3-25-2021

Evaluation of Parametric and Nonparametric Statistical Models in Wrong-way Driving Crash Severity Prediction

Sajidur Rahman Nafis
Florida International University, snafi002@fiu.edu

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

EVALUATION OF PARAMETRIC AND NONPARAMETRIC STATISTICAL MODELS IN WRONG-WAY DRIVING CRASH SEVERITY PREDICTION

A dissertation submitted in partial fulfillment of

the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

CIVIL ENGINEERING

by

Sajidur Rahman Nafis

2021

To: Dean John L. Volakis
College of Engineering and Computing

This dissertation, written by Sajidur Rahman Nafis, and entitled Evaluation of Parametric and Nonparametric Statistical Models in Wrong-Way Driving Crash Severity Prediction, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

_____
Albert Gan

_____
Mohammed Hadi

_____
Xia Jin

_____
B. M. Golam Kibria

_____
Priyanka Alluri, Major Professor

Date of Defense: March 25, 2021

The dissertation of Sajidur Rahman Nafis is approved.

_____
Dean John L. Volakis
College of Engineering and Computing

_____
Andrés G. Gil
Vice President for Research and Economics Development
and Dean of the University Graduate School

Florida International University, 2021

DEDICATION

To my parents, Dr. Monzur Rahman and Ms. Azra Karim, my siblings Tausif and Faria, and my wife, Tarannum Islam, for their unconditional love and support.

ACKNOWLEDGMENT

Foremost, I thank the Almighty God, I could not have accomplished this work without His blessings.

I would like to convey my immense gratitude and appreciation to my supervisor Dr. Priyanka Alluri for her dedication and unmeasurable support throughout my research. She has been an excellent supervisor. This research would not have been possible without her guidance and constant supervision. She has always shared her exceptional engineering knowledge with me and inoculated the meaning of good values of life in me.

I would like to express my heartfelt appreciation to my committee members, Dr. Albert Gan, Dr. Mohammed Hadi, Dr. Xia Jin, for advising me whenever I needed, showing interest in my research work, and never hesitating to provide insight and guidance throughout this journey.

Sincere appreciation goes to Dr. B. M. Golam Kibria, who taught me statistics. He always made himself available for my thesis-related discussion and provided constructive criticism and suggestions.

I am appreciative of the collaborative and supportive work of everyone who assisted me in my dissertation. Special thanks to Dr. Wensong Wu, Dr. Wanyang Wu, Dr. Yan Xiao. I would also like to thank my long-time friend, Dr. Arif Istiaque Shuvo, Dr. Tarique Hasan Khan, and Mr. Henrick Haule for supporting and guiding me throughout my journey. Sincere appreciation to Ms. Tarannum Islam for assisting me whenever I needed it. It is my pleasure to thank my research laboratory colleagues and research co-authors for challenging me to work hard and strive for excellence.

I would like to recognize the financial support from Florida International University's Department of Civil and Environmental Engineering and the Florida Department of Transportation. Their support helped me accomplish my doctoral degree and prepare me for my professional journey ahead.

To my family and friends, thank you for always loving, encouraging, and cheering on me on each milestone that I achieved.

ABSTRACT OF THE DISSERTATION

EVALUATION OF PARAMETRIC AND NONPARAMETRIC STATISTICAL MODELS IN WRONG-WAY DRIVING CRASH SEVERITY PREDICTION

by

Sajidur Rahman Nafis

Florida International University, 2021

Miami, Florida

Professor Priyanka Alluri, Major Professor

Wrong-way driving (WWD) crashes result in more fatalities per crash, involve more vehicles, and cause extended road closures compared to other types of crashes. Although crashes involving wrong-way drivers are relatively few, they often lead to fatalities and serious injuries. Researchers have been using parametric statistical models to identify factors that affect WWD crash severity. However, these parametric models are generally based on several assumptions, and the results could generate numerous errors and become questionable when these assumptions are violated. On the other hand, nonparametric methods such as data mining or machine learning techniques do not use a predetermined functional form, can address the correlation problem among independent variables, display results graphically, and simplify the potential complex relationship between the variables.

The main objective of this research was to demonstrate the applicability of nonparametric statistical models in successfully identifying factors affecting traffic crash severity. To achieve this goal, the performance of parametric and nonparametric statistical

models in WWD crash severity prediction was evaluated. The following parametric methods were evaluated: Logistic Regression (LR), Ridge Regression (RR), Least Absolute Shrinkage and Selection Operator (LASSO), Linear Discriminant Analysis (LDA), and Gaussian Naïve Bayes (GNB). The following nonparametric methods were evaluated: Random Forests (RF), Decision Trees (DT), and Support Vector Machine (SVM). The evaluation was based on sensitivity, specificity, and prediction accuracy. The research also demonstrated the applicability of nonparametric supervised learning algorithms on crash severity analysis by combining tree-based data mining techniques and marginal effect analysis to show the correlation between the response and the predictor variables.

The analysis was based on 1,475 WWD crashes that occurred on arterial road networks from 2012-2016 in Florida. The results showed that nonparametric models provided better prediction accuracy on predicting serious injury compared to parametric models. By conducting prediction accuracy comparison, contributor variables' marginal effect analysis, variable importance evaluation, and crash severity pattern recognition analysis, the nonparametric models have been demonstrated to be valid and proved to serve as an alternative tool in transportation safety studies.

The results showed that head-on collisions, weekends, high-speed facilities, crashes involving vehicles entering from a driveway, dark-not lighted roadways, older drivers, and driver impairment are important factors that play a crucial role in WWD crash severity on non-limited access facilities. This information may assist researchers and safety engineers in identifying specific strategies to reduce the severity of WWD crashes on arterial streets.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

AADT	Annual Average Daily Traffic
ACS	American Community Survey
ADOT	Arizona Department of Transportation
AHC	Agglomerative Hierarchical Clustering
AM	Access Management
ARBM	All Roads Base Map
BAC	Blood Alcohol Content
CALTRANS	California Department of Transportation
CARS	Cash Analysis Reporting System
CART	Classification and Regression Trees
CFX	Central Florida Expressway Authority
CG	Comparison Group
CV	Connected Vehicle
CVS	Connected Vehicle System
DHSMV	Department of Highway Safety and Motor Vehicles
DMS	Dynamic Message Signs
DS	Descriptive Statistics
DOT	Department of Transportation
DR	Decision Rules
DT	Decision Tree
DUI	Driving Under the Influence

