Technical Report Documentation Page

1. Report No. ABC-UTC-2016-C3-FIU03-Final	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle **WORK ZONE SAFETY ANALYSIS, INVESTIGATING BENEFITS FROM ACCELERATED BRIDGE CONSTRUCTION (ABC) ON ROADWAY SAFETY**		5. Report Date December 2021 6. Performing Organization Code
7. Author(s) Seyedmirsajd Mokhtarimousavi, Dr. Islam Mantawy, and Atorod Azizinamini (https://orcid.org/0000-0001-7627-6757)		8. Performing Organization Report No. March 2020 to December 2021
9. Performing Organization Name and Address Department of Civil and Environmental Engineering Florida International University 10555 West Flagler Street, EC 3680 Miami, FL 33174		10. Work Unit No. (TRAIS) 11. Contract or Grant No. Enter the correct number: 69A3551747121
12. Sponsoring Organization Name and Address Accelerated Bridge Construction University Transportation Center Florida International University 10555 W. Flagler Street, EC 3680 Miami, FL 33174	US Department of Transportation Office of the Assistant Secretary for Research and Technology And Federal Highway Administration 1200 New Jersey Avenue, SE Washington, DC 201590	13. Type of Report and Period Covered Final Report (January 2020- December 2020) 14. Sponsoring Agency Code
15. Supplementary Notes Visit www.abc-utc.fiu.edu for other ABC reports.		

16. Abstract

Identifying the factors associated with crash severity and frequency in work zone locations is of important value to roadway safety. In addition, the significant loss of workers' lives and injuries resulting from work zone crashes indicates the emergent need for a comprehensive and in-depth investigation of work zone crash mechanisms. The cost of work zone crashes is another issue that should be taken into account as work zone crashes impose millions of dollars on society each year. Applying innovative construction methods like Accelerated Bridge Construction (ABC) dramatically decreases on-site construction duration and thus improves roadway safety. This safe and cost-effective procedure for building new bridges or replacing/rehabilitating existing bridges in just a few weeks instead of months or years may prevent crashes and avoid injuries as a result of work zone presence. To this end, this study focuses on three major areas: crash severity at construction work zones with worker presence, crash frequency at bridge locations, and assessment of the associated costs to calculate the contribution of safety to the benefit-cost ratio of ABC as compared to conventional methods. The findings of this study provide in-depth investigation of work zone crash contributing factors in conjunction with the results from statistical and machine learning models, which can provide a more comprehensive interpretation of crash severity/frequency outcomes. Besides, the results illustrated that the safety benefits of ABC implementation consist of a considerable portion of its benefit-cost ratio as compared to on-site methods.

17. Key Words Work Zone Safety, Construction worker safety, Crash Severity, Crash Frequency, Mixed Logit Model, Support Vector Machine, Cuckoo Search Optimization Algorithm, Artificial Bee Colony Optimization Algorithm, ABC.		18. Distribution Statement No restrictions.	
19. Security Classification (of this report) Unclassified.	20. Security Classification (of this page) Unclassified.	21. No. of Pages Total of 164 pages	22. Price

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

(this page is intentionally left blank)

Work zone safety analysis, investigating benefits from accelerated bridge construction (ABC) on roadway safety

Final Report

October 2021

Principal Investigator: Dr. Islam Mantawy

Department of Civil and Environmental Engineering
Florida International University

Author
Dr. Seyedmirsajad Mokhtarimousavi

Sponsored by
Accelerated Bridge Construction University Transportation Center

A report from
Department of Civil and Environmental Engineering
Florida International University
10555 West Flagler Street, EC 3680
Miami, FL 33174
Phone: 305-348-2824 / Fax: 305-348-2802
https://cee.fiu.edu/

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Program. However, the U.S. Government assumes no liability for the contents or use thereof.

TABLE OF CONTENTS

CHAPTER PAGE

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

FIGURE PAGE

ABBREVIATIONS AND ACRONYMS

AADT	Annual Average Daily Traffic
ABC	Accelerated Bridge Construction
ABC	Artificial Bee Colony
ADT	Average Daily Traffic
ARTBA	American Road and Transportation Builders Association
BCA	Benefit-Cost Analysis
BCR	Benefit-Cost Ratio
BPNN	Back-Propagation Neural Network
CART	Classification and Regression Trees
CF	Crash Frequency
CFS	Correlation-based Feature Selector
DBN	Deep Belief Network
DF	Degree of Freedom
DUI	Driving Under the Influence
FHWA	Federal Highway Administrations
GA	Genetic Algorithm
HSIS	Highway Safety Information System
INJ	Injury
INV	Involved
LE	Law Enforcement
MAD	Mean Absolute Deviation
MGORP	Mixed Generalized Ordered Response Probit

MLP	Multi-Layer Perceptron
MMUCC	Model Minimum Uniform Crash Criteria
MSE	Mean Squared Error
MSPE	Mean Squared Predictor Error
MUTCD	Manual on Uniform Traffic Control Devices
NHTSA	National Highway Traffic Safety Administration
NN	Neural Network
OP	Ordered Probit
PDO	Property Damage Only
RBFNN	Radial Basis Function Neural Network
RCI	Roadway Characteristics Inventory
RLS	Recurrent Least Squares
RMSE	Root Mean Square Errors
S4	Signal Four
SD	Standard Deviation
SEV	Severity
SFP	Safety Performance Function
SPMT	Self-propelled Modular Transporters
SVM	Support Vector Machine
SVR	Support Vector Regression
Veh	Vehicle
WOA	Whale Optimization Algorithm
WZ	Work Zone

ACKNOWLEDGMENTS

This project was supported by the Accelerated Bridge Construction University Transportation Center (ABC-UTC at www.abc-utc.fiu.edu) at Florida International University (FIU). The authors would like to acknowledge the ABC-UTC support.

The author would like to extend special appreciation to the ABC-UTC and the U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology for funding this project.

The author would like to thank Dr. Mohammed Hadi, department of civil and environmental engineering, Florida International University; this work would not have been possible without the valuable support, motivation and encouragement of him.

The author would like to thank the Research Advisory Panel members: Ahmad Abu-Hawash (Iowa DOT), James Corney (Utah DOT), Romeo Garcia (FHWA), and Bruce Johnson (Oregon DOT).

CHAPTER 1
INTRODUCTION

1.1. Research Background and Motivation

According to statistics from the American Road and Transportation Builders Association (ARTBA), Florida is among the top three states with the highest work zone fatal crashes, with a total of 67, 73, and 71 fatal crashes resulting in 73, 80, and 76 fatalities from 2015 to 2017, respectively (ARTBA 2018). Work zone crashes constitute approximately 1.55% of the total crashes (i.e., 2,112,783), with 9,142 injury crashes between the years 2015 and 2017 in Florida (S4A 2018). Figure 1-1 shows the details of the work zone crashes by year in Florida.

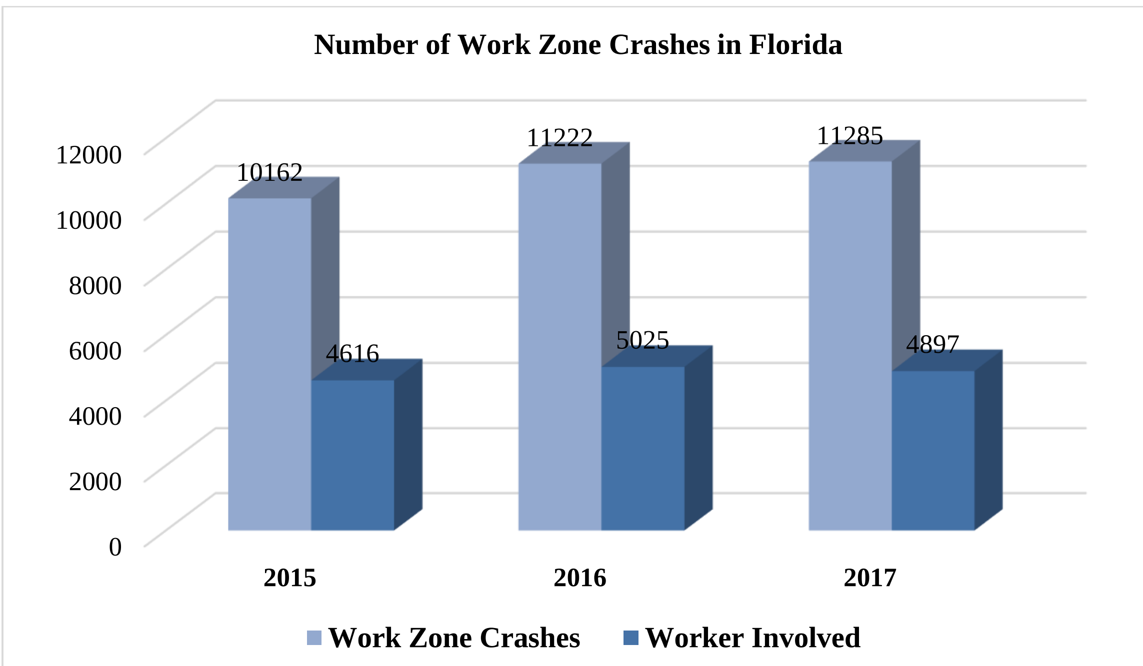


Figure 1-1 Work Zone Crash in FL.

Another important aspect of work zone crashes that needs to be considered by decision makers is worker safety. Among the total number of work zone crashes in 2017 in the state of Florida, around 43.4% were associated with worker presence, in which 16 workers were killed. The worker fatalities in 2017 are 33.3% and 45.45% higher compared to 2016 and 2015, respectively (ARTBA 2018). The significant loss of workers' lives and injuries resulting from work zone crashes indicates the emergent need for a comprehensive and in-depth investigation of work zone crash mechanisms.

The costs of crashes and associated costs are other adverse effects of construction work zones. As mentioned in (Mohan and Gautam 2000), according to the US Department of Transportation's (DOT) National Highway Traffic Safety Administration (NHTSA), the direct costs of work zone crashes in 1997, including 658 fatalities, 36,000 injuries and 52,000 property damage-only crashes, was $5.74 billion in the United States. It has been reported in many previous studies that there is a meaningful deference in crash severity and crash rates, with and without work zone presence conditions (Mallela and Sadavisam 2011). Moreover, in a recent review from Yang et al. on work zone safety analysis and modeling, it was stated that over 85% of previous studies regarding work zone crash frequency demonstrate an obvious increase in crash frequencies during work zone operations (Yang *et al.* 2015). Hence, in order to prevent imposing millions of dollars on society each year, it is necessary to investigate the possible causes associated with crash severity and frequency to improve work zone safety.

According to the Manual on Uniform Traffic Control Devices (MUTCD), bridge construction/reconstruction is categorized as a long-term stationary work zone since the construction duration is typically three days. This long roadway occupancy and its related components, such as lane closures, lane width reductions, changes in road geometry, and the presence of construction workers, increase the crash occurrence risk. Accelerated

Bridge Construction (ABC) is an advanced method of project delivery with the aim to reduce the on-site bridge construction timeline without losing bridge quality (Ralls 2007). This innovative bridge construction method can be employed for either constructing new bridges or the replacement of existing bridges with a significantly lower amount of traffic disruption during implementation.

To date, there is no such study that assesses the roadway safety enhancement aspect of the ABC implementation method. In addition, most of the existing safety research has focused on the traveling public and not on worker safety. To fill this gap, at first glance, this study seeks to identify the contributing factors that affect the severity of work zone crashes associated with worker presence and crash frequency at bridge construction work zone locations. Moreover, it provides quantitative evidence of the benefits that can be obtained through the ABC implementation compared to conventional on-site bridge construction from a roadway safety point of view.

1.2. Research Objectives

Taking the above-mentioned information into consideration, this research aims to identify the contributing factors that affect crash severity and frequency of work zone crashes through a combination of the results from the conventional statistical models and machine learning techniques. This can provide a more comprehensive interpretation of work zone crash severity and frequency outcomes. The analysis sheds light on the internal probability patterns of crash contributing factors, as well as their overall impacts. In addition, it seeks to assess the impact of ABC implementation to enhance work zone safety through a benefit-cost analysis, which has not yet been investigated and documented.

In this regard, different data sources, such as crash records, project information and layouts, roadway geometric features and traffic data, were combined to develop enhanced prediction models. In order to identify factors affecting work zone crash severity, a three-year period of statewide crash data was obtained from the Florida Signal Four Analytics (S4A) tool for worker-involved work zone crashes. The most significant contributing factors in terms of crash severity were investigated using logistic regression and Support Vector Machine (SVM) modeling frameworks for daytime and nighttime conditions separately. In addition, likelihood ratio tests were conducted to examine the overall stability of model estimates across time periods.

Identifying factors affecting work zone crash severity is important; however, factors affecting crash frequency also need to be studied by considering individual work zone location. Since this study seeks to assess the impacts of ABC implementation to enhance work zone safety, in order to identify contributing factors for crash occurrences under work zone conditions, a number of 60 bridge locations were selected in Miami-Dade County. Crash frequency models were then developed through a Negative Binomial regression technique and Support Vector Regression (SVR).

Taking the above-mentioned into consideration, the primary objectives of this project are three-fold:

1. Provide descriptive statistics analysis of work zone crash characteristics.
2. Model and analyze crash severity and frequency characteristics associated with construction work zones.

3. Assess the costs related to crash occurrence and crash severity due to work zone presence for both the traveling public and construction crew, together with a benefit-cost analysis to investigate the benefits of implementing ABC.

1.3. Project Outline

In this project, work zone crashes are investigated through the modeling of crash severity and crash frequency using the aforementioned data, statistical models and data mining techniques.

This chapter presents a general framework, research background and motivation, research objectives, and the project's organization.

Chapter 2 includes a comprehensive review of previous research on modeling crash severity and frequency with a focus on work zone safety. The literature review includes reviewing the crash risk prediction models in terms of statistical and machine learning methods.

Chapter 3 presents a detailed descriptive analysis of work zone crash severity from a large sample of work zone crashes in Florida. A mixed logit modeling framework was employed to determine the statistically significant crash severity contributing factors. In addition, likelihood ratio tests were conducted to examine the overall stability of model estimates across time periods. In order to explore the nonlinear relationship between crash severity outcomes by time of day and for prediction performance comparison purposes, a Support Vector Machine (SVM) model was also employed. Results for the binary level severity modeling were provided in terms of different binary and categorical variables. Finally, using results from the SVM variable impact analysis, a heat map was created on a typical work zone layout to visualize the critical locations of work zone configuration for worker safety.

Chapter 4 provides a descriptive analysis of work zone crashes that occurred in the bridge construction locations. Crash frequency in bridge locations were modeled using a negative binomial regression model and support vector regression technique. Finally, the impact of potential contributing factors on crash occurrence were identified and investigated using the models' results.

Chapter 5 evaluates the costs associated with crash occurrence and crash severity in work zone locations. Then, the benefits of ABC implementation through a monetary assessment of the potential avoided crashes to support the decision-making process of highway construction projects. Moreover, the ABC method will be evaluated if the benefits outweigh the costs in the project's life-cycle.

Chapter 6 concludes the major findings of the project from each section and provides recommendation for future research.

CHAPTER 2
LITERATURE REVIEW

In the current section, numerous prior studies from different points of view have been reviewed. In this chapter, a detailed literature review related to crash severity and crash frequency was conducted and organized in statistical methods, machine learning methods, and the corresponding work zone related studies.

2.1. Modeling Crash Severity

2.1.1. Statistical Methods

Statistical models are the primary method used in traffic crash severity analysis, and regression models are the most common techniques used to identify the relationship between dependent and independent variables. Previous works have shown that either modeling approach, such as binary logit and binary probit models (Haleem and Abdel-Aty 2010, Ahangari *et al.* 2018, Mokhtarimousavi *et al.* 2019), ordered response models (Ye and Lord 2014, Ghasemzadeh *et al.* 2018, Haghighi *et al.* 2018, Mokhtarimousavi *et al.* 2020a), multinomial logistic regression model (Islam and Mannering 2006, Ye and Lord 2014, Mokhtarimousavi 2019), random parameter models (Anderson and Hernandez 2017, Mokhtarimousavi *et al.* 2021a) , or more advanced models such as latent class models (Behnood *et al.* 2014, Behnood and Mannering 2016), and generalized models such as (Eluru *et al.* 2008, Osman *et al.* 2018), can adequately predict injury severity.

Although the application of different statistical modeling approaches has been documented in previous studies, the non-mixed models have an inherent shortcoming in that they do not account for unobserved heterogeneity commonly present in the crash data

(Chen and Tarko 2014, Behnood and Mannering 2016, Anderson and Hernandez 2017, Mamdoohi *et al.* 2018, Sharifi *et al.* 2020). An extensive review of severity analysis was conducted in a recent review paper by Mannering and Bhat (Mannering and Bhat 2014). This publication contains additional details of the methodological frontier in the crash severity analysis.

These discrete outcome models have evolved to consider several severity contributors to severity levels of the crashes as dependent variables. These severity levels have been aggregated for crash level, driver level, occupant level, or vehicle level of severity. For example, after applying the Ordered Probit (OP) model on a 5-year crash dataset extracted for Washington State, contributing factors affecting injury severity at work zone crashes under adverse weather conditions were investigated by Ghasemzade and Ahmed (Ghasemzadeh *et al.* 2018). They found that variables such as weather, lighting conditions, rural principal arterials roadway type, driving under the influence (DUI), and traveling during peak hours are among the most important factors influencing the severity of crashes at work zones. In another study by Osman et al. (Osman *et al.* 2018), a Mixed Generalized Ordered Response Probit (MGORP) modeling framework was used to investigate the impacts of contributing factors in different work zone configurations on passenger-car crash injury severity. Lane closure, lanes shift, crossover, work on shoulder or median, and intermittent/mobile work zones are considered for this study. Crashes during weekends, partial control of access, roadways classified as rural, crashes during evening times, and curved roadways were found as the factors that increase the likelihood of severe outcomes for the occupants of passenger cars across all work zones.

Ozturk et. al in (Ozturk *et al.* 2015) compared the crash severity in work zone crashes to non-work zone crashes for crash records between 2004 and 2010 in the state of New Jersey. Utilizing binary logit models, they found that some crash types such as overturn and angle have higher impacts on work zone crash severity than non-work zone conditions. Also, their results revealed that DUI has a higher impact on work zone crash severity. In addition, lightweight vehicles are more prone to be involved in more severe crashes in work zones than non-work zone locations. Utilizing five years of crash data from 2013 and 2017 in Miami-Dade County, Mokhtarimousavi et al. (Mokhtarimousavi *et al.* 2019), employed a mixed binary logit to investigate work zone crash severity. Four variables, including work zone type lane closures, crashes that took place between 4:00 p.m. and 8:00 p.m., clear weather condition, and alcohol-related crashes, were found to be statistically significant with a heterogeneous impact on crash severity. They found that crashes that occurred under conditions of lane closure, afternoon peak, clear weather condition, and alcohol consumption were all less likely to result in an injury crash. Work zones may sometimes negatively affect the transit services, for instance, the transit signal priority which operates based on certain rules and guidelines (Ali *et al.* 2017, Ali *et al.* 2018).

Three separate logistic regression models were developed in (Weng and Meng 2011) to study driver casualty risk in the construction, maintenance and utility work zones on public roads within the 51 U.S. states between 2001 and 2006. They found that five risk factors are associated with increased driver casualty risk for all three work zone types. Road alignment, truck involvement, most harmful event, vehicle age and notification time were variables that increased the risk of being in more severe crashes, while traffic control devices and restraint use were associated with reduced driver casualty risk. Li and Bai in (Li and Bai 2008) used a logistic regression technique to identify the significant risk factors for work zone crash severity in Kansas highway work zones. They found that risk factors such as poor light condition, truck involvement, having only two travel lanes, and high-speed limit are associated with high risk levels in work zone crashes.

Although some variables were found to have the same impacts on crash severity in work zone locations (either increasing the risk of being in more severe crashes or decreasing the risk) in previous studies, mixed impacts were found as well. For instance, the impact of the number of lanes in (Weng and Meng 2011) versus the findings of (Li and Bai 2008), or driving under the influence (i.e., alcohol/drug) in (Qi *et al.* 2005) versus the findings of (Mokhtarimousavi *et al.* 2019), and so forth.

2.1.2. Machine Learning Methods

Machine Learning techniques (MLs) have recently been widely applied in transportation studies (Tabibi *et al.* 2016, Mahmoudzadeh *et al.* 2019, Nezafat *et al.* 2019, Parsa *et al.* 2019a, Mahmoudzadeh and Wang 2020, Parsa *et al.* 2020, Taghipour *et al.* 2020), including traffic injury severity analysis (Li *et al.* 2012, Yu and Abdel-Aty 2013, Chen *et al.* 2016, Alkheder *et al.* 2017, Mokhtarimousavi 2019, Mokhtarimousavi *et al.* 2019, Mousavi *et al.* 2019b, Mokhtarimousavi *et al.* 2020a).

Even though conventional statistical models have been widely applied for crash injury severity analysis, the results of these statistical models may be biased for their two major limitations: pre-assumption of data distribution, and consideration of a linear form of utility functions (Li *et al.* 2012, Zeng and Huang 2014). Compared to statistical methods that provide good indications of the likelihood, MLs have been frequently applied to provide more accurate prediction models due to their ability to deal with more complex functions. Different methods have been employed to solve classification problems such as studying injury severity in safety analysis, including Support Vector Machine (SVM) (Li *et al.* 2012, Chen *et al.* 2016, Mokhtarimousavi *et al.* 2019, Kitali *et al.* 2020), Artificial Neural Network (ANN) (Delen *et al.* 2006, Rezaie Moghaddam *et al.* 2011, Zeng and Huang 2014, Alkheder *et al.* 2017), K-Nearest Neighbor (KNN) (Beshah and Hill 2010, Iranitalab and Khattak 2017), and Classification and Regression Trees (CART) (Kashani and Mohaymany 2011, Chang and Chien 2013, Ghasemzadeh and Ahmed 2017).

In recent studies that compared the prediction performance of MLs to conventional statistical models, it was demonstrated that MLs provide either superior or comparable prediction performance results (Li *et al.* 2012, Zeng and Huang 2014, Alkheder *et al.* 2017, Iranitalab and Khattak 2017, Mokhtarimousavi 2019, Mokhtarimousavi *et al.* 2019, Mokhtarimousavi *et al.* 2020a). Previous model comparisons have been performed with a number of statistical models such as Ordered Probit (OP) (Li *et al.* 2012, Alkheder *et al.* 2017, Ghasemzadeh and Ahmed 2017), Ordered Logit (OL) (Zeng and Huang 2014), Multinomial Logit (MNL) (Iranitalab and Khattak 2017, Mokhtarimousavi 2019), and Binary Mixed Logit (BMXL) (Mokhtarimousavi *et al.* 2019).

Crash severity was investigated by Li et al., in (Li *et al.* 2012) through the application of SVM on individual crash data collected at 326 freeway diverge areas. While it was shown that the SVM provides more accurate predictions of crash severity outcomes compared to the OP model, a sensitivity analysis was performed to extract the impacts of external factors. Although the results of the SVM model was consistent with the OP model for several variables, SVM produced more reasonable results for two of the contributors, including exit ramp and shoulder width of the freeway mainline. In another study and based on two-year crash data gathered in New Mexico, SVM was utilized to investigate driver injury severity patterns in rollover crashes (Chen *et al.* 2016). In this study, the significant variables were first identified through an application of CART model, and then, after incorporating the significant variables in SVM, the model prediction performance was evaluated. It was shown that while the cubic SVM classifier outperforms the medium Gaussian RBF SVM classifier, aggregating a multi-categorical response variable into a binary response variable will also improve the model prediction performance. In addition, it was found that alcohol or drug consumption is the most significant cause of drivers being involved in more severe rollover crashes, while using a seatbelt is the most effective way to protect drivers.

Applying an ANN model trained by the Whale Optimization Algorithm (WOA) by Mokhtarimousavi et al. in (Mokhtarimousavi *et al.* 2020a), injury severity in vehicle-pedestrian crashes was explored by day-of-week. The purpose of the research was to investigate the contributing factors through statistical and WOA-ANN models, where the statistical models deal with likelihood estimation and the relative probability is calculated through the ANN models. Feed-forward and cascade forward backpropagation training algorithms were applied to train the base ANN model, and a Cross-Validation (CV)-based experimental design was used to simultaneously obtain the best training algorithm with the corresponding optimal number of hidden layers. Moreover, they statistically tested if weekday and weekend crashes should be modeled separately. They found that there is a substantial and statistically significant difference between the estimated parameters across days of the week. In addition, their results revealed the internal probability patterns of crash variables for weekday and weekend models, as well as their overall impacts.

Chang and Chien in (Chang and Chien 2013) proposed a CART model to study driver injury severity in truck-involved accidents. After analysis of twenty-one predictor variables, they found that variables, including drinking-driving, seatbelt use, vehicle type, collision type, etc., are among the key determinants of injury severity outcomes for truck accidents. In another study, Wei et al. (Wei *et al.* 2017) adapted a CART model to study the severity of work zone crashes under different lighting conditions. They studied Tennessee work zone crashes during 2003–2015 under daylight, dark-lighted, and dark-not-lighted conditions. They found that the higher the number of lanes in the daylight condition resulted in a higher number of severe crashes, while a lower number of severe crashes occurred at night. The found the same results for driving under influence of drugs and alcohol.

In the study done by Weng et al. (Weng and Meng 2011), a tree-based logistic regression approach was adopted to assess a work zone casualty risk on highway work zone crashes between 2004 and 2008 in Michigan. The results demonstrated that the proposed approach provided more accurate prediction results and outperformed the pure decision tree model. Another hybrid machine learning and statistical model called the probit-decision tree model was recently applied in (Ghasemzadeh and Ahmed 2017) to explore the weather-related work zone crash severity. Using a 5-year period (2010-2014) of Highway Safety Information System (HSIS) data, work zone crashes were extracted from nine states, including California, Illinois, Maine, Michigan, Minnesota, North Carolina, Ohio, Utah and Washington. The results revealed that variables indicated a curve at the crash location, number of motor vehicles involved, presence of traffic control devices such as stop, signal, yield signs, land use (crash occurred in urban area), crash type, work zone activity (if the work zone has been active or not when the crash occurred) and lighting condition were among the most important contributing factors on work zone weather-related crash severity.

2.2. Modeling Crash Frequency

2.2.1. Statistical Methods

Crash frequency, which is the number of crashes occurring in a specific location during a specific time period of interest, has been investigated by different statistical models over the years. These models include Poisson regression (Miaou 1993, Qi *et al.* 2013, Ye *et al.* 2018), negative binomial (Khattak *et al.* 2002, Qi *et al.* 2013, Alluri *et al.* 2017, Mousavi *et al.* 2019a, Ulak *et al.* 2020), Poisson-lognormal (Park and Lord 2007, Ma *et al.* 2008, El-Basyouny and Sayed 2009b), zero-inflated (Lord *et al.* 2005, Dong *et al.* 2014, Raihan *et al.* 2019), random-effects (Aguero-Valverde 2013, Ma *et al.* 2017), random parameters models (Anastasopoulos and Mannering 2009, El-Basyouny and Sayed 2009a), Gaussian mixture model (Mansourkhaki *et al.* 2017), and so forth. An extensive review of methodological alternatives for crash frequency analysis can be found in a review paper by Lord and Mannering (Lord and Mannering 2010). Please see this publication for more details on existing models for analyzing crash-frequency data, including the advantages and disadvantages of each approach and the methodological frontier in crash frequency analysis.

In the context of work zone safety, the crash counts in a work zone location at a specific period of time is denoted as the work zone crash frequency. Among all of the modeling approaches for analyzing cash frequency, the negative binomial is the most frequently applied model. Khattak et al. in (Khattak *et al.* 2002) studied the effect of work zone duration on crash frequency. Using California freeway work zones data and applying negative binomial models, the impact of work zone duration for both injury and non-injury crashes in the pre-work zone and during-work zone periods was analyzed. They found that frequencies increased with increasing work zone duration, length, and average daily traffic.

A truncated Poisson model and negative binomial models were used in (Qi *et al.* 2013) to identify the factors that influence the frequency of rear-end crashes in work zones locations. For this analysis, 6,095 work zone crashes occurred in New York State from 1994 to 2001, including 2,481 rear-end crashes. Their analysis demonstrated that work zones controlled by flaggers and work zones with alternating one-way traffic are more prone to have more rear-end crashes compared to those controlled by arrow boards. In addition, among all work zone types, work zones for capacity and pavement improvements are associated with more rear-end crashes. Ozturk et al. in (Ozturk *et al.* 2013) developed negative binomial regression models for daytime and nighttime conditions using 120 construction work zone crash records from 60 work zone locations in New Jersey between 2004 and 2010. They found that project duration, work zone length, and traffic volume were among the most important factors that increase work zone crash frequency.

In a study done by Chen and Tarko (Chen and Tarko 2014), using work zone data obtained from the project engineer's survey with the Indiana road inventory, various work zone design and traffic management features such as lane shift, lane split, and detour were studied. As a result, a random parameter negative binomial model was developed, and the results were further validated with a fixed parameters negative binomial model with random effect. From the methodological points of view, they stated that the convenient fixed parameters negative binomial models may be more practical than the random parameters models. In another study with a focus on law enforcement, a random-effect negative binomial model was developed in (Chen and Tarko 2012). A number of 72 work zones on state-maintained freeway and non-freeway roads in the state of Indiana between 2008 and 2010 were studied. Some temporal variations in the risk of crash frequency were found in the results. For example, it was shown that the crash frequency was 24% higher between November and December, and it was also 20% higher in the months of May, June, and July. Also, they showed that there was a 41.5% reduction in the work zone crash frequency when police enforcement was involved.

2.2.2. Machine Learning Methods

Although statistical models have been utilized to analyze vehicle crash frequency for many years, Machine Learning techniques (MLs) have recently received attention among traffic safety professionals. Considering the same limitations mentioned earlier in this chapter regarding conventional statistical models, which may lead to erroneous estimations of crash frequency likelihood, MLs do not require any pre- assumption of an underlying relationship between the response variable and predictors (i.e., independent variables). A number of Crash Frequency (CF) models were developed for prediction crash frequency, and among them, CART (Chang and Chen 2005), SVM (Li *et al.* 2008, Qu *et al.* 2013, Dong *et al.* 2015) and ANN (Jafari *et al.* 2015, Huang *et al.* 2016, Zeng *et al.* 2016, Pan *et al.* 2017) are the most applied models in the literature.

Chang and Chen (Chang and Chen 2005) used a CART model to analyze two-year crash data for National Freeway 1 in Taiwan. They treated the crash frequency (i.e., non-negative integers) as a classification problem by classifying them into 0 to 3 and 4 or more observed accidents and found that daily traffic volume and precipitation are two key variables for freeway accident frequencies. They compared the prediction performance of the proposed CART model with a negative binomial regression model and concluded that CART is a promising tool to analyze freeway accident frequencies.

SVM and negative binomial regression models were applied in (Li *et al.* 2008) to predict 122 motor vehicle crashes that occurred during a 5-year period in Texas. They tested the developed models on different fitting set sizes and compared their prediction performance through the Mean Absolute Deviation (MAD) and Mean Squared Predictor Error (MSPE) evaluation criteria. They demonstrated that SVM models predict crash data more effectively and accurately than traditional NB models. They further compared the results with a Back-Propagation Neural Network (BPNN) documented in previous studies and found that SVM provides compatible prediction results. In addition, a sensitivity analyses on the variables ADT and right-shoulder width were performed. It was shown that while right-shoulder width has a quadratic functional form, an increase in right-shoulder width results in a decrease in crash frequency. Contradicting results were found for ADT.

In the study done by Que et al. (Qu *et al.* 2013), the application of SVM to predict real-time freeway sideswipe crashes in the Milwaukee, Wisconsin was presented. They implemented the SVM models with three different kernel functions with significant features as inputs. Then, the devolved models were compared with the multi-layer perceptron (MLP) artificial neural network. They found that while the SVM provides compatible prediction accuracy results with MLP-ANN, it can better identify sideswipe crashes at higher false alarm rates. In other study done by Dong et al. (Dong *et al.* 2015), the potential application of SVM to predict zone-level crashes was studied. In order to handle high-dimension spatial data, Correlation-based Feature Selector (CFS) was utilized to evaluate candidate contributing factors prior to fitting the models. Using the data from Hillsborough County in Florida, the results of SVM models were compared with the Bayesian spatial model. The results showed that while SVM is able to take spatial proximity into account, it showed better prediction performance compared to the Bayesian spatial model. The best model results were obtained through the SVM implemented with the RBF kernel and setting the 10% of the whole dataset as the testing data. The mean predicted probabilities were considered as a criterion with sensitivity analysis to explore the impacts of explanatory variables on crash occurrence.

Jafari et al. (Jafari *et al.* 2015) used the ANN model to predict road traffic death rates for 178 countries across the world. A Genetic Algorithm (GA) model was then applied to optimize the ANN parameters, while the model's prediction results were verified by conducting a five-fold cross-validation. Road traffic death rate was aggregated into three classes, 0-9, 10-19, and over 20 deaths per 100, 000 population, and Root Mean Square Error (RMSE) was considered as the model fitting criteria. It was shown that the proposed GANN is able to predict road traffic deaths with a satisfactory RMSE, and the advantages of using GA over gradient search techniques were highlighted. Huang et al. in (Huang *et al.* 2016) investigated the nonlinear relationships between crash frequency and the relevant risk factors on road segments in Hong Kong using a Radial Basis Function Neural Network (RBFNN) model. They optimized the RBFNN model first by a K-means cluster algorithm to determine the centers of RBFs. Then, a Recurrent Least Squares (RLS) algorithm was applied to estimate the basis and weights between the output node and RBFs. The prediction performance of the developed model was compared with the traditional NB and Back-Propagation Neural Network (BPNN) models. The results revealed that most of the contributing factors, including AADT, speed limit, and lane-changing opportunity, have nonlinear relationships with crash frequency, while length and rainfall have positive impacts on crash frequency.

Zhang et al. in (Zeng *et al.* 2016) also utilized a Neural Network (NN) model to explore the nonlinear relationship between crashes that occurred on 211 road segments in Hong Kong, along with risk factors. A network structure optimization algorithm was proposed to avoid the over-fitting issue, and a rule extraction method is proposed to manage the black-box nature of the NN model. They found that variables such as AADT and segment length have a positive impact on crash frequency, while higher speed limits resulted in a lower number of crashes. In addition, comparing the prediction results with the NB model showed that the proposed model outperforms the traditional NB model.

A Deep Belief Network (DBN) was utilized in (Pan *et al.* 2017) to predict crashes from different highways and regions in order to develop a global road Safety Performance Function (SFP). Three crash data sets from 2000 to 2008 were used and aggregated into different highway classes, number of lanes, access control, and region (Ontario (ON) province, Colorado (CO) State and Washington (WA) State), and were divided into urban and rural subgroups. Their results showed that a single DBN model can be trained globally with multiple datasets, while its prediction performance is comparable to the traditional NB model. They concluded that instead of developing several local models separately through the traditional statistical models like NB, applying the proposed model will significantly reduce the modelling works. More importantly, it was shown that DBN has the flexibility to make use of new crash data, which will be a very tedious process if NB is employed.

CHAPTER 3
CRASH SEVERITY MODELING OF WORK ZONE CRASHES

3.1. Introduction

In recent years, work zones have been a high-priority issue in traffic safety analysis, which has gained increased attention among transportation safety analysts and decision makers (Adomah *et al.* 2021). In addition, advances in intelligent transportation system technologies have shown promising benefits to address work zone safety concerns. For instance, the safety benefit of Reduced Speed Work Zone Warning (RSZW) that increase the safety of drivers and construction workers in Work Zones have been recently investigated by (Arafat *et al.* 2020) and (Hadi *et al.* 2019a). However, up to date these applications are still in the early stages of deployment.

The environmental and geometric characteristics of work zones make them prone to crash occurrence or increasing crash severity (Garber and Zhao 2002, Khattak *et al.* 2002). According to statistics from the American Road and Transportation Builders Association (ARTBA), there were 710 crashes that resulted in 799 fatalities in work zone locations in U.S. roadway networks in 2017. Florida is among the top three states for the highest number of work zone crashes, with an average number of 76 fatalities resulting from 71 fatal crashes in 2017 (ARTBA 2018). Worker safety is another important aspect of work zone crashes, which has been rarely discussed within work zone crash severity literature (Yang *et al.* 2015). Hadi et al. in (Hadi *et al.* 2019b) utilized the Florida ITS Evaluation tool (FITSEVAL) to assess the mobility, safety, environmental, and user-cost benefits of Smart Work Zone applications. The study results showed that the Smart Work Zone systems are easy to use and seem to provide reasonable safety benefits.

Worker fatalities in 2017 show a 33% increase compared to 2016 (16 deaths in 2017 vs. 12 in 2016). The significant loss of workers' lives and injuries resulting from work zone crashes must be properly addressed so that a comprehensive and in-depth investigation of work zone crash mechanisms can be conducted.

Despite the recent efforts to investigate crash severity, worker presence and its impact on injury severity in work zone crashes is still unexplored. A better understanding of work zone crash characteristics can enhance roadway safety, not only for road users, but also, for construction crews.

From a logistics perspective, work zone activities can occur during nighttime hours to reduce the adverse impacts on traffic operations and complaints by the traveling public (Srinivasan *et al.* 2011, Nafis *et al.* 2019). However, this requires further attention to worker safety due to the more hazardous work conditions at night. Although the number of work zone crashes that occurred during the daytime, involved workers, and resulted in an injury in Florida in 2016 were higher than the ones that occurred during the nighttime (76.32% vs. 23.68% respectively), they shared the same number of fatalities, of which 34 people were killed in total (S4A 2018). The lower traffic volumes during the nighttime hours increases driver maneuverability and yields higher operating speeds, which increases safety risks for the construction crew. The visibility of drivers and workers at night is another issue that can affect the relative daytime and nighttime work zone crash risk and severity (Arditi *et al.* 2007, Li and Bai 2009, Srinivasan *et al.* 2011).

