB, and ASCE, and serves as a Program Evaluator for ABET.

Babin Manandhar, EIT

Babin Manandhar is an Engineer in Training (EIT) and a research engineer at the Howard University Transportation Research and Data Center. He has a Master of Engineering degree in Civil and Environmental Engineering from Howard University.

Adam Gatiba, PE

Adam Gatiba is a Professional Engineer who has worked for various engineering consulting firms, assisting federal, state, and local agencies in providing solutions to transportation engineering problems. He also supports the HUTRC in numerous data-driven transportation engineering research projects. Adam holds a Master of Science degree in Civil Engineering from Howard University.

MTI FOUNDER

Hon. Norman Y. Mineta

MTI BOARD OF TRUSTEES

Directors

MINETA TRANSPORTATION INSTITUTE