FARS	Fatality Analysis Reporting System
FB	Full Bayes
FDOT	Florida Department of Transportation
FGDL	Florida Geographic Data Library
FN	False Negative
FP	False Positive
FPL-LR	Fifth's Penalized-Likelihood Logistic Regression
GIS	Geographic Information System
GLM	Generalized Linear Regressions
GOL	Generalized Ordered Logit
GPS	Global Positioning System
GNB	Gaussian Naïve Bayes
IDOT	Illinois Department of Transportation
IIRPM	Internally Illuminated Raised Pavement Marker
ITS	Intelligent Transportation Systems
KDE	Kernel Density Estimation
LASSO	Least Absolute Shrinkage and Selection Operator
LED	Light-Emitting Diode
LDA	Linear Discriminant Analysis
LR	Logistics Regression
MCA	Multiple Correspondence Analysis
ML	Machine Learning
MUTCD	Manual on Uniform Traffic Control Devices

NB	Naïve Bayes
NHTSA	National Highway Traffic Safety Administration
NTSB	National Transportation Safety Board
OOB	Out-Of-Bag
PDO	Property Damage Only
PDP	Partial Dependence Plot
RCI	Roadway Characteristics Inventory
RF	Random Forest
RIDOT	Rhode Island Department of Transportation
ROC	Receiver Operator Curve
RR	Ridge Regression
RRFB	Rectangular Rapid Flashing Beacon
S&PM	Signing and Pavement Markings
SVM	Support Vector Machine
TMC	Transportation Management Center
TN	True Negative
TP	True Positive
TSC	Traffic Safety Culture
TSM&O	Transportation Systems Management & Operations
TTI	Texas Transportation Institute
TWLTL	Two-way Left-Turn Lane
TxDOT	Texas Department of Transportation

UBR	Unified Base-map Repository
V2I	Vehicle-to-Infrastructure
V2V	Vehicle-to-Vehicle
WWD	Wrong-Way Driving

CHAPTER 1
INTRODUCTION

1.1 Background

A wrong-way driving (WWD) incident involves a vehicle traveling opposite to the legal flow of traffic on a direction-separated highway, freeway or arterial, or access ramp (NTSB, 2012). Annually, WWD crashes result in about 350 fatalities nationwide, making up 3% of all crashes that occur on high-speed divided highways (NTSB, 2012). Although crashes involving wrong-way drivers are relatively few, they often lead to severe head-on collisions. As such, the fatality rate in WWD incidents is much higher compared to other crashes, often causing death or incapacitating injuries (Zhou et al., 2012). WWD fatal crash rate was found to be 1.34 fatalities per fatal crash, while the rate was found to be 1.10 fatalities per fatal crash for all types of crashes (Pour-Rouholamin et al., 2016).

The majority of previous studies concerning WWD crashes have focused on freeways. This could be potentially because they draw more media attention, involve more vehicles, cause extended freeway closures, and result in more fatalities per fatal crash. Although WWD crashes on limited-access facilities receive more attention, WWD crashes are more frequent on arterial streets compared to freeways (Ponnaluri, 2018), requiring special attention. Moreover, the characteristics and the analysis procedures of WWD crashes on arterials might be different from the analysis of WWD crashes on freeways.

Mitigating WWD crashes on arterials is complicated because there are multiple access points along with arterial facilities. In other words, there are many possible locations where a driver may enter the facility the wrong-way, and it is difficult to have some type

of WWD countermeasure(s) at each of these access points. Furthermore, preventing WWD crashes becomes more difficult as they are rare and random.

In previous studies, descriptive statistics, generalized linear regression models, and parametric statistical techniques have often been used to analyze WWD crashes and identify influential factors. These studies play a pivotal role in the development of countermeasures for WWD crashes. However, a major limitation with linear regression models is that they use a linear relationship between WWD crash severity and the influential variables, leading to inaccurate injury severity estimations (Mussone et al., 1999). Parametric techniques such as generalized linear regressions (GLMs) make assumptions about the dependent and independent variables, which may not always be correct. Data mining techniques can be resourceful in this case. Unlike common regression models, data mining techniques have the ability to identify and explain the complex patterns associated with crash risk without having a functional form and predetermined assumptions (Kashani et al., 2011). Although data mining techniques have been used in many studies, their applications on WWD crash severity analysis are very limited. In addition, while most previous studies address WWD crashes on freeways, the analysis of WWD crashes on arterials or non-limited access facilities has not been conducted in-depth.

1.2 Problem Statement

According to the National Transportation Safety Board (NTSB), WWD is defined as “vehicular movement along a travel lane of a roadway in a direction opposing the legal flow of traffic on high-speed divided highways or access ramps.” (NTSB, 2012). About 3% of all crashes that occur on high-speed divided highways involve wrong-way drivers,

and most of these crashes result in fatal or severe injuries (NTSB, 2012). For instance, Zhou et al. (2016) reported that each WWD fatal crash results in 1.4 fatalities and 2.1 incapacitating injuries. Pour-Rouholamin et al. (2016) analyzed WWD crashes from the Fatality Analysis Reporting System (FARS) database for ten years, from 2004 to 2013, in the U.S. and found that on average, the 265 fatal WWD crashes that occurred on controlled-access highways resulted in 355 fatalities, at a rate of 1.34 deaths per WWD fatal crash. This rate is quite high compared to the fatality rate of 1.10 for all other crash types on controlled-access highways. In addition to WWD crashes on freeways, there is a large portion of WWD crashes that occur on the arterial road network. WWD crashes on freeways, although relatively more severe, constitute only a small fraction of all WWD crashes on the state highway system (Ponnaluri & Heery, 2016). WWD crashes on arterials are more frequent. The possibility of a WWD crash on arterials was found to be 2.3 times higher than on freeways (Ponnaluri, 2018). For instance, from 2011-2015, of the 6,888 WWD crashes that occurred on the public road network in Florida, only 4% (i.e., 281 crashes) occurred on freeways, while the remaining 96% (i.e., 6,607 WWD crashes) occurred on non-limited access facilities (Alluri et al., 2019). These statistics warrant the need to analyze and address WWD crashes on non-limited access facilities.