With that in mind, and considering that most of the existing safety research has focused on the traveling public and not on worker safety, this chapter seeks to identify the contributing factors that affect the severity of work zone crashes associated with worker presence by time-of-day. To the best of the authors' knowledge, this is a very first attempt to analyze the severity of work zone crashes associated with worker presence by time-of-day through discrete choice and supervised machine learning models.

In this chapter, crash severity outcomes of work zone crashes involving workers is investigated by time-of-day. Preliminary insight into potential significant variables was obtained through the application of a Random Forest (RF) analysis by ranking candidate variables according to their relevant importance. A mixed logit modeling framework was then applied to determine statistically significant crash severity contributing factors. In addition, likelihood ratio tests were conducted to examine the overall stability of the model's estimates across time periods. In order to explore the nonlinear relationship between crash severity outcomes by time-of-day, as well as to compare the effects to that of the logit model and to assess prediction performance, a Support Vector Machine (SVM) model was also employed. A Cuckoo Search (CS) metaheuristic algorithm was then utilized to tune SVM parameters with the goal of enhancing the prediction performance and, as a result, improve inference on variable effects. Variable impact analysis was also performed by taking into account the black-box characteristic of the SVM and comparing it to the effects of variables indented through the logit modeling framework.

As for the analysis of this chapter, crash records are obtained from the Florida Signal Four Analytics tool (S4A 2018), which is a statewide interactive, web-based geospatial crash analytical tool.

3.2. Descriptive Statistics for Crash Severity

A three-year period of statewide crash data was collected from January 1, 2015 to December 31, 2017. Crashes that occurred in work zone areas with worker presence were then extracted from the crash records. The dataset contained a total of 2,113,678 crash records, with 1.55% of the crashes occurring in work zones (i.e., 32,669 occurred in work zones). Out of the total number of work zone crashes, 44.50% were associated with worker presence. The crash severity levels are classified into three levels as: no injury or property damage only (PDO), injury which includes possible injury, non-incapacitating and incapacitating injuries, and fatality. The purpose of conducting a descriptive analysis of work zone-related crash severity is to provide an initial view of data distribution within work zone crash severity levels.

Figure 3-1 demonstrates the distribution of work zone crashes involving workers per year by severity levels.

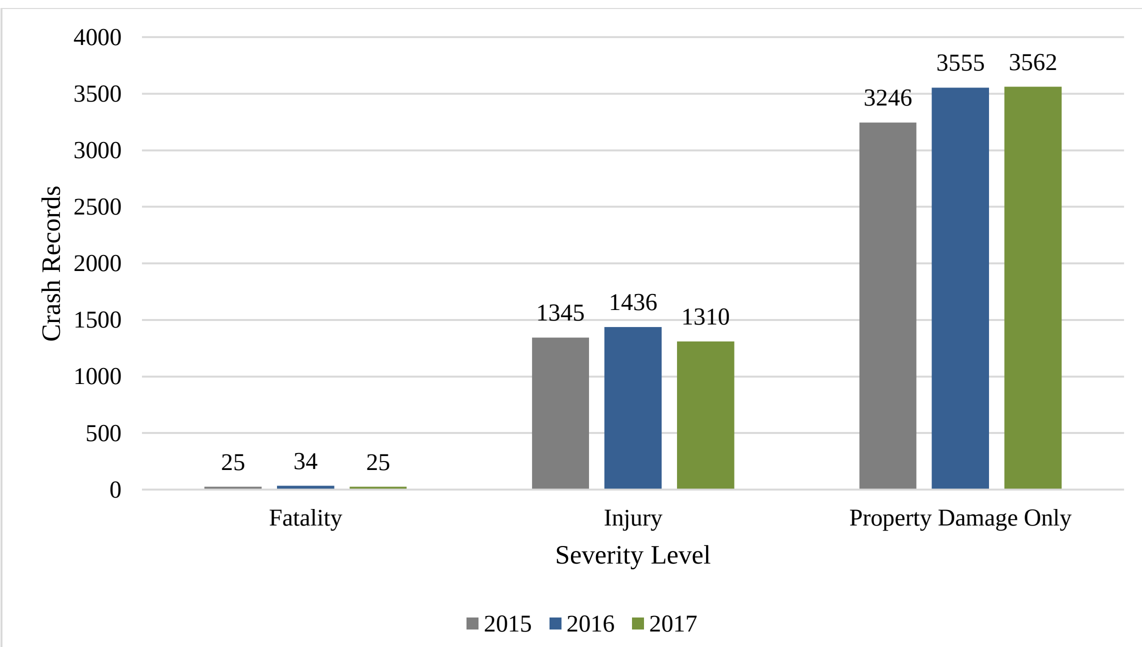


Figure 3-1 FL WZ Crash Severity Statistics (2015-2017)

In terms of the total number of crashes, 71.3% of the work zone crashes were PDO, 28.1% of the work zone crashes were injury, and 0.6% of the work zone crashes were fatal crashes. As shown in Figure 3-1, an unusual increment was observed for the number of work zone crashes in 2016 for each category. However, as illustrated in Figure 3-2, the general trends of fatal and injury crashes, work zone crashes and total crashes has followed an increasing trend in last nine years in the state of Florida.

Although the low percentages of work zone crashes may not seem alarming at first glance, according to the statistics, during 2015 and 2017, the average injury and fatality at work zone crashes is 5% higher than the amount of fatalities in road crashes not involving work zones (28.68% and 23.68%, respectively).

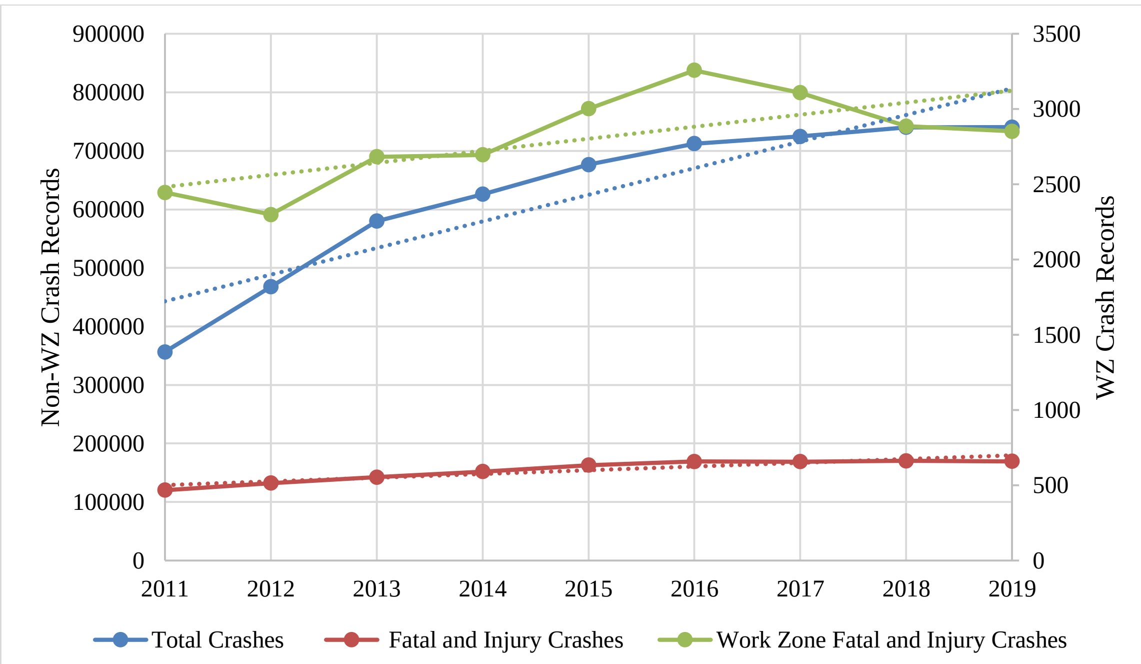


Figure 3-2 FL Crash Severity Statistics (2011-2019)

3.2.1. Severity Distribution by Time of Occurrence

Figure 3-3 illustrates the relationship between work zone crash severity and time of crash occurrence in the format of a pie chart. The time of day was divided into four categories, namely morning peak (6:00–10:00 a.m.), daytime non-peak (10:00 a.m.–4:00 p.m.), afternoon peak (4:00–8:00 p.m.), and nighttime (8:00 p.m.–6:00 a.m.). As shown in Figure 3-3, crashes that occurred during nighttime period were the most severe crashes with the highest rates of injury and fatality compared to other periods. In contrast, statistical results show that daytime non-peak was the safest time period in regard to injury and fatal crashes. In addition, most frequent work zone crashes involving workers occurred during the daytime off-peak periods. This makes sense, as it matches the construction time schedule.

A chi-square test was also performed to evaluate the association between time of day variables and crash severity outcomes. With a chi-square value of χ^2 = 31.386 and degree of freedom (df) =6, the test results show that work zone crash severity with a 95% level of confidence is significantly associated with time of crash occurrence.

Figure 3-3 Severity Distributions by Time of the Day

3.2.2. Severity Distribution by Crash Attributes

Work zone crashes, which is one of the most important attributes of crashes, has been aggregated by crash type to investigate injury severity in previous studies such as rear-end crashes in (Khattak 2001). This crash type has been recorded as 13 different types in the S4 crash database, namely angle, animal-related, bicycle, head-on, left turn, off-road, pedestrian, rear-end, right turn, rollover, sideswipe, and other and unknown crashes. The severity distribution of work zone crashes by work zone type is illustrated in Figure 3-4. As shown in this figure, rear-end crashes are more likely to be severe based on severity proportions. They consist of 26.2% and 55.2% of fatal and injury crashes, respectively. In addition, Pearson chi-square test results (χ^2= 1240.00 and df=24) demonstrate a significant correlation between crash type and crash severity outcomes.

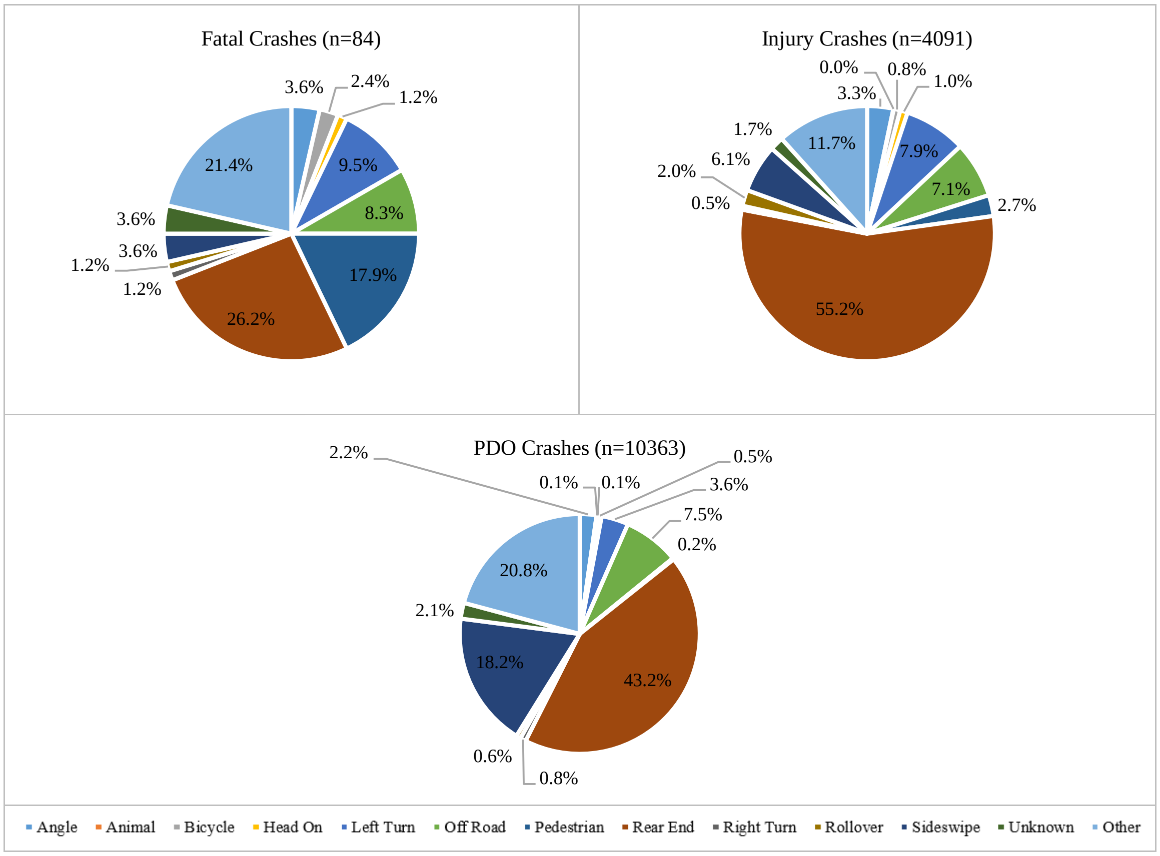


Figure 3-4 Severity Distributions by Crash Type

The crash type patterns during 24 hours shown in Figure 3-5 demonstrates that rear-end crashes were the most frequent crash types occurring more frequently during two time periods, between 6:00 a.m. to 5:00 p.m., and 8:00 p.m. to 12:00 a.m.. This emphasizes the importance of investigating nighttime work zone crashes involving workers for the sake of construction worker safety. It also indicates that although the number of rear-end crashes are high during daytime and nighttime time periods, the corresponding contributing factors may share different impacts as environmental and traffic conditions are different during day and night, visibility and traveling speed for instance.

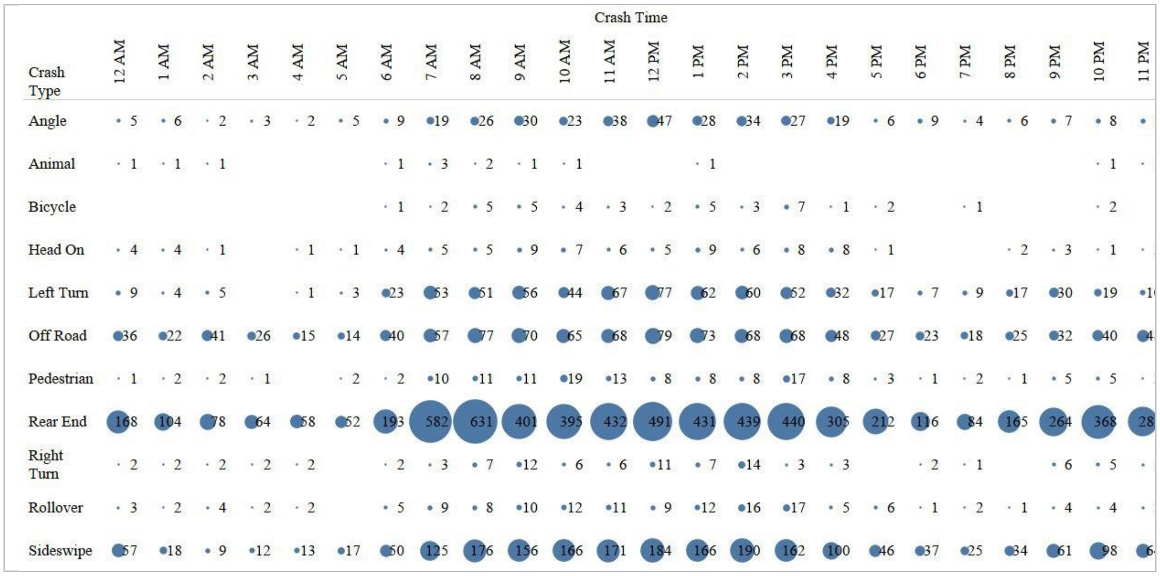


Figure 3-5 Work Zone Crash Type Pattern Over 24 Hours

The number of vehicles involved in a crash is another important consideration of crash attributes. The higher the number of vehicles are involved in a crash, in the higher the direct and indirect costs of the crash for both crash partners and the traveling public. In the present study, the number of vehicles involved in the crash was considered to be single vehicle and multi-vehicle involved (i.e., more than one vehicle involved in the crash event). The crash severity distributions by number of vehicles are shown in Figure 3-6.

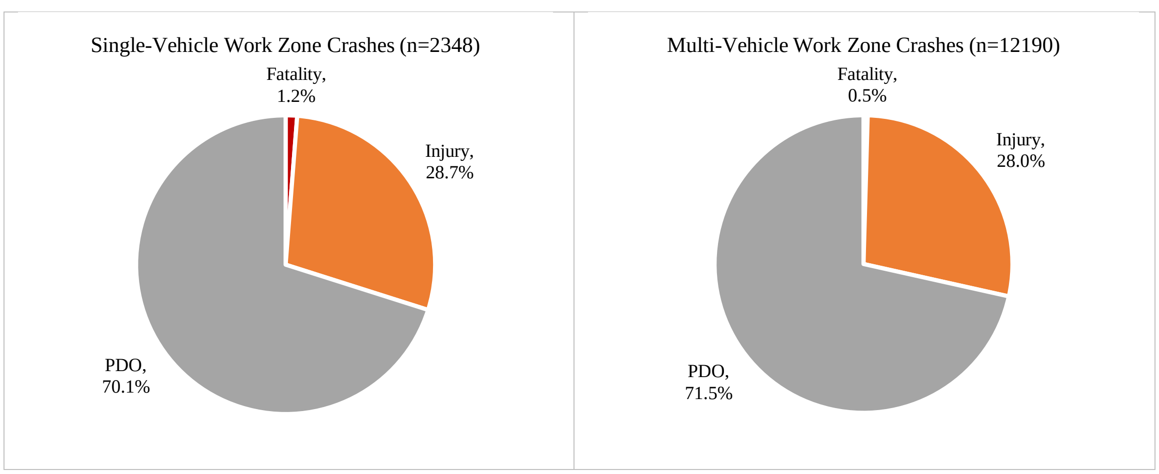
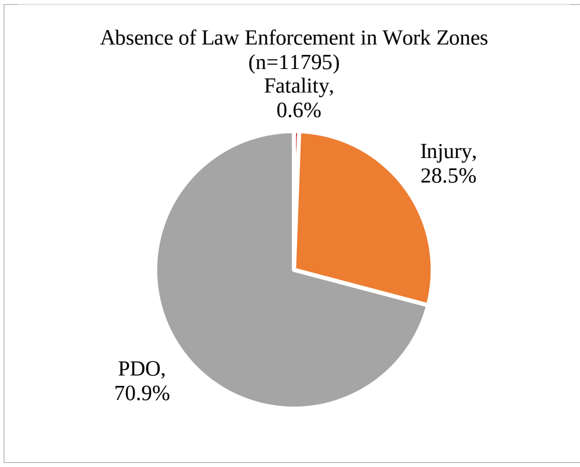


Figure 3-6 Severity Distributions by Vehicles Involved

Based on the proportions observed, single vehicle crashes were more severe than multi-vehicle crashes. Although they shared almost the same percentage of injury crashes, the percentage of single-vehicle fatal crashes is over two times as much as multi-vehicle crashes. The Pearson's chi-squared test result showed that with a χ^2=21.75 and df=2, the crash severity and number of vehicles in crash are significantly correlated within the 95% level of significance.

The presence of law enforcement and its effectiveness in preventing drivers who are inattentive or who exhibit irresponsible behavior has been widely studied in (Kamyab *et al.* 2003). The negative impact of law enforcement's presence on crash severity was also investigated by Raub et al. in (Raub *et al.* 2001, Mokhtarimousavi *et al.* 2020b). The law enforcement factor has been recorded in a crash dataset as whether law enforcement is available or not. Figure 3-7 shows that with the presence of law enforcement at construction work zone sites, the portion of fatality and injury crashes are slightly lower than without law enforcement, 1 and 2 percent, respectively. Moreover, the results from the Pearson chi-square test (χ^2= 4.659 and df=2) showed that the *P* value (0.973) is greater than 0.05; therefore, the null hypothesis shows that the association that existed between the law enforcement presence and crash severity was rejected.



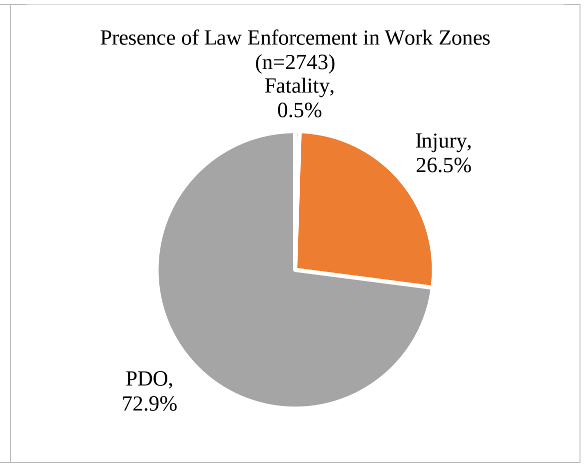


Figure 3-7 Severity Distributions by Law Enforcement

3.2.3. Severity Distribution by Environmental Conditions

In this research, variables including weather, surface and light conditions were studied in terms of crash severity as environmental conditions.

The variables that represent the weather conditions were originally categorized into eight groups; however, 99.54% of work zone crashes occurred during three weather conditions, namely clear, cloudy, and rainy, as shown in Figure 3-8.

Clear weather was the riskiest condition for work zone crashes involving workers, in which the highest number of fatality crashes occurred. As for injury crashes, the majority of crashes occurred during cloudy weather conditions (29.1%). The results of the chi-squared test with χ^2= 3.43 and df=4 resulted in the P value being equal to 0.487 (which is greater than 0.05); thus, the null hypothesis was rejected.

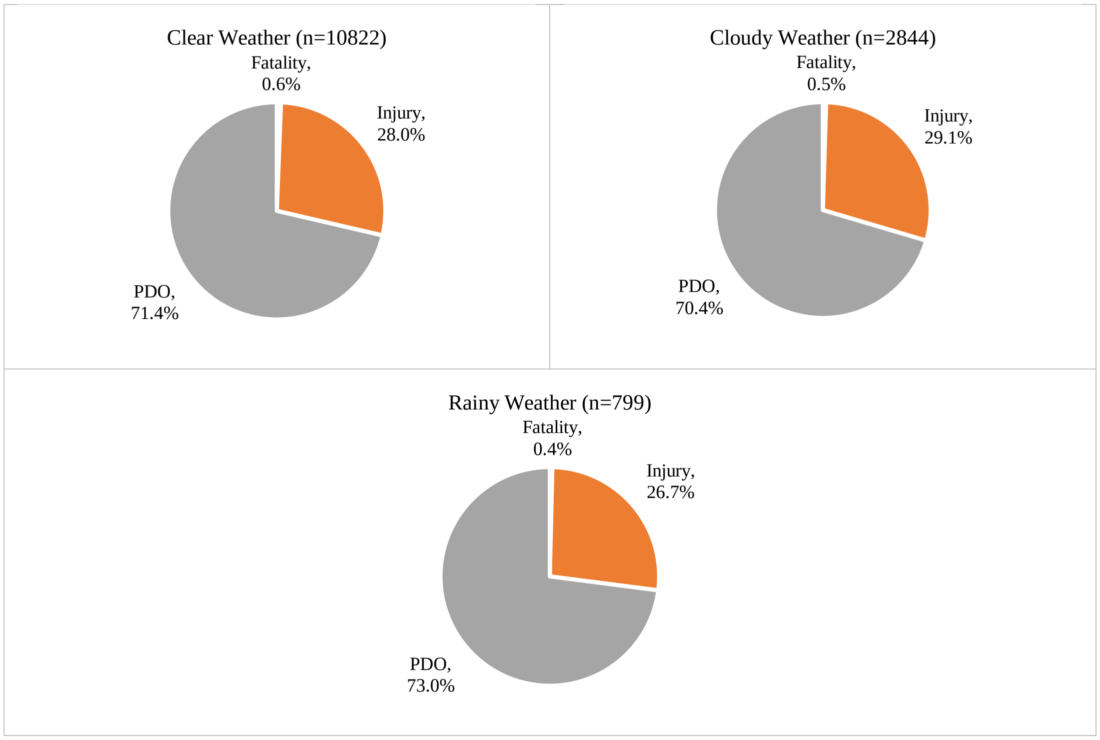


Figure 3-8 Severity Distributions by Weather Condition

Similarly, road surface condition was originally categorized into seven conditions; however, 99.08% of work zone crashes occurred during three weather conditions, namely clear, cloudy, and rainy, as shown below in Figure 3-9.

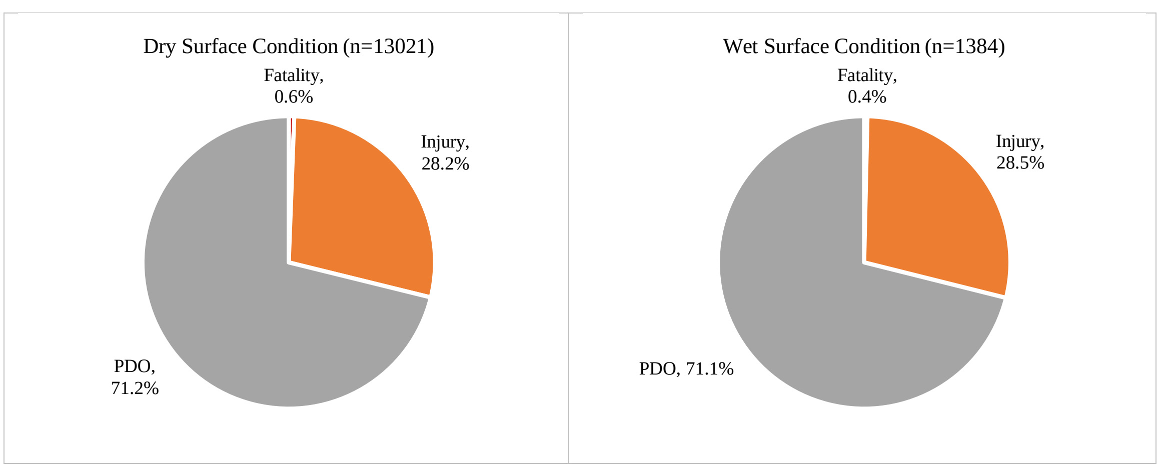


Figure 3-9 Severity Distributions by Road Surface Condition

Results have shown that a portion of fatality crashes was 50% greater on dry roadway surfaces; however, injury crashes were slightly higher on wet surfaces. Similar relationships, such as weather condition, were observed between road surface and crash severity. The results of the chi-squared test with χ^2= 1.266 and df=2 showed that the obtained *P* value (0.531) was greater than 0.05; thus, the hypothesis of having an association between road surface conditions and crash severity was rejected.

The last considered environmentally-related variable is the light condition. In order to see the effect of light on crash severity, the time of crash occurrence was divided into two time periods. Crashes that occurred between 6:00 a.m. to 7:59 p.m. were considered daytime crashes, and those that occurred between 8:00 p.m. to 5:59 a.m. were considered nighttime crashes. As presented in Figure 3-10, the work zone fatality rate in nighttime crashes was two times higher than those occurring during daytime conditions. In addition, results from the chi-squared test with χ^2=23.621, df=2, and the *P* value equal to 0.00 indicated that an association exists between the light conditions and crash severity at the 95% level of significance.

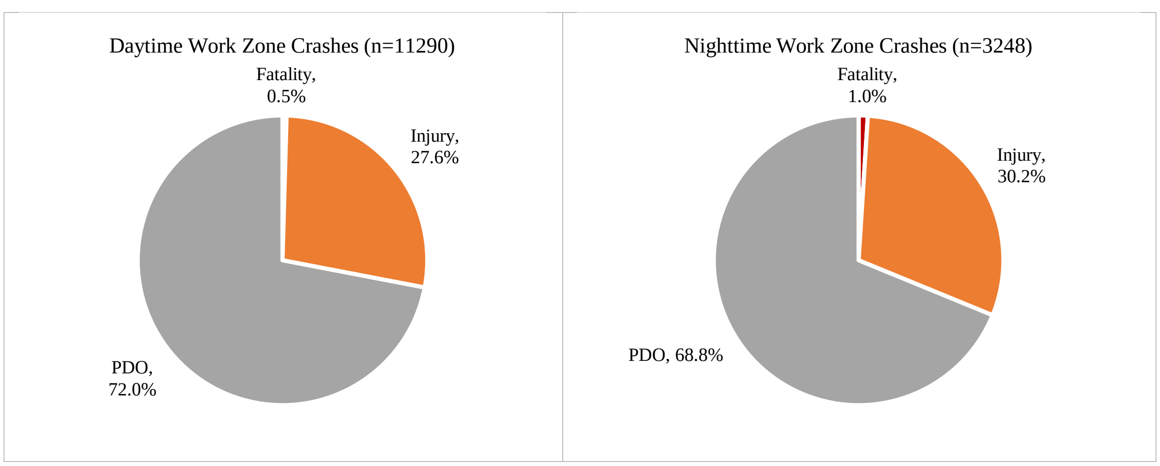


Figure 3-10 Severity Distributions by Light Condition

3.2.4. Severity Distribution by Driver Characteristics

The variables related to the driver's characteristics considered in this study include driving under the influence (DUI) of drugs and alcohol.

The work zone crash severity distribution in terms of alcohol and drug involvement is shown in Figure 3-11. As can be seen, although there is a huge difference between the number crashes in which the drivers were and were not correlated with DUI, the portion of fatality crashes were five and over six times higher for alcohol and drug involvement, respectively.

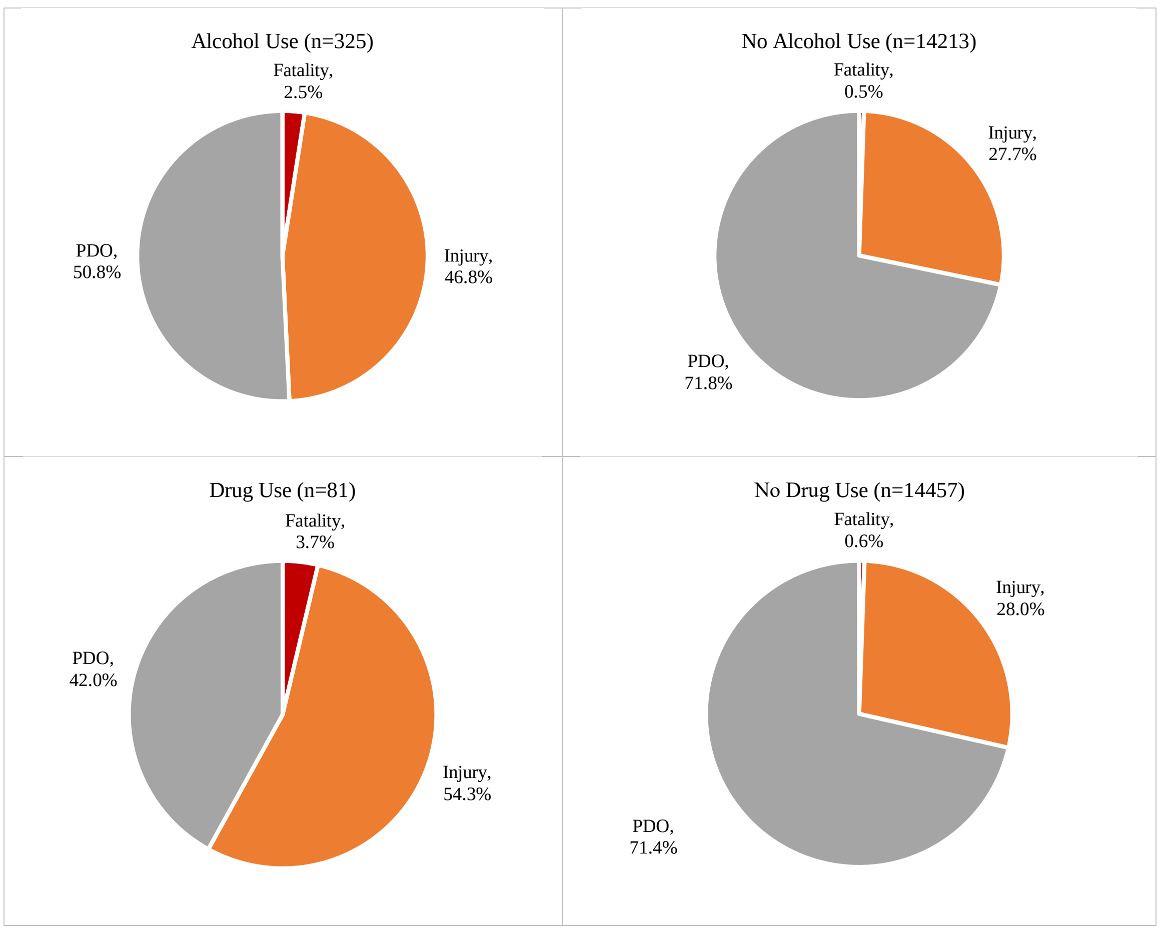


Figure 3-11 Severity Distributions by DUI

The Pearson chi-squared tests (χ^2=81.038, df=2 for alcohol use and χ^2=43.430, df=2) with both P values equal to 0.00 (less than 0.05) had results that showed that there was a significant correlation between work zone crash severity and the DUI condition.

3.2.5. Severity Distribution by Work Zone Characteristics

Work zone characteristics in this study were considered the variables that indicated where the crash was located in the work zone location and work zone type.

According to the National Work Zone Safety Information Clearinghouse, the main work zone components, in the order of entering the work zone location, include: before the first work zone warning sign, advance warning area, transition area, activity area, and termination area. Figure 3-12 demonstrates the severity distributions by crash location in work zone areas. As can be seen, the leading locations in work zones with the highest rates of fatality and injury crashes are activity area, transition area, advanced warning area, before the first work zone warning sign, and termination. The Pearson chi-squared test result (χ^2= 27.656, df=8, and P= 0.001) indicates that there is a significant association between crash location in work zone and severity for the work zones crashes.

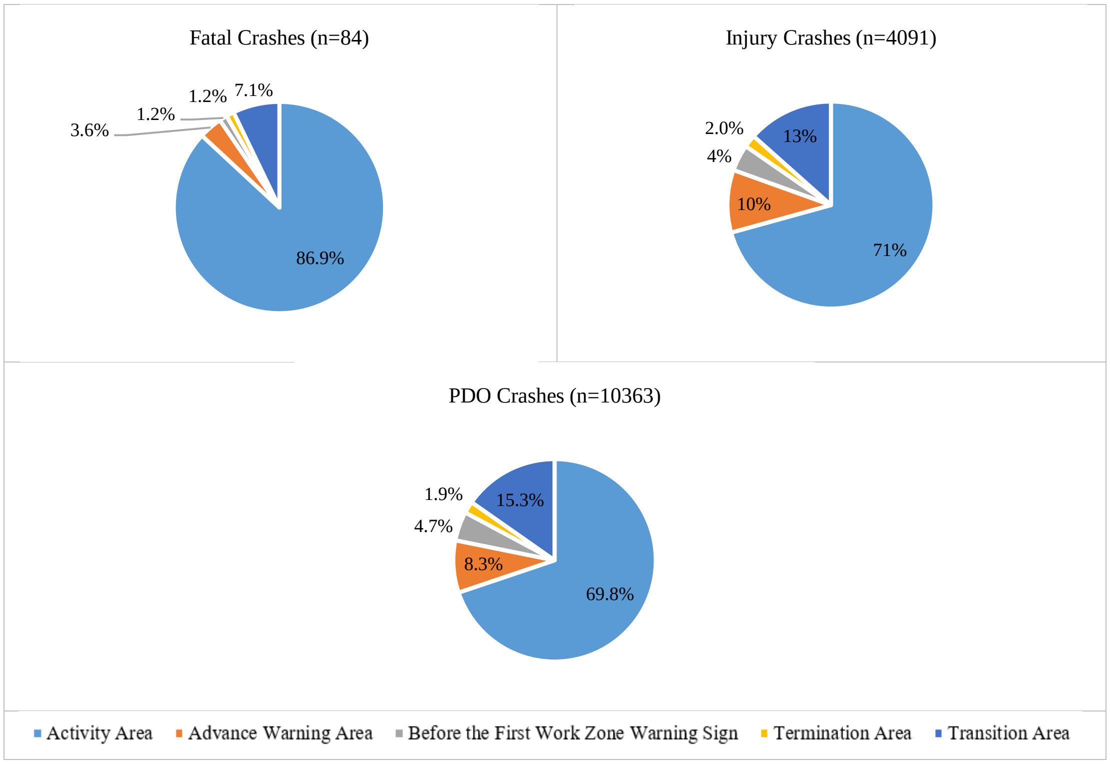


Figure 3-12 Severity Distributions by Crash Location in WZ

Work zone was recorded as five types in the crash dataset, namely intermittent or moving work zones, lane closure, lane shift/crossover, work on shoulder or median, and other types. Looking at the severity distributions by work zone types illustrated in Figure 3-13, it is clear that among the work zone types, work on shoulder or median and lane closer were the work zone types with the highest proportion of fatality and injury crashes.