Mitigating WWD crashes is challenging, especially on the arterial network. This is because there are multiple access points along arterial roadways. In addition to this, another challenge in WWD crash analysis is that WWD crash data on arterials are heterogeneous in nature due to the variations in the roadway geometry and less homogeneous road sections. Moreover, missing values for some of the factors can make WWD crash data analysis extremely difficult. Since the crash data are heterogeneous by nature (Karlaftis &

Tarko, 1998), certain critical relationships useful for influencing the cause of crashes may remain hidden if they are not sectioned in subsets (Depaire et al., 2008). Parametric models such as generalized linear models (GLMs) are the most popular models favored by analysts and many researchers as they produce easily interpretable functional forms by establishing a quantitative relationship between the response variable and the explanatory features (PSECS, 2017). However, GLMs are based on several assumptions, and the results could generate numerous errors and become questionable when these assumptions are violated (Zheng, 2018). Acuna and Rodrigues (2004) concluded that missing data samples affect statistical-based algorithms, and nonparametric classifiers perform better than parametric classifiers for datasets with missing values (Acuna & Rodriguez, 2004). If these issues are not addressed well, the presence of heterogeneity in the dataset may lead to biased results (Karlaftis & Tarko, 1998). The authors found that models based on clustered heterogeneous datasets yield improved and accurate results compared to the models based on the pooled heterogeneous dataset. Researchers have recently been using data mining techniques to understand the factors contributing to crash severity (Kuhnert et al., 2000; Sohn & Shin, 2001; Chang & Wang, 2006; Kashani et al., 2011; Pakgohar et al., 2011). Data mining procedures use artificial intelligence and statistical analyses to extract interpretable knowledge from databases. Unlike the regular regression models, data mining techniques have the ability to identify and explain the intricate patterns associated with crash risk without needing to use a functional form (Kashani et al., 2011). Although data mining is a powerful technique, it is often overlooked by transportation researchers due to difficulty in interpreting its results. Data mining techniques could yield results with higher accuracy but lack interpretability and act as a black box (Shmueli, 2010; Zheng, 2018). For instance,

when using black-box machine learning (ML) algorithms such as the Support Vector Machine (SVM) or Random Forests (RF), it is hard to comprehend the relations between predictors and the model outcome. Enhancing data mining techniques' interpretability could increase their acceptance in transportation safety research (Cortez & Embrechts, 2011).

1.3 Research Goal and Objectives

The objectives of this research are to: (a) demonstrate the applicability of nonparametric data mining techniques by comparing the prediction performance of parametric and nonparametric statistical models, and (b) identify factors that affect serious WWD crash injuries on arterial roadways using nonparametric techniques.

First, the prediction accuracy of five parametric models and three nonparametric are evaluated in arterial WWD crash severity analysis. All the models are also compared based on classification sensitivity and specificity. In addition, the marginal effect of the nonparametric model is demonstrated to show the applicability of nonparametric techniques in predicting the correlation between predictor-response features. Next, the combination of the following three data mining models is used to identify the pattern of the influential contributing factors that affect the arterial WWD crash injury severity: Agglomerative Hierarchical Clustering (AHC), Random Forests (RF), and Decision Tree (DT).

1.4 Dissertation Organization

The remaining chapters of this dissertation are organized as follows:

- Chapter 2 entails a comprehensive synthesis of the literature on WWD crash analysis. The chapter discusses the existing studies on WWD crashes on limited access facilities and previous study methods used in predicting WWD crash severity and frequency. It discusses the existing studies on WWD crashes on arterials. The chapter also covers statewide studies on WWD mitigation and countermeasure implementation. In the end, the chapter includes a discussion on nonparametric methodologies used in crash data analysis.
- Chapter 3 discusses the methodologies used to achieve the research objectives.
- Chapter 4 discusses the data used to achieve the research goal in this research. Specially, the chapter discusses, in detail, the types of data used, data sources, data collection strategy, and data processing steps.
- Chapter 5 presents the analyses and discusses the results. The descriptive statistics are first discussed, followed by the comparison of all the prediction methods implemented in this research. Finally, the factors contributing to the WWD crash severity are discussed.
- Chapter 6 concludes this dissertation by presenting a summary of this research, discussion, contributions, and recommendations for future research.

CHAPTER 2
LITERATURE REVIEW

This chapter provides a comprehensive literature review on the following five topics: (a) WWD crashes on limited-access facilities; (b) WWD crashes on non-limited access facilities; (c) state-of-the-practice in WWD mitigation; (d) WWD countermeasures; and (e) existing methodologies in the context of safety. Section 2.1 focuses on the different risk factors, causes, patterns, and contributing factors associated with WWD crashes on limited-access facilities. Section 2.2 discusses studies on WWD crashes on non-limited access facilities. Section 2.3 discusses national practices in mitigating WWD incidents. Section 2.4 includes studies on traditional and innovative WWD countermeasures. Finally, Section 2.5 includes a discussion on nonparametric methodologies and the data mining techniques used in crash data analysis.

2.1 Existing Studies on WWD Crashes on Limited-access Facilities

WWD crashes have a higher propensity to result in fatal and severe injuries. In the United States, WWD crashes result in 300-400 fatalities every year (Moler, 2002). Wrong-way drivers on freeways pose a serious risk to the safety of themselves and other motorists. Although crashes involving wrong-way drivers are relatively few, they often lead to severe head-on collisions. As such, the fatality rate in WWD incidents is much higher compared to other crashes.

Mitigation of WWD incidents has therefore been on the national front, with particular emphasis being given to identifying practical and proven countermeasures. However, before implement countermeasures, it is required to understand the causes of

WWD crashes and factors influencing the frequency and severity of these crashes. The factors influencing WWD crashes are divided into the following three broad categories:

- Demographic and socioeconomic factors
- Roadway geometric factors
- Temporal factors

2.1.1 Demographic and Socioeconomic Factors

WWD incidents were found to be affected by several demographic and socioeconomic factors, including age, gender, socioeconomic background, etc. Table 2-1 summarizes the results from several studies that evaluated the impact of demographic factors on WWD incidents. For each study, the table also provides the specific demographic factors identified, the study period, the study region, and the analysis approach.