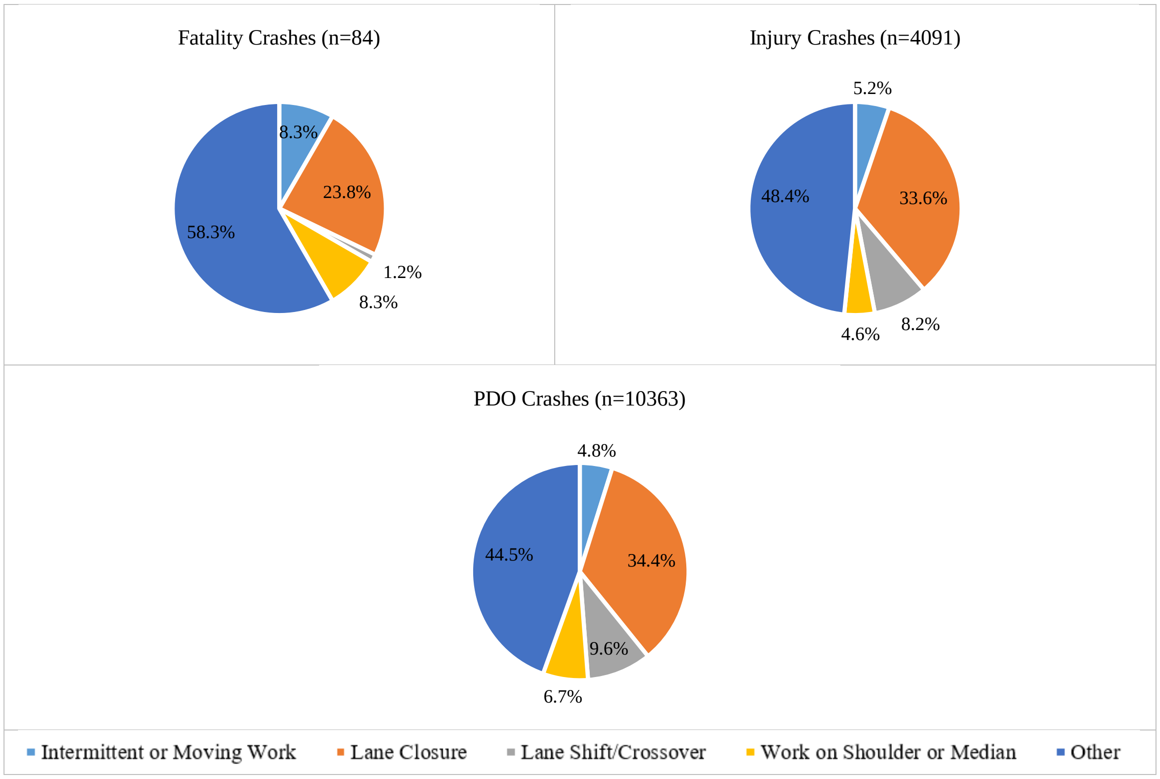


Figure 3-13 Severity Distributions by Work Zone Type

Results from the chi-squared test indicated that with a= χ^2=51.858, df=8, and the P value equal to 0.00, there is an association that exists between work zone type and crash severity at the 95% level of significance.

3.3. Methodology

This section will detail the methodologies used throughout the study of the factors that contributed to the severity of work zone crashes. These methods include logistic regression, random forest (RF), parameter transferability, SVM, and the CS algorithm used to tune the SVM parameters. The premise behind the application of the logit model and the corresponding parameter transferability test is two-fold. First, the logit model is used to determine significant factors that contribute to work zone crash severity. Second, the parameter transferability test statistically confirms whether the contributing factors to work zone crash severity are different by daytime and nighttime conditions. The identified variables in the logit model are then investigated in detail using the enhanced SVM approach.

3.3.1. Binary Mixed Logit

In studying crash severity, mixed logit models (MXLs), also called random parameters logit models, are the most popular among all econometric methods (Haleem *et al.* 2015, Behnood and Mannering 2017, Seraneeprakarn *et al.* 2017, Mokhtarimousavi *et al.* 2019). MXLs are used in safety analyses to estimate the relationship between explanatory variables and crash severity while considering the presence of unobserved heterogeneity. Allowing parameters to differ across observations, MXLs address the limitations of fixed parameter modeling approaches by accounting for heterogeneous effects and correlation with unobserved factors in crash data, which results in more reliable parameter estimates (Washington *et al.* 2010, Cerwick *et al.* 2014). For example, although it may be possible to estimate various crash characteristics and environmental characteristics based on crash data, there are several data items that influence crash occurrence and severity that are difficult to collect and are not normally available. The application of a random parameters model attempts to account for these unobservables and the resulting unobserved heterogeneity, which if not accounted for, can result in erroneous parameter estimates.

For the current work, the binary mixed logit modeling framework (BMXL) is utilized. In this model, the estimated probability is considered the integral of the standard logit probability over its corresponding parameter density (Ye and Lord 2014). The traditional binary logit model structure, in which the probability that an outcome (i.e., crash injury) takes on the value of 1, is shown in Equation (3-1):

$$P_n(i) = \frac{e^{(\boldsymbol{\beta}_i \boldsymbol{X}_{in})}}{1 + e^{(\boldsymbol{\beta}_i \boldsymbol{X}_{in})}} \tag{3-1}$$

where $P_n(i)$ is the logit probability of crash n resulting in crash severity i, $\boldsymbol{X_{in}}$ is a vector of observable characteristics (i.e., variables shown in Table 1), and $\boldsymbol{\beta_i}$ is a vector of parameters to be estimated for crash severity i.

By extending Equation (3-1) to include the estimation of random parameters (i.e., a mixed model), a model with a mixing distribution is now defined as (McFadden and Train 2000, Train 2009, Washington *et al.* 2010):

$$P_n(i \mid \Omega) = \int \frac{e^{(\boldsymbol{\beta}_i \boldsymbol{X}_{in})}}{1 + e^{(\boldsymbol{\beta}_i \boldsymbol{X}_{in})}} f(\beta \mid \Omega) d\beta \tag{3-2}$$

where $P_n(i \mid \Omega)$ is the mixed logit probability (weighted average of the MNL probabilities) with weights determined by the density function of β, $f(\beta \mid \Omega)$. The density function of β, $f(\beta \mid \Omega)$, is conditional on distributional parameter $\boldsymbol{\Omega}$, where $\boldsymbol{\Omega}$ represents a vector of parameters (mean and variance) to be estimated. The distribution of Ω is specified by the analyst, and in most cases, is specified to be normally distributed (Mannering and Bhat 2014, Mannering *et al.* 2016). In the end, by the addition of $f(\beta \mid \Omega)$, β can now account for crash-specific variations of the effects of observable characteristics X_{in} on crash severity i probabilities.

Lastly, to assess the impact of an explanatory variable on the outcome probability of crash severity i, marginal effects are computed. Considering that all variables used in the modeling procedure are indicator variables, marginal effects are calculated as (Greene 2018):

$$ME_{X_{ink}}^{P_n(i)} = Prob\left[P_n(i) = 1 \mid X_{(X_{ink})}, X_{ink} = 1\right] - Prob\left[P_n(i) = 1 \mid X_{(X_{ink})}, X_{ink} = 0\right] \quad (3\text{-}3)$$

3.3.2. Parameter Transferability

As stated previously, the next step is to determine if daytime and nighttime crashes need to be analyzed independently. To do this, a parameter transferability test is conducted. The parameter transferability test statistically tests if the estimated parameters in work zone crash severity models are significantly different between daytime and nighttime conditions. This is accomplished through a log-likelihood ratio test that follows a chi-square distribution with degrees of freedom equal to the number of estimated parameters, as computed in Equation (3-4) (Washington *et al.* 2010):

$$x^2 = -2\left[LL\left(\beta_{MX1_{MX2}}\right) - LL(\beta_{MX1})\right] \quad (3\text{-}4)$$

where $LL\left(\beta_{MX1_{MX2}}\right)$ is the log-likelihood at convergence of model MX_1 based on using the time-period data for model MX_2, and $LL(\beta_{MX1})$ is the log-likelihood at convergence of model MX_1. Suppose that the model for daytime crashes is fit using the data from nighttime crashes and vice-versa, hence, the original log-likelihood values are used to calculate the chi-square statistics. Finally, by considering the degrees of freedom (the number of estimated parameters in the model using the other model's data), the significance is determined. This log-likelihood ratio test examines the null hypothesis that daytime and nighttime crashes should be modeled together and that their contributing factors, or parameter estimates, are not statistically different. Therefore, this work seeks to determine whether or not this hypothesis is rejected.

3.3.3. Support Vector Machine (SVM)

Upon determining the significant contributing factors through the logit model and the results from the parameter transferability test, the SVM model is applied to capture crash severity patterns among all explanatory variables. SVM is a non-parametric supervised learning classification model introduced and developed by Vapnik et al. in the 1990s (Boser *et al.* 1992, Vapnik 1998). Based on the statistical learning theory and structural risk minimization, the SVM algorithm aims to find (n-1) dimensional separating hyperplanes (one hyperplane in binary classification problems), while simultaneously maximizing the distances of the nearest data points to the decision boundary (i.e., the margin). The hyperplane can be written as a set of points, $\boldsymbol{x}$**,** as illustrated in Equation (3-5):

$$y(x) = \boldsymbol{w}.\boldsymbol{x} + b = 0 \tag{3-5}$$

where hyperplane $y(x) = 0$ defines a decision boundary in the feature space, while the parameters of $\boldsymbol{w}$ (a normal vector) b (bias) are determined through the learning procedure. In order to find the optimal separating hyperplane, given a training set of explanatory variables and severity outcomes pairs (x_i, y_i), the SVM algorithm solves the quadratic optimization problem shown in Equation (3-6) (Bottou and Lin 2007):

$$\min Q(w, b, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} \xi_i \tag{3-6}$$

$$\text{Subject to,} \forall_i \; y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0,$$

where ϕ is a feature vector, ξ are slack variables measuring the misclassification errors, and C represents a control (or penalty) variable for large and small margin violations.

Ultimately, the SVM contains a subset of points of the two classes (crash severity outcomes) called support vectors. Along with the support vectors are a corresponding set of weights $\boldsymbol{w}$ (one for each feature), also called alpha, on an optimal hyperplane in which the parameter bias defines the distance to the origin of the hyperplane. Furthermore, transformation into a higher-dimensional space for data which are not linearly separable in the original space is implemented by introducing the following kernel function: $K\left(x_i, x_j\right) \equiv \phi\,(x_i)^T \phi\,(x_j)$. Although several kernels have been developed and applied to SVM modeling, the Gaussian Radial Basis Function (RBF) is one of the more commonly used kernel functions. It has demonstrated better results in related works (Yu and Abdel-Aty 2014, Mokhtarimousavi *et al.* 2019), and thus was used for crash severity analysis in this study. The RBF kernel is defined as:

$$K_{Gaussian}(\boldsymbol{x}_i \times \boldsymbol{x}_j) = exp\left(-\frac{\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2}{2\sigma^2}\right) \tag{3-7}$$

Equation (3-7) illustrates the sigma selection process of the Gaussian RBF kernel for classifications where σ is considered to be 0.4. In this context, σ is the parameter that controls the width of the Gaussian.

Traditionally, non-heuristic algorithms such as grid-search and gradient descent were applied to set SVM parameters (Chapelle *et al.* 2002, Keerthi 2002, Wang *et al.* 2005). These methods, however, are vulnerable to local optimum and cannot guarantee convergence to a global optimum (Mokhtarimousavi *et al.* 2014, Mokhtarimousavi *et al.* 2015, Mokhtarimousavi *et al.* 2018, Shamshiripour and Samimi 2019, Mokhtarimousavi *et al.* 2021b). On the other hand, biologically-inspired metaheuristics such as the Genetic Algorithm (GA), Artificial Bee Colony (ABC), Particle Swarm Optimization (PSO), Fruit Fly Optimization Algorithm (FOA), etc., are more likely to result in finding the global optimum solution compared to the traditional aforementioned methods (Shen *et al.* 2016, Taghiyeh and Xu 2016, Mokhtarimousavi *et al.* 2018).

Since the prediction performance of SVM in safety analysis can be significantly enhanced by tuning the model parameters (Mokhtarimousavi *et al.* 2019), the CS, a powerful metaheuristic algorithm for global optimization, was employed to tune the SVM parameters. A critical SVM parameter is b, which is the bias term. This term allows the SVM to pass the origin in order to determine a separating hyperplane with the maximum margin. Without the bias, the SVM will always go through the origin of the feature space. Another critical parameter is alpha, which is a vector of weights from which the hyperplane is formed. Another parameter that will be tuned using the CS-SVM is the number of support vectors.

3.3.4. Cuckoo Search (CS) Optimization Algorithm

The final step, as stated previously, is the application of the CS algorithm. The CS algorithm is a swarm-intelligence metaheuristic algorithm developed by Xin-she Yang and Suash Deb in 2009 (Yang and Deb 2009), and in the case of the current work, was used to tune the SVM parameters. This nature-inspired metaheuristic mimics the breeding behavior of a specific bird family called “cuckoo.” In order to understand the cuckoo’s unique breeding behavior and how the algorithm employs this factor to find a global optimal solution, two concepts need to be explained. These two concepts will be discussed below.

A. The Cuckoo’s reproduction strategy

The cuckoo follows a unique reproduction system called “brood parasitism.” This strategy makes them dependent on other birds to hatch their eggs. The female cuckoo tries to find the nest of another species that recently laid eggs so that it will lay and hide its own eggs. If the eggs are identified by the host bird, they may either be thrown away or the host bird will abandon the nest and make a new one.

B. Lévy Flights

In CS, a cuckoo searches for a new nest via Lévy flights, which is a forward-step technique that resembles movement by birds and animals. Lévy flights essentially follow a random process to search for food because the next step is based on the current location and the transition probability to the next location. This random walk is derived from a Lévy distribution with an infinite variance and mean. Such behavior was applied to different optimization algorithms, and the results demonstrated its superiority and capability over other distributions, specifically in CS (Yang 2010).

The procedure of CS algorithms to find global optimum solutions is based on three main rules (Yang and Deb 2009):

- Each cuckoo dumps eggs on a randomly selected nest.
- The best nest with the highest quality eggs (i.e., solutions) will be passed over to the next generation.
- For a fixed number of available host nests, the egg laid by a cuckoo can be discovered by the host bird with the probability $p \in [0,1]$.

When choosing a new random nest (i.e., generating new solutions), a Lévy flight is performed, as follows:

$$X_i^{t+1} = X_i^t + \alpha \oplus L\acute{e}vy\ (\lambda) \tag{3-8}$$

where X_i^{t+1} is a new solution and α ($\alpha > 0$) is the step size associated with the scales of the problem. The product $\oplus$ means entry-wise multiplications. Lastly, the Lévy (λ) follows the Lévy distribution with an infinite variance and infinite mean (Yang and Deb 2009):

$$L\acute{e}vy\ (\lambda) \sim u = t^{-\lambda} \tag{3-9}$$

3.3.5. Random Forest (RF)

The final technique applied in the current study is random forest (RF). RF, which was developed by (Breiman 2001), is essentially a collection of Decision Trees (DTs) and an ensemble machine learning technique that uses a bagging algorithm to generate multiple random decision trees to perform class predictions of each predictor. The RF functionality offers unbiased estimates of the classification error, which is also robust against over-fitting problems as those found in DTs (Shi and Abdel-Aty 2015, Taghiyeh *et al.* 2020). Variable selection is an important feature of RF, that as mentioned earlier, has been frequently used in safety studies. It was utilized in this study for the purpose of finding the important variables highly related to the response variable for the purpose of interpretation. This helps to achieve a more efficient statistical model estimate. Variable selection in RF is based on detecting the interactions between variables through a tree growing procedure, and recording the prediction error rates on out-of-bag (OOB) data (observations that are not used in training set) before and after random permutation of the predictor variable.

In this study, using R, RF was developed to screen the importance of each indicator variable to be used in mixed logit modeling procedures. The most important variables were selected by monitoring the increase of prediction errors when OOB data (i.e., 30% of the training samples that are not used in the tree growth) were permuted for that variable, while all others were left unchanged (Liaw and Wiener 2002).

3.4. Empirical Setting and Data

In the present study, crash records are obtained from the Florida Signal Four Analytics tool (S4A 2018), which is a statewide interactive, web-based geospatial crash analytical tool. A three-year period of statewide crash data was collected from January 1, 2015 to December 31, 2017. Crashes that occurred in work zone areas with worker presence were then extracted from the crash records. The dataset contained a total of 2,112,783 crash records, with 1.55% of the crashes occurring in work zones (i.e., 32,750 occurred in work zones). Out of the total number of work zone crashes, 37.48% were associated with the worker presence. After data cleaning, a total of 12,042 usable crash records were identified from 2015 to 2017, in which there were 64 fatal crashes, 3,476 injury crashes, and 8,502 no-injury crashes. Due to the low frequency of fatal crashes, fatal and injury crashes were combined to create one severity level referred to as "Fatality/Injury." The other considered severity level is property damage only (PDO) or "No Injury." According to the average times for sunset and sunrise conditions for the state of Florida (Timeanddate), two time periods, from 6:00 to 19:59 and 20.00 to 05:59, were considered for daytime and nighttime conditions. The frequency of the dependent and independent variables in the utilized dataset are shown in Table 3-1.

Table 3-1 Variable Definition and Data Description

Variable Decs	Variable Name	Crash Severity Levels						Total
		Fatality		Injury		No Injury		
		Percent	Freq.	Percent	Freq.	Percent	Freq.	
Crash Severity	SEV	0.53%	64	28.87%	3,476	70.60%	8,502	12,042
Crash-Level								
Crash Time	TOD							
Daytime	DAYT	56.25%	36	74.97%	2,606	76.89%	6,537	9,179
Nighttime	NIGHTT	43.75%	28	25.03%	870	23.11%	1,965	2,863
Crash Type	CRSHTYP							
Backed Into	CRSHTBI	1.56%	1	0.66%	23	2.40%	204	228
Left Entering	CRSHTLE	4.69%	3	5.58%	194	2.62%	223	420
Left-Rear	CRSHTLR	1.56%	1	1.78%	62	0.91%	77	140
Off Road	CRSHTOR	10.94%	7	7.77%	270	8.07%	686	963
Parked Vehicle	CRSHTPV	7.81%	5	2.93%	102	4.54%	386	493
Pedestrian	CRSHTPDS	20.31%	13	2.42%	84	0.15%	13	110
Rear-End	CRSHTRE	32.81%	21	61.25%	2,129	49.59%	4,216	6,366
Right Angle	CRSHTRA	4.69%	3	3.57%	124	2.38%	202	329
Rollover	CRSHTROLO	1.56%	1	2.19%	76	0.62%	53	130
Same Direction Sideswipe	CRSHTSDS	3.13%	2	6.39%	222	20.17%	1,715	1,939
Single Vehicle	CRSHTSV	10.94%	7	5.47%	190	8.55%	727	924

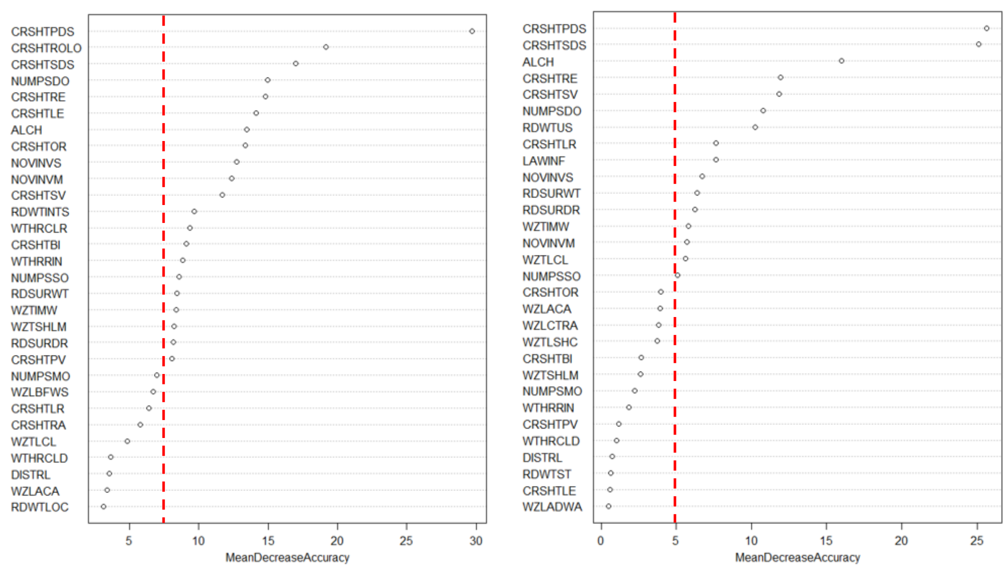
Table 3-1 Variable Definition and Data Description

Road Surface Condition	RDSURF							
Dry	RDSURDR	93.75%	60	90.22%	3,136	90.41%	7,687	10,883
Wet	RDSURWT	6.25%	4	9.78%	340	9.59%	815	1,159
Weather Condition	WETHR							
Clear	WTHRCLR	79.69%	51	74.31%	2,583	74.82%	6,361	8,995
Cloudy	WTHRCLD	15.63%	10	20.40%	709	19.44%	1,653	2,372
Rain	WTHRRIN	4.69%	3	5.29%	184	5.74%	488	675
Road Sys Identifier	RDWTYP							
County	RDWTCNT	7.81%	5	10.33%	359	10.41%	885	1,249
Interstate	RDWTINTS	40.63%	26	36.39%	1,265	37.36%	3,176	4,467
Local	RDWTLOC	12.50%	8	10.70%	372	13.57%	1,154	1,534
State	RDWTST	28.13%	18	29.80%	1,036	27.44%	2,333	3,387
Turnpike/Toll	RDWTTRNT	4.69%	3	3.42%	119	3.93%	334	456
U.S.	RDWTUS	6.25%	4	9.35%	325	7.29%	620	949
Number of Vehicle Involved in Crash	NOVINV							
Single Vehicle	NOVINVS	60.94%	39	83.86%	2,915	83.75%	7,120	10,074
Multi Vehicle	NOVINVM	39.06%	25	16.14%	561	16.25%	1,382	1,968
Veh-Level								
Number of Passengers	NUMPAS							
Driver Only	NUMPSDO	60.94%	39	55.93%	1,944	68.27%	5,804	7,787
Single Occupant	NUMPSSO	7.81%	5	19.79%	688	12.22%	1,039	1,732
Multi Occupant	NUMPSMO	31.25%	20	24.28%	844	19.51%	1,659	2,523
Alcohol Related	ALCH							
Yes		89.06%	57	96.14%	3,342	98.32%	8,359	11,758
No		10.94%	7	3.86%	134	1.68%	143	284
Distraction Related	DISTRL							
Yes		89.06%	57	79.49%	2,763	82.73%	7,034	9,854
No		10.94%	7	20.51%	713	17.27%	1,468	2,188
Work Zone								
Type of Work Zone	WZTYP							
Intermittent or Moving Work	WZTIMW	7.81%	5	5.41%	188	4.70%	400	593
Lane Closure	WZTLCL	29.69%	19	35.53%	1,235	37.07%	3,152	4,406
Lane Shift/Crossover	WZTLSHC	1.56%	1	8.72%	303	10.23%	870	1,174
Work on Shoulder or Median	WZTSHLM	60.94%	39	50.35%	1,750	47.99%	4,080	5,869
Crash Location in Work Zone	WZLOC							
Activity Area	WZLACA	90.63%	58	70.11%	2,437	68.71%	5,842	8,337
Advance Warning Area	WZLADWA	4.69%	3	9.67%	336	8.75%	744	1,083
Before the First Work Zone Warning Sign	WZLBFWS	1.56%	1	4.09%	142	4.76%	405	548
Termination Area	WZLCTRA	0%	0	1.99%	69	1.75%	149	218
Transition Area	WZLTRA	3.13%	2	14.15%	492	16.02%	1,362	1,856
Law Enforcement in Work Zone	LAWINF							
Yes		84.38%	54	81.33%	2,827	79.22%	6,735	9,616
No		15.63%	10	18.67%	649	20.78%	1,767	2,426

3.5. Model Estimation Results

3.5.1. Variable Importance

The importance of variables was estimated by considering the mean decrease accuracy index, which is the decrease in model accuracy from permuting the values in each feature for explaining the target variable (i.e., severity levels). In order to obtain the sufficient number of trees to reach relatively stable results, the OOB error rate was monitored against a various number of trees, and the minimum rates were achieved using 500 trees for both daytime and nighttime conditions. The final results for variable importance ranking are shown in Figure 3-14.



Daytime	Nighttime

Figure 3-14 Variable Importance Ranking Using Random Forest

In order to choose the most important covariates affecting severity in daytime and nighttime conditions, a cut-off value of 7 and 5 were considered, respectively. This led to selecting 21 candidate variables for daytime and 16 for nighttime to fit the mixed logit models.

3.5.2. Binary Mixed Logit Model Results

Separate models have been generated for work zone crashes that involved workers. The mixed logit estimation results for the daytime and nighttime periods with corresponding marginal effects are presented in Table 3-2 and Table 3-3, respectively. In the estimation results, the random parameters were selected considering the statistically significant standard deviations for the normal distribution. In addition, the marginal effects were used to illustrate the injury-severity probability change due to a one-unit change in the explanatory variables.

Table 3-2 Daytime Mixed Logit Model Estimation Results

Variable	Coefficient	Std. Error	*t*-statistic	Marginal Effects
Constant	-3.516	0.321	-10.96	
(Std. Dev. Of Normally Distributed Random Parameter)	*(18.027)*	*(0.802)*	*(22.48)*	
Work Zone Characteristics				
Work Zone Type (1 if Intermittent or Moving Work, 0 Otherwise)	2.388	0. 289	8.25	0.001
Work Zone Type (1 if Work on Shoulder or Median, 0 Otherwise)	1.172	0.132	8.85	0.001
Crash Characteristics				
Crash Type (1 if Rear-End, 0 Otherwise)	-2.475	0.252	-9.80	-0.001
(Std. Dev. Of Normally Distributed Random Parameter)	*(5.681)*	*(0.273)*	*(20.78)*	
Crash Type (1 if Pedestrian Related, 0 Otherwise)	21.683	1.197	18.19	0.008
Crash Type (1 if Same Direction Sideswipe, 0 Otherwise)	-11.330	0.570	-19.89	-0.004
Crash Type (1 if Left Entering, 0 Otherwise)	2.793	0.074	7.48	0.001
Crash Type (1 if Rollover, 0 Otherwise)	11.001	0.701	15.71	0.004
Crash Type (1 if Single Vehicle, 0 Otherwise)	-2.113	0.325	-6.51	-0.001
Crash Type (1 if Backed-Into, 0 Otherwise)	-13.34	0.843	-15.81	-0.005
Crash Type (1 if Parked Vehicle, 0 Otherwise)	-7.082	0.494	-14.33	-0.003
Alcohol Related (1 if Yes, 0 if No)	6.836	0.676	10.11	0.003
Environmental Characteristics				
Weather Condition (1 if Rainy, 0 if No)	-1.268	0.247	-5.12	-0.001
Number of Vehicle Involved (1 if Multiple Vehicles, 0 if No)	2.490	0.320	7.77	0.001
Number of Passengers (1 if Single Occupant, 0 if No)	-1.636	0.198	-8.27	-0.001
Number of Passengers (1 if Driver Only, 0 if No)	-5.302	0.283	-18.73	-0.002
Model Summary*				

Table 3-2 Daytime Mixed Logit Model Estimation Results

Number of Observations	9,179
Log-Likelihood at Zero	-5509.21
Log-Likelihood at Convergence	-5120.87
Overall Prediction Accuracy	62.37%
Sensitivity	34.63%
Specificity	73.58%
AUC	0.668

*Analysis of Binary Choice Model Predictions Based on Threshold = 0.5000

Table 3-3 Nighttime Mixed Logit Model Estimation Results

Variable	Coefficient	Std. Error	*t*-statistic	Marginal Effect
Constant	-0.3251	0.116	-2.79	
(Std. Dev. Of Normally Distributed Random Parameter)	*(1.378)*	*(0.769)*	*(17.92)*	
Work Zone Characteristics				
Law Enforcement in Work Zone (1 Y, 0 Otherwise)	-0.638	0.088	-7.20	-0.095
Crash Characteristics				
Crash Type (1 if Pedestrian Related, 0 Otherwise)	6.182	1.035	5.97	0.923
Crash Type (1 if Single Vehicle, 0 Otherwise)	-1.030	0.176	-5.83	-0.154
Crash Type (1 if Left-Rear, 0 Otherwise)	1.007	0.384	2.62	0.150
Crash Type (1 if Rear-End, 0 Otherwise)	0.552	0.109	5.04	0.082
(Std. Dev. Of Normally Distributed Random Parameter)	*(1.020)*	*(0.084)*	*(12.10)*	
Crash Type (1 if Same Direction Sideswipe, 0 Otherwise)	-2.959	0.350	-8.44	-0.442
(Std. Dev. Of Normally Distributed Random Parameter)	*(3.379)*	*(0.408)*	*(8.28)*	
Alcohol Related (1 if Yes, 0 if No)	1.429	0.180	7.91	0.213
(Std. Dev. Of Normally Distributed Random Parameter)	*(2.798)*	*(0.310)*	*(9.01)*	
Number of Passengers (1 if Driver Only, 0 if No)	-1.171	0.100	-11.63	-0.175
(Std. Dev. Of Normally Distributed Random Parameter)	*(3.112)*	*(0.153)*	*(20.28)*	
Model Summary*				
Number of Observations	2,863			
Log-Likelihood at Zero	-1780.77			
Log-Likelihood at Convergence	-1621.53			
Overall Prediction Accuracy	61.37%			
Sensitivity	38.42%			
Specificity	71.86%			
AUC	0.693			

*Analysis of Binary Choice Model Predictions Based on Threshold = 0.5000

3.5.3. Model Temporal Stability Test Results

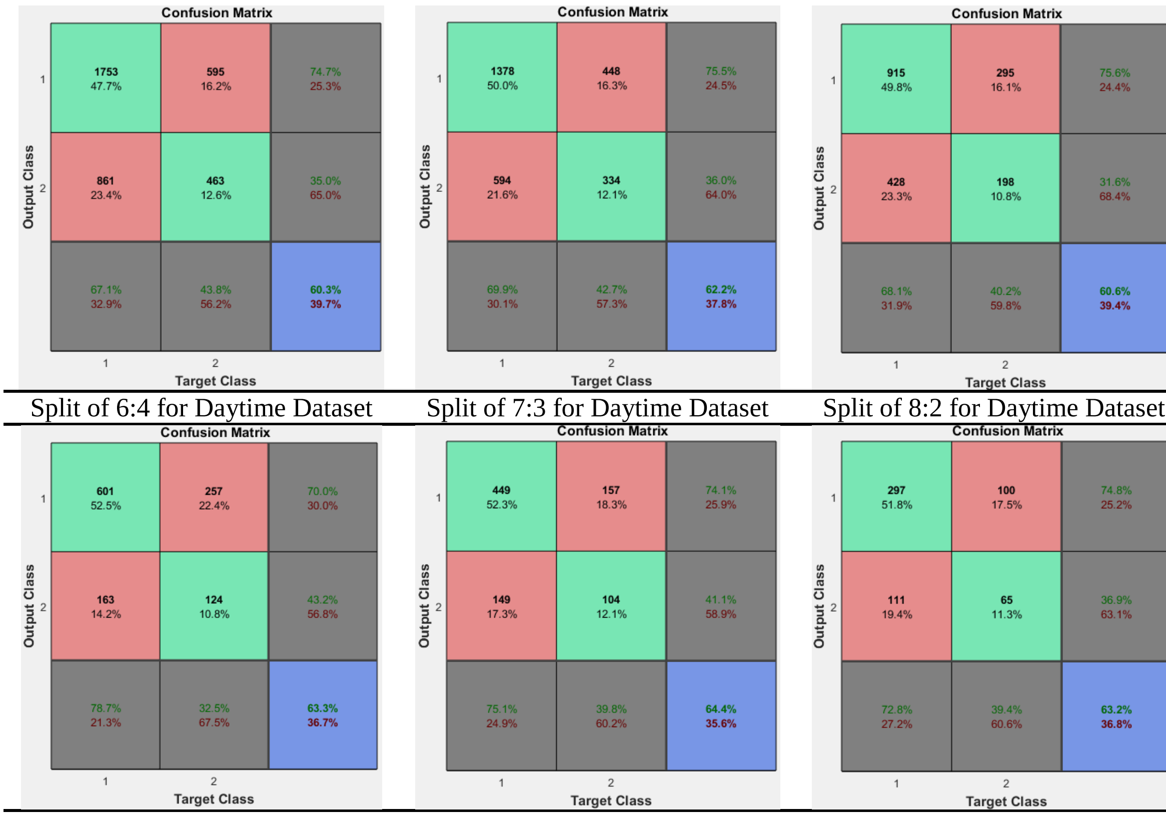
In regard to examining the temporal stability of model estimates across time periods (daytime vs. nighttime), applying Equation (3-4) results in a chi-square statistic of 4,136.78

and 7,928.2, with the corresponding degrees of freedom of 16 and 9 for MX1 and MX2, respectively. The significant difference between daytime and nighttime models suggests that work zone crashes that involve workers need to be modeled separately for safety analysis, with well over 99% confidence. This indicates that a single model, including daytime and nighttime crashes for the given data, would be incorrect. In other words, parameter estimates are statistically different for daytime and nighttime crash estimation and are not transferable between daytime and nighttime crashes. This finding is in line with a number of recent safety analysis studies that demonstrated separate injury-severity models that need to be estimated for different time periods (Behnood and Mannering 2015, Anderson and Dong 2017).

3.5.4. SVM Results

In this study, SVM models with RBF kernel function were coded in the MATLAB R2018b programming environment. To better assess the model prediction performance, the entire daytime and nighttime datasets were randomly separated into three sub-datasets (a training set and a testing set) with a ratio of 6:4 (i.e., 60% for training and 40% for testing), 7:3, and 8:2. Preliminary performance test results illustrated in a format of confusion matrices in Figure 3-15 reveal that SVM models with the split of 7:3 performed better in both daytime and nighttime models. This split ratio was therefore selected for further model prediction performance improvement through the application of CS metaheuristic optimization in parameter tuning.

Considering the fact that the performance of metaheuristic algorithms is also significantly influenced by the proper tuning of their parameters, a Taguchi's robust design method was used to obtain the best parameters of the CS algorithm (for detailed information regarding the Taguchi method, readers are referred to (Peace 1993)). In performing the Taguchi test, a number of 1,000 iterations, population size of 100, step size equal to 0.1 and discovery rate equal to 0.6 were utilized when using CS algorithms.



Split of 6:4 for Daytime Dataset	Split of 7:3 for Daytime Dataset	Split of 8:2 for Daytime Dataset
Split of 6:4 for Nighttime Dataset	Split of 7:3 for Nighttime Dataset	Split of 8:2 for Nighttime Dataset

Figure 3-15 SVM Confusion Matrices of Different Data Splits

In the assessment of the prediction performance of the CS-SVM models, the following criteria were calculated:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \qquad (3\text{-}10)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100\% \tag{3-11}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \times 100\% \tag{3-12}$$

Where the parameters in the equations refer to true positive (TP), true negative (TN), false positive (FP), and false negative (FN) counts. Accuracy measures the overall effectiveness of a classifier by its percentage of correct predictions. Sensitivity shows the effectiveness of a classifier to identify positive labels, while specificity illustrates this for negative labels. In addition, the area under the receiver operating characteristic curve (AUC) criterion, which is recognized as one of the best measures to evaluate two-class classification models, is calculated as proposed in (Bradley 1997). This metric reflects the model performance based on the True Positive Rate (TPR) and the False Positive Rate (FPR) parameters at all classification thresholds, which in fact is the number of times a failure is ranked below a non-failure in the list. The detailed classification results of the final CS-SVM model are included in Table 3-4.

Table 3-4 Results of CS-SVM Models

CS-SVM	Confusion Matrices			Accuracy	Sensitivity	Specificity	AUC
Daytime Model		*No Injury*	*Injury*	84.00%	98.71%	48.71%	0.8811
	No Injury	1917 (69.6%)	416 (15.1%)				
	Injury	25 (0.9%)	395 (14.3%)				
Nighttime Model		*No Injury*	*Injury*	89.40%	97.32%	71.37%	0.9033
	No Injury	580 (67.6%)	75 (8.7%)				
	Injury	16 (1.9%)	187 (21.8%)				

3.5.5. Variable Impact Analysis

In order to quantify the contribution of the explanatory variables to the probability distribution of work zone crash severities, a two-stage sensitivity analysis was conducted. This method was recently adopted in SVM safety studies to identify the relationships between crash injury severity and various explanatory variables (Li *et al.* 2012, Yu and Abdel-Aty 2013, 2014, Chen *et al.* 2016, Khoda Bakhshi and Ahmed 2020). In this method, the value of each explanatory variable is replaced with a user-defined value (the same value is used for all input variables), while the others remain unchanged (Esmaeilzadeh and Mokhtarimousavi 2020). Then, the corresponding probabilities of the severity outcomes (No Injury and Fatality/Injury in this study) before and after this perturbation are simulated in CS-SVM models and recorded. The results are shown in Tables 3-5 and 3-6 for daytime and nighttime models, respectively.