Table 2-1: Demographic Factors Affecting WWD Crashes

Demographic Factors	Study Period	State	Method	Reference
Impaired driver	1967–1970	Texas	DS	Friebele et al., 1971
Intoxicated drivers; Urban areas	1983–1987	California	DS	Copelan, 1989
Male drivers; Drivers less than 34 years old; Intoxicated drivers; Urban areas	1997–2000	Texas	DS	Cooner et al., 2004a
Alcohol-related; Younger drivers; Older drivers; Interstate routes; Rural areas	2000–2005	North Carolina	DS	Braam, 2006
Younger drivers; Intoxicated drivers; Older drivers; Female drivers	2003–2005	Switzerland	DS	Scaramuzza & Cavegn, 2007
Older drivers; Younger drivers; Inexperienced drivers; Intoxicated drivers	1996–1998	Netherlands	DS	SWOV, 2009
Intoxicated drivers; Older drivers; Male drivers; Passenger cars; Non-Hispanic and native Americans	1990–2004	New Mexico	CG	Lathrop et al., 2010
Intoxicated drivers; Younger drivers; Older drivers; Male drivers	2005–2009	Michigan	DS	Morena & Leix, 2012
Younger drivers (16–24 years); Male drivers; Impaired drivers	2007–2011	Texas	DS	Finley et al., 2014
Older drivers; Younger drivers; Dementia	2005–2009	Japan	DS	Xing, 2014
Older drivers; Younger drivers; Male drivers; Local drivers; Intoxicated drivers; Urban areas; Passenger cars; Single-occupant vehicles	2004–2009	Illinois	DS	Zhou et al., 2015
Older drivers; Intoxicated drivers; Physically impaired drivers; Driver residency distance (local drivers); Vehicles older than 15 years; Months of March, May, and November	2009–2013	Alabama	FPL-LR	Pour-Rouholamin et al., 2014
Older drivers; Intoxicated drivers; Local drivers; Driving older vehicles; Passenger cars; Single-occupant vehicles; Unlicensed drivers	2009–2012	France	LR	Kemel, 2015
Older drivers; Intoxicated drivers; Physically impaired drivers; Driver residency distance (local drivers); Vehicles older than 15 years	2009–2013	Alabama	GOL	Pour-Rouholamin & Zhou, 2016
Impaired drivers; Younger drivers	2009–2013	Florida	DS	FDOT, 2015
Driver age; Driver gender; Driver condition (eyesight, fatigue, illness, seizure, epilepsy); Intoxicated drivers; Urban areas; Vehicle use	2003–2010	Florida	LR	Ponnaluri, 2016
Urban areas; Driver impairment; Male drivers	2004-2014	Arizona	DS	Simpson & Bruggeman, 2015
driver age, driver condition, roadway surface conditions, and lighting conditions	04−13 IL 09-13Al	Alabama	MCA	Jalayer et al., 2018a
Older drivers; Impaired drivers; Urban areas (frequent WWD crashes); Rural areas (severe WWD crashes)	2009-2013	Alabama	FPL-LR	Zhang et al., 2017

Note: DS: Descriptive Statistics; CG: Comparison Group; GOL: Generalized Ordered Logit; FPL-LR: Firth's Penalized-Likelihood Logistic Regression; LR: Logistic Regression; MCA: Multiple Correspondence Analysis; RPOPM: Random-Parameters Ordered Probit Model.

2.1.2 Roadway Geometric Factors

In addition to demographic and socioeconomic factors, roadway geometric factors also affect WWD incidents. Table 2-2 summarizes the results from several studies that evaluated the impact of roadway geometric factors on WWD incidents. For each study, the table also provides the specific roadway geometric factors analyzed, the study period, the study region, and the analysis approach.

Table 2-2: Roadway Geometric Factors Affecting WWD Crashes

Geometric Factors	Study Period	State	Method	Reference
Entrance by exit ramp; Diamond interchange; Partial interchange; Less than 1,000 feet of sight distance; Improper signing	1967–1970	Texas	DS	Friebele et al., 1971
Interchanges with short sight distance; Partial cloverleaf interchanges; Half and full diamond interchanges; Trumpet interchanges; Slip ramps; Buttonhook ramps; Scissors exit ramp; Left-side exit ramp; Five-legged intersections near exit ramps	1983–1987	California	DS	Copelan, 1989
Left-side exit ramps; One-way street transitioned into freeway	1997–2000	Texas	DS	Cooner et al., 2004a
Two-quadrant parclo interchanges; Full diamond interchanges	2000–2005	North Carolina	DS	Braam, 2006
Parclo interchanges; Trumpet interchanges; Tight diamond interchanges	2005–2009	Michigan	DS	Morena & Leix, 2012
Type of interchange; Making a U-turn on the carriageway	2005–2009	Japan	DS	Xing, 2014
Type of interchange	2004–2009	Illinois	DS	Zhou et al., 2015
Roadway condition	2009–2013	Alabama	FPL-LR	Pour-Rouholamin et al., 2014
The distance from the ramp median to the left-turn stop line on a crossroad	2004–2013	Illinois	DS	Wang et al., 2017

Note: DS: Descriptive Statistics; CG: Comparison Group; GOL: Generalized Ordered Logit; FPL-LR: Firth's Penalized-Likelihood Logistic Regression; LR: Logistic Regression; MCA: Multiple Correspondence Analysis; RPOPM: Random-Parameters Ordered Probit Model.

2.1.3 Temporal Factors

Table 2-3 summarizes the results from several studies that evaluated the impact of temporal factors on WWD incidents. For each study, the table also provides the specific demographic factors identified, the study period, the study region, and the analysis method.

Table 2-3: WWD Studies on Temporal Factors

Temporal Factors	Study Period	State	Method	Reference
Darkness; Time of the day	1983–1987	California	DS	Copelan, 1989
Early morning hours	1997–2000	Texas	DS	Cooner et al., 2004a
Time of day (midnight to 5:59 a.m.); Months of February and June	2000–2005	North Carolina	DS	Braam, 2006
Time of day; Lighting condition	2003–2005	Switzerland	DS	Scaramuzza and Cavegn, 2007
Darkness; Month of November; Non-Hispanic and native Americans	1990–2004	New Mexico	CG	Lathrop et al., 2010
Darkness; Time of the day (late night and early morning)	2005–2009	Michigan	DS	Morena and Leix, 2012
Time of day (7:00 p.m. to 12:00 p.m.)	2007–2011	Texas	DS	Finley et al., 2014
Darkness; Time of day (4:00 p.m. to 10:00 p.m.)	2005–2009	Japan	DS	Xing, 2014
Weekends; Darkness; Time of day (midnight to 5:00 a.m.)	2004–2009	Illinois	DS	Zhou et al., 2015
Time of day (evening and afternoon); Months of March, May, and November	2009–2013	Alabama	FPL-LR	Pour-Rouholamin et al., 2014
Darkness	2009–2012	France	LR	Kemel, 2015
Time of day (evening and afternoon); Months of March, May, and November	2009–2013	Alabama	GOL	Pour-Rouholamin and Zhou, 2016
Months of January through April, June, and July; Weekends; Darkness	2009–2013	Florida	DS	FDOT, 2015
Time of day; Darkness	2003–2010	Florida	LR	Ponnaluri, 2016a
Night-time; Weekends	2004-2014	Arizona	DS	Simpson et al., 2015
Darkness	2009–2012	France	LR	Kemel, 2015
Dark roadways with no lighting	2009-2013	Alabama	DS; FPL-LR	Zhang et al., 2017

Note: DS: Descriptive Statistics; CG: Comparison Group; GOL: Generalized Ordered Logit; FPL-LR: Firth's Penalized-Likelihood Logistic Regression; LR: Logistic Regression; MCA: Multiple Correspondence Analysis; RPOPM: Random-Parameters Ordered Probit Model.

2.2 Existing Studies on WWD Crashes on Non-limited Access Facilities

To date, there has been a lot of research on addressing WWD on freeways (Copelan, 1989; Cooner et al., 2004a; Braam, 2006; Lathrop et al., 2010; Finley et al., 2014; Rogers et al., 2016; Alluri et al., 2018a), while there are very few studies that analyzed WWD incidents on non-limited access facilities. Although WWD crashes on limited-access facilities get more attention, WWD crashes are more frequent on arterial streets compared to freeways. In 1973, Vaswani conducted one of the first studies that analyzed and compared WWD crashes on arterials and freeways (Vaswani, 1973). When WWD crashes on limited-access facilities were considered, the fatality and injury rates were found to be 0.47 and 1.18, respectively. On the other hand, the fatality and injury rates of all crashes on limited-access facilities were found to be 0.016 and 0.42, respectively. When WWD crashes on non-limited access facilities were considered, the fatality and injury rates were found to be 0.22 and 1.03, respectively. On the other hand, the fatality and injury rates of all crashes on non-limited access facilities were found to be 0.016 and 0.42, respectively. On non-limited access facilities, the study concluded that the fatality and injury rates of WWD crashes were about 2.8 times and 2.2 times, respectively, more than the fatality and injury rates in non-WWD crashes.

More recently, Ponnaluri (2016b) compared the probabilities of WWD crashes on freeways and arterials. Ponnaluri first surveyed the transport professional groups and the general road users. WWD crash data was next analyzed, and finally, the crash data analysis results were compared to the survey results. Based on 60 survey responses (30 each from the transport professional group and the general road users), WWD crashes on arterials were found to be two times more frequent than WWD crashes on freeways (odds ratio:

2.16). Ponnaluri (2016b) next analyzed 2003-2010 WWD crash data in Florida using binomial logistic regression. The analysis was based on 999,456 crash records. The data analysis showed that the odds ratio of a WWD crash was 2.29 on arterial roadways (i.e., non-limited access facilities) compared to freeways (i.e., limited-access facilities); these statistics were found to be similar to the survey results. Ponnaluri (2016b) concluded that WWD crashes are more frequent on non-limited access facilities; however, fatal WWD crashes are more frequent on limited-access facilities. In general, the higher proportion of fatal and severe injury crashes on freeways could be attributed to high speeds (Elvik, 2013).

Ponnaluri (2018) extended his previous work on WWD by conducting a more comprehensive evaluation of WWD crashes on arterials and freeways. The main goal of the study was to compare the WWD crashes on arterial corridors with the WWD crashes on freeways and highlight the need for analyzing WWD crashes on arterials. The analysis was based on 999,456 crashes that occurred on both arterials and freeways, of which only 3,823 crashes (i.e., 3.84%) were categorized as WWD crashes. Ponnaluri (2018) used a stepwise regression model to identify statistically significant covariates at a 5% significance level. Males were found to be 1.3 times more prone to WWD crashes than females. However, exposure was not considered in the analysis. Younger drivers aged 21-40 years were found to be more likely to get involved in WWD crashes; older drivers aged over 80 years were also found to be prone to WWD crashes, especially on freeways. The likelihood of WWD crashes on arterials was found to increase when the driver is not from Florida (i.e., tourists). Consistent with other WWD studies, this research also showed that WWD fatalities are higher for intoxicated drivers. Alcohol-related fatal WWD crashes were found to be more prominent on weekends, especially on Saturdays. As expected,

WWD crashes were found to be more frequent between 6 pm and 6 am; adequate street lighting could potentially help reduce WWD crashes.

2.3 State-of-the-Practice in WWD Mitigation

WWD crashes have a higher propensity to result in fatal and severe injuries. As such, several states and federal organizations have been working hard to mitigate WWD crashes. A majority of the efforts focused on identifying contributing factors and developing effective countermeasures. Numerous states, including Florida, Texas, California, Illinois, and Arizona, have become pioneers in mitigating WWD incidents.

2.3.1 National Effort

WWD mitigation has been on the national front, with particular emphasis being given to identifying practical and proven countermeasures. These countermeasures could be divided into three broad categories:

- countermeasures that address WWD driver-related factors
- countermeasures that improve highway geometric conditions
- countermeasures that provide WWD navigation alerts on vehicles

2.3.1.1 Driver

A majority of at-fault drivers involved in WWD crashes are either alcohol/drug-impaired or are older drivers. This observation was influenced by the fact that seven out of the nine WWD drivers investigated by NTSB in 2012 had Blood Alcohol Content (BAC) $\geq$ 0.15 (NTSB, 2012). For alcohol-impaired drivers, the NTSB report recommended considering passive safety devices such as the use of alcohol ignition interlock devices and new in-vehicle alcohol detection technologies. Considering the fact that older drivers are

over-represented in fatal WWD crashes, the report also recommended countermeasures focusing on older driver safety. More specifically, NTSB suggested that each individual state in the U.S. develop comprehensive highway safety programs for older drivers that incorporate the program elements outlined in the NHTSA Highway Safety Program Guideline No. 13 - Older Driver Safety.

2.3.1.2 Highways

Improving geometric highway conditions is one of the proven ways to mitigate WWD crashes. The most common initiating event for WWD on controlled-access facilities is entering the mainline traffic lanes from an exit ramp. NTSB (2012) specifically emphasized the use of highway signage and traffic control devices that are designed to direct motorists onto controlled-access highway entrance ramps and discourage wrong-way movement onto ramp exits. These countermeasures aim at addressing factors that may influence WWD crashes due to road geometrics resulting in poor visibility, inadequate traffic control, lack of positive signing, and absence of street lighting. The report also recommended using reduced sign heights, adding red reflective tape to vertical posts, and using oversized wrong-way signs for enhanced visibility. Additionally, the report suggested some countermeasures to mitigate WWD crashes caused by drivers entering the highway using exit ramps. These recommendations include illuminating wrong-way signs which flash when a wrong-way vehicle is detected and installing a second set of wrong-way signs at the exit ramp farther upstream from the crossroads. Other recommendations include the use of channelized striping to guide drivers onto the on-ramp.

2.3.1.3 Vehicle Safety Systems

Providing navigation system alerts that inform drivers of wrong-way movements onto controlled-access highway exit ramps before they reach mainline traffic could enhance safety. As such, using wrong-way navigation alerts on vehicles and emerging technology following vast progress made on in-vehicle technology. These systems will rely on the use of the vehicle's navigation system, combined with the Global Positioning System (GPS). However, "for wrong-way navigation alert systems to be reliable and effective, GPS providers must follow consistent human factors policies in messaging and alerting" (NTSB, 2012).

2.3.2 Florida

FDOT has been a pioneer in addressing the WWD issue. FDOT has begun tackling this issue from several fronts. The Department has focused on developing a policy-specific framework emphasizing continual consultation, coordination, and communication. FDOT has also developed strategic and coordinated research efforts tackling all the issues with WWD incidents and assisting the agencies with developing an implementation strategy to mitigate WWD incidents.