Table 3-5 CS-SVM Daytime Variable Impact Analysis

Variable	Severity		Variable	Severity	
	No Injury	Fatality/Inj		No Injury	Fatality/Inj
Crash-Level Variables			Dry	0.862	0.138
Crash Type			Wet	0.806	0.194
Backed Into	0.829	0.171	**Vehicle-Level Variables**		
Left Entering	0.777	0.223	*Number of Passengers*		
Left-Rear	0.785	0.215	Driver Only	0.883	0.117
Off-Road	0.792	0.208	Single Occupant	0.821	0.179
Parked Vehicle	0.827	0.173	Multi Occupant	0.776	0.224
Pedestrian	0.767	0.233	*Alcohol Related*		
Rear-End	0.849	0.151	Yes	0.785	0.215
Right Angle	0.798	0.202	No	0.867	0.133
Rollover	0.771	0.229	*Distraction Related*		
Same Direction Sideswipe	0.863	0.137	Yes	0.818	0.182
Single Vehicle	0.819	0.181	No	0.867	0.133
Weather Condition			**Work Zone Variables**		
Clear	0.857	0.143	*Type of Work Zone*		
Cloudy	0.825	0.175	Intermittent or Moving Work	0.787	0.213
Rain	0.808	0.192	Lane Closure	0.838	0.162
Road Sys Identifier			Lane Shift/Crossover	0.811	0.189
County	0.828	0.172	Work on Shoulder or Median	0.854	0.146
Interstate	0.843	0.157	*Crash Location in Work Zone*		
Local	0.838	0.162	Activity Area	0.861	0.139
State	0.829	0.171	Advance Warning Area	0.814	0.186
Turnpike/Toll	0.813	0.187	Before the First Work Zone Warning Sign	0.806	0.194
U.S.	0.793	0.207	Termination Area	0.795	0.205
Number of Vehicle Involved in Crash			Transition Area	0.827	0.173
Single Vehicle	0.801	0.199	*Law Enforcement in Work Zone*		
Multi Vehicle	0.857	0.143	Yes	0.813	0.187
Road Surface Condition			No	0.864	0.136

Table 3-6 CS-SVM Nighttime Variable Impact Analysis

Variable	Severity		Variable	Severity	
	No Injury	Fatality/Inj		No Injury	Fatality/Inj
Crash-Level Variables			Dry	0.804	0.196
Crash Type			Wet	0.782	0.218
Backed Into	0.787	0.213	**Vehicle-Level Variables**		
Left Entering	0.774	0.226	*Number of Passengers*		
Left-Rear	0.778	0.222	Driver Only	0.830	0.170
Off-Road	0.792	0.208	Single Occupant	0.777	0.223
Parked Vehicle	0.790	0.210	Multi Occupant	0.750	0.250
Pedestrian	0.764	0.236	*Alcohol Related*		
Rear-End	0.768	0.232	Yes	0.739	0.261
Right Angle	0.776	0.224	No	0.814	0.186
Rollover	0.780	0.220	*Distraction Related*		
Same Direction Sideswipe	0.819	0.181	Yes	0.786	0.214
Single Vehicle	0.795	0.205	No	0.802	0.198
Weather Condition			**Work Zone Variables**		
Clear	0.805	0.195	*Type of Work Zone*		
Cloudy	0.784	0.216	Intermittent or Moving Work	0.765	0.235
Rain	0.777	0.223	Lane Closure	0.813	0.187
Road Sys Identifier			Lane Shift/Crossover	0.797	0.203
County	0.795	0.205	Work on Shoulder or Median	0.760	0.240
Interstate	0.810	0.190	*Crash Location in Work Zone*		
Local	0.756	0.244	Activity Area	0.806	0.194
State	0.778	0.222	Advance Warning Area	0.772	0.228
Turnpike/Toll	0.810	0.190	Before the First Work Zone Warning Sign	0.788	0.212
U.S.	0.758	0.242	Termination Area	0.756	0.244
Number of Vehicle Involved in Crash			Transition Area	0.797	0.203
Single Vehicle	0.795	0.205	*Law Enforcement in Work Zone*		
Multi Vehicle	0.793	0.207	Yes	0.823	0.177
Road Surface Condition			No	0.763	0.237

3.6. Discussions

A total of 23 indicator variables were found to be significant throughout the daytime and nighttime logit models, with five variables being significant in both models. Of the five variables found to be significant in both models, three have heterogeneous effects on crash severity outcomes in the nighttime model: sideswipe crashes in the same direction, alcohol consumption, and driver-only involvement. Although these variables were found to have heterogeneous effects in the nighttime model, their effects in the daytime model were homogeneous.

To facilitate the discussion, the contributing factors according to the daytime and nighttime crash severity models and their effects will be discussed separately. Discussion of the contributing factors will be followed by a comparison of the results from both models.

3.6.1. Daytime Crash Severity Models

As for random parameters, rear-end crash type is the only variable found to have a normally distributed estimated random parameter. Model estimates show that the parameter for rear-end crash type has an estimated mean of -2.475 and an estimated standard deviation of 5.681. These estimates indicate that the estimated parameter mean is greater than zero for 33.15% of crashes and less than zero for 66.85% of crashes. That is, 33.15% of rear-end crashes are more likely to result in an injury, and 66.85% are less likely to result in an injury. As stated by (Wang *et al.* 1996), rear-end crashes increased significantly in work zone locations. In addition, rear-end crashes were found to be the prominent crash type in work zones (Srinivasan *et al.* 2007). Sudden stops, following too closely while drivers are distracted due to cell phone use, and distraction with worker presence or work zone equipment are all factors more likely to be the reported cause for rear-end crashes (Osman *et al.* 2018). Therefore, the significance of this variable in a work zone context is anticipated. As for the heterogeneous effects on crash severity, a potential reason may stem from the differences in speed limits, driver compliance, and location where the crash occurred. Specifically, work zones often have a lower speed limit (i.e., speed drop). Subsequently, if a rear-end crash occurs at lower speeds, a rear-end crash in which no injury is sustained can be expected. However, if drivers are distracted at the start of the work zone or do not comply with the lower speed limit, rear-end crashes will occur at higher speeds, resulting in the likelihood of a more severe crash involving injuries.

The abovementioned results are consistent with the CS-SVM (referred to as SVM for the remainder of this section) daytime output. Based on the variable impact analysis of the SVM model, it was found that drivers are more likely to suffer severe and fatal injuries in pedestrian-related, rollover, and left-entering work zone crashes, but less likely in backed-into, same direction sideswipe, parked vehicle, single vehicle, and rear-end crashes. Rear-end crashes and sideswipe crashes in the same direction have the lowest probabilities of fatality/injury crashes with 0.151 and 0.137, respectively. On the other hand, with 4.48% and 2.69% lower probabilities when compared to pedestrian-related and rollover crashes, left-entering crashes are among the top three crash types that result in more severe crashes. Figure 3-16 illustrates the effects of different crash types sorted from those with the highest to lowest impacts, on fatality/injury crashes obtained from the daytime SVM model.

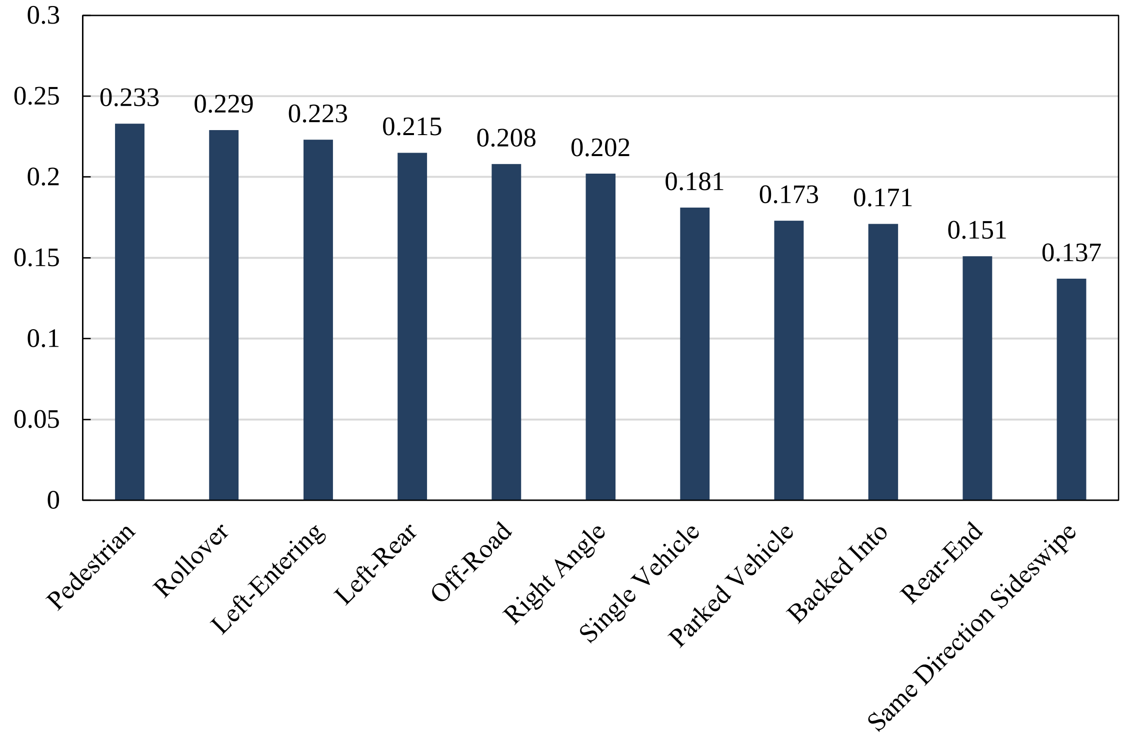


Figure 3-16 Effect of Crash Type on SEV, CS-SVM Daytime Model

3.6.2. Nighttime Crash Severity Models

A total of eight indicator variables were found to be significant in the nighttime model, where four of the eight variables were found to have heterogeneous impacts on crash severity outcomes. As previously stated, sideswipe crashes in the same direction, alcohol consumption, and driver-only involvement factors were found to be significant in both daytime and nighttime models. However, in the nighttime model, these three factors were found to have normally distributed random parameters. With a mean of -2.959 and standard deviation of 3.379, 19.06% (greater than zero) of sideswipe crashes in the same direction in worker-involved work zone crashes are more likely to result in fatality/injury crashes. Simultaneously, 80.94% (less than zero) of sideswipe crashes in the same direction are associated with crashes in which no injury was sustained. The heterogeneous nature may have linked with the unobservables related to nighttime conditions, such as the level of lighting present in the work zone or the ability to see reflective vests worn by workers.

The second variable to have a normally distributed estimated random parameter is the indicator for drivers under the influence of alcohol. With a mean of 1.429 and standard deviation of 2.798, 30.48% of crashes involving a driver under the influence of alcohol are less likely to result in severe injury crashes, and 69.52% are more likely. With a higher likelihood of alcohol consumption during nighttime hours, the significance of this variable is expected (Yasmin *et al.* 2014). The heterogeneous nature is consistent with findings from previous works. For example, (Xie *et al.* 2012) found that driving under the influence increases the likelihood of a no-injury crash.

The majority of work, however, found alcohol to increase the likelihood of a severe injury crash (Kockelman and Kweon 2002, Qi *et al.* 2005, Bai and Li 2007, Harb *et al.* 2008, Morgan and Mannering 2011, Xiong *et al.* 2014, Chen *et al.* 2015). This is in agreement with the results from the present work that found that the majority have an increase in the severe injury likelihood. In addition, these results are in-line with the results of the developed SVM model, where the SVM model shows a 40.32% higher probability of a severe crash if the driver was driving under the influence of alcohol.

The driver-only indicator variable is the last indicator to have a normally distributed estimated random parameter, with a mean of -1.171 and a standard deviation of 3.112. This indicates that 35.34% (greater than zero) of crashes are associated with only one driver in the vehicle and are more likely to result in a severe injury, whereas 64.66% of only crashes with one driver (less than zero) are less likely to result in severe crashes. To be more specific, the majority of driver-only vehicles involved in work zone crashes with workers were less likely to result in severe injury crashes. Although a proportion of single occupant crashes increases the likelihood of a more severe crash, the majority of single occupant crashes decreases the likelihood. It has been shown that the increase in the number of occupants increases work zone crashes (Ozturk 2014, Osman *et al.* 2018). It was found that there was a 95 percent level of significance correlation between the number of occupants involved and work zone crash severity (Ozturk 2014). The increase in likelihood may be capturing unobservable characteristics related to the driver, which are not included in the data. The same results obtained with the SVM model show that multi-occupant vehicles have 12.10% and 47.05% higher probabilities of being in severe crashes compared to single-occupant and driver-only conditions in work zone locations. This may be attributed to distracted driving (i.e., a driver's attention can be diverted away from the driving task) as a result of distractions by vehicle occupant/occupants. This may affect a driver's ability to safely perform the driving task.

The final indicator in the nighttime model with an estimated random parameter is the indicator for rear-end crashes. Like the daytime model, rear-end crashes were again heterogeneous for the nighttime model. The mean of 0.552 and the standard deviation of 1.020 indicates that while 29.42% of the distribution is less than zero, the majority of rear-end crashes are associated with a higher probability of crash severity at 70.58%. This parameter may be capturing unobservables related to weather conditions. For instance, it was found by (Qi *et al.* 2005) that there is a correlation between weather conditions and the severity of rear-end crashes in work zone accidents. This is in-line with the SVM variable impact analysis results that indicate that drivers are more likely to suffer from fatality/injury in work zone crashes in rainy weather conditions with an increase in the probability to 0.223. This is 3.24% and 14.35% higher than when in cloudy and clear weather conditions, which is equal to 0.216 and 0.195, respectively. A possible explanation for this finding may be the reduction of braking capacity due to wet and slippery road surface conditions, low visibility, lighting glare, and lack of alertness (Abaza *et al.* 2017). This parameter may also be attempting to capture unobservables related to driver-specific information, such as perception reaction time, visual acuity, bone mass, etc. (Mannering *et al.* 2016).

Other significant variables, such as pedestrian-related, single vehicle, and left-rear crash types, have significant impacts on the probability of a crash resulting in an injury. The analysis of marginal effects shows that pedestrian-related and left-rear crashes have a 0.923 and 0.150 higher probability of resulting in severe crashes. On the other hand, single vehicle involvement in work zone crashes have a 0.154 lower probability of resulting in a severe crash, based on the marginal effect analysis. This result is not only consistent with the SVM output that shows a lower probability of single vehicle involvement compared to multi-vehicle in severe injuries, but is also consistent with the findings of previous studies on work zone injury severity (Katta 2013, Dias 2015).

Previous studies related to injury severity of work zone crashes have lacked the consideration of variables related to the presence of law enforcement. The results from this study illustrate that law enforcement had a significant effect on crash severity and decreases the likelihood of severe crashes. To be specific, the BMXL model indicates that for a one-unit increase in law enforcement (in other words, going from absence to presence of law enforcement), we expect a 0.095 decrease in the probability of the dependent variable severity, holding all other independent variables constant. The same conclusion can be inferred from the SVM results. It was shown that of the crash occurrences in nighttime work zones while workers are present, the absence of law enforcement is associated with approximately a 34.0% higher probability of severe injuries. This result illustrates that the use of proper temporary traffic control (i.e., stationary enforcement, circulating enforcement, etc.) is essential to warning drivers that they are approaching a work zone location, especially where workers are present.

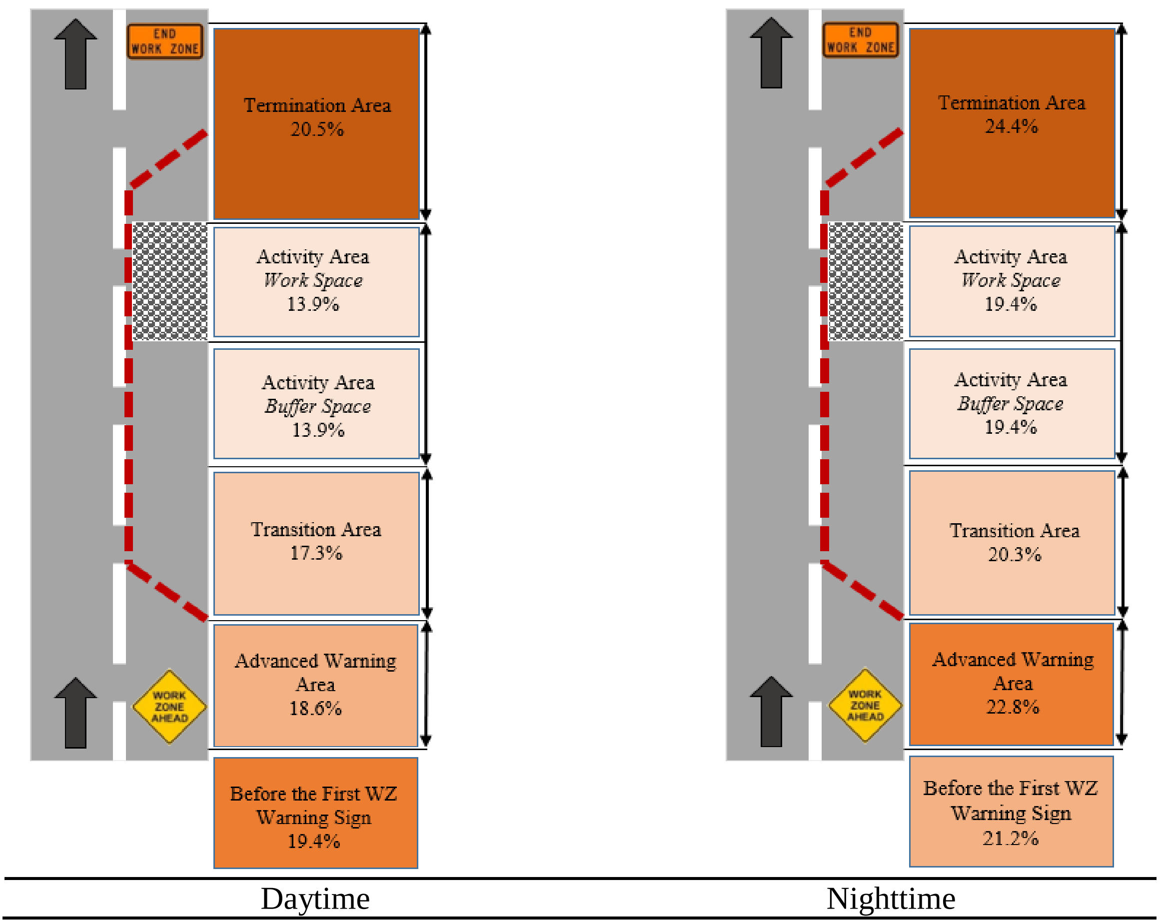


Figure 3-17 Critical Locations in Work Zone[1]

[1] The darker the color, the higher probability of severe crashes in that location.

Finally, using the results from the SVM variable impact analysis, a heat map was created on a typical work zone layout to visualize the critical locations of work zone configuration for worker safety. The heat map is shown in Figure 3-17.

As observed in Figure 3-17, the termination area is the most critical location that increases the likelihood of severe crashes in both daytime and nighttime work zones. This area, in terms of impact on severity, is followed by the area before the first work zone sign in the daytime and the advance warning area in the nighttime work zones. This finding may be attributed to a driver's intention to speed as they are exiting the work zone area, which is consistent with the findings of (Osman *et al.* 2018). The effects of speed variation on crash severity and frequency was also recently investigated in (Kamrani *et al.* 2018, Arvin *et al.* 2019b, a, Parsa *et al.* 2019b), and their results demonstrated that higher speed volatility is associated with a higher likelihood of crash occurrence.

3.7. Summary and Conclusions

In this chapter, the contributing factors of crash severity in work zone crashes that involved workers was investigated to determine the most important factors and their corresponding impacts on crash severity. Florida work zone crashes between 2015-2017 were the focus of the analysis. Descriptive analysis was used to test the relationships between crash severity and numerous variables. Significant contributing factors were included within the binary level severity models for both daytime and nighttime crashes. The findings of the study on work zone crash severity, crash severity modeling and analysis are summarized below.

Awareness of worker safety in construction work zone-related crashes represents a significant concern in roadway safety since it causes worker casualties. Different studies have been conducted to investigate the crash characteristics of nighttime and daytime construction activities, while the statistical reasons for these characteristics have not been known in the past. In addition, worker presence and its impact on severity in work zone crashes has remained unexplored. To address this gap in research, this study was undertaken to empirically examine the crash severity contributing factors by time-of-day for worker-involved work zone crashes. First, in order to facilitate the modeling procedures, random forest models were initially developed to select influential explanatory variables associated with crash severity. Assuming that the Florida crash data used for the current study was susceptive to heterogeneity, the potential candidate variables were then estimated utilizing a mixed logit modeling framework.

Likelihood ratio tests were conducted to examine the overall temporal stability of model estimates for severity outcomes (No Injury and Fatality/Injury). Marginal effects of each explanatory variable were also considered to investigate the effects of individual parameter estimates on work zone injury-severity probabilities. The results of a parameter transferability test demonstrated significant temporal instability among parameter estimates, which implies that worker-involved work zone crashes need to be modeled separately by time-of-day.

Due to the limitations of parametric models, such as the pre-assumption of data distribution and linear form of utility functions, which may not necessarily be applicable for crash data, non-parametric SVM models were also utilized to predict the entire set of explanatory variables in both models[2]. Since the prediction performance of SVM classification can be significantly enhanced by tuning its hyper-parameters, a higher level of performance is achieved by employing the CS metaheuristic optimization algorithm in SVM parameter tuning. When comparing the model performance, CS-SVM produced a higher percentage of correct prediction of the severity levels by 35.04% for daytime and 38.81% for nighttime compared to the SVM models, which were also higher than those produced by the BMXL model by 62.37% and 61.37%, respectively. This implies the ability to apply SI optimization techniques in SVM parameter selection to achieve higher prediction performance.

Although the prediction accuracy is the most intuitive measure of assessing classification models, higher prediction accuracy is not the only advantage of the proposed SVM models over the binary mixed logit models. Aside from prediction accuracy, other prediction metrics were considered for the models' goodness-of-fit comparison. For instance, the value of the AUC metric for the CS-SVM daytime model by 0.8811 and the nighttime model by 0.9033 is also substantially higher than that of the BMXL models by 0.6680 and 0.6926, respectively. These improvements may also be associated with consideration of the nonlinearity between the explanatory variables and severity outcomes, which is in line with the funding of previous studies such as (Yu and Abdel-Aty 2014) and

---

[2] It should be noted that some indicator variables have been excluded from the input variables for the BMXL models due collinearity (collinearity also implies correlation). This is a common issue in parametric statistical modeling approaches to estimating the relationship between crash variables and severity, which may reduce the total estimation accuracy. Applying machine learning techniques can relieve this issue as the geometric feature of variables is considered (i.e., distances between data points) and not just the linear relationship, which may lead to better prediction outcomes.

(Chen *et al.* 2016). In addition, conducting a two-stage sensitivity analyses (i.e., data perturbation and before-after comparison), the effects of each explanatory variable on the probability distribution of crash severity outcomes has been quantified. The results obtained from here demonstrate that driver alcohol involvement, rainy weather condition, wet road surface, multi-occupant for vehicle occupancy, and distraction are the most significant causes of fatalities/injuries in work zone crashes involving workers in both daytime and nighttime models. In terms of the variables, which are the number of vehicle-involved and law enforcement indicators, a mixed effect was found between daytime and nighttime conditions.

Non-parametric models like SVMs lack the ability to recognize significant variables affecting the response variable (outcome). On the other hand, the results from statistical methods do not show where the variable effect stands among all of the variables within each category. Taking this into consideration, the integration of the traditional statistical model and machine learning technique results enhance the understanding of work zone crash characteristics to interpret the effects of work zone presence on crash severity outcome. In addition, this research is based on a three-year statewide work zone crash dataset, where a sufficient number of crashes were considered in the modeling frameworks. This may eventually lead to valuable comparative information about these types of crash characteristics and provide safety experts and decision makers with the ability to prioritize the work zone operations based on different temporal, environmental, and geospatial conditions toward roadway user and worker safety.

In addition to previous works, (Li *et al.* 2012, Yu and Abdel-Aty 2013, Chen *et al.* 2016) investigated the application of SVM models for crash injury severity analysis.

Though the researchers pointed out several limitations of kernel function selection or the appropriate split of training and testing datasets, their paper sheds more light on parameter tuning. Realizing the importance of the parameter tuning process on the prediction performance of ML models is a distinguished line of research that has extended the application of swarm intelligence algorithms. It has been shown that incorporating metaheuristic optimization in SVM parameter tuning can significantly enhance the prediction performance of this supervised learning method. Although this line of research is very promising, the amount of studies that address this issue is still relatively scarce. In a bid to contribute to this growing body of knowledge to achieve higher model classification performance, future investigation can focus on applying different features and parameter selection techniques on different machine learning methods.

From a statistical modeling perspective, the ability of a model to accurately predict outcomes is just as important as its ability to explain causal factors, and current traffic safety literature lacks such a discussion. Thus, a deeper examination of model outcomes is necessary in traffic safety analysis to avoid any misunderstanding of the impact of contributing factors. Investigating different methods to evaluate variable importance when predicting the target variable to improve statistical model prediction power deserves more serious consideration for future research. Moreover, investigation of the similarities and differences of risk factors in work zone crash severities with or without worker presence by time of day may be of interest for future studies.

CHAPTER 4
CRASH FREQUENCY MODELING OF WORK ZONE CRASHES

4.1. Introduction

Construction work zones are one of the top priorities for transportation safety analysts, as they pose a huge challenge to roadway safety. Placing construction machinery on blocked travel lanes while construction crews are working, and changes in driving characteristics such as speed, lane changing maneuvers, etc., make the environmental and geometric characteristics of work zones prone to crash occurrence. In order to come up with the strategies to minimize the adverse effects of construction work zones, studying the risk factors and how work zone safety is affected by them is an area in need of greater research. Crash characteristics differ from location to location, as well as over time, along with varying features of participants at fault, environmental and geometrical conditions, and social factors.

It is often inevitable to establish work zones on roadways for construction activities such as bridge construction, bridge repair, or rehabilitation activities. In addition, although there would be a number new bridges being constructed (as the current chapter of this study is focused on), during the next decade in the state of Florida, there is a strong indication that emphasis will be placed on maintenance and rehabilitation of the existing bridges rather than on the construction of new ones. According to American Road & Transportation Builders Association (ARTBA) (ARTBA 2020) and Federal Highway Administration (FHWA) National Bridge Inventory (NBI) (NBI 2020), out of 12,518 bridges in Florida, 361 are classified as structurally deficient, which accounts for 2.9% of the total bridges in this state. Moreover, vehicles are restricted to crossing over 965 bridges due to loading restrictions, which may lead to an inefficient traffic flow in roadway networks.

Although the impact of work zone presence on crash frequency has been investigated and has shown increasing crash rates in previous studies such as (Khattak *et al.* 2002, Ullman *et al.* 2008, Jin and Saito 2009, Ozturk 2014), work zone type specifications have not been investigated yet, and the risk factors associated with work zone crash frequency at bridge locations are not fully understood. With this in mind, this study is focused on investigation of the contributing factors that affect crash frequency at bridge-related construction work zones. To this end, a Negative Binomial (NB) regression model and a Support Vector Regression (SVR) model were developed for modeling work zone crash frequency.

In this regard, a unique dataset was created, including work zone crashes that occurred in 60 bridge locations in Miami-Dade County. The dataset used for frequency analysis was integrated with crash data (three years of work zone crashes), road inventory data (Annual Average Daily Traffic (AADT), truck AADT, posted speed limit, roadway function classification, etc.), bridge geometric specification (median type, shoulder type, median width, shoulder width, surface width, curve indicator, number of lane), bridge location specifications (intersection, ramp, and horizontal curve indicators), and work zone related data (percentage of law enforcement and workers involvement).

The remainder of this chapter is organized as follows: the descriptive statistics of crash frequency for all work zone crashes involving workers that occurred in Florida between 2015 to 2017 are presented in the next section, followed by the details of the modeling frameworks and model specifications. The data used for the selected bridge locations in Miami-Dade County between 2015 and 2017 is explained in Section 4-4, and then the model estimation and research results are discussed in Section 4-5. A discussion is provided in Section 4-6, and finally, the key findings, research outcomes, and concluding remarks are summarized in Section 4-7.

4.2. Descriptive Statistics for Crash Frequency

In this section, all work zone crashes involving workers were included in the descriptive analysis. The Florida work zone crash frequency and related parameters for the years 2015 to 2017 were described in the following order: crash frequency by temporal variables, crash attributes, environmental condition, driver characteristics, and work zone characteristics.

4.2.1. Frequency Distribution by Temporal Variables

The annual crashes between 2015 and 2017 for both work zone crashes and the crashes that involved workers are shown in Figure 1-1. The crash trend shows that the number of work zone accidents in Florida increased year by year, from 10,162 accidents in 2015 to 11,285 in 2017 (i.e., 11% increase). Although the number of worker-involved work zone crashes increased by 409 from 2015 to 2016, they decreased by 128 from 2016 to 2017, which shows an almost 7% increase from 2015 to 2017. The temporal distribution of work zone crashes over the studied three-year period, which includes monthly, day of week, and hourly distributions, are shown in Figure 4-1 to 4-3, respectively.

As shown in Figure 4-1, March, April, and August shared the highest number of crashes with 1,428, 1,310, and 1,291 respectively, while the minimum number of crashes observed in January, September, and December were 1,036, 1,080, and 1,089. The lower number of crashes in December and January are excepted, as there are fewer working days, and thus, fewer active work zones during these two months of the year.

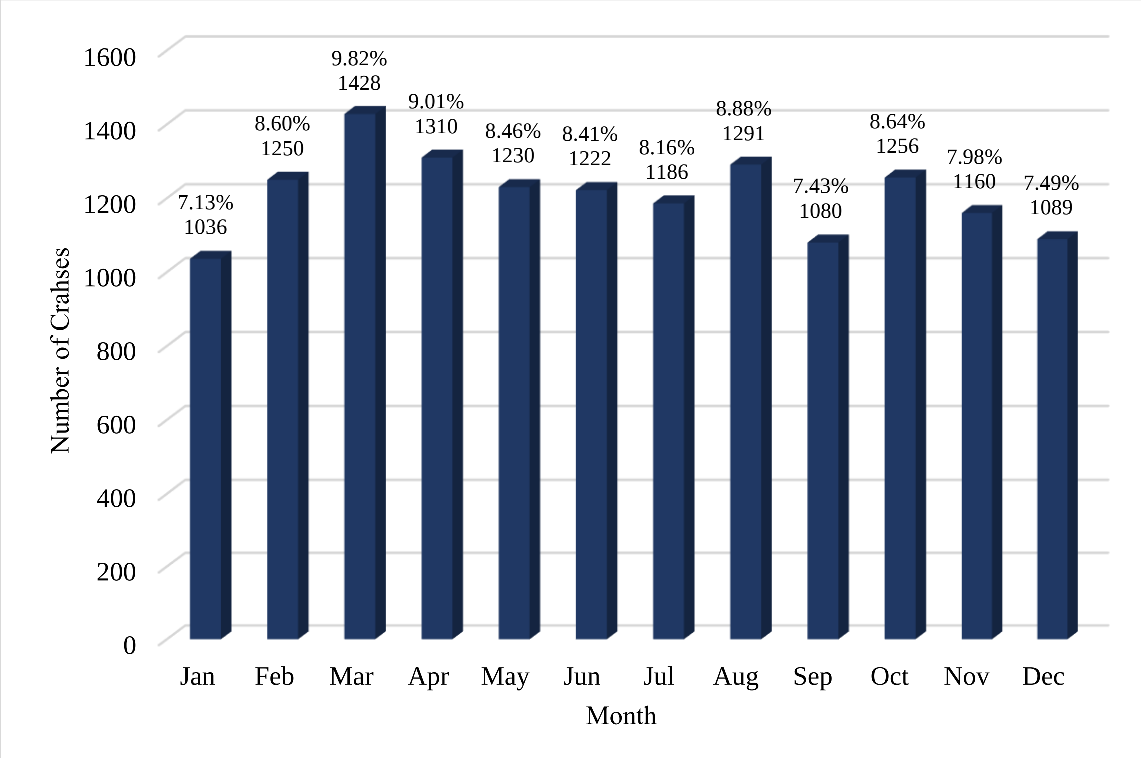


Figure 4-1 Monthly Distribution of Work Zone Crashes

The same patterns are observed in the daily distribution shown in Figure 4-2, which depict the lowest number of crashes that occurred during weekends. Tuesday, Wednesday, and Thursday were the among the days with highest number of work zone crashes in Florida in a three-year period.

The number of work zone crashes by time of day illustrated in Figure 4-3 demonstrates that morning peak (crashes occurring between 6:00–10:00 a.m.), daytime non-peak (crashes occurring between 10:00 a.m.–4:00 p.m.), afternoon peak (crashes occurring between 4:00–8:00 p.m.), and nighttime (crashes occurring between 8:00 p.m.–6:00 a.m.), consisted of 24.2%, 43.4%, 10.0%, and 23.3%, respectively.

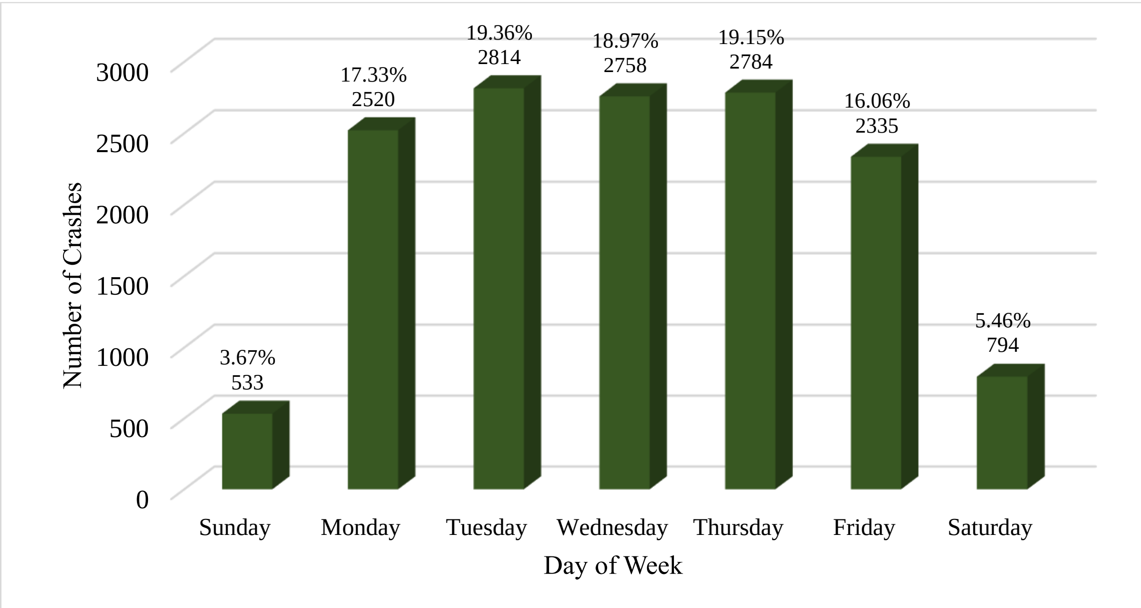


Figure 4-2 Daily Distribution of Work Zone Crashes

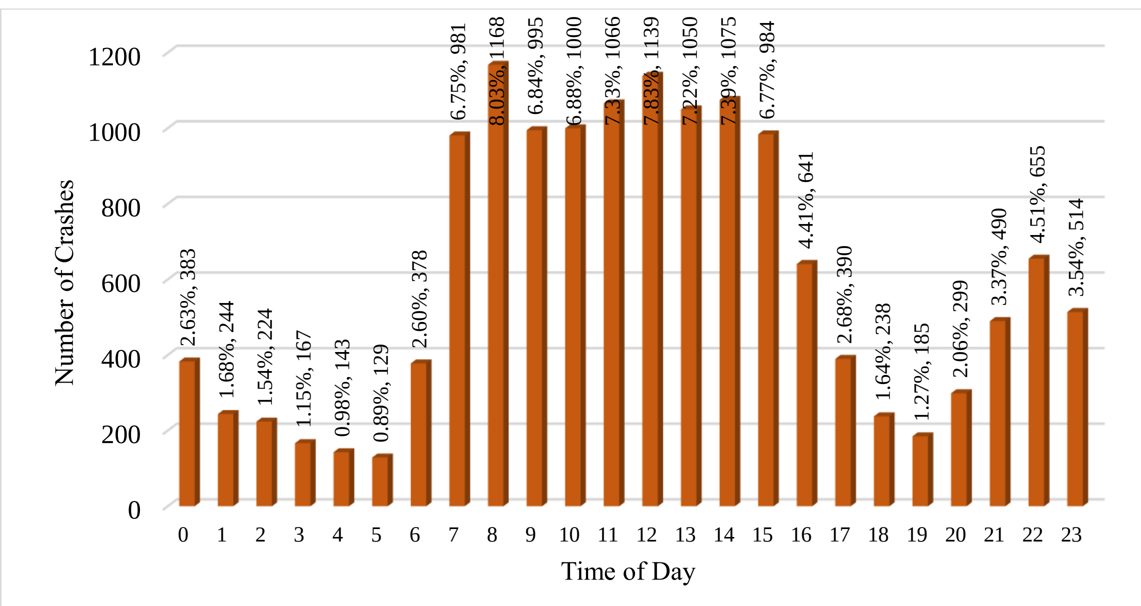


Figure 4-3 Hourly Distribution of Work Zone Crashes

This data indicated that the number of morning peak and nighttime time periods shared almost the same number of work zone crashes involving workers. This shed lights on the importance of nighttime work zones in traffic safety analysis.

4.2.2. Frequency Distribution by Crash Attributes

In line with the previous literature (Chambless *et al.* 2002, Qi *et al.* 2013, Ozturk *et al.* 2014) and as illustrated in Figure 4-4, rear-end crashes stand as the most frequent crash type in the work zone crashes involving workers by 46.4% of total crashes.