Figure 2-1 represents the FDOT's framework with the backdrop of leadership-supported institutionalization to strategize road safety improvements. This policy-oriented framework aims to "address WWD incidents in a systematic manner and propose a systemic discipline for transforming policy objectives to actionable outcomes." (Ponnaluri, 2016a)

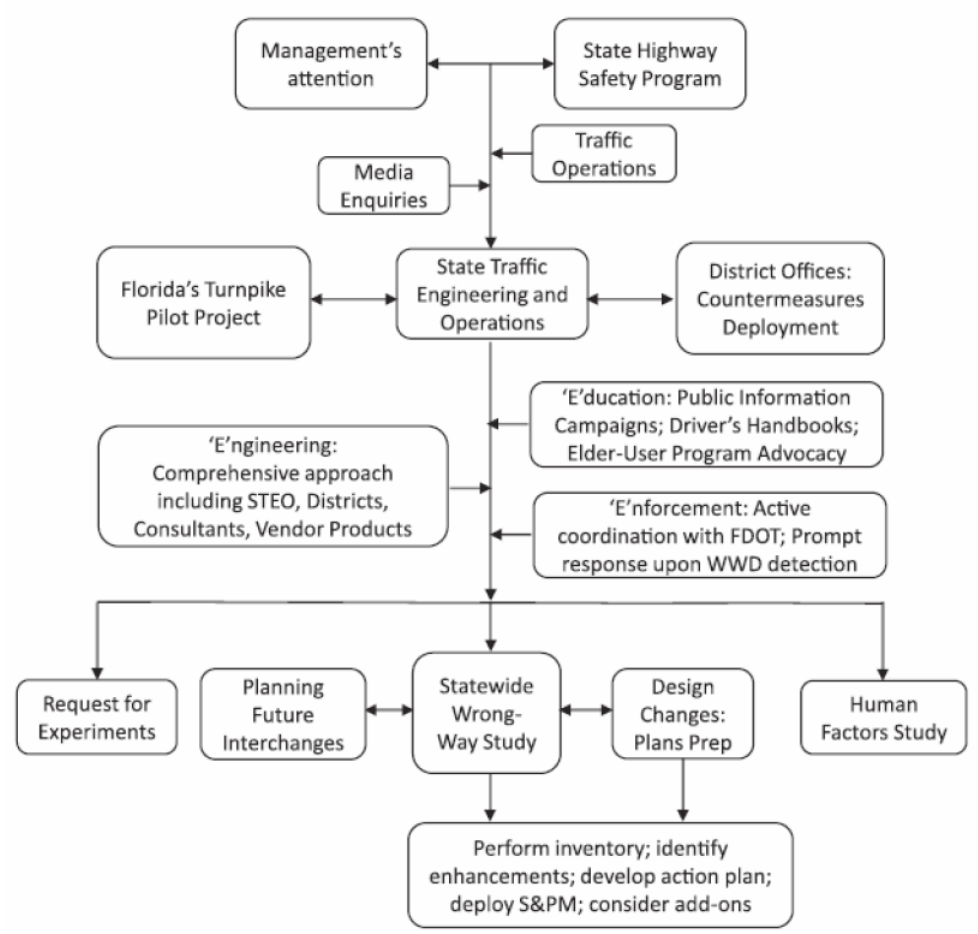


Figure 2-1: FDOT Framework to Mitigate WWD Incidents (Ponnaluri, 2016a)

In 2015, FDOT completed a statewide WWD crash study to understand the factors contributing to WWD crashes (Kittleson & Associates, 2015). In the same year, Boot et al. (2015) conducted a human factors study to understand the role of human cognition in the driver decision-making process. On the deployment front, FDOT Districts have deployed the following pilot countermeasures at WWD incident locations across the state:

- Newly-developed signing and pavement marking standards (FDOT Plans Preparation Manual, Figures 7.1.1. and 7.1.2)
- Detection-triggered Red Rectangular Rapid Flashing Beacons (Red-RRFBs)

- Detection-triggered light-emitting diode (LED) lights around “WRONG-WAY” signs
- Red flush-mount Internally Illuminated Raised Pavement Markers (IIRPMs)
- Detection-triggered blank-out signs that flash “WRONG-WAY.”
- Delineators along off-ramps
- Detection-triggered Wigwag flashing beacons

Most recently, the pilot countermeasures were compared, and a combination of countermeasures was recommended for future deployment consideration (Lin et al., 2017). In addition to the Engineering countermeasures, FDOT has also focused on the other 3Es, i.e., Education, Enforcement, and Emergency Response. For example, FDOT considers July as WWD Awareness Month and works on educating the public regarding tips to follow to avoid being involved in WWD crashes. The “StayRightatNight” campaign urges drivers to avoid a crash with a wrong-way driver and has generated significant interest in social media (DHSMV, 2016).

2.3.3 California

The California Department of Transportation (Caltrans) has been researching and identifying effective WWD countermeasures since the early 1960s (Tamburri, 1965). Several studies have focused on improving the signage, pavement markings, and geometric roadway design where low-mounted DO NOT ENTER signs mounted together with WRONG-WAY signs were recommended (Tamburri, 1965; Doty and Ledbetter, 1965; Rinde, 1978). In addition, Caltrans’s WWD monitoring program was recommended for

identifying locations for WWD crash investigation. WWD crash rate was significantly reduced in California after implementing the research results in the 1970s and 1980s.

2.3.4 Texas

In the early 1970s, researchers at the Texas Transportation Institute (TTI) surveyed the state and local highway engineers and law enforcement personnel in an attempt to qualitatively determine the nature of WWD crashes in Texas (Friebele et al., 1971). In 2003, the Texas Department of Transportation (TxDOT) sponsored a WWD research following several severe WWD crashes across the state. The major findings from the research called for the use of reflectorized wrong-way arrows on exit ramps, lowered DO NOT ENTER and WRONG-WAY signs mounted together on the same sign support, and the development of a field checklist for wrong-way entry problem locations (Cooner et al., 2004a; Cooner et al., 2004b).

Since alcohol was a contributing factor in over one-third of all WWD crashes in Texas, researchers designed and conducted two nighttime closed-course studies to determine where alcohol-impaired drivers look in the forward driving scene, provide insights into how alcohol-impaired drivers recognize and read signs, and finally assess the conspicuity of selected WWD countermeasures from the perspective of alcohol-impaired drivers (Finley et al., 2014). The study findings indicated that alcohol-impaired drivers might tend to look less to the left and right and more at the pavement in front of the vehicle. In addition, researchers confirmed that alcohol-impaired drivers do not actively search the forward driving scene as much as non-impaired drivers. Instead, alcohol-impaired drivers concentrate their glances in a smaller area within the forward driving scene. Researchers

also confirmed that drivers at higher BAC levels took longer to locate signs and must be closer to a sign before they can identify the background color and read the legend. Since alcohol-impaired drivers tend to look more at the pavement in front of the vehicle, researchers recommended that wrong-way arrows should be installed and maintained on all exit ramps on controlled-access highways.

The study also conducted a focus group discussion to obtain motorists' information regarding the design of WWD warning messages on Dynamic Message Signs (DMS). Overall, the majority of the focus group participants thought that the warning message is supposed to have the word DANGER instead of WARNING, WRONG-WAY DRIVER instead of ONCOMING VEHICLE. They also recommended the provision of location information and the approximate time (Finley et al., 2014).

2.3.5 Illinois

In the 1980s, the Illinois Department of Transportation (IDOT) experimented with sensors embedded in the roadway to detect wrong-way traffic movement, which, if activated, would lower a signal arm across the road and initiate a DMS to alert to existing traffic about the WWD hazard ahead (Finley et al., 2014). More recently, Zhou et al. (2012) developed a new method that involved ranking high wrong-way crash locations based on the weighted number of wrong-way entries (Zhou et al., 2012). The study further developed promising, cost-effective countermeasures to reduce WWD incidents and their associated crashes. In May 2014, the Illinois Center for Transportation and the IDOT published guidelines for reducing WWD crashes on freeways. The Illinois Guidebook contains information on several common countermeasures (e.g., signs and pavement markings),

advanced technologies, geometric elements, and related considerations, and enforcement and education strategies (Zhou et al., 2012). The guidebook also contains a Wrong-way Entry Field Inspection Checklist and WWD Road Safety Audit prompt list. However, the guidebook does not provide specific recommendations regarding the appropriate WWD countermeasures and mitigation methods based on specific site conditions.

Wang et al. (2018) identified and addressed the current limitation of 3Es (Engineering, Education, and Enforcement) in the context of WWD incidents, and recommended three strategies: Connected Vehicle System (CVS), Access Management (AM), and Traffic Safety Culture (TSC). As the CVS is in the developing process, the authors focused more on the latter two, which are practice-ready. The TSC addresses intentional driver behaviors and includes those strategies that address social and cultural behaviors such as alcohol consumption, seatbelt usage, etc. For example, using 'Designated Driver strategy' to address driver impairment, where a person refrains from alcohol on social occasions or gathering in order to drive his/her companions who consumed alcoholic beverages. On the other hand, the AM strategies address both intentional and unintentional behaviors. They work with the regulations and design of road and infrastructure geometry. For instance, the following measures could be taken to stop intentional wrong-way drivers originating from roadside services: prohibiting left turns from service area by channelizing driveways, indicating drivers of the next U-turn by adding more signs, and blocking the driveway near divided highways when other access areas from an adjacent intersecting road exist.

2.3.6 Arizona

The Arizona Department of Transportation (ADOT) invested a $3.7 million project in 2017 to construct a first-in-the-nation WWD thermal detection system along a 15-mile stretch on I-17 in Phoenix, Arizona (ADOT, 2017). This project is implemented following the end of the Proof of Concept phase whose objectives were to determine the viability of existing detector systems to identify the entry of wrong-way vehicles onto the highway systems using the following five different technologies: microwave sensors, Doppler radar, video imaging, thermal sensors, and magnetic sensors (Simpson 2013). The system is designed to take a three-phase approach when a wrong-way vehicle is detected: alerting wrong-way drivers so they can self-correct, warning right-way drivers, and at the same time notifying law enforcement.

Additionally, larger and lowered "WRONG-WAY" and "DO NOT ENTER" signs have been installed on hundreds of freeway ramps and overpasses in Phoenix and on rural highways. Considering the fact that more than half of the WWD crashes in Arizona were due to impaired driving, ADOT understands that engineering and enforcement measures can only reduce the risk but can't entirely prevent wrong-way driving (Simpson and Bruggeman, 2015). Thus, ADOT has started the "*Drive Aware*" safety campaign that aims at helping motorists minimize the risk of being in a crash with a wrong-way vehicle. Specifically, the campaign details what drivers should do if they encounter a wrong-way vehicle, see an overhead sign warning of an oncoming wrong-way vehicle, and general tips that will keep drivers safer.

2.4 Wrong-Way Driving Countermeasures

Table 2-4 summarizes the possible reasons for a WWD incident (Zhou et al., 2012). As can be inferred from the table, the WWD crash contributing factors could be categorized into six categories: traffic violation, impaired judgment, inattention, insufficient knowledge, infrastructure deficiency, and others. The following sections discuss the traditional countermeasures and the existing and emerging technologies that could be deployed to address the WWD issue. Several states, including Florida, have deployed Intelligent Transportation Systems (ITS) technologies and Transportation Systems Management & Operations (TSM&O) strategies at off-ramps and freeway mainlines to mitigate WWD incidents in real-time. However, very few strategies, if any, have been deployed along arterials. This section, therefore, focuses primarily on the WWD mitigation strategies on limited-access facilities.

Table 2-4: WWD Crash Contributing Factors (Zhou et al., 2012)

Category	Description
Traffic violation	Driving Under the Influence (DUI) Intentional reckless driving Suicide Test of courage Escaping from a crime scene Avoiding traffic congestion
Impaired judgment	Older adult driver Physical illness Drivers with a psychiatric problem
Inattention	Careless, absent-mindedness, distraction Falling asleep at the wheel Inattention to informational signpost
Insufficient knowledge	Unfamiliar with the roadway infrastructure Lack of understanding of how to use the highway Loss of bearing
Infrastructure deficiency	Insufficient lighting Heavy vegetation Insufficient field of view
Others	Inclement weather

2.4.1 Traditional Countermeasures

Several traditional countermeasures to mitigate WWD incidents have been deployed over the past few decades. Signing and Pavement Markings (S&PM) have traditionally been used to deter WWD events. In 1967, California took the initiative and started using cameras to detect WWD incidents (Tamburri, 1969). A few years later, in 1973, California began to lower the height of "Do Not Enter" signs and "Wrong-way" signs; and also began to display both the signs together on the same post. This strategy has resulted in a significant reduction in WWD incidents; WWD incidents decreased from 50-60 to 2-6 per month in the areas where the aforementioned WWD signs were installed (Leduc, 2008). In 1988, Campbell and Middlebrooks evaluated the effectiveness of lowering the "Wrong-way" signs posted near exit ramps in Atlanta, Georgia. The authors found that many wrong-way drivers corrected before entering the freeway, and within a month, WWD maneuvers were reduced up to 97% (Campbell & Middlebrooks, 1988). North Texas Tollway Authority also evaluated the effectiveness of lower "Wrong-way" and "Do Not Enter" signs by lowering the signs and putting them 2 feet above the ground at 28 (out of 142) exit ramps in their jurisdiction. Finley et al. (2014) stated that the effectiveness of the lower signs could be accurately determined if the entire freeway corridor was equipped with lower signs.