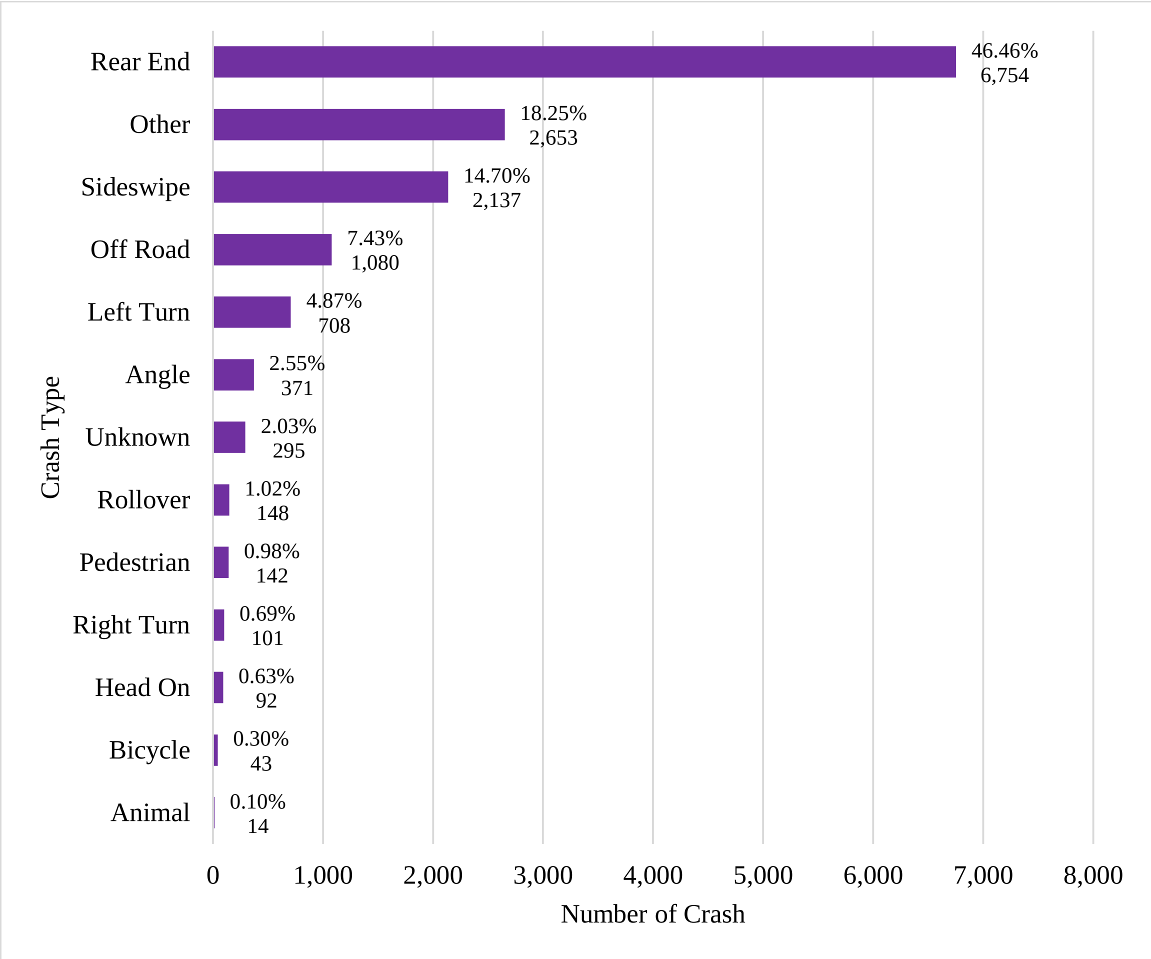


Figure 4-4 Work Zone Crashes Divided by Crash Types

The second crash attributes considered in the current study were the total number of crashes divided by the number of vehicles involved (i.e., single versus multi-vehicle), as shown in Figure 4-5. In this study, crashes with only one vehicle involved was considered a single-vehicle crash, and work zone crashes involving more than one vehicle were considered multi-vehicle work zone crashes.

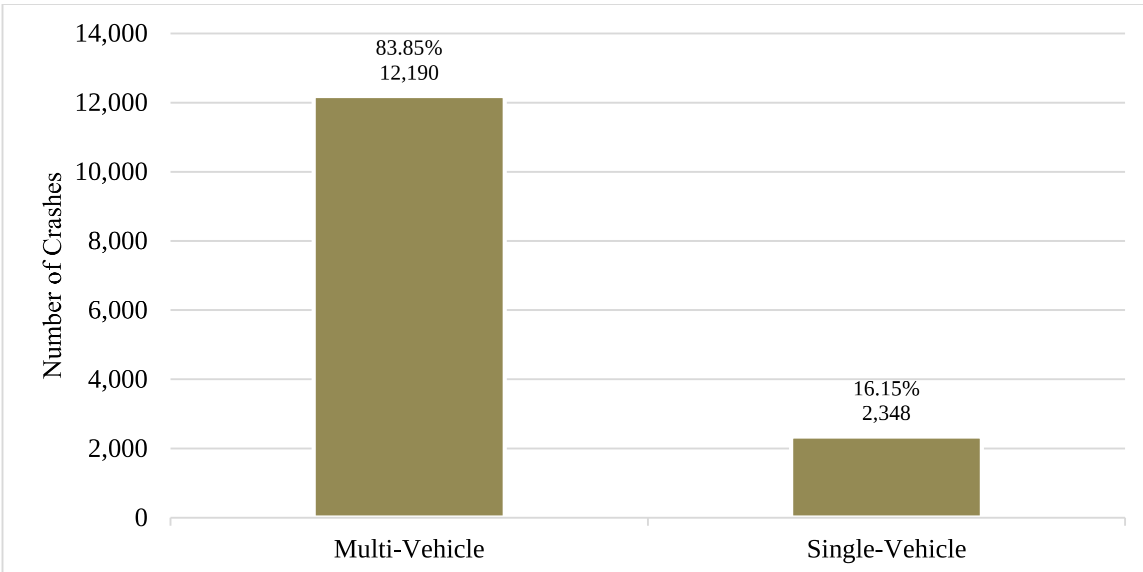


Figure 4-5 WZ Crashes Divided by Number of Vehicle Involved

Although results in the previous sections showed that single vehicle crashes were more severe than multi-vehicle crashes, the results in Figure 4-5 demonstrate that multi-vehicle crashes consisted of almost 84% of total work zone crashes involving workers, which is over five times higher than single vehicle crashes.

The impact of law enforcement and its effect on work zone crash frequency has been studied previously by Chen and Tarko in (Chen and Tarko 2012). Their results showed that police enforcement presence resulted in a 41.5% reduction in the frequency of work zone crashes. It was shown in Figure 4-6 that there is a significant difference in the number of crashes between the presence and absence of law enforcement. Over 80% of work zone crashes occurred when there was no law enforcement present.

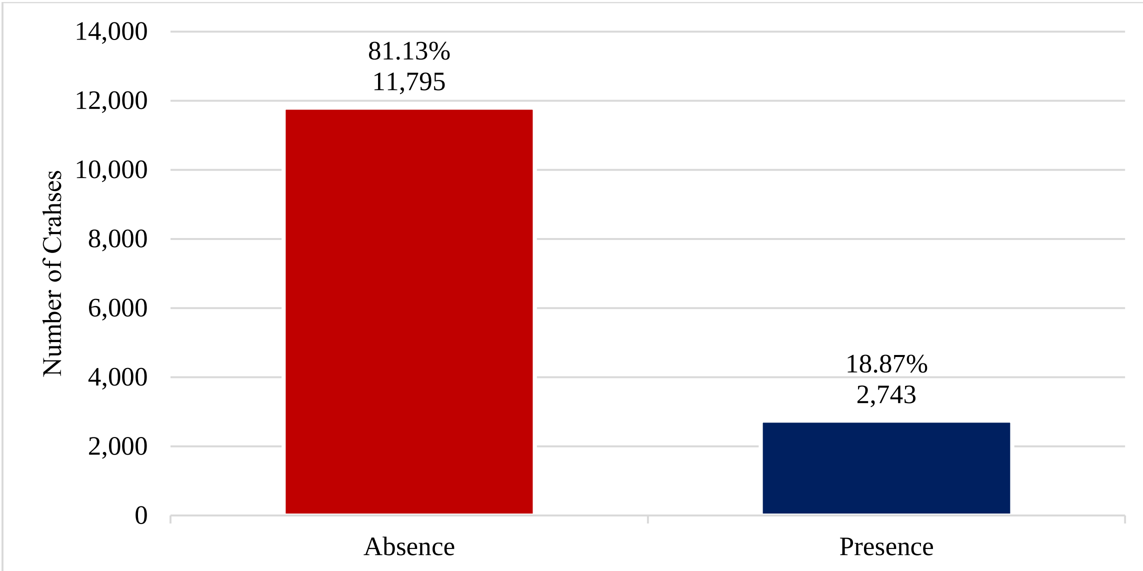


Figure 4-6 WZ Crashes Divided by Law Enforcement Presence

4.2.3. Frequency Distribution by Environmental Conditions

The variables related to environmental conditions considered in this study include weather and light conditions. The relationship between the number of work zone crashes and weather condition is shown in Figure 4-7. Due to a very low number of crashes in some weather conditions (the sum of all is less than 1%), weather conditions such as billowing sand, soil, and dirt, and fog, smog, and smoke, severe crosswinds, sleet, hail, and freezing rain, were excluded from work zone crash frequency analysis. The very low number of crashes under these types of weather conditions are expected because construction activities are prohibited in these situations. As is apparent, 74.82% of total crashes occurred during clear weather conditions, which is reasonable due to construction work activity requirements.

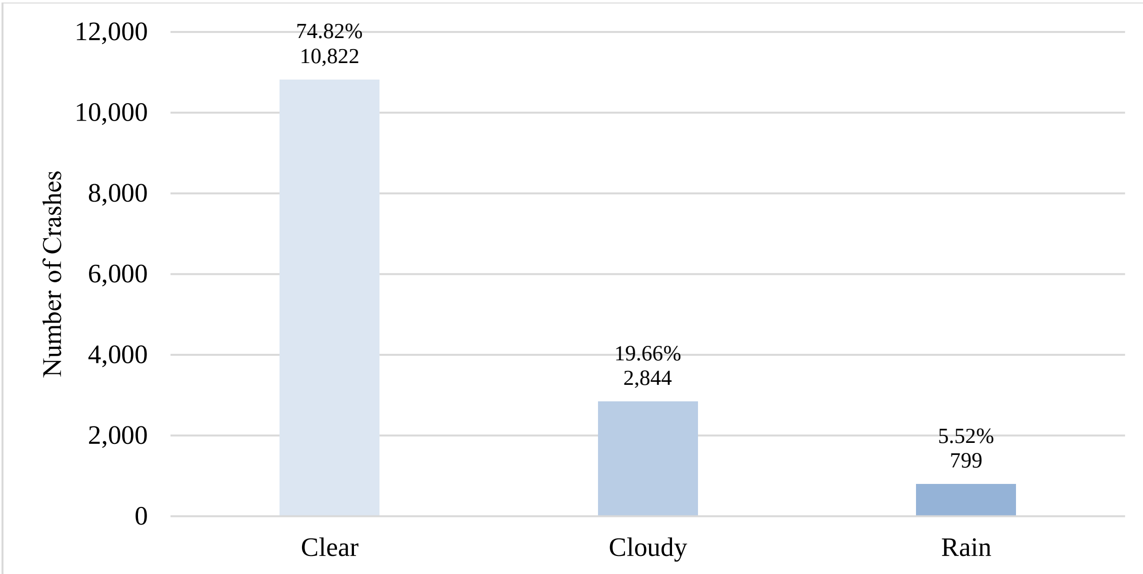


Figure 4-7 WZ Crashes Divided by Weather Condition

In addition, investigating the relationship between weather conditions and crash types, as shown in Table 4-8, demonstrates that while rear-end has the highest portion of crash type in all weather conditions, the crash ratio for all crash types decreased during clear weather to cloudy and rainy weather conditions.

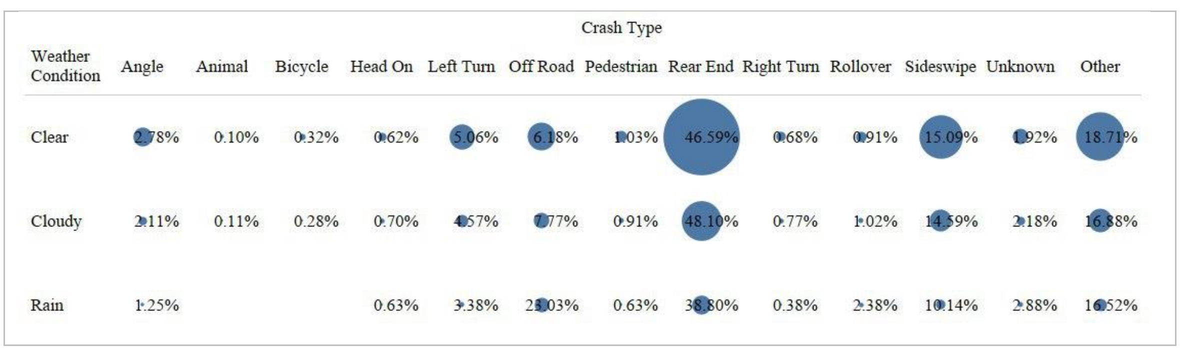


Figure 4-8 WZ Crash Type by Weather Condition

In terms of roadway surface conditions, as is apparent in Figure 4-9, over 90% of work zone crashes involving workers occurred on dry surfaces. The observed lower number of work zone crashes on wet surface conditions is reasonable due to the fact that drivers tend to be more cautious and lower their speed on wet surfaces.

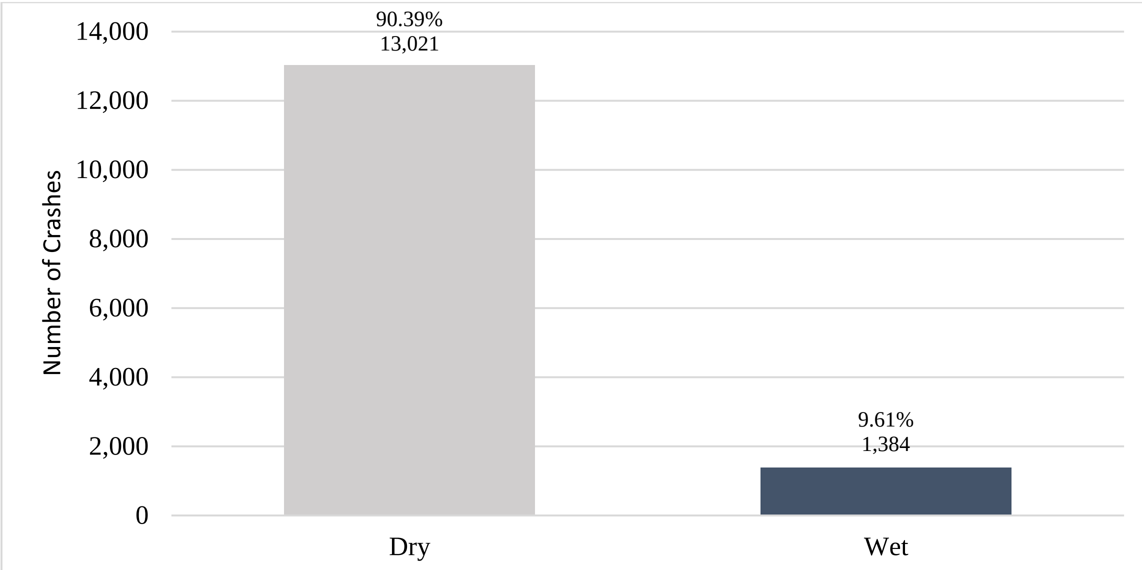


Figure 4-9 WZ Crashes Divided by Surface Condition

4.2.4. Frequency Distribution by Driver Characteristics

The total number of work zone crashes divided by DUI, as shown in Figure 4-10, shows that the number of crashes consisted of 2.2% and only 0.6% of the total work zone crash records with alcohol and drug impairment. Although the portion of DUI in the total crash records is insignificant, the previous analysis showed that alcohol consumption is significantly associated with crash severity.

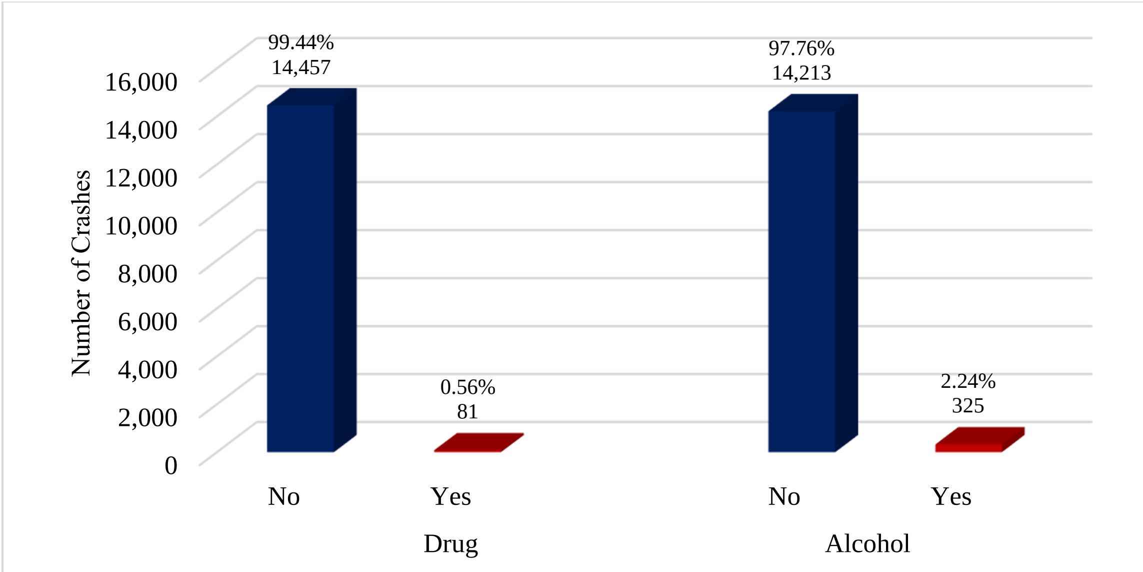


Figure 4-10 WZ Crashes Divided by DUI Condition

4.2.5. Frequency Distribution by Work Zone Characteristics

Work zone area is categorized into five main segments, as stated earlier. The crash frequency was investigated for the sub-locations and is shown in Figure 4-11. As observed, over 70% of work zone crashes involving workers were located in the activity area. This is reasonable, as most of the construction activities are placed in the activity area, and thus, more construction crews are present. The same pattern of crash frequency within work zone locations was found in previous studies such as (Garber and Zhao 2002) and (Ozturk *et al.* 2013). The transition area with 2,138 crashes stands as the second location in work zones with a share of 14.71% of total crashes. This is followed by 1,267 crashes (8.72%) in the advance warning area.

The last investigated variable in terms of crash frequency is work zone type. As shown in Figure 4-12, among the five types of work zone activities, the majority of the work zone crashes occurred at work, on a shoulder or a median (45.8%), or a lane closure

(34.1%), while intermittent or moving work zones stand as the safest work zone type by 5% of total crashes. The same statistics found in (Dias 2015) for work zone crashes occurred in Kansas.

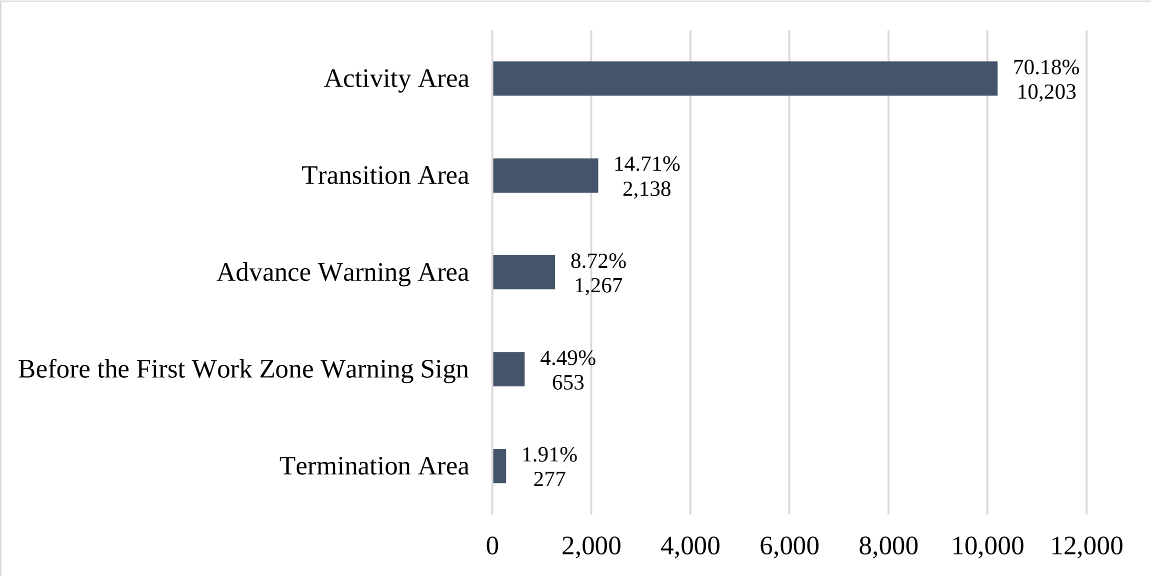


Figure 4-11 WZ Crashes Divided by Location in WZ

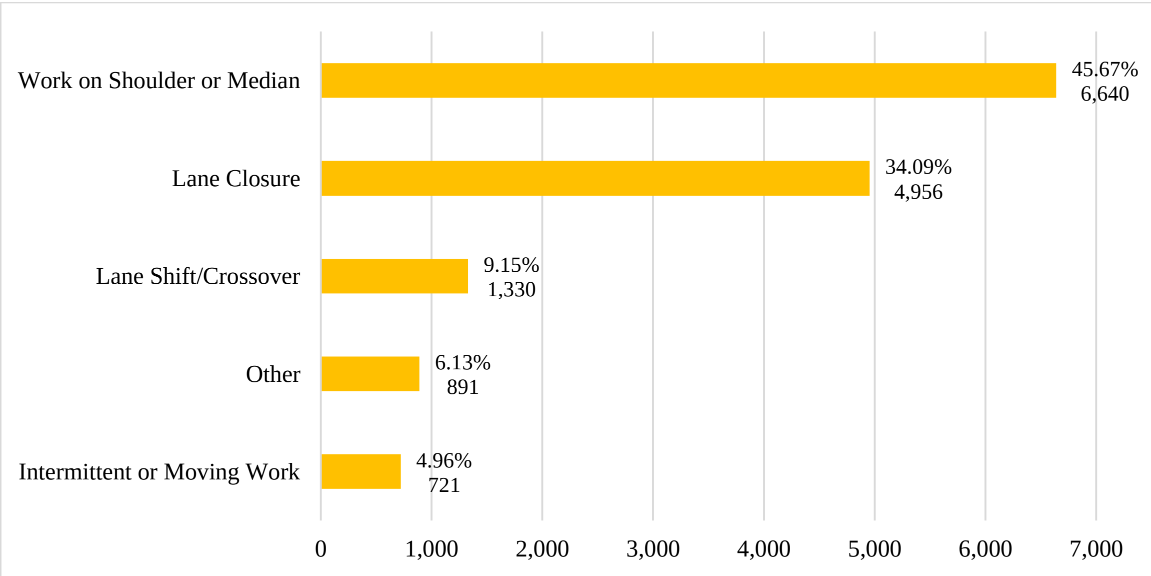


Figure 4-13 WZ Crashes Divided by WZ Type

4.3. Methodology

While the previous chapters of this study focused on the factors that contributed to work zone crash severity and were then investigated and analyzed, this section studies the variables that impact the number of crashes at work zone locations.

Among all statistical modeling approaches used to capture the relationship between the number of crashes occurring on specific roadway segments over a time period and selected contributing factors, the Negative Binomial (NB) regression model has been frequently applied. As the main feature of the NB approach, it can handle the overdispersion characteristic of crash-frequency data (i.e., when the variance exceeds the mean of the crash counts), which is commonly available in crash frequency datasets. Thus, it is applied in the current study.

The application of machine learning techniques in crash frequency analysis has been recently gained attention among traffic safety researchers. From all modeling approaches, the Support Vector Regression (SVR) model is one of most applied models. Like other non-parametric approaches and compared to traditional parametric models, it does not need a pre-assumption of data distribution and can usually provide a better statistical fit than traditional statistical models. With this in mind, an SVR modeling approach was also applied in this research to study crash frequency. Since machine learning models have always been criticized for working like a black-box, in which the impact of independent variables on the response variable cannot be explored, a sensitivity analysis was also performed to reveal the impact of selected contributing factors on work zone crash frequency.

An extensive review of crash frequency modeling and analysis was conducted by Lord and Mannering (Lord and Mannering 2010), Mannering and Bhat (Mannering and Bhat 2014), and Yang et al. (Yang *et al.* 2015). For additional details of the methodological frontier in crash frequency analysis, readers are referred to the aforementioned publications.

4.3.1. Negative Binomial (NB) regression

A Negative Binomial (NB) regression model was used in this study to model crash frequency, where the response variable is the total crash count (nonnegative integer) for a given period of time. The NB regression model is derived from the Poisson regression model. Its principle elements are shown below in Equation (4-1) (Washington *et al.* 2010):

$$P(y_i) = \frac{EXP(-\lambda_i)\lambda_i^{y_i}}{y_i!} \tag{4-1}$$

where $P(y_i)$ is the probability of work zone location i having y_i observed number of accidents per defined period of time. The λ_i is the Poisson parameter for work zone work zone location i, which is equal to work zone location $i's$ expected number of crashes as a function of explanatory variables, as shown in Equation (4-2):

$$\lambda_i = EXP(\beta X_i) \tag{4-2}$$

The Poisson regression model's interpretation is based on the implicit assumption of that the variance is equal to the mean, and thus would come with an analysis error where this assumption has been violated. The NB model however, counts for overdispersion through an error term ε_i, as expressed below in Equation (4-3):

$$\lambda_i = EXP(\beta X_i + \varepsilon_i) \tag{4-3}$$

The $EXP(\varepsilon_i)$ is a Gamma-distributed disturbance term, which allows the variance to differ from the mean in the crash frequency model, as follows (Washington *et al.* 2010):

$$VAR[y_i] = E[y_i]\big[1 + \alpha E[y_i]\big] = E[y_i] + \alpha E[y_i]^2 \tag{4-4}$$

where α refers to the overdispersion parameter and the overdispersion rate is:

$$E(Y_i) = \lambda_i; \qquad\qquad Var(y_i)/\lambda_i = 1 + \alpha\lambda_i \tag{4-5}$$

If α found to significantly differ from zero, the negative binomial model will be a good fit to be applied in such a count data (as in this study), otherwise a Poisson model should be used. The negative binomial distribution form is illustrated in Equation (4-6).

$$P(y_i) = \frac{\Gamma((1/\alpha) + y_i)}{\Gamma(1/\alpha) + y_i!} \left(\frac{1/\alpha}{(1/\alpha) + \lambda_i}\right)^{1/\alpha} + \left(\frac{\lambda_i}{(1/\alpha) + \lambda_i}\right)^{y_i} \tag{4-6}$$

where $\Gamma(\cdot)$ is a gamma function. The likelihood function is derived from the formulation (4-6), as shown in Equation (4-7) (Washington *et al.* 2010):

$$L(\lambda_i) = \Pi_i \frac{\Gamma((1/\alpha) + y_i)}{\Gamma(1/\alpha) + y_i!} (\frac{1/\alpha}{(1/\alpha) + \lambda_i})^{1/\alpha} + (\frac{\lambda_i}{(1/\alpha) + \lambda_i})^{y_i} \tag{4-7}$$

4.3.2. Support Vector Regression (SVR)

The concepts of SVM presented in Chapter 3 can also become applicable for regression problems (i.e., when the response variable is continuous).

In the context of supervised Machine Learning (ML), SVR is a generalization form of SVM, in which it attempts to estimate the mapping multivariate function from the input variables (set of explanatory variables) to a continuous-valued output variable (number of crashes at each work zone location). This generalization is accomplished by introducing the concept of ε-tube, which is an ε-insensitive region around the multivariate function (Awad and Khanna 2015). The graphical demonstration of on-dimensional SVR algorithms are shown in Figure 4-14.

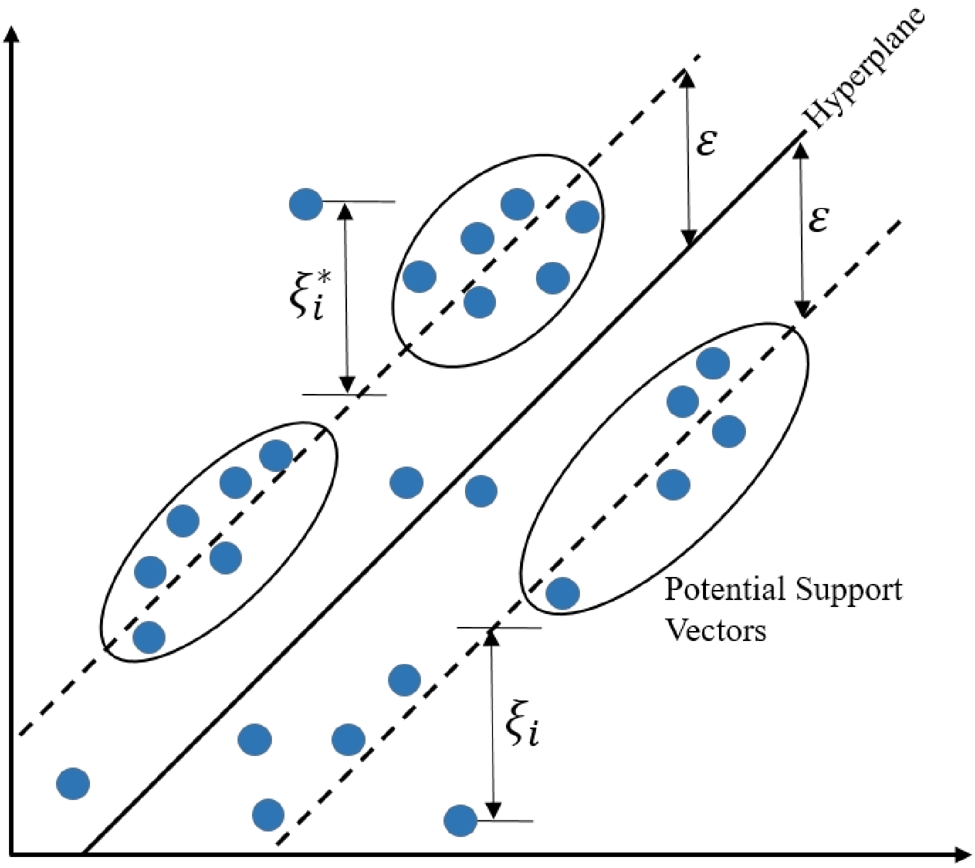


Figure 4-14 A Geometrical Perspective of a Linear SVR

Considering a regression function, as defined in Equation (4-8), that is trained on a crash dataset X, where $X = \{u_i, v_i; \quad i = 1, \ldots, n\}$ with u_i considered the crash contributing factors (input vectors), v_i is the number of observed crashes (linked targets), and w is the vector of coefficients. A function $g(u)$ will be used to illustrate the relationship between crash frequency and the number of contributing factors.

$$g(u) = w.u + b \tag{4-8}$$

In SVM regression, understanding how the algorithm model dataset is achieved occurs through the application of the loss function. Different loss functions such as linear, quadratic, and exponential have been applied to SVR models in literature; however, the standard Vapnik's ε-insensitive loss function is applied in the current study, as shown in Equation (4-9);

$$L_{\varepsilon}\big(v, g(u)\big) = \begin{cases} 0 & \text{for } |v - g(u)| \leq \varepsilon \\ |v - g(u)| - \varepsilon & \text{otherwise} \end{cases} \tag{4-9}$$

where the $L_{\varepsilon}\big(v, g(u)\big)$ demonstrates the deviation of the estimated function from the observed function (Deka 2014).

Consider the regression function presented in Equation (4-8), where $w \in X$ and X is the input space, b is the bias term ($b \in R$) that determines the margin of hyperplane from support vectors, and $(w.u)$ is a dot product of vectors w and u, and the SVR formulates the function approximation as an optimization problem. The optimization problem aimed to find the best tube (i.e., narrowest with more flatness), while minimizing the prediction error (i.e., the distance between the observed and the predicted outputs). In the optimization problem, the flatness can be achieved by minimizing the norm $\|w\|^2$, while the error/outliers is controlled through the slack variables (ξ_i and ξ_i^*) by which the deviation of the training sample outside of the ε-insensitive zone is evaluated and is penalized via the parameter C in the estimated function. The formulation of the optimization problem is shown in Equation (4-10) (Deka 2014, Awad and Khanna 2015):

$$\min_{w,b,\xi_i,\xi_i^*} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$$

$$\text{subject to} \quad \begin{aligned} & v_i - (w.u_i + b) \leq \varepsilon + \xi_i \\ & (\mathrm{w}.u_i + b) - v_i \leq \varepsilon + \xi_i \\ & \xi_i + \xi_i^* \geq 0, \quad i = 1,2,\dots,n \end{aligned} \tag{4-10}$$

As in classification problems (i.e., when the response variable is discrete), in order to deal with nonlinearity, the SVR model is also characterized by the use of kernels. Let introduce $\varphi(u)$ be a non-linear function to map u_i into a higher dimensional feature space. Thus, the decision function $g(u)$ will be reformed by:

$$g(w,b) = w.\phi(u) + b \tag{4-11}$$

Similarly, the optimization problem can be generalized for nonlinear regression problems, as shown in Equation (4-12):

$$\min_{w,b,\xi_i,\xi_i^*} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$$

$$\text{subject to} \quad \begin{aligned} & v_i - (w.\phi(u_i) + b) \leq \varepsilon + \xi_i \\ & (w.\phi(u_i) + b) - v_i \leq \varepsilon + \xi_i \\ & \xi_i + \xi_i^* \geq 0, \quad i = 1,2,\dots,n \end{aligned} \tag{4-12}$$

Finally, the estimated function of the nonlinear SVR with incorporation of the kernel function as $K\big(u(i), u(j)\big) = (\phi(u_i), \phi(u_j))$ is expressed as follows:

$$g(u) = \sum_{i,j=1}^{l} (\alpha_i^* - \alpha_i) K(u_i . u) + b \tag{4-13}$$

where α_i^* and α_i are Lagrange multipliers which lie between 0 and the value of the penalty parameter C. The nonlinear SVR with Vapnik's ε-insensitive loss function adapted from (Deka 2014) and (Yu *et al.* 2006) is shown in Figure 4-15.

The kernel function is the most significant component of SVR models by which the model prediction performance can be highly impacted. An appropriate selection of the kernel function leads the model to better transform data points from low dimensional to a higher dimensional data space. It can also better cope with potential non-linear relationship between dependent and independent variables. Different kernel functions have been developed and applied to SVR (and SVM) models in the literature, including linear, polynomial, Gaussian (or RBF), and Sigmoid kernel functions; however, as discussed in the previous chapter, RBF was applied in this study.

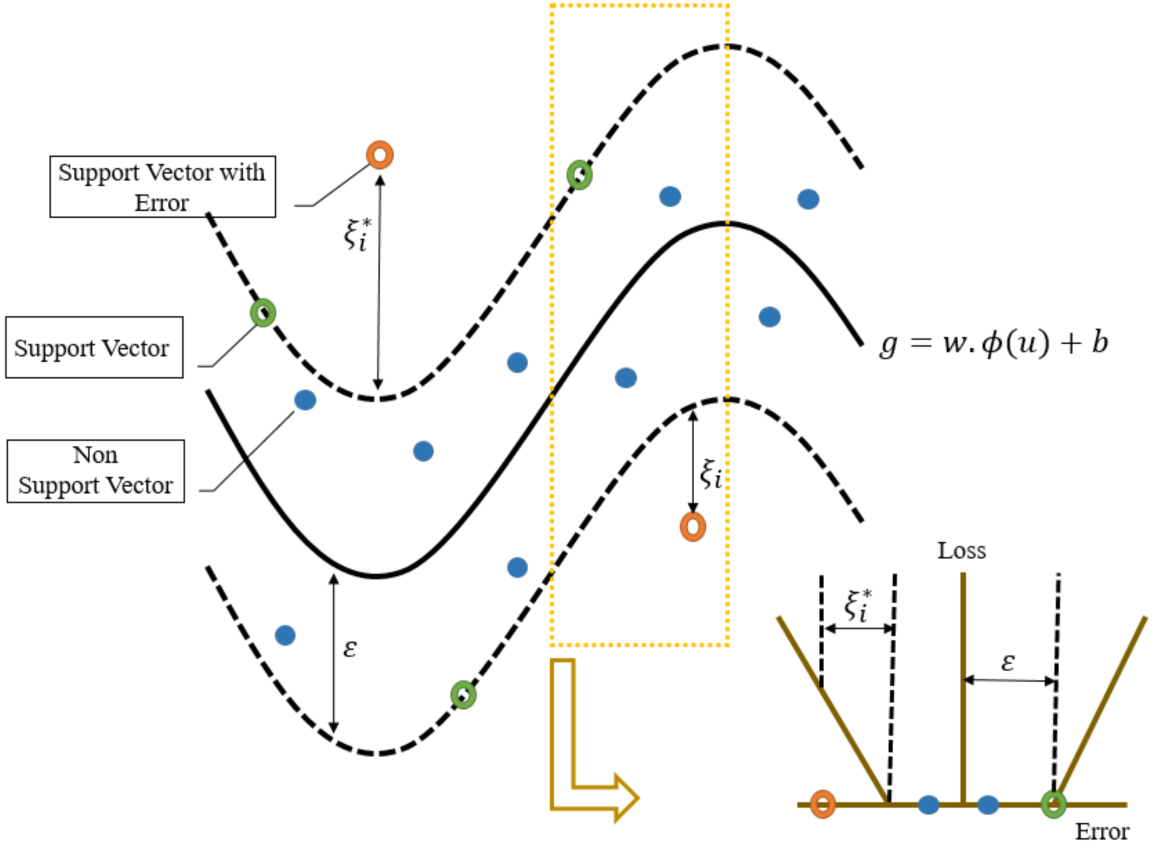


Figure 4-15 A Geometrical Perspective of a Non-Linear SVR

Like SVM models, SVR prediction performance is highly dependent on tuning the model's hyper-parameters. To this end, the Artificial Bee Colony (ABC) optimization algorithm was employed to tune SVR hyper-parameters. The parameters impact the perdition performance of a non-linear SVR model, include the following:

- *Set of support vectors*: A matrix corresponding to support vectors in the normalized data space, which attempts to find the right level of the slack variable.
- *Alpha*: A vector of weights from which the hyperplane is formed.
- *The term bias*: It allows the SVM model to pass the origin in order to come up with a separating hyperplane with the maximum margin.

4.3.3. Artificial Bee Colony (ABC)

The Artificial Bee Colony (ABC) algorithm is another swarm intelligence-based metaheuristic algorithm in the area of optimization proposed by Karaboga (Karaboga 2005). The ABC mimics the intelligent foraging behavior of honey bees. In the ABC model, the initial population of artificial bees is categorized in three main groups to execute different tasks: employed bees, onlooker bees, and scout bees. The employed bees search for food sources, evaluate the quality of the food sources and keep the locations of high-quality sources in their memory. Once they are back in their hive, they share the memorized food source information to other bees by performing a dance (i.e., waggle dance), and the higher quality the food source, the longer the dance. The onlooker bees' duties are to explore rich food sources (considering the dance time of employed bees), while the other food sources around the hive will be randomly explored by the scout bees. After each food source is explored by the employed bees and is then consumed by the onlooker bees, the scout bees begin finding new food sources by making random searches. In other words, the number of employed bees is equal to number of food sources, and each food source is considered a possible solution for the optimization problem.

In summary, the ABC model structure for an optimization problem consists of four main phases: initialization phase, employed bees phase, onlooker bees phase and scout bees phase. The explained tasks are employed in the ABC algorithm in each phase to explore the search space of an optimization problem in order to find optimum solutions.

In the first phase of the ABC algorithm, the food sources are randomly produced in the search space, through Equation (4-13), while the quality of each food source is evaluated by its fitness function in the employed bees' phase. The procedure of improving each self-solution obtained by employed bees in the second phase is illustrated in Equation (4-14) (Kiran *et al.* 2015).

$$X_i^j = X_{low}^j + r_i^j \times \left(X_{up}^j X_{low}^j\right) \quad i = 1,2,\dots,N \; and \; j = 1,2,\dots,D \tag{4-13}$$

Where, X_i is the *j*th dimension of the *i*th solution in the search space (i.e., each solution X_i is a D-dimensional vector), X_{low} and X_{up} represent the upper and lower bounds for the *j*th dimension, and r_i is a random number between 0 and 1.

$$N_i^j(t+1) = X_i^j(t) + r \times \left(X_i^j(t) - X_k^j(t)\right) \;\; i = 1,2,\dots,N, i \neq k \; and \; j \in \{1,2,\dots,D\} \tag{4-14}$$

The procedure of mimicking employed bees' behaviors to find better food sources in the vicinity of memorized food sources is illustrated in Equation (4-14). Where $X_i^j(t)$ is the food source in mind (i.e., ith solution), $N_i^j(t+1)$ is the candidate food source in the neighborhood of $X_i^j(t)$ at iteration t. $X_k^j(t)$ is the food source, which is randomly selected for the jth dimension of the ith solution. In each iteration, the newly found solution is compared to the previously found solution. If it is better, it will be memorized; otherwise, the next iteration will be abandoned.

In the second phase, the selection of food sources by onlooker bees is based on how rich the food source is (i.e., fitness value) probabilistically. The computational formula is illustrated as follows (Kiran *et al.* 2015):

$$\rho_i(t) = \frac{fit\ (t)}{\sum_{n=1}^{N} fit_n(t)} \tag{4-15}$$

Where $\rho_i(t)$ is the chance of the ith solution's selection by an onlooker bee. In the next step, the selected solution found by the employed bee will be improved by the onlooker bee, as explained in Equation (4-14). Then, the evaluation of the solution will be repeated again.

In the last phase, at the end of every cycle, the employed bee, whose solution cannot be improved after a predefined number of trials (called "limit"), converts to a scout bee. Also, the worst solution is abandoned, and the bee convert becomes a scout bee. The random search process for exploring new food sources begins again with the converted scout.

4.4. Empirical Setting and Data

A three-year monitoring period from 2015 to 2017 was carried out on a number of bridge locations in Miami-Dade County. During the period of observation, crash data were collected from the S4 crash database for crash records that were marked as work zone-related crashes. These crashes were extracted 300-350 ft. (adjusted) from upstream and downstream of bridge locations through overlaying analysis in ArcGIS tool. In addition to crash records, traffic flow condition, roadway and bridge geometric design features were also taken into account in order to create a unique crash dataset for performing crash frequency analysis.

In order to have a crash dataset that contains sufficient information in predictor variables to model crash frequency, bridge locations were selected from the point of view of incorporating a wide range of bridge geometric features, such as length, number of lanes, and span length on different road functional classifications.

Finally, 60 bridge locations that also matched FDOT's District 6 construction activities information were selected for the crash frequency analysis. The locations of the selected bridges in Miami-Dade County are marked and illustrated in Figure 4-16.

Figure 4-16 Locations of the Selected Bridges

Roadway characteristics, including functional class, number of lanes, type of road, etc., were extracted from FDOT's Roadway Characteristics Inventory (RCI) and RCI GIS shapefiles. The historical traffic flow related variables, including annual average daily traffic (AADT) and annual average daily truck volume were extracted along roadways from the Florida Traffic Online website and were averaged for the three-year period.

Considering that a bridge is a specific road segment, it may share different roadway characteristics, such as median and shoulder types/widths, compared to those available for roadway segments. This information for some specific roadway segments such as bridge locations may not be even available in the database. In addition, bridge surface width may not to be the same as traveling lanes along roadways.

With this in mind and to avoid bias, a GIS tool was applied to manually measure the geometric features of each bridge location. In addition, due to bridge construction activities, some bridge geometric features may change across the time period. Hence, using the construction activities information from FDOT's District 6, the impact of construction on bridge-specific variables were monitored via Google Earth Pro toll. Thus, the bridges that had traffic going through during the construction period was selected. An example of bridge median and surface measurements is shown in Figure 4-17, followed by construction activities on a bridge location in Figure 4-18.

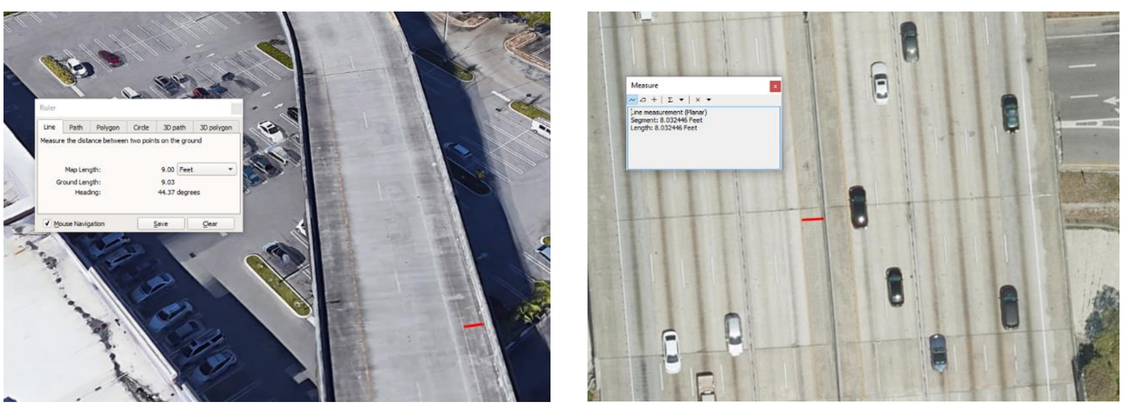
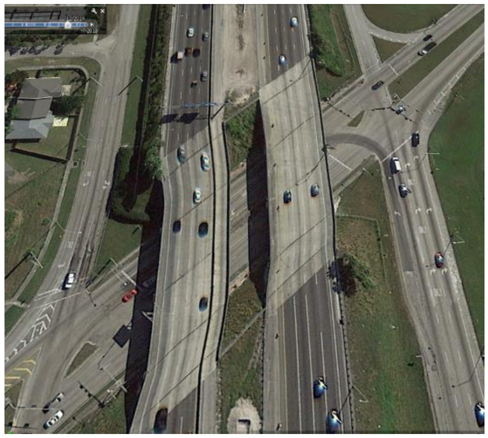


Figure 14-17 Bridge Shoulder and Median Width Measurements



12-2014 1-2016

3-2017 12-2017

Figure 4-18 A Bridge Construction Activity Over Time

Since work zone crashes in both directions were considered at bridge locations (i.e., not a directional investigation of crash frequency), the information regarding shoulder and median upstream and downstream of the bridge were excluded from further analysis.

The outside shoulders are available in the RCI database in ten categories; however, shoulder types in bridges only included five of the defined types, as illustrated in Figure 4-19. Median types were also defined in the RCI database in ten categories, but the selected bridges were categorized as six types: barrier wall, paved median, paved with barrier other than guardrail, paved with guardrail, raised traffic separator, and no median.

Curb & Gutter

Paved with Guardrail

Paved with Barrier Wall

Curb & Gutter

Paved with Warning Device

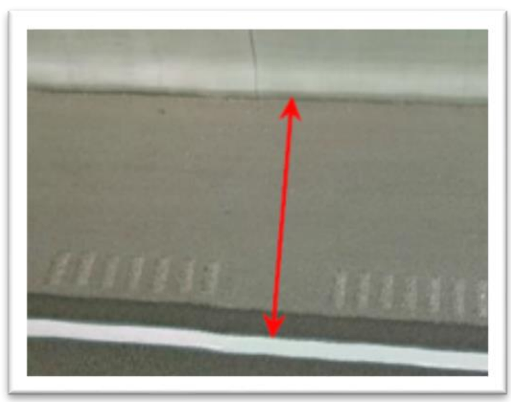
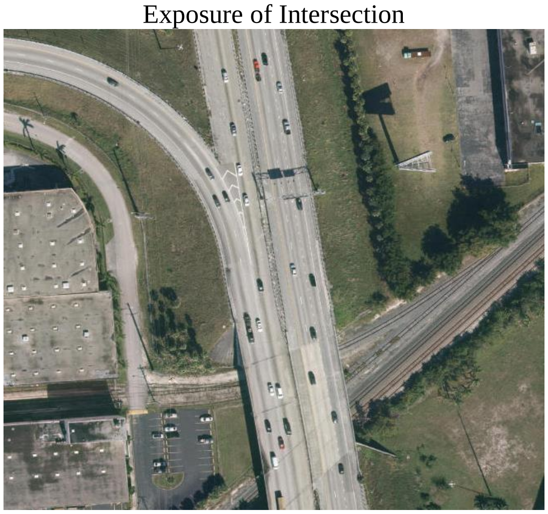


Figure 4-19 Bridge Shoulder Types

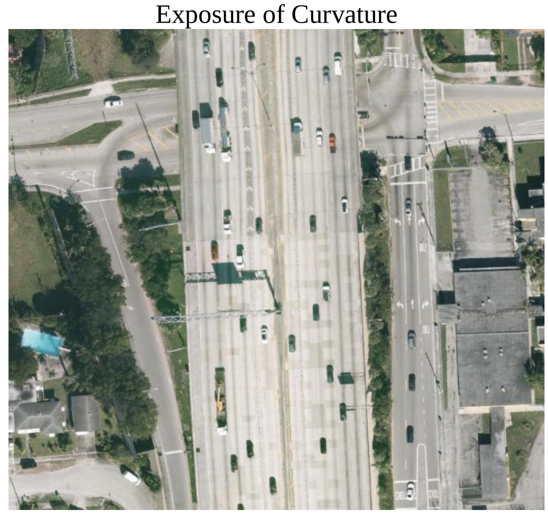
Four bridge exposure variables, including horizontal curve, ramp, intersection and express lane, were considered for the crash frequency dataset. The indicator variable "Express lane" indicates whether it is available on the bridge location, while the rest of variables were considered if they existed on the bridge location or 300-350 ft. (adjusted) upstream or downstream of bridge location. The examples of bridge exposure variables are shown in Figure 4-20.

Exposure of Intersection

Exposure of Curvature



Exposure of Ramp



Exposure of Express lane

Figure 4-20 Bridge Exposures Variables

The final variable definitions used in the frequency models are shown in Table 4-1, followed by a Summary Statistics displayed in Table 4-2.

Table 4-1 Variable Definition for Frequency Model

Variable Name	Definition	Variable Type	
WZCRSH	Number of work zone crashes occurred on the bridge deck or 300-350 upstream/downstream of the bridge	Response variable- Continuous	
BRLNG	Bridge span, measured by subtracting ending milepost and beginning milepost	Continuous	
WRKINV	Percentage of crashes involved workers	Continuous	
LWENINV	Percentage of crashes in which law enforcement were present	Continuous	
WZTYP	Type of work zone	1. Intermittent or Moving Work 2. Lane Closure 3. Lane Shift/Crossover 4. Work on Shoulder or Median	Categorical
AADT	Annual average daily traffic volume along roadway	Continuous	
TRKAADT	Annual average daily truck volume along roadway	Continuous	
NUMLN	Number of lanes on the bridge	Continuous	
HCRIND	If horizontal curve present on the bridge or 300-350 upstream/downstream of the bridge	Indicator	
RMPIND	If ramp present on the bridge or 300-350 upstream/downstream of the bridge	Indicator	
INTIND	If intersection present on the bridge or 300-350 upstream/downstream of the bridge	Indicator	
EXLIND	If express lane present on the bridge	Indicator	
PSPD	Posted speed limit	Continuous	
RODFUN	Roadway functional classification	1. Major Collector 2. Minor Arterial 3. Arterial-Freeways and Expressways 4. Arterial-Interstate 5. Arterial-Other	Categorical
BRSRWTH	Bridge surface width	Continuous	
BRMDTYP	Bridge median type	1. Barrier Wall 2. No median 3. Paved Median 4. Paved with Barrier other than Guardrail 5. Paved with Guardrail 6. Raised Traffic Separator	Categorical
BRMWTH	Bridge width of median	Continuous	
BRSHLDT	Bridge shoulder type	1. Curb & Gutter 2. Paved with Barrier Wall 3. Paved with Guardrail 4. Paved with Warning Device 5. Raised Curb	Categorical
BSHLWTH	Bridge width of shoulder	Continuous	

Table 4-2 Summary Statistics for Work Zone Crash Data

Variable	Minimum	Maximum	Mean	SD
Total Crash Records (N=60)	1	43	8.43	11.63
Bridge Length (ft)	53	2450	371.72	468.85
% of Worker Involvement	0	100	56.97	38.42
% of Law Enforcement	0	100	37.78	36.95
Work Zone Type	1	4	2.67	0.98
AADT/1000	6.4	245.7	107.75	82.81
Truck AADT/1000	0.33	35.16	6.10	6.063
Number of Lane	1	6	3.08	1.109
Horizontal Curvature Indicator (Yes=1, No=0)	0	1	0.22	0.42
Ramp Indicator (Yes=1, No=0)	0	1	0.38	0.49
Intersection Indicator (Yes=1, No=0)	0	1	0.35	0.48
Express Lane Indicator (Yes=1, No=0)	0	1	0.20	0.40
Posted Speed Limit (mph)	15	65	47.67	11.95
Road Functional Classification	1	5	3.23	1.155
Bridge Surface Width (ft)	12	72	36.43	13.36
Bridge Median Type	1	6	2.467	1.74
Bridge Median Width (ft)	0	53	8.45	9.45
Bridge Shoulder Type	1	5	2.23	0.10
Bridge Shoulder width (ft)	0	14	6.77	3.60

4.5. Model Estimation Results

In the current study, as in severity analysis, the statistical model's estimations were undertaken using NLOGIT and Econometric Software version 6, and the MATLAB R2018b programming environment was used to implement machine learning models.

Four evaluation metrics were measured in order to assess the prediction performance of the developed NB and SVR models, as well as for comparison purposes, as proposed in (Oh *et al.* 2003) and applied in previous transportation safety-related literature (Li *et al.* 2008, Gu *et al.* 2018).

The R-squared statistic (also known as the coefficient of determination) is the goodness-of-fit, which is computed from the predictions to actual values and measures how close the data are to the fitted regression line. Mean Absolute Deviation (MAD) measures how close the predictions are to the actual number of crashes. Mean Squared Error (MSE) is the average of the sum of the squares of the difference between the predicted number of crashes and the observed ones, and Root Mean Squared Error (RMSE) is simply the square root of the MSE metric.

These above-mentioned metrics are described below in Equations (4-16) to (4-19);

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4-16}$$

$$MAD = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{4-17}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{4-18}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{4-19}$$

where $\bar{y}$ stands for the average value of bridge related work zone crashes, $\hat{y}_i$ is the predicted number of crashes at bridge location i and y_i is the observed number of work zone crashes.

The model's results are presented separately in the following subsections.

4.5.1. Negative Binomial (NB) Regression Model Results

In this section, in order to prove if a negative binomial distribution is an appropriate fit for our crash data, we first checked for the presence of "overdispersion." In other words, we checked if the response variable follows the Poisson distribution (i.e., as the null hypothesis). In this regard, a non-parametric 1-sample K-S test was performed on SPSS. The results shown in Table 4-3 indicate that considering crash counts with the mean of 8.433 and the standard deviation of 11.626, the null hypothesis can be rejected with well over 95% confidence. Thus, the negative binomial is an appropriate approach for modelling the random variation of the number of work zone crashes, as there is clear evidence of overdispersion presence.

Table 4-3 One-Sample Kolmogorov-Smirnov Test

One-Sample Kolmogorov-Smirnov Test		
		Number of WZ crashes (2015-2017)
N		60
Poisson Parameter[a,b]	Mean	8.433
Most Extreme Differences	Absolute	.556
	Positive	.556
	Negative	-.187
Kolmogorov-Smirnov Z		4.306
Asymp. Sig. (2-tailed)		.000
a. Test distribution is Poisson. b. Calculated from data		

The estimated parameters from the NB modeling results are shown in Table 4-4 and were used to investigate the relationship between the contributing factors as independent variables and crash frequency of bridge-related work zone crashes. The comparison between actual crash frequencies and predictions of the NB model on the basis of bridge locations is shown in Figure 4-21.

The McFadden Pseudo R-squared value of the 0.08 form model summary in Table 4-4 indicates a reasonable model fit and is based on the Chi squared value. The model is significant at 1% confidence level. The positive value of Alpha, which is significant with well over 99% confidence level, also demonstrates that the data is overdispersed. As for interpretations of the results, a positive sign of the estimated parameters implies increased crash frequency with an increase in the value of the independent variable. For a one-unit change in the predictor variable, the difference in the logs of expected counts of the response variable is expected to change by the respective regression coefficient, given the other predictor variables in the model are held constant[3].

Table 4-4 Results of NB Model and Marginal Effects

Variable	Coefficient	Std. Error	*t*-Statistic	Marginal Effect
BRMDTYP4|	1.081**	0.497	2.18	12.439
LWENINV|	-1.028**	0.504	-2.04	-8.790
BRSRWTH|	0.029***	0.010	2.87	0.247
HCRIND|	-0.892**	0.434	-2.06	-5.612
BRMDTYP6|	-0.665*	0.402	-1.65	-4.315
Constant	1.267***	0.397	3.19	
α	0.782***	0.295	2.65	
Summary statistics				
Number of observations	60			
Log-Likelihood at Convergence	-175.691			
Log-Likelihood at Zero	-190.529			
McFadden Pseudo R-squared	0.08			
R-squared	0.180			
MAD	5.521			
MSE	50.495			
RMSE	7.105			

****, **, *, are Significance at 99%, 95%, 90% confidence levels*

[3] For more detail on model explanation please see "NEGATIVE BINOMIAL REGRESSION | STATA ANNOTATED OUTPUT" https://stats.idre.ucla.edu/stata/output/negative-binomial-regression/

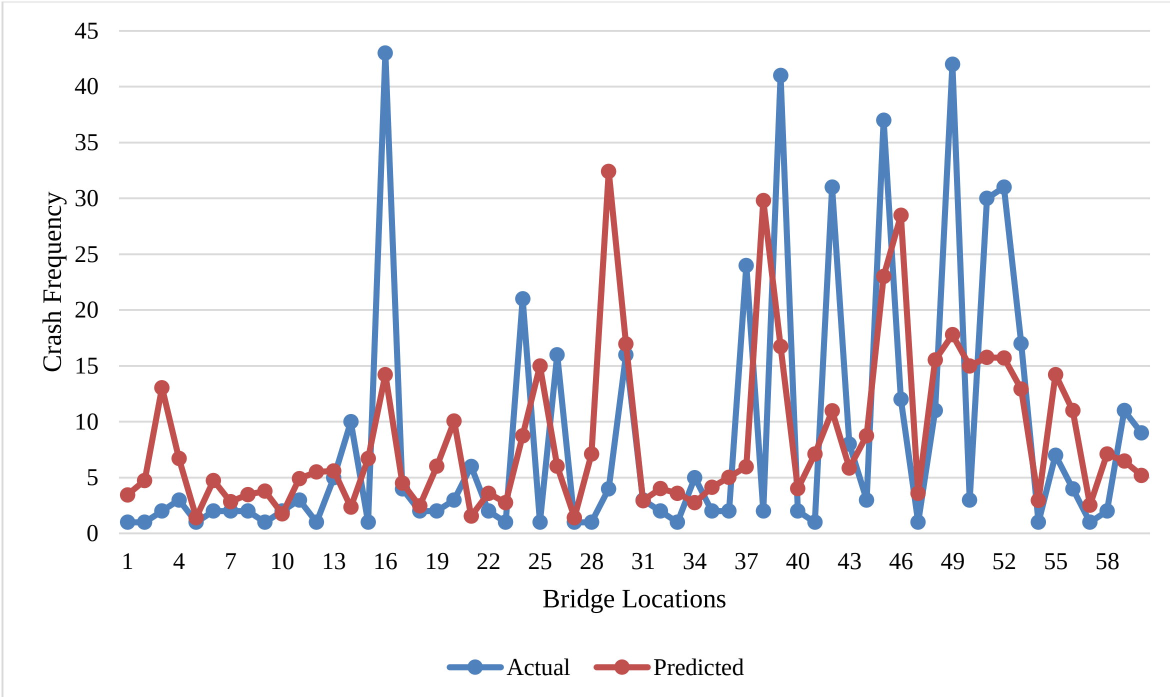


Figure 4-21 NB Model, Actual Vs. Predicted Crash Frequency

4.5.2. Support Vector Regression (SVR) Model Results

In this study, as it was mentioned earlier, the SVR model with the RBF kernel function was implemented in the MATLAB R2018b programming environment.

Three different data splits of 60, 70, and 80 for training and testing sets were considered first to assess the SVR model prediction's abilities, as well as to select the model to be trained (i.e., in order to tune the algorithm hyper parameters) by ABC algorithm. The selected data set was then used for comparison purposes with the NB model, as well as to implement the sensitivity analysis in order to explore the impacts of contributing factors on bridge-related crash frequency.

To this end, the entire dataset was randomly separated into three sub-datasets as stated, and the R-squared statistic (coefficient of determination) was considered a criterion to select the initial model. The results are illustrated in Figure 4-22.

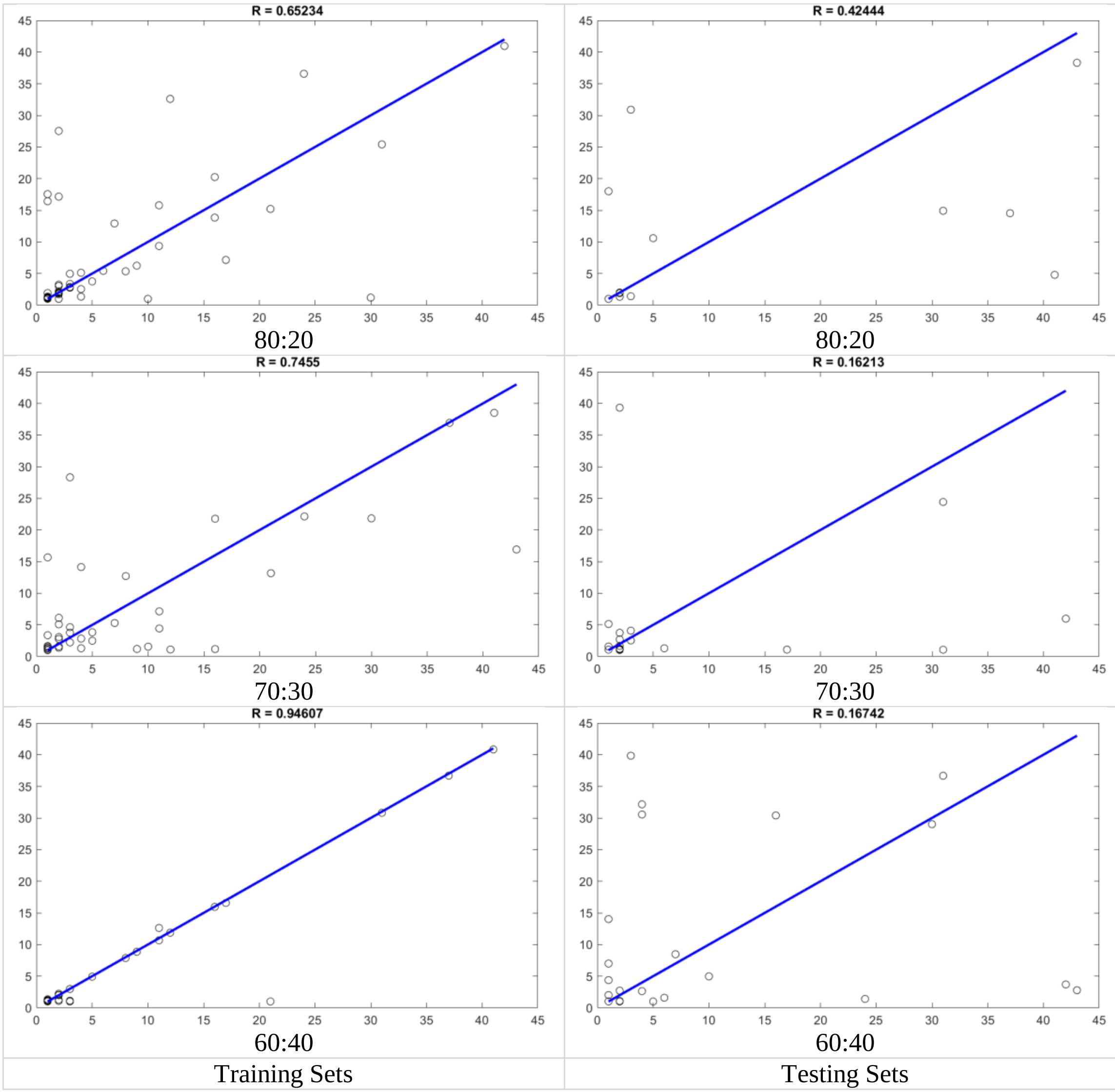


Figure 4-22 R-Squared Statistics on Different Data Splits

Preliminary performance test results reveal that the SVR model with the split of 8:2 performed better than the other data splits (i.e., higher R-squared value). Thus, these results were considered for further model prediction performance improvement through the application of ABC metaheuristic optimization in parameter tuning.

As explained in the previous chapter for crash severity analysis, the performance of metaheuristic algorithms is also considerably influenced by the proper tuning of parameters, and a Taguchi's robust design method was used to obtain the best parameters of the ABC algorithm (for detailed information regarding the Taguchi method, readers are referred to (Peace 1993)). In performing the Taguchi test, all of the combinations of parameter settings were examined, along with the best achieved combination considering the S/N ratio plot. Finally, a number of 1,000 iterations, colony size of 100, and limit of search and scouts of 10 were utilized when using ABC algorithms to train the SVR model. The prediction results of SVR models, which were implemented on the entire dataset, is summarized in Table 4-5, and the output of ABC-SVR is provided in Figure 4-23.

Table 4-5 Results of SVR Models

Model	R-squared	MAD	MSE	RMSE
SVR	0.324	5.666	106.117	10.301
ABC-SVR	0.542	4.249	64.086	8.005

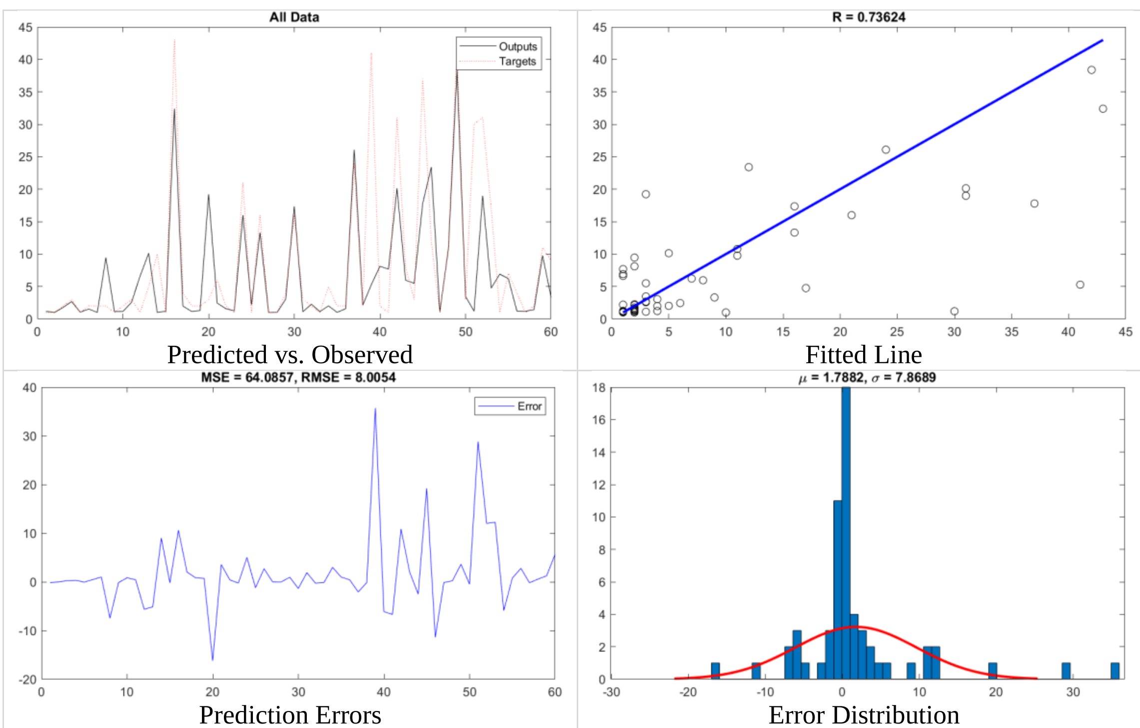


Figure 4-23 ABC-SVR Model Output

Since the contributing factors in a crash dataset comes with different units and magnitudes, and the SVR model basically works with the geometric features of data points, for the purpose of performing sensitivity analysis, the data for each variable needed to be normalized. This may also improve the prediction performance of the SVR model. With this in mind, the final ABC-SVR model was fitted again with a normalized dataset to extract the impact of contributing factors on bridge-related work zone crash frequency. The data normalization was accomplished through the following equation (Li *et al.* 2008):

$$xn_i = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} \qquad (4\text{-}20)$$

where x_i is the vector representing the independent variables.

As in other machine learning models, SVR works like a black-box, in which the impact of explanatory variables has not been understood, as they do not have a specific functional form like traditional statistical models. To this end, the method originally proposed by Fish and Blodgett in (Fish and Blodgett 2003), was conducted on the ABC-SVR to explore the impacts of each explanatory variable on work zone crash frequency.

The sensitivity analysis consisted of recording variation from the response variable (crash frequency) for different values of independent variables (crash contributing factors), one at the time. These variations, which are also normalized, are for a continuous variable that lies between its mean and standard deviation (plus and minus), within a reasonable interval. Categorical variables vary among all of the categories, except for the reference variable, since it keeps all other variables unchanged.

The sensitivity results for the variables were found to be statistically significant in the NB model, including bridge median type, presence of law enforcement, bridge surface width, and horizontal curve, which is shown in Figure 4-24.

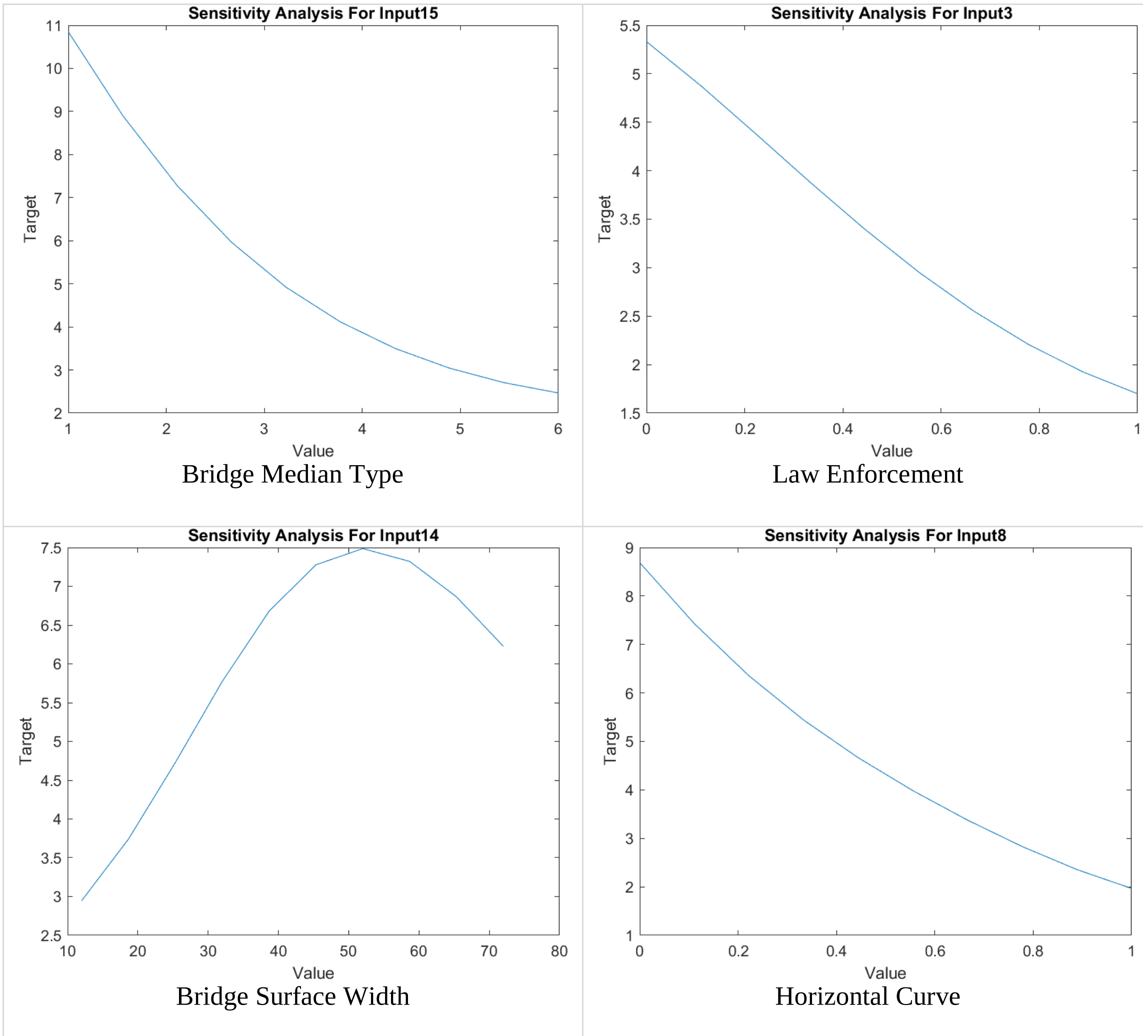


Figure 4-24 ABC-SVR Sensitivity Analysis

4.6. Discussions

Out of a total of 34 indicator variables, the impact of five variables were found to significantly affect the work zone crashes that occurred in bridge locations in the NB model. The coefficient of the variables that were not found statistically significant was eliminated in the model output. As presented in Table 4-4, the greater the coefficient, the higher the likelihood of having a work zone crash at bridge locations. In addition, the marginal effect was also computed to approximate how much crash frequency is expected to change (either decrease or increase) for a unit change of the explanatory variables.

The interpretation of the obtained results from both modeling approaches are discussed below:

As for bridge median type indicators, a paved median with a barrier other than a guardrail is associated with a higher probability of crash occurrence at bridge locations. The positive coefficient value indicates that the crash risk on a paved median with barriers is higher than other types of bridge median by 12.439. It was also shown that raised traffic separators on the median results in a 4.315 decrease in the likelihood of crash occurrence at bridge locations. The results from the ABC-SVR sensitivity analysis also shows that the probability of median type 4 is 56.91% higher than type 6, and 3.85 compared to 2.466. In other words, bridges with concrete barrier medians are more probable to results in a crash than a bridge with guardrail barriers and raised traffic separators. This is consistent with the results found in (Montella 2010), which demonstrated that concrete barriers, when compared with steel barriers, were more expected to result in severe crashes. In another study, it was shown that 91.98% of cross medians and median barrier crashes occurred on the roadways with concrete barriers, while just 1.08% of crashes occurred on roadways with guardrails (Chitturi *et al.* 2011).

As found in work zone crash severity models, law enforcement was found to have statistically significant contributing factors to crash frequency. According to the negative binomial results, a marginal effect of -8.790 indicated that law enforcement presence had a significant impact in preventing work zone crashes on bridge locations. The ABC-SVR model results also demonstrate that, by one unit change in the law enforcement variable (i.e., going from presence to absence of law enforcement), the mean predicted probability of a crash occurring at a bridge location will increase by 211.76%, or from 1.7 to 5.3. This result is in line with the previous studies that revealed the positive impact of law enforcement on crash reduction, including work zone crashes (Chen and Tarko 2012), alcohol-impaired driving crashes (Fell *et al.* 2014), motor-vehicle crashes (Redelmeier *et al.* 2003), etc.

The horizontal curve indicator as a road character at upstream or downstream of the bridge was found to be statistically significant in the NB model. The marginal effect shows that 5.612 is less likely to result in a bridge-related work zone crash. In the study conducted by Eftekharzadeh and Khodabakhshi (Eftekharzadeh and Khodabakhshi 2014) and Khoury et al. (Khoury *et al.* 2019), it was shown that drivers tend to reduce their traveling speed when driving on a horizontal curve. This may result in having more control over the vehicle, thus reducing the number of accidents at locations with horizontal curves. The ABC-SVR variable impact results also show that the presence of a horizontal curve results in a 341.69% decrease in the probability of crash occurrence.

Bridge surface width is the last variable found to have a statistically significant impact on crash frequency, indicating that with the increase of surface width, a higher number of crashes can be expected. It should be noted that although surface width and number of lanes are highly correlated in many cases, they are different in some cases, including if the bridges are located at intersections or a ramp exists on the bridge. The marginal effect indicates that for a one-unit change in surface width, a crash is more probable to occur by 0.247. A mixed effect was found in the literature that investigated the impact of road surface width on crash frequency. For example, it was found in (Qin 2012) and (Ma *et al.* 2008) that surface width contributes to a fewer number of crashes; however, a nonlinear relationship was found in (Das and Abdel-Aty 2011).

The ABC-SVR results demonstrated that the relationship between crash frequency and surface width has a quadratic functional form, which is consistent with the findings of (Das and Abdel-Aty 2011). It is interesting to observe that bridge crash counts reach a maximum when the surface width is approximately 50 ft. A higher or lower value of surface width results in a decrease in the number of crashes on bridge locations.

4.7. Summary and Conclusions

This chapter investigated the relationship between work zone presence on bridge locations and crash occurrence. A detailed descriptive analysis was performed for the total work zone crashes in Florida, while statistical and machine learning approaches were utilized to model crash frequency at 60 bridge locations in Miami-Dade County.

According to the results from the descriptive analysis, it was found that the number of work zone crashes and worker-involved work zone crashes experienced an 11% and 7% increase from 2015 to 2017, respectively. In addition, most of the work zone crashes occurred between 7:00 a.m. to 3:00 p.m.

From a modeling perspective, a comparison between the prediction performance of the SVR and NB models showed that SVR predicted bridge crashes more effectively and accurately than traditional NB models. An interesting finding of this study was that a nonlinear relationship was observed from the ABC-SVR results for the bridge surface width. This cannot be captured through any conventional statistical models like the NB, as they are restricted on the linear relationship between crash frequency and explanatory variables. This is a fundamental limitation of such statistical models, which makes machine learning models a more promising tool for modeling crash frequency.

CHAPTER 5
BENEFIT-COST ANALYSIS OF ABC IMPLEMENTING

5.1. Introduction

Roadway safety benefit-cost analysis is a critical component used to enhance traffic safety on transportation networks. Work zones are essential components of highway renovation, technological upgrading, and maintaining and improving roadway systems. This may, however, have negative impacts, such as a decrease in roadway capacity, an increase in traffic congestion, and a new set of traffic safety concerns.

As defined in the Highway Capacity Manual (HCM) (HCM 2010), work zone is a segment of highway that impinges on the number of traveling lanes as a result of construction, maintenance, or utility work activities. Work zone road user costs (WZ RUC) is defined as "the additional costs borne by motorists and the community at-large as a result of work zone activity (Mallela and Sadavisam 2011)." Different costs are associated with work zone presence, which can be mainly divided into the mobility, safety, and reliability categories. There are costs from these categories that have monetized impacts, such as: costs associated with travel delay, crash costs, vehicle operational cost (VOC), emission costs, and the impacts of nearby projects, each of which need to be taken into account for work zone design and implementation plans.

Bridge construction is defined as long-term stationary work zones that can result in either increasing the risk of being in a crash or being in a more severe crash. From the perspective of all stake holders, the timely completion of a construction project is of great importance. Any inadvertent delays in projects can increase the cost exponentially and cause nuisance to the public. A reappraisal of factors which affect the on-site construction time mostly includes the traditional methodologies which causes delays in planning and scheduling of projects. Although such methodologies have been successfully implemented for infrastructure projects for decades, new techniques have been developed which can reduce the construction time. One such methodology is the Accelerated Bridge Construction (ABC) which provides a framework for fast delivery of the projects. Generally, it uses precast elements of the bridge fabricated on site or away, moved to the bridge location and installed in place (Farhangdoust and Mehrabi 2019). Besides the expedited construction, the ABC also reduces the labor man-hours which helps improve worker safety. From its inception, the confidence in ABC techniques have improved over the years as a result of successful implementation and improved performance of ABC built bridges.

The new developments in ABC are a result of extensive experimental and analytical studies on the performance of these techniques. The component and full-scale testing of various ABC components have revealed emulative performance to field cast construction (Sadeghnejad *et al.* 2019, Farhangdoust and Mehrabi 2020, Sadeghnejad *et al.* 2020). These improvements are mainly attributed to the use of materials, such as UHPC, which have superior material and mechanical properties. In many instances, the ABC methods exceed the performance metrics as prescribed in the specifications. A number of ABC techniques have been used in the construction industry including design of connections, development of new structural members and repair of existing bridge components (Azizinamini *et al.* 2019, Rehmat *et al.* 2019a, b, Farzad *et al.* 2020). As a result, industry professionals and transportation officials are in the process of incorporating ABC codes to their design and construction specification. These codes will encourage industry usage and improve confidence level in ABC.

As an innovative construction method, ABC dramatically decreases on-site construction duration, and thus, may also have roadway safety benefits (i.e., can be considered a highway safety improvement project) (Mehrabi and Farhangdoust 2019). To determine the economic benefits of its safety improvements, crash costs can be utilized to quantify the impacts of crashes reduced by ABC implementation. To this end, and within the context of this study, the WZ RUC computation process is based on the assessment of the monetized components of crash costs resulting from work zone activities at bridge locations. This issue will be investigated in this chapter in the following sections.

5.2. Work Zone Crash Cost

Crash costs are most often reported by crash severity, which is basically reported using injury scales such as KABCO, as explained in the previous section. There are several methodologies employed to calculate the unit cost of crashes, which are not limited to, but include: crash costs by KABCO, injury scale translators, costs by crash type, estimates for cost components, and so forth (Harmon *et al.* 2018). Although there are differences between other approaches and each method provides some pros and cons over the others, there is not necessarily a preferred method.

In the context of work zone safety, the associated crash costs is a function of the expected change in crash rate/frequency due to the presence of work zones (Mallela and Sadavisam 2011). Considering the data limitations of the current study, in which the beginning and ending date/time of work zone activities (i.e., the exact duration of work zone presence) and work zone length were not available, an annual average of a three-year crash cost estimation was performed. In addition, since the ABC implementation aims to end the work zone activity, this study seeks to evaluate the crash costs (frequency and severity) associated work zone presence, regardless of work zone type and durations. The estimated costs considered in this study include:

1) Estimated vehicle damage, including the property and vehicle damages recorded in the crash report.
2) Cost per equivalent KABCO crash severity level using the Florida crash cost method based on the 1994 and 2013 USDOT guidance with state-specific adaptations.

According to the Model Minimum Uniform Crash Criteria (MMUCC) (NHTSA 2011) and Federal Highway Administration's (FHWA) KABCO Injury Classification Scale and Definitions for Florida (FHWA), the KABCO scaled crash injury definitions are as follows:

- **Fatal Injury (K):** Stands for any injury that results in a death within a 30-day period after the crash occurred.
- **Incapacitating Injury (A):** Stands for a serious injury other than a fatality, such as disabling injuries including broken bones, severed limbs, etc. These injuries usually require hospitalization and transport to a medical facility.
- **Non-incapacitating Evident Injury (B):** Stands for minor injury and non-disabling injuries that are evident at the scene of the crash, such as lacerations, scrapes, bruises, etc.

- **Possible Injury (C):** Stands for any injury reported or claimed, which is not a fatal, incapacitating (serious injury), or non-incapacitating (minor injury).
- **No Injury/PDO (O):** Stands for a situation in which a person received any bodily harm from the motor vehicle crash.

The abovementioned severity scales will be used in the following section in this chapter to convert and estimate the monetary value of different levels of crash severity for crash cost analysis.

5.2.1. Descriptive Statistics of Work Zone Crash Costs

Work zone–related crash records from the Florida Signal Four Analytics tool (S4A 2018) database were extracted, as well as the overlaying of the bridge locations information from the ArcGIS tool as the input for crash cost estimation. Since the Florida method for calculating crash unit cost is based on the KABCO crash severity scale, severity levels were converted to KABCO scales (which will be discussed in the following section). The detailed statistics of the number of participants in each of the five levels in the KABCO scale for work zone crashes occurred in the 60 bridge locations in Miami-Dade County, as shown in Figure 5-1.

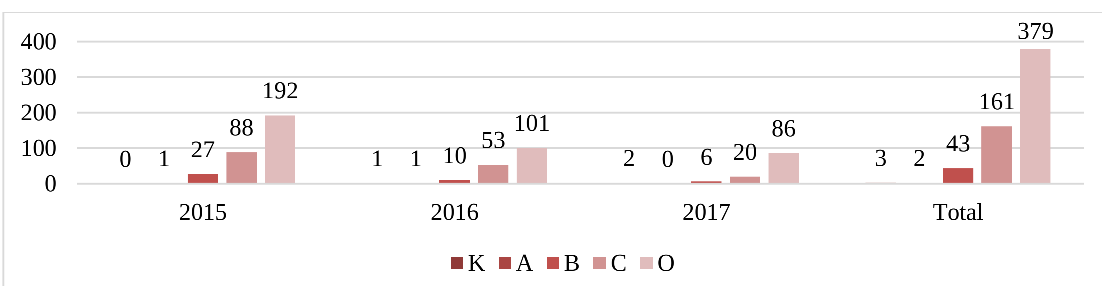


Figure 5-1 Number of Crash Participants in KABCO Scale

In the crash database, the estimated damage is specified as the monetary value of damaged properties and vehicles in the crash. The annual distribution of estimated damage is shown in Figure 5-2, followed by the estimated damage by work zone type in Figure 5-3.

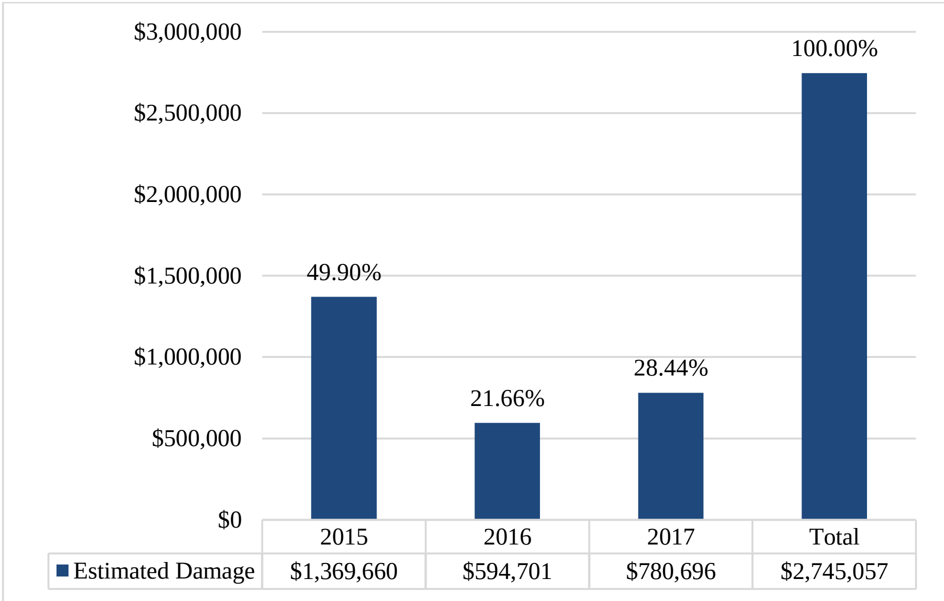


Figure 5-2 Annual Distribution of Estimated Damage

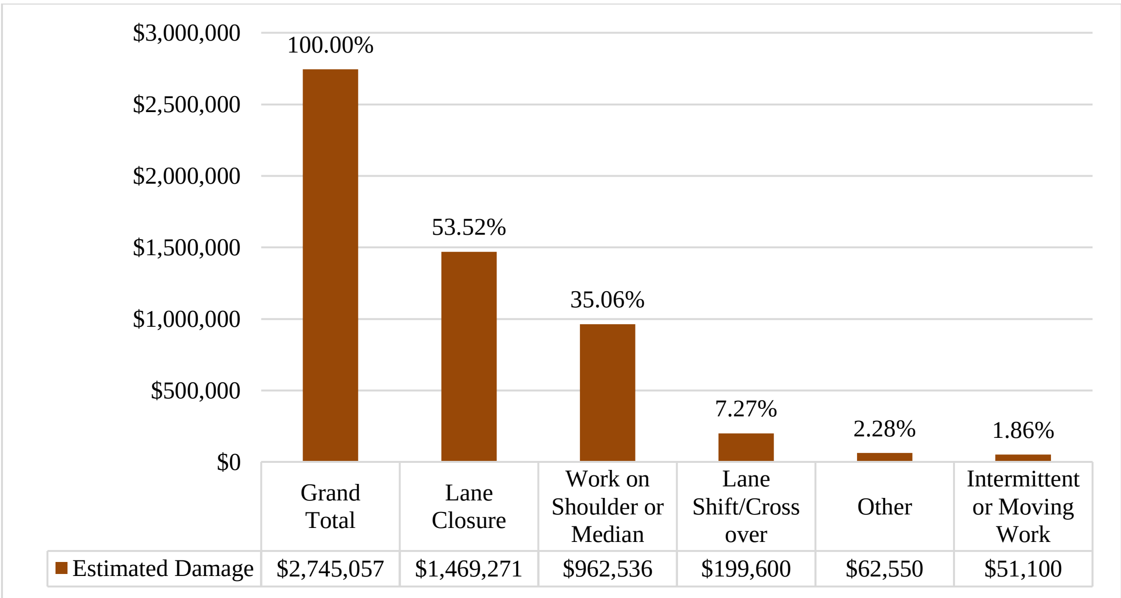


Figure 5-3 Distribution of Estimated Damage by WZ Type

As shown in Figure 5-2, estimated vehicle and property damages in 2015 is 130% and 75.45% higher than that in 2016 and 2017, respectively. Also, as demonstrated in Figure 5-3, lane closure and work on shoulder or median types of work zone stands among the most destructive work zone types by 53.52% and 35.06% of the total amount of estimated damages. In addition, Figure 5-4 demonstrates that the activity area in work zones resulted in higher crash costs. This is quite intuitive since the construction machinery is located in the activity area, and thus, may result in higher property damage costs and vehicle damages than other locations in the work zone.

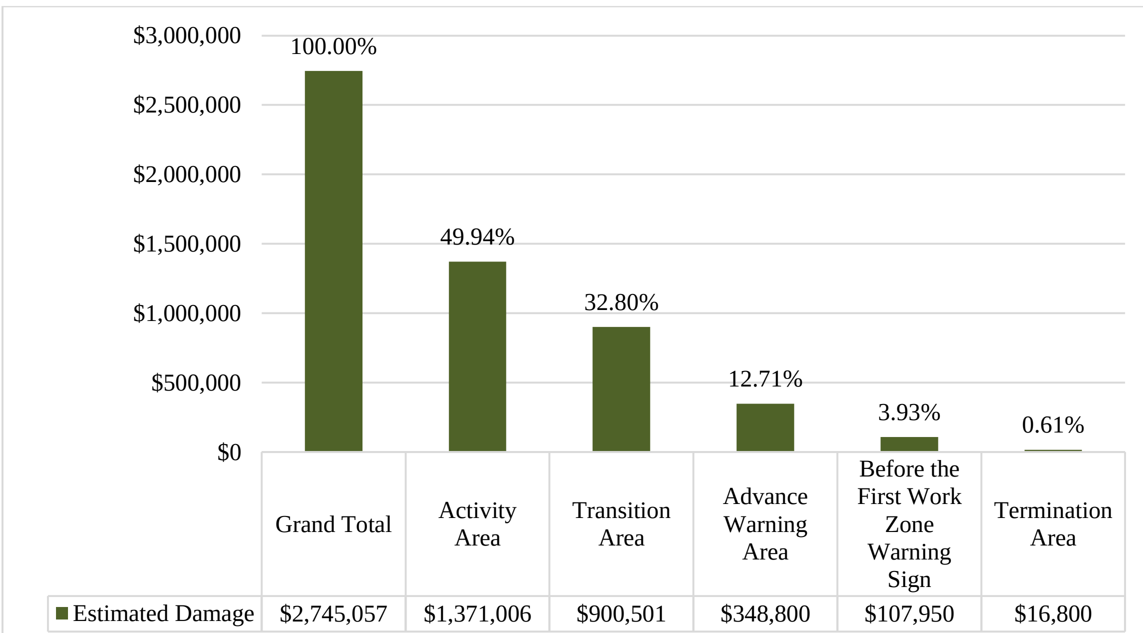


Figure 5-4 Distribution of Damage by Location in WZ

Daytime and nighttime crashes share different impacts of contributing factors in terms of crash severity. This can be also seen in Figure 5-5, which shows that since nighttime crashes were more severe than daytime crashes, they share greater vehicle and property damage costs, which is 68.74% more than daytime conditions to be exact. The same conclusion can be drawn for estimated crash cost by weather condition. As shown in Figure 5-6, the crash costs associated with clear weather condition are significantly higher than the costs associated with cloudy and rainy weather conditions, with 369% and 959.27%, respectively.

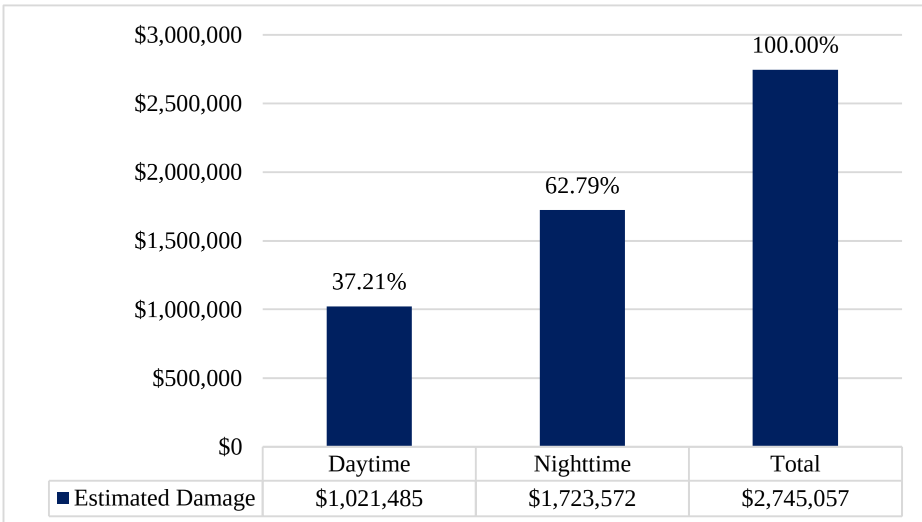


Figure 5-5 Distribution of Damage by Light Cond.

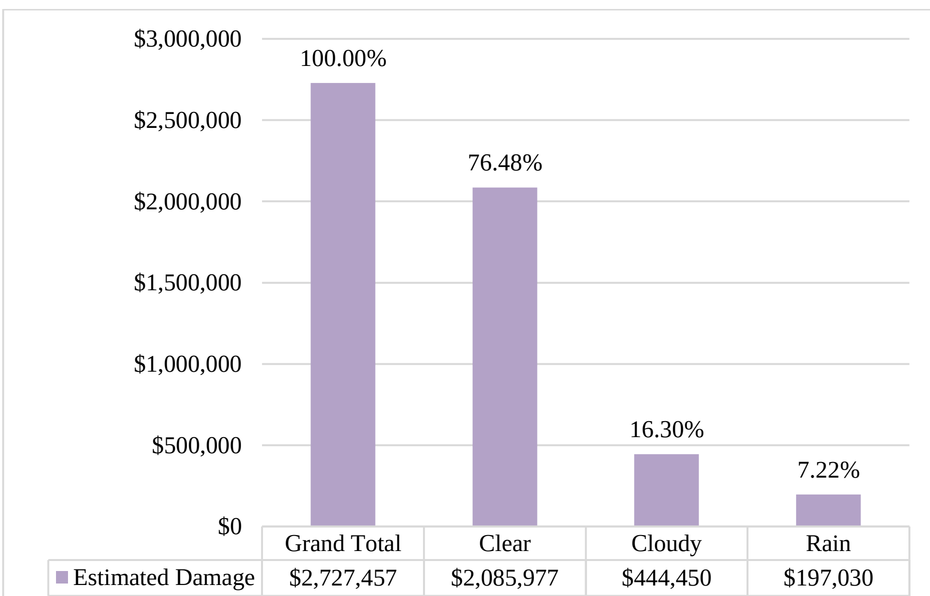


Figure 5-6 Distribution of Damage by Weather Cond.

In terms of crash type, since rear-end crashes were the most severe and frequent crash type in the general work zone crashes, they also consist of the highest associated vehicle and property costs among all crash types in bridge-related crashes (with a share of 59.02%).

Crashes involving workers were associated with higher costs than situations where no workers were present. As shown in Figure 5-7, worker-involved work zone crashes at bridge locations consisted of 65.61% of the total crash costs, which is 90.78% higher than the non-worker-involved crashes (34.39% lower).

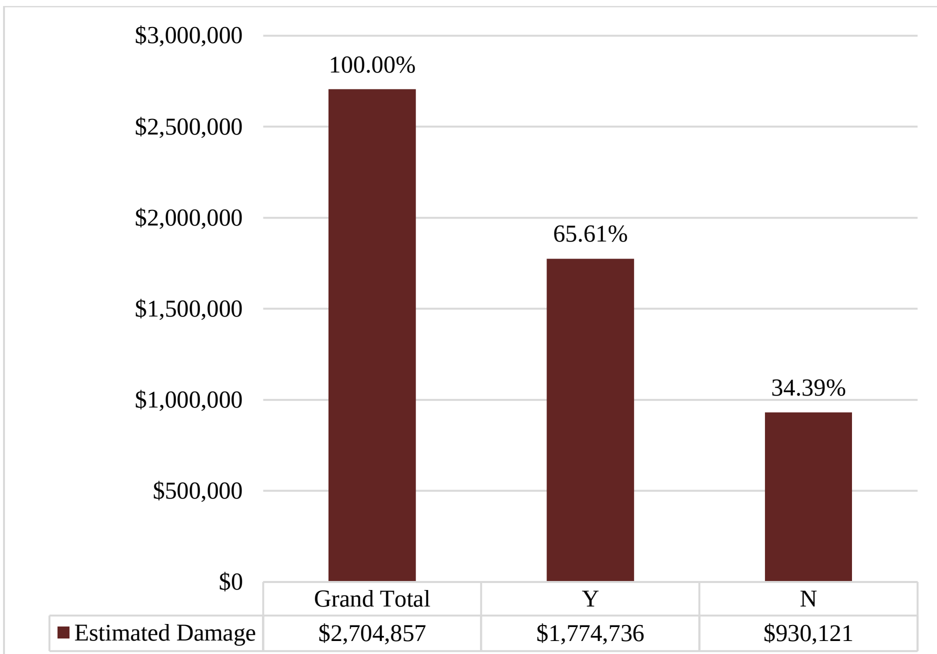


Figure 5-7 Distribution of Damage by Worker Presence

As shown for general work zone crash severity and frequency, law enforcement (LE) availability at work zones result in less severe and a lower number of crashes. Crash records that indicated the presence of law enforcement were less costly than those with an absence of enforcement.

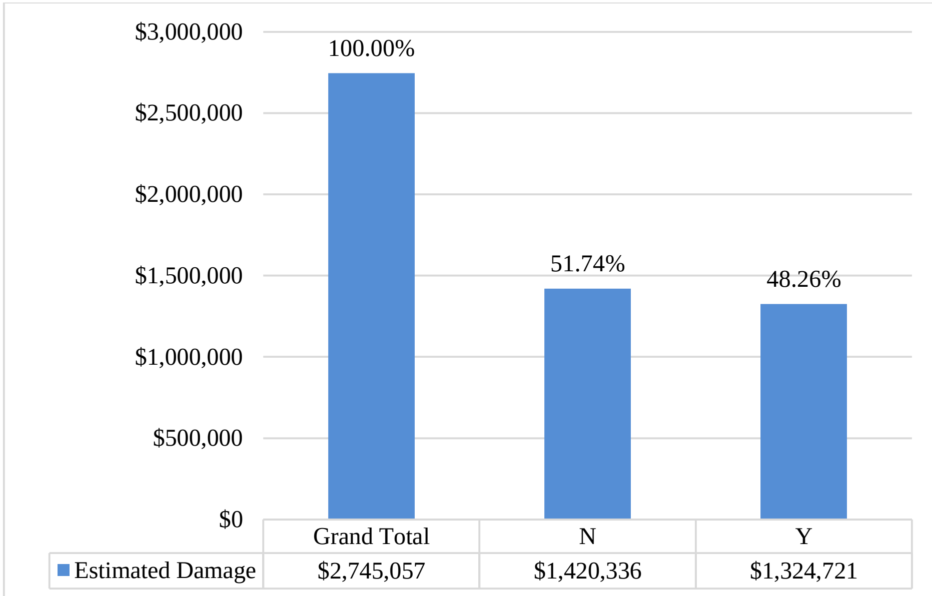


Figure 5-8 Distribution of Damage by LE Presence

5.2.2. Monetary Terms of Crash Severity Levels

The first step to calculating crash costs associated with work zone crashes at bridge locations was to extract the number of participants in each severity level of the KABCO scales, which is presented in Figure 5-1. This analysis will help to capture the distribution of injury severities for each crash severity by taking into account the number of participants involved. In terms of the PDO crashes, as mentioned earlier, instead of using the monetary value to calculate associated crash costs, the estimation from police reports are considered. This may result in having a more reliable crash cost estimate as different construction work zones may share different types of work and necessary machinery.

When calculating a crash benefit-cost ratio (BCR) in the state of Florida, a set of crash unit costs, in which they convert USDOT Maximum Abbreviated Injury Scale (MAIS) person-injury unit costs into comprehensive KABCO crash unit costs, is used

(Mallela and Sadavisam 2011). These crash unit costs are presented in Table 5-1 (Harmon *et al.* 2018).

Table 5-1 Florida DOT crash unit costs for BCA (2013 dollars)

Severity Level	Comprehensive Crash Unit Cost
Fatal Injury (K)	$10,100,000
Incapacitating Injury (A)	$818,636
Non-incapacitating (B)	$163,254
Possible Injury (C)	$99,645
No Injury/PDO (O)	$6,500

According to the Bureau of Labor Statistics, Consumer Price Index, the U.S. dollar value in 2015, 2016, and 2017 experienced an inflation rate of 1.74%, 3.03%, and 5.22% compared to 2013 (BLS 2020). These values then experienced an inflation rate of 8.17%, 6.83%, and 4.60% compared to 2020. Hence, these conversion rates should be taken into account when calculating crash costs.

Considering the observed annual crash frequency divided by KABCO severity levels and corresponding dollar values for 2015, 2016, and 2017 to the present 2020 values, the associated unit costs of work zone crashes at bridge locations are calculated and summarized in Table 5-2.

Table 5-2 WZ Comprehensive Crash Unit Cost

Severity	Frequency	Crash Unit Cost	Injury Cost	Equivalent $ Value*	Total Cost
2015					
K	0	$10,100,000	0	1.74%	0
A	1	$818,636	$818,636	1.74%	$832,880
B	27	$163,254	$4,407,858	1.74%	$4,484,555
C	88	$99,645	$8,768,760	1.74%	$8,921,336
O	192	N/A	$1,369,660**	N/A	$1,369,660
Total 2015					$15,608,431
Value equivalent of total in 2015 to 2020 (8.17% inflation rate)					$16,883,640
2016					
K	1	$10,100,000	$10,100,000	3.03%	$10,406,030
A	1	$818,636	$818,636	3.03%	$843,441
B	10	$163,254	$1,632,540	3.03%	$1,682,006
C	53	$99,645	$5,281,185	3.03%	$5,441,205

Table 5-2 WZ Comprehensive Crash Unit Cost

O	101	N/A	$594,701**	N/A	$594,701
Total 2016					$18,967,383
Value equivalent of total in 2016 to 2020 (6.83% inflation rate)					$20,262,855
2017					
K	2	$10,100,000	$20,200,000	5.22%	$21,254,440
A	0	$818,636	$0	5.22%	$0
B	6	$163,254	$979,524	5.22%	$1,030,655
C	20	$99,645	$1,992,900	5.22%	$2,096,929
O	86	N/A	$780,696**	N/A	$780,696
Total 2017					$25,162,720
Value equivalent of total in 2017 to 2020 (4.60% inflation rate)					$26,320,205
Three-Year Total					$63,466,700
Three-Year Average (present value)					$21,155,567

*as compared to 2013

**total estimated damage from crash reports

5.3. ABC Implementation Costs

Considering that construction costs such as operation, materials, machinery, labor, etc. may differ from location to location, and the crash cost analysis was conducted based on Florida crash statistics, an ABC project information implemented in the state of Florida was utilized for the ABC cost analysis. The ABC project information was extracted from the ABC-UTC website titled ABC Project & Research Databases (ABC-UTC 2020).

Graves Avenue over the I-4 project was located in the city of Orlando in central Florida and was built in 2006. The existing bridge was a two-lane four-span concrete beam bridge with a dimension of 215 feet long and 30 feet wide, and was built in 1958. This was replaced with a 286-foot long and 59-foot wide bridge with the span length of 143 feet through the ABC method to accommodate the widening of the Interstate 4 Highway from four lanes to six lanes in 2006. This bridge was constructed through the Self-propelled Modular Transporters (SPMT) method to move the bridge spans. The comparison of reduction time and associated costs of the ABC method and conventional method is shown in Table 5-3.

Table 5-3 Comparison of ABC and Conventional Method

Method	Mobility Impact (Lane Closure)	Implementation Costs	Present $ Value*
ABC	4 days	$28, 168,175	$35,824,285
Conventional	32 days	$27, 600,000	$35,101,680
Difference	28 days	$568,175	$722,605

*as compared to 2006 (27.18% inflation rate)

In this study, the value of crash costs (i.e., number of crashes and associated costs) that can be saved as a result of shortening the construction duration through the ABC implementation process is considered a safety benefit of the ABC method. This value over the difference of implementation costs (i.e., the additional costs associated with ABC) compared to the conventional construction method will illustrate the safety benefits of ABC over its surplus expenses, as shown in Equation (5-1).

$$Safety\ Benefit = \frac{X * Annual\ average\ crash\ cost\ for\ conventional\ method\ per\ bridge}{Cost\ of\ ABC\ Implementation - Cost\ of\ Conventional} \tag{5-1}$$

Where the X is the number of days reduced in the work zone duration. Using the information provided in Table 5-3, ABC reduced the lane closure period (work zone presence) by 28 days (from 32 to 4 nights), and thus, considering that the calculated crash costs is for 60 bridge locations and normalizing it into 32 days of lane closure, Equation (5-1) can be written as follows:

$$Safety\ Benfits\ = \frac{28 * (\frac{\$21{,}155{,}567}{60 * 32})}{\$722{,}605} = 0.427 = 42.7\% \tag{5-2}$$

It was shown that with the safety benefit of $308,518, which was a result of cutting the lane closure duration by 28 nights, the safety benefits of the Florida project consisted of 42.7% of the total ABC implementation costs.

5.4. Summary and Conclusions

This chapter analyzed the bridge-related work zone crash-associated costs and the roadway safety benefits that can be obtained by utilizing ABC compared to conventional bridge construction methods. To achieve this objective, first, the costs associated with work-zone crashes at the selected bridge locations in Miami-Dade County were analyzed. Then, considering that the selected bridges were constructed through the conventional bridge construction methods, an ABC project implemented in Florida was selected. Different data sources and manuals such as crash data, police reports, the ABC-UTC project and research databases, Highway Capacity Manual (HCM), crash costs for highway safety analysis, work zone road user costs, etc. were utilized to conduct the benefit cost analysis.

Results of the benefit cost analysis illustrated that a portion of the roadway safety benefits of ABC implementation is equal to almost 43% of its associated costs, which can be saved right after bridge insulations. This is a considerable share of ABC-associated costs, which can also be higher when the conventional methods take longer and need more lane closures. On the other hand, the considered project was installed through the SPMT method, so different ABC methods may result in different times of implementation.

It should be noted that this chapter focused on the roadway safety benefits of ABC compared to the conventional construction method, while there are other transportation benefits that may result from cutting work zone duration, such as delay, emission, and user costs. Thus, in order to view all of the ABC benefits, each factor needs to be taken into account separately.

CHAPTER 6
CONCLUSION

The first part of this project focused on the impacts of work zone presence on the traveling public and construction crews and were investigated in the context of crash severity. Second, the impacts of work zone presence on bridge locations were studied, as defined by crash frequency. Finally, the costs associated with work zone-related crashes at bridge locations were analyzed, and benefits from ABC implementation methods were calculated. To this end, different descriptive statistics, statistical modeling approaches, and machine learning techniques together with metaheuristic optimization algorithms were developed and utilized. The following list illustrates the contributions of this project to the body of transportation knowledge:

- Applying machine learning techniques for work zone crash analysis.
- Incorporating Artificial Intelligence (AI) for enhancing the prediction performance of the developed machine learning models.
- Performing sensitivity analysis to deal with the black-box nature of the proposed machine learning models.
- In-depth investigation of contributing factors in conjunction with the results from statistical and machine learning models to provide a more comprehensive interpretation of crash severity/frequency outcomes.

6.1. Crash Severity at Work Zones

Work zones are critical locations in roadway networks in which different crash risk factors are involved compared to general traffic crashes. Worker safety is one of the main concerns of transportation safety analysts when planning and designing a work zone. In this project, the risk factors associated with work zone crashes involving workers were examined through binary level logistic regression and support vector machine classification models. Different models for daytime and nighttime crashes were devolved.

A total of 9,179 and 2,863 crash records for daytime and nighttime work zone crashes from 2015 through 2017 were used in both crash severity models. While the statistically significant crash severity contributing factors were determined through the mixed logit modeling framework, the nonlinear relationship between crash severity outcomes by time-of-day were explored by an SVM model trained by the Cuckoo Search (CS) metaheuristic optimization algorithm. Likelihood ratio tests were also conducted to examine the overall stability of the models' estimates across time periods.

Results demonstrated that while there is a significant temporal instability among parameter estimates for daytime and nighttime models, driver alcohol involvement, rainy weather condition, wet road surface, multi-occupant for vehicle occupancy, and distraction are the most significant causes of fatalities/injuries in work zone crashes involving workers in both models. For the variables, which are the number of vehicle-involved and law enforcement indicators, a mixed effect was found between daytime and nighttime conditions. It was also shown that different risk factors were involved in work zone critical locations between daytime and nighttime conditions.

From modeling points of view, when comparing model performance, the CS-SVM produced a better prediction performance compared to the SVM and BMXL models. The modeling results also shed light on the ability of SI optimization techniques in the SVM parameter selection to achieve a higher prediction performance.

6.2. Crash Frequency at Work Zones

Since bridges are specific segments of roadways and share different geometric characteristics, the study of crash frequency contributing factors is of importance, not only for transportation safety analysis, but also for bride geometric design. To this end, this project focused on examining the relationship between work zone presence and crash occurrence at bridge locations through a detailed descriptive analysis and by developing crash frequency models.

Using multiple data sources such as crash records, roadway geometric features, and traffic data, a crash dataset containing 60 bridge locations and associated work zone crashes from 2015 to 2017 was created to develop predictive models. While a detailed descriptive analysis was provided to illustrate the percentage distribution of crash frequency based on months of year, day of week, time of day, crash type, number of vehicles involved, weather condition, location at work zone, work zone type, and so forth, the risk factors were examined through the developed predictive models.

Incorporating 18 explanatory variables make a considerable set of contributing factors for bridge related work zone crash frequency analysis, including: work zone related features such as percentage of workers involved, percentage of law enforcement, and work zone type; bridge geometric characteristics including bridge length, surface width, median type and width, shoulder type and width; roadway characteristics such as road functional classification, ramp, intersection, and express lane existence; traffic conditions indicators such as AADT and truck AADT.

The analysis of crash frequency through the NB model revealed that five explanatory variables, including paved median with barrier other than guardrail, raised traffic separators, law enforcement, horizontal curve indicator, and bridge surface width, had a statistically significant impact on crash frequency at bridge locations. In addition, a nonlinear relationship was observed from the ABC-SVR results between the bridge surface width and number of crashes.

From a modeling point of view, it was shown that the developed ABC-SVR model, compared to the SVR and NB models, results in significantly better prediction performance, and thus, more reliable model inference.

6.3. ABC Benefit-Cost Analysis

It has been well documented in the literature that vehicles traveling through roadways with work zones have a higher chance of being in an accident. Work zone crash cost is another important aspect of work zone safety that needs to be considered in the decision-making process. As shown through the detailed descriptive analysis and crash cost calculations, bridge-related work zone crashes account for the annual cost of $21,155,567 from 2015 to 2017. An investigation into the ABC-associated costs from a case study in Florida revealed that this consists of almost 43% of the total ABC implementation costs.

DOTs developed their own decision-making process to assess the viability of ABC technologies and determine the effects of ABC on the overall cost of the bridge. Although the impact of delay and delay-related user costs were incorporated into the developed ABC decision matrices and considered a benefit of ABC over conventional bridge construction methods, the impacts of ABC on roadway safety are overlooked. In addition, there is a lack of consideration for worker safety when assessing the ABC projects. These benefits can be achieved through the modeling and analysis of work zone related crashes through the detailed descriptive and crash analysis. While the descriptive analysis reveals the general crash trends, the significant contributing factors that impact the crash frequency and severity can be extracted from associated models. Hence, there is an emergent need for a comprehensive statewide and in-depth investigation of work zone crash mechanisms to be ultimately incorporated into the decision-making process, since it is a significant portion of the ABC implementation costs.

6.4. Study Limitations and Future Works

Data limitations is a common issue when studying work zone crashes. The work zone crash analysis will only be as accurate as the applied data. There is a lack of work zone-specific information in police reports and accordingly in crash data, including whether or not the crash occurred at an active work zone, or the type of work zone activity such as construction, maintenance, or utility work activities was missing. Moreover, construction project profiles were missing in the work zone operation data, such as time frames of work zone durations. To avoid potential bias, only the crashes marked as work zone related crashes at bridge locations have been included in the crash frequency and cost analysis of this project. Hence, we made sure that the considered crashes were certainly work zone related ones; however, there may be other crashes that occurred as a result of work zone presence, but these were not marked in police reports.

Future research can focus on providing more detailed work zone related data such as work zone activities and duration for pre-work zone and during construction time analysis. The study of daytime and nighttime work zone scheduling and their economic impacts are recommendations for future research.

While the models developed in this research may not be transferable to other locations, the practical aspects of the proposed methodology can be applied using local work zone crash data from other areas.

In addition, considering that the work zone impacts can be studied from different points of view including safety, mobility, and environmental impacts, multi-criteria decision making analysis would be a great addition to this research. This can help decision making process and assist stakeholders to consider preference information when making a decision on bridge construction strategies.

REFERENCES

Abaza, O.A., Chowdhury, T.D., Arafat, M., 2017. Comparative analysis of skid resistance for different roadway surface treatments. American Journal of Engineering and Applied Sciences 10 (4), 890-899.

Abc-Utc, Project database. Accelerated Bridge Construction University Transportation Center (ABC-UTC), https://abc-utc.fiu.edu/resources/project-research-databases/, May, 2020.

Adomah, E., Khoda Bakhshi, A., Ahmed, M.M., 2021. Safety impact of connected vehicles on driver behavior in rural work zones under foggy weather conditions. In: Proceedings of the Transportation Research Board 100th Annual Meeting Transportation Research Board, Washington, D.C.

Aguero-Valverde, J., 2013. Full bayes poisson gamma, poisson lognormal, and zero inflated random effects models: Comparing the precision of crash frequency estimates. Accident Analysis & Prevention 50, 289-297.

Ahangari, S., Jeihani, M., Jarju, A., 2018, Drivers' behavior analysis under rainy weather conditions using a driving simulator.

Ali, M., Zerpa, L.A., Kaisar, E.I., Masters, K.O., 2018, Guidance for identifying corridor conditions that warrant deploying transit signal priority. No. 18-05194.

Ali, M.S., Kaisar, E.I., Hadi, M., 2017. Guidance for identifying corridor conditions that warrant deploying transit signal priority and queue jump. In: Proceedings of the 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), pp. 657-662. DOI:10.1109/MTITS.2017.8005595.

Alkheder, S., Taamneh, M., Taamneh, S., 2017. Severity prediction of traffic accident using an artificial neural network. Journal of Forecasting 36 (1), 100-108.

Alluri, P., Raihan, M.A., Saha, D., Wu, W., Huq, A., Nafis, S., Gan, A., 2017, Statewide analysis of bicycle crashes. Florida Department of Transportation (FDOT), No. BDV29-977-23.

Anastasopoulos, P.C., Mannering, F.L., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. Accident Analysis & Prevention 41 (1), 153-159.

Anderson, J., Hernandez, S., 2017. Roadway classifications and the accident injury severities of heavy-vehicle drivers. Analytic Methods in Accident Research 15, 17-28.

Anderson, J.C., Dong, S., 2017. Heavy-vehicle driver injury severity analysis by time of week: A mixed logit approach using hsis crash data. Institute of Transportation Engineers. ITE Journal 87 (9), 41-49.

Arafat, M., Iqbal, S., Hadi, M., 2020. Utilizing an analytical hierarchy process with stochastic return on investment to justify connected vehicle-based deployment decisions. Transportation Research Record, 0361198120929686.

Arditi, D., Lee, D.-E., Polat, G., 2007. Fatal accidents in nighttime vs. Daytime highway construction work zones. Journal of Safety Research 38 (4), 399-405.

Artba, American road & transportation builders association. National Work Zone Safety Information Clearinghouse, https://www.workzonesafety.org/, December, 10, 2018.

Artba, National bridge inventory: Florida. American Road & Transportation Builders Association https://artbabridgereport.org/state/profile/FL, April, Accessed April, 10, 2020.

Arvin, R., Kamrani, M., Khattak, A.J., 2019a, Examining the role of speed and driving stability on crash severity using shrp2 naturalistic driving study data.

Arvin, R., Kamrani, M., Khattak, A.J., 2019b. How instantaneous driving behavior contributes to crashes at intersections: Extracting useful information from connected vehicle message data. Accident Analysis & Prevention 127, 118-133.

Awad, M., Khanna, R., 2015. Support vector regression. Efficient learning machines. Springer, pp. 67-80.

Azizinamini, A., Rehmat, S., Sadeghnejad, A., 2019. Enhancing resiliency and delivery of bridge elements using ultra-high performance concrete as formwork. Transportation Research Record 2673 (5), 443-453.

Bai, Y., Li, Y., 2007, Determining major causes of highway work zone accidents in kansas, phase ii. Kansas Department of Transportation, Final Report: K-TRAN: KU-06-1.

Behnood, A., Mannering, F., 2017. Determinants of bicyclist injury severities in bicycle-vehicle crashes: A random parameters approach with heterogeneity in means and variances. Analytic methods in accident research 16, 35-47.

Behnood, A., Mannering, F.L., 2015. The temporal stability of factors affecting driver-injury severities in single-vehicle crashes: Some empirical evidence. Analytic methods in accident research 8, 7-32.

Behnood, A., Mannering, F.L., 2016. An empirical assessment of the effects of economic recessions on pedestrian-injury crashes using mixed and latent-class models. Analytic methods in accident research 12, 1-17.

Behnood, A., Roshandeh, A.M., Mannering, F.L., 2014. Latent class analysis of the effects of age, gender, and alcohol consumption on driver-injury severities. Analytic Methods in Accident Research 3, 56-91.

Beshah, T., Hill, S., 2010. Mining road traffic accident data to improve safety: Role of road-related factors on accident severity in ethiopia. In: Proceedings of the 2010 AAAI Spring Symposium Series.

Bls, Consumer price index (cpi). U.S. Bureau of Labor Statistics, https://www.bls.gov/news.release/cpi.toc.htm, May, 2020.

Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers. In: Proceedings of the Proceedings of the fifth annual workshop on Computational learning theory, pp. 144-152.

Bottou, L., Lin, C.-J., 2007. Support vector machine solvers. Large scale kernel machines 3 (1), 301-320.

Bradley, A.P., 1997. The use of the area under the roc curve in the evaluation of machine learning algorithms. Pattern recognition 30 (7), 1145-1159.

Breiman, L., 2001. Random forests. Machine learning 45 (1), 5-32.

Cerwick, D.M., Gkritza, K., Shaheed, M.S., Hans, Z., 2014. A comparison of the mixed logit and latent class methods for crash severity analysis. Analytic Methods in Accident Research 3, 11-27.

Chambless, J., Ghadiali, A.M., Lindly, J.K., Mcfadden, J., 2002. Multistate work-zone crash characteristics. Institute of Transportation Engineers. ITE Journal 72 (5), 46.

Chang, L.-Y., Chen, W.-C., 2005. Data mining of tree-based models to analyze freeway accident frequency. Journal of safety research 36 (4), 365-375.

Chang, L.-Y., Chien, J.-T., 2013. Analysis of driver injury severity in truck-involved accidents using a non-parametric classification tree model. Safety science 51 (1), 17-22.

Chapelle, O., Vapnik, V., Bousquet, O., Mukherjee, S., 2002. Choosing multiple parameters for support vector machines. Machine learning 46 (1-3), 131-159.

Chen, C., Zhang, G., Qian, Z., Tarefder, R.A., Tian, Z., 2016. Investigating driver injury severity patterns in rollover crashes using support vector machine models. Accident Analysis & Prevention 90, 128-139.

Chen, C., Zhang, G., Tarefder, R., Ma, J., Wei, H., Guan, H., 2015. A multinomial logit model-bayesian network hybrid approach for driver injury severity analyses in rear-end crashes. Accident Analysis & Prevention 80, 76-88.

Chen, E., Tarko, A.P., 2012. Analysis of crash frequency in work zones with focus on police enforcement. Transportation research record 2280 (1), 127-134.

Chen, E., Tarko, A.P., 2014. Modeling safety of highway work zones with random parameters and random effects models. Analytic methods in accident research 1, 86-95.

Chitturi, M.V., Ooms, A.W., Bill, A.R., Noyce, D.A., 2011. Injury outcomes and costs for cross-median and median barrier crashes. Journal of safety research 42 (2), 87-92.

Das, A., Abdel-Aty, M.A., 2011. A combined frequency–severity approach for the analysis of rear-end crashes on urban arterials. Safety Science 49 (8-9), 1156-1163.

Deka, P.C., 2014. Support vector machine applications in the field of hydrology: A review. Applied soft computing 19, 372-386.

Delen, D., Sharda, R., Bessonov, M., 2006. Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. Accident Analysis & Prevention 38 (3), 434-444.

Dias, I.M., 2015. Work zone crash analysis and modeling to identify factors associated with crash severity and frequency. Ph.D. Dissertation. Kansas State University.

Dong, C., Richards, S.H., Clarke, D.B., Zhou, X., Ma, Z., 2014. Examining signalized intersection crash frequency using multivariate zero-inflated poisson regression. Safety science 70, 63-69.

Dong, N., Huang, H., Zheng, L., 2015. Support vector machine in crash prediction at the level of traffic analysis zones: Assessing the spatial proximity effects. Accident Analysis & Prevention 82, 192-198.

Eftekharzadeh, S., Khodabakhshi, A., 2014. Safety evaluation of highway geometric design criteria in horizontal curves at downgrades. International Journal of Civil Engineering 12 (3), 326-332.

El-Basyouny, K., Sayed, T., 2009a. Accident prediction models with random corridor parameters. Accident Analysis & Prevention 41 (5), 1118-1123.

El-Basyouny, K., Sayed, T., 2009b. Collision prediction models using multivariate poisson-lognormal regression. Accident Analysis & Prevention 41 (4), 820-828.

Eluru, N., Bhat, C.R., Hensher, D.A., 2008. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. Accident Analysis & Prevention 40 (3), 1033-1054.

Esmaeilzadeh, E., Mokhtarimousavi, S., 2020. Machine learning approach for flight departure delay prediction and analysis. Transportation Research Record, 0361198120930014.

Farhangdoust, S., Mehrabi, A., 2019. Help-evaluation of nondestructive evaluation methods for applicability to precast concrete deck panels. In: Proceedings of the ASNT Research Symposium 2019, Orange County, CA, pp. DOI: 10.32548/RS.2019.004.

Farhangdoust, S., Mehrabi, A., 2020. Non-destructive evaluation of closure joints in accelerated bridge construction using a damage etiology approach. Applied Sciences 10 (4), 1457.

Farzad, M., Sadeghnejad, A., Rastkar, S., Moshkforoush, A., Azizinamini, A., 2020. A theoretical analysis of mechanical and durability enhancement of circular reinforced concrete columns repaired with uhpc. Engineering Structures 209, 109928.

Fell, J.C., Waehrer, G., Voas, R.B., Auld-Owens, A., Carr, K., Pell, K., 2014. Effects of enforcement intensity on alcohol impaired driving crashes. Accident Analysis & Prevention 73, 181-186.

Fhwa, Kabco injury classification scale and definitions. Federal Highway Administrations (FHWA), https://safety.fhwa.dot.gov/hsip/spm/conversion_tbl/pdfs/kabco_ctable_by_state.pdf, 2020.

Fish, K.E., Blodgett, J.G., 2003. A visual method for determining variable importance in an artificial neural network model: An empirical benchmark study. Journal of Targeting, Measurement and Analysis for Marketing 11 (3), 244-254.

Garber, N.J., Zhao, M., 2002. Distribution and characteristics of crashes at different work zone locations in virginia. Transportation Research Record 1794 (1), 19-25.

Ghasemzadeh, A., Ahmed, M.M., 2017, A probit-decision tree approach to analyze effects of adverse weather conditions on work zone crash severity using second strategic highway research program roadway information dataset.

Ghasemzadeh, A., Hammit, B.E., Ahmed, M.M., Young, R.K., 2018. Parametric ordinal logistic regression and non-parametric decision tree approaches for assessing the impact of weather conditions on driver speed selection using naturalistic driving data. Transportation research record, 0361198118758035.

Greene, W.H., 2018. Econometric analysis (8th edition), Pearson, New York, NY.

Gu, X., Li, T., Wang, Y., Zhang, L., Wang, Y., Yao, J., 2018. Traffic fatalities prediction using support vector machine with hybrid particle swarm optimization. Journal of Algorithms & Computational Technology 12 (1), 20-29.

Hadi, M., Iqbal, M.S., Wang, T., Xiao, Y., Arafat, M., Afreen, S., 2019a, Connected vehicle vehicle-to-infrastructure support of active traffic management. FDOT, Office of Research and Development.

Hadi, M., Xiao, Y., Iqbal, M.S., Wang, T., Arafat, M., Hoque, F., 2019b, Estimation of system performance and technology impacts to support future year planning. FDOT, Office of Research and Development.

Haghighi, N., Liu, X.C., Zhang, G., Porter, R.J., 2018. Impact of roadway geometric features on crash severity on rural two-lane highways. Accident Analysis & Prevention 111, 34-42.

Haleem, K., Abdel-Aty, M., 2010. Examining traffic crash injury severity at unsignalized intersections. Journal of safety research 41 (4), 347-357.

Haleem, K., Alluri, P., Gan, A., 2015. Analyzing pedestrian crash injury severity at signalized and non-signalized locations. Accident Analysis & Prevention 81, 14-23.

Harb, R., Radwan, E., Yan, X., Pande, A., Abdel-Aty, M., 2008. Freeway work-zone crash analysis and risk identification using multiple and conditional logistic regression. Journal of Transportation Engineering 134 (5), 203-214.

Harmon, T., Bahar, G.B., Gross, F.B., 2018, Crash costs for highway safety analysis.

Hcm, 2010, Highway capacity manual. National Research Council, Transportation research board, Washington, DC.

Huang, H., Zeng, Q., Pei, X., Wong, S., Xu, P., 2016. Predicting crash frequency using an optimised radial basis function neural network model. Transportmetrica A: transport science 12 (4), 330-345.

Iranitalab, A., Khattak, A., 2017. Comparison of four statistical and machine learning methods for crash severity prediction. Accident Analysis & Prevention 108, 27-36.

Islam, S., Mannering, F., 2006. Driver aging and its effect on male and female single-vehicle accident injuries: Some additional evidence. Journal of safety Research 37 (3), 267-276.

Jafari, S.A., Jahandideh, S., Jahandideh, M., Asadabadi, E.B., 2015. Prediction of road traffic death rate using neural networks optimised by genetic algorithm. International journal of injury control and safety promotion 22 (2), 153-157.

Jin, T.G., Saito, M., 2009. Spatial and temporal analyses of crashes at two work zones with different traffic control devices: Barrels and jersey barriers. In: Proceedings of the Transportation Research Board 37 88th Annual Meeting, Washington, DC.

Kamrani, M., Arvin, R., Khattak, A.J., 2018. Extracting useful information from basic safety message data: An empirical study of driving volatility measures and crash frequency at intersections. Transportation Research Record 2672 (38), 290-301.

Kamyab, A., Mcdonald, T., Storm, B., Anderson-Wilk, M., 2003, Effectiveness of extra enforcement in construction and maintenance work zones. Midwest States Smart

Work Zone Deployment Initiative, Mid-America Transportation Center, University of Nebraska-Lincoln, MwSWZDI Year 4 Technology Evaluation #1.

Karaboga, D., 2005, An idea based on honey bee swarm for numerical optimization. Erciyes University, Engineering Faculty, Computer Engineering Department, Technical report-tr06.

Kashani, A.T., Mohaymany, A.S., 2011. Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models. Safety Science 49 (10), 1314-1320.

Katta, V., 2013. Development of crash severity model for predicting risk factors in work zones for ohio. Master's Thesis. University of Toledo.

Keerthi, S.S., 2002. Efficient tuning of svm hyperparameters using radius/margin bound and iterative algorithms. IEEE Transactions on Neural Networks 13 (5), 1225-1229.

Khattak, A.J., 2001. Injury severity in multivehicle rear-end crashes. Transportation Research Record 1746 (1), 59-68.

Khattak, A.J., Khattak, A.J., Council, F.M., 2002. Effects of work zone presence on injury and non-injury crashes. Accident Analysis & Prevention 34 (1), 19-29.

Khoda Bakhshi, A., Ahmed, M.M., 2020. Utilizing black-box visualization tools to interpret non-parametric real-time risk assessment models. Transportmetrica A: Transport Science, 1-27.

Khoury, J., Amine, K., Abi Saad, R., 2019. An initial investigation of the effects of a fully automated vehicle fleet on geometric design. Journal of Advanced Transportation 2019.

Kiran, M.S., Hakli, H., Gunduz, M., Uguz, H., 2015. Artificial bee colony algorithm with variable search strategy for continuous optimization. Information Sciences 300, 140-157.

Kitali, A.E., Mokhtarimousavi, S., Kadeha, C., Alluri, P., 2020. Severity analysis of crashes on express lane facilities using support vector machine model trained by firefly algorithm. Traffic Injury Prevention 22 (1), 79-84. https://doi.org/10.1080/15389588.2020.1840563.

Kockelman, K.M., Kweon, Y.-J., 2002. Driver injury severity: An application of ordered probit models. Accident Analysis & Prevention 34 (3), 313-321.

Li, X., Lord, D., Zhang, Y., Xie, Y., 2008. Predicting motor vehicle crashes using support vector machine models. Accident Analysis & Prevention 40 (4), 1611-1618.

Li, Y., Bai, Y., 2008. Development of crash-severity-index models for the measurement of work zone risk levels. Accident Analysis & Prevention 40 (5), 1724-1731.

Li, Y., Bai, Y., 2009. Highway work zone risk factors and their impact on crash severity. Journal of Transportation engineering 135 (10), 694-701.

Li, Z., Liu, P., Wang, W., Xu, C., 2012. Using support vector machine models for crash injury severity analysis. Accident Analysis & Prevention 45, 478-486.

Liaw, A., Wiener, M., 2002. Classification and regression by randomforest. R news 2 (3), 18-22.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transportation research part A: policy and practice 44 (5), 291-305.

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

Ma, J., Kockelman, K.M., Damien, P., 2008. A multivariate poisson-lognormal regression model for prediction of crash counts by severity, using bayesian methods. Accident Analysis & Prevention 40 (3), 964-975.

Ma, Z., Zhang, H., Steven, I., Chien, J., Wang, J., Dong, C., 2017. Predicting expressway crash frequency using a random effect negative binomial model: A case study in china. Accident Analysis & Prevention 98, 214-222.

Mahmoudzadeh, A., Firoozi Yeganeh, S., Ochoa, B., Wang, X.B., Tippy, J., 2019, A practical approach for evaluating and scheduling a campus bus system based on clustering. Transportation Research Board (TRB), No. 19–030, 2019.

Mahmoudzadeh, A., Wang, X.B., 2020. Cluster based methodology for scheduling a university shuttle system. Transportation Research Record 2674 (1), 236-248.

Mallela, J., Sadavisam, S., 2011, Work zone road user costs: Concepts and applications. United States. Federal Highway Administration, DTFH61-06-D-00004.

Mamdoohi, S., Song, Z., Sharifi, M.S., Nasr-Isfahani, H., 2018, Estimation of standard pedestrian equivalent factors for heterogeneous pedestrian stream containing individuals with disabilities.

Mannering, F.L., Bhat, C.R., 2014. Analytic methods in accident research: Methodological frontier and future directions. Analytic methods in accident research 1, 1-22.

Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Analytic methods in accident research 11, 1-16.

Mansourkhaki, A., Karimpour, A., Sadoghi Yazdi, H., 2017. Non-stationary concept of accident prediction. In: Proceedings of the Proceedings of the Institution of Civil Engineers-Transport, pp. 140-151.

Mcfadden, D., Train, K., 2000. Mixed mnl models for discrete response. Journal of applied Econometrics 15 (5), 447-470.

Mehrabi, A., Farhangdoust, S., 2019. Ndt methods applicable to health monitoring of abc closure joints. In: Proceedings of the 27th Research Symposium-The American Society for Non-destructive Testing (ASNT), Orlando, Fl, pp. 26-29.

Miaou, S.-P., 1993, The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. Oak Ridge National Lab., TN (United States).

Mohan, S.B., Gautam, P., 2000. Cost of highway work zone injuries. In: Proceedings of the Construction Congress VI: Building Together for a Better Tomorrow in an Increasingly Complex World, pp. 1196-1207.

Mokhtarimousavi, S., 2019. A time of day analysis of pedestrian-involved crashes in california: Investigation of injury severity, a logistic regression and machine learning approach using hsis data. Institute of Transportation Engineers. ITE Journal 89 (10), 25-33.

Mokhtarimousavi, S., Anderson, J.C., Azizinamini, A., Hadi, M., 2019. Improved support vector machine models for work zone crash injury severity prediction and analysis. Transportation Research Record 2673 (11), 680-692. DOI:10.1177/0361198119845899.

Mokhtarimousavi, S., Anderson, J.C., Azizinamini, A., Hadi, M., 2020a. Factors affecting injury severity in vehicle-pedestrian crashes: A day-of-week analysis using random parameter ordered response models and artificial neural networks. International Journal of Transportation Science and Technology 9 (2), 100-115. DOI:10.1016/j.ijtst.2020.01.001.

Mokhtarimousavi, S., Anderson, J.C., Hadi, M., Azizinamini, A., 2021a. A temporal investigation of crash severity factors in worker-involved work zone crashes: Random parameters and machine learning approaches. Transportation Research Interdisciplinary Perspectives 10, 100378.

Mokhtarimousavi, S., Azizinamini, A., Hadi, M., 2020b. Severity of worker-involved work zone crashes: A study of contributing factors. In: Proceedings of the International Conference on Transportation and Development 2020, pp. 47-59.

Mokhtarimousavi, S., Hadi, M., Sadeghvaziri, E., Azizinamini, A., 2021b. Evolutionary training approaches for machine learning models for analyzing classification imbalanced crash datasets. In: Proceedings of the Transportation Research Board 100th Annual Meeting, Washington, DC. , pp. TRBAM-21-03834.

Mokhtarimousavi, S., Rahami, H., Kaveh, A., 2015. Multi-objective mathematical modeling of aircraft landing problem on a runway in static mode, scheduling and sequence determination using nsga-ii. Iran University of Science & Technology 5 (1), 21-36.

Mokhtarimousavi, S., Rahami, H., Saffarzadeh, M., Piri, S., 2014. Determination of the aircraft landing sequence by two meta-heuristic algorithms. International Journal of Transportation Engineering 1 (4), 271-284. DOI:10.22119/IJTE.2014.6299.

Mokhtarimousavi, S., Talebi, D., Asgari, H., 2018. A non-dominated sorting genetic algorithm approach for optimization of multi-objective airport gate assignment problem. Transportation Research Record 2672 (23), 59-70. DOI:10.1177/0361198118781386.

Montella, A., 2010. A comparative analysis of hotspot identification methods. Accident Analysis & Prevention 42 (2), 571-581.

Morgan, A., Mannering, F.L., 2011. The effects of road-surface conditions, age, and gender on driver-injury severities. Accident Analysis & Prevention 43 (5), 1852-1863.

Mousavi, S.M., Marzoughi, H., Parr, S.A., Wolshon, B., Pande, A., 2019a. A mixed crash frequency estimation model for interrupted flow segments. International

Conference on Transportation and Development 2019. https://ascelibrary.org/doi/10.1061/9780784482575.008.

Mousavi, S.M., Zhang, Z., Parr, S.A., Pande, A., Wolshon, B., 2019b. Identifying high crash risk highway segments using jerk-cluster analysis. International Conference on Transportation and Development 2019. https://ascelibrary.org/doi/10.1061/9780784482575.012. pp. 112–123.

Nafis, S.R., Alluri, P., Jung, R., Ennemoser, R., Gan, A., 2019. A comprehensive review of states' existing practices in using communication technologies to increase public involvement for transportation projects. Transportation Research Board 98th Annual MeetingTransportation Research Board (19-01674).

Nbi, National bridge inventory Federal Highway Administration (FHWA) https://www.fhwa.dot.gov/bridge/nbi.cfm, Accessed April, 10, 2020.

Nezafat, R.V., Sahin, O., Cetin, M., 2019. Transfer learning using deep neural networks for classification of truck body types based on side-fire lidar data. Journal of Big Data Analytics in Transportation 1 (1), 71-82.

Nhtsa, 2011, Model minimum uniform crash criteria 4th edition. National Highway Traffic Safety Administration, USDOT.

Oh, J., Lyon, C., Washington, S., Persaud, B., Bared, J., 2003. Validation of fhwa crash models for rural intersections: Lessons learned. Transportation Research Record 1840 (1), 41-49.

Osman, M., Paleti, R., Mishra, S., 2018. Analysis of passenger-car crash injury severity in different work zone configurations. Accident Analysis & Prevention 111, 161-172.

Ozturk, O., 2014. Investigating impact of work zones on crash frequency, severity and traffic. Ph.D. Dissertation. Rutgers, The State University of New Jersey-New Brunswick.

Ozturk, O., Ozbay, K., Yang, H., 2014. Estimating the impact of work zones on highway safety. In: Proceedings of the Transportation Research Board 93rd Annual Meeting.

Ozturk, O., Ozbay, K., Yang, H., 2015. Investigating the impact of work zones on crash severity by compar-ative analysis. In: Proceedings of the Journal of Transportation Research Board, TRB's 94th Annual Meeting, Washington, DC, pp. 1-18.

Ozturk, O., Ozbay, K., Yang, H., Bartin, B., 2013. Crash frequency modeling for highway construction zones. In: Proceedings of the Transportation Research Board 92nd Annual Meeting, pp. 13-4555.

Pan, G., Fu, L., Thakali, L., 2017. Development of a global road safety performance function using deep neural networks. International journal of transportation science and technology 6 (3), 159-173.

Park, E.S., Lord, D., 2007. Multivariate poisson-lognormal models for jointly modeling crash frequency by severity. Transportation Research Record 2019 (1), 1-6.

Parsa, A.B., Chauhan, R.S., Taghipour, H., Derrible, S., 2019a. Applying deep learning to detect traffic accidents in real time using spatiotemporal sequential data. arXiv preprint arXiv:1912.06991.

Parsa, A.B., Movahedi, A., Taghipour, H., Derrible, S., Mohammadian, A.K., 2020. Toward safer highways, application of xgboost and shap for real-time accident detection and feature analysis. Accident Analysis & Prevention 136, 105405.

Parsa, A.B., Taghipour, H., Derrible, S., Mohammadian, A.K., 2019b. Real-time accident detection: Coping with imbalanced data. Accident Analysis & Prevention 129, 202-210.

Peace, G.S., 1993. Taguchi methods: A hands-on approach, Addison Wesley Publishing Company, New York, NY.

Qi, Y., Srinivasan, R., Teng, H., Baker, R., 2013. Analysis of the frequency and severity of rear-end crashes in work zones. Traffic injury prevention 14 (1), 61-72.

Qi, Y., Srinivasan, R., Teng, H., Baker, R.F., 2005, Frequency of work zone accidents on construction projects.

Qin, X., 2012. Quantile effects of causal factors on crash distributions. Transportation research record 2279 (1), 40-46.

Qu, X., Wang, W., Wang, W., Liu, P., 2013. Real-time freeway sideswipe crash prediction by support vector machine. IET Intelligent Transport Systems 7 (4), 445-453.

Raihan, M.A., Alluri, P., Wu, W., Gan, A., 2019. Estimation of bicycle crash modification factors (cmfs) on urban facilities using zero inflated negative binomial models. Accident Analysis & Prevention 123, 303-313.

Ralls, M.L., 2007. Accelerated bridge construction. ASPIRE.[Online Article] 1620.

Raub, R.A., Sawaya, O.B., Schofer, J.L., Ziliaskopoulos, A., 2001. Enhanced crash reporting to explore work zone crash patterns. Paper No. 01 166.

Redelmeier, D.A., Tibshirani, R.J., Evans, L., 2003. Traffic-law enforcement and risk of death from motor-vehicle crashes: Case-crossover study. The Lancet 361 (9376), 2177-2182.

Rehmat, S., Sadeghnejad, A., Azizinamini, A., 2019a. Connection between concrete filled tube (cft) columns and prefabricated elements using uhpc. In: Proceedings of the International Interactive Symposium on Ultra-High Performance Concrete.

Rehmat, S., Sadeghnejad, A., Azizinamini, A., 2019b. Prefabricated uhpc structural formwork for cap beams. In: Proceedings of the International Interactive Symposium on Ultra-High Performance Concrete.

Rezaie Moghaddam, F., Afandizadeh, S., Ziyadi, M., 2011. Prediction of accident severity using artificial neural networks. International Journal of Civil Engineering 9 (1), 41-48.

S4a, 2018. Signal four analytics. The GeoPlan Center, Department of Urban & Regional Planning, University of Florida.

Sadeghnejad, A., Rehmat, S., Mantawy, I.M., Azizinamini, A., 2020. Comparative study of cyclic and shake table tests for simple for dead load and continuous for live load steel bridge system in seismic area. Transportation Research Record, 0361198120921853.

Sadeghnejad, A., Taghinezhadbilondy, R., Azizinamini, A., 2019. Seismic performance of a new connection detail in an sdcl steel bridge system. Journal of Bridge Engineering 24 (10), 04019094.

Seraneeprakarn, P., Huang, S., Shankar, V., Mannering, F., Venkataraman, N., Milton, J., 2017. Occupant injury severities in hybrid-vehicle involved crashes: A random parameters approach with heterogeneity in means and variances. Analytic Methods in Accident Research 15, 41-55.

Shamshiripour, A., Samimi, A., 2019. Estimating a mixed-profile mdcev: Case of daily activity type and duration. Transportation Letters 11 (6), 289-302.

Sharifi, M.S., Song, Z., Esfahani, H.N., Christensen, K., 2020. Exploring heterogeneous pedestrian stream characteristics at walking facilities with different angle intersections. Physica A: Statistical Mechanics and its Applications 540, 123112.

Shen, L., Chen, H., Yu, Z., Kang, W., Zhang, B., Li, H., Yang, B., Liu, D., 2016. Evolving support vector machines using fruit fly optimization for medical data classification. Knowledge-Based Systems 96, 61-75.

Shi, Q., Abdel-Aty, M., 2015. Big data applications in real-time traffic operation and safety monitoring and improvement on urban expressways. Transportation Research Part C: Emerging Technologies 58, 380-394.

Srinivasan, R., Ullman, G., Finley, M., Council, F., 2011. Use of empirical bayesian methods to estimate crash modification factors for daytime versus nighttime work zones. Transportation research record 2241 (1), 29-38.

Srinivasan, S., Carrick, G., Zhu, X., Heaslip, K., Washburn, S., 2007, Analysis of crashes in freeway work-zone queues: A case study. Southeastern Transportation Center, University of Tennessee, Knoxville, Tennessee, Final Report: 07-UF-R-S3.

Tabibi, M., Koloushani, M., Fatemi, A., 2016. Introducing automatic control method of speed limit signs safety standards before horizontal curves via mobile mapping system. Amirkabir Journal of Civil Engineering 48 (3), 301-3014.

Taghipour, H., Parsa, A.B., Mohammadian, A.K., 2020. Application of machine learning techniques in short-term travel time prediction using multiple data sources. engrxiv.org.

Taghiyeh, S., Lengacher, D.C., Handfield, R.B., 2020. Loss rate forecasting framework based on macroeconomic changes: Application to us credit card industry. Expert Systems with Applications, 113954.

Taghiyeh, S., Xu, J., 2016. A new particle swarm optimization algorithm for noisy optimization problems. Swarm Intelligence 10 (3), 161-192.

Timeanddate, Sunrise and sunset in florida, united states. Timeanddate.com, https://www.timeanddate.com/astronomy/usa/florida, Accessed Jannuary 10, 2019.

Train, K.E., 2009. Discrete choice methods with simulation Cambridge university press.

Ulak, M.B., Ozguven, E.E., Karabag, H.H., Ghorbanzadeh, M., Moses, R., Dulebenets, M., 2020. Development of safety performance functions for restricted crossing u-turn intersections. Journal of Transportation Engineering, Part A: Systems 146 (6), 04020038.

Ullman, G.L., Finley, M.D., Bryden, J.E., Srinivasan, R., Council, F.M., 2008, Traffic safety evaluation of nighttime and daytime work zones. NCHRP Report, Project 17-30.

Vapnik, V., 1998. Statistical learning theory, John Wiley & Sons, New York, NY.

Wang, J., Hughes, W.E., Council, F.M., Paniati, J.F., 1996. Investigation of highway work zone crashes: What we know and what we don't know. Transportation Research Record 1529 (1), 54-62.

Wang, J., Wu, X., Zhang, C., 2005. Support vector machines based on k-means clustering for real-time business intelligence systems. International Journal of Business Intelligence and Data Mining 1 (1), 54-64.

Washington, S.P., Karlaftis, M.G., Mannering, F., 2010. Statistical and econometric methods for transportation data analysis (2nd edition), Chapman and Hall/CRC, Boca Raton, FL.

Wei, X., Shu, X., Huang, B., Taylor, E.L., Chen, H., 2017. Analyzing traffic crash severity in work zones under different light conditions. Journal of advanced transportation 2017.

Weng, J., Meng, Q., 2011. Analysis of driver casualty risk for different work zone types. Accident Analysis & Prevention 43 (5), 1811-1817.

Xie, Y., Zhao, K., Huynh, N., 2012. Analysis of driver injury severity in rural single-vehicle crashes. Accident Analysis & Prevention 47, 36-44.

Xiong, Y., Tobias, J.L., Mannering, F.L., 2014. The analysis of vehicle crash injury-severity data: A markov switching approach with road-segment heterogeneity. Transportation research part B: methodological 67, 109-128.

Yang, H., Ozbay, K., Ozturk, O., Xie, K., 2015. Work zone safety analysis and modeling: A state-of-the-art review. Traffic injury prevention 16 (4), 387-396.

Yang, X.-S., 2010. Suash,"" engineering optimisation by cuckoo search",". International Journal of Mathematical Modelling and Numerical Optimisation 1 (4), 330-343.

Yang, X.-S., Deb, S., 2009. Cuckoo search via lévy flights. In: Proceedings of the 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC), pp. 210-214.

Yasmin, S., Eluru, N., Bhat, C.R., Tay, R., 2014. A latent segmentation based generalized ordered logit model to examine factors influencing driver injury severity. Analytic methods in accident research 1, 23-38.

Ye, F., Lord, D., 2014. Comparing three commonly used crash severity models on sample size requirements: Multinomial logit, ordered probit and mixed logit models. Analytic methods in accident research 1, 72-85.

Ye, X., Wang, K., Zou, Y., Lord, D., 2018. A semi-nonparametric poisson regression model for analyzing motor vehicle crash data. PloS one 13 (5).

Yu, P.-S., Chen, S.-T., Chang, I.-F., 2006. Support vector regression for real-time flood stage forecasting. Journal of Hydrology 328 (3-4), 704-716.

Yu, R., Abdel-Aty, M., 2013. Utilizing support vector machine in real-time crash risk evaluation. Accident Analysis & Prevention 51, 252-259.

Yu, R., Abdel-Aty, M., 2014. Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. Safety science 63, 50-56.

Zeng, Q., Huang, H., 2014. A stable and optimized neural network model for crash injury severity prediction. Accident Analysis & Prevention 73, 351-358.

Zeng, Q., Huang, H., Pei, X., Wong, S., Gao, M., 2016. Rule extraction from an optimized neural network for traffic crash frequency modeling. Accident Analysis & Prevention 97, 87-95.