The Manual on Uniform Traffic Control Devices (MUTCD) recommended several countermeasures for addressing the WWD issue, such as pavement arrows for wrong-way, colored edge lines on exit ramps, red reflective raised pavement markers, etc.; and these countermeasures have been widely used (Cooner et al., 2004b). In addition, Texas and Virginia installed raised pavement markers at off-ramps (Athey Creek Consultants, 2016).

Virginia Highway and Transportation Research Council evaluated the effectiveness of raised pavement markers in effectively correcting the wrong-way drivers' actions; the markers were considered to be effective in 94% of the cases (Shepard, 1976). Researchers from TTI wanted to see how these traditional pavement markings and wrong-way signs affected intoxicated drivers' behavior. Their research indicated that impaired drivers look straight ahead on the pavement and tend to look left and right less. However, intoxicated drivers do not recognize the lowered "Wrong-way" signs, and these are less effective on them (Finley et al., 2017). Getting intoxicated drivers' attention is challenging; however, some measures such as enlarging the sign, incorporating flashing LED lights, and adding red retroreflective tape on signposts can assist drivers under the influence. One thing to be noted is that intoxicated drivers need to be closer to the LED signs compared to the traditional regular signs to read them clearly (Finley et al., 2014; Finley et al., 2017).

2.4.2 Existing and Emerging Technologies

The traditional WWD countermeasures that recommend changes to roadway signage and pavement marking improvements, although often effective, do not prevent all WWD incidents. More rigorous and active WWD detection and mitigation methods are required to: (a) alert wrong-way drivers as soon as they turn the wrong-way; (b) warn right-way drivers of a possible wrong-way driver; and (c) inform law enforcement officials, Transportation Management Centers (TMCs), and first responders in real-time about wrong-way drivers.

More recently, in addition to the traditional WWD countermeasures, ITS technologies and TSM&O strategies are increasingly being deployed to tackle the WWD

issue. ITS technologies can alert wrong-way drivers in real-time using detection-triggered Wrong-way signs, etc. Right-way drivers can be alerted using the existing ITS technologies, such as DMS and LED signs (Finley et al., 2016). In some cases, multiple technologies are combined together to prevent WWD incidents. For instance, the Washington State Department of Transportation (WSDOT) notifies wrong-way drivers using a combination of video cameras, LEDs, and flashers (Cooner et al., 2004b; Finley et al., 2016).

Advanced technology-based countermeasures such as detection-triggered LEDs surrounding Wrong-way Signs and red-RRFBs, have played a crucial role in reducing WWD incidents. Many states, such as Florida, Wisconsin, and Texas, have been using LED signs to alert wrong-way drivers (Finley et al., 2014; Sandt et al., 2017). Wrong-way signs with LED border illumination were examined in San Antonio, Texas. Researchers observed an approximate 35% reduction in 911 calls per month related to WWD incidents on the roadway corridors installed with these countermeasures (Venglar & Fariello, 2014). In Florida, red-RRFB Wrong-way signs have been installed at several off-ramps across the state; these devices were found to work successfully in detecting and alerting the wrong-way driving vehicles, providing an opportunity for the wrong-way drivers to turn around and correct themselves (Finley et al., 2014; Sandt et al., 2017).

As can be inferred from the above discussion, several states, including Florida, have deployed ITS technologies and TSM&O strategies at off-ramps and freeway mainline to mitigate WWD incidents in real-time. The following sections discuss some of these technologies that have been deployed to detect and respond to WWD incidents in real-time.

Thermal Camera System

A thermal camera detection system is a promising technology that alerts when wrong-way drivers enter a roadway. ADOT was the first in the nation to use this technology in combating wrong-way driving. The detection system is activated when it detects a wrong-way vehicle entering the roadway, and then the system immediately alerts the wrong-way driver. In addition, the system sends notifications to and alerts the public safety department. ADOT immediately processes the alert and sends a warning to the road users via message boards. Currently, Arizona invested $4 million in this system consisting of 90 thermal cameras along 15 miles of I-17. According to the ADOT officials, this system has detected more than 45 WWD vehicles in the past year (U.S. News, 2019). This system has also resulted in quicker response times for the law enforcement officials and the first responders. At the end of 2018, FDOT has also taken the initiative to add new software to the existing traffic cameras on the Howard Frankland bridge over Tampa Bay to detect WWD events (Trimble, 2018).

Radar Detection

A significant reduction in WWD can be achieved with the deployment of early warning systems. Wrong-way drivers can be actively warned using accurate radar detectors and active warning systems. After the radar detects a WWD vehicle, alert systems such as flashing beacons, audible alerts, and/or DMS can alert wrong-way drivers. This type of system can be used in combination with CCTV cameras installed in both directions to visually verify WWD events.

In 2015, the Rhode Island Department of Transportation deployed radar technology at 24 locations statewide to detect and warn wrong-way drivers and caution right-way drivers by displaying messages on the DMS in real-time (RIDOT, 2015). A study in Florida by Ozkul and Lin found that about 66 of 78 (i.e., 85%) wrong-way drivers corrected themselves after they noticed the alert from the radar (Ozkul & Lin, 2017). Similarly, New Mexico developed a directional traffic sensor system to alert wrong-way drivers (Cooner et al., 2004; Finley et al., 2016).

Integrated On-road Detection, Tracking, and Notification System

An effective strategy to detect, alert, and mitigate WWD incidents in real-time includes a combination of technologies and countermeasures. A couple of agencies have lead an effort to develop, implement, and test an integrated on-road detection, tracking, and notification system to address WWD incidents. The United Civil Group Corporation, on behalf of ADOT, has developed an integrated conceptual methodology to detect wrong-way drivers, alert the wrong-way driver, track the WWD vehicle, immediately inform the DOT and the law enforcement agencies, and warn the right-way drivers. The study also generated a systematic deployment plan and guidelines to address WWD incidents (Simpson & Bruggeman, 2015).

More recently, the Texas Department of Transportation (TxDOT) conducted a study to create an integrated, comprehensive system to detect and alert wrong-way drivers. For creating an integrated WWD mitigation system, the authors generated a step-by-step conceptual operation, designed functional requirements, and developed a system designed for a Connected Vehicle (CV) to counter WWD. The system was developed to identify

WWD incidents, notify DOT and law enforcement agencies, and alert the right-way vehicles along the corridor. Prior to its deployment, the authors recommended testing the concept on a testbed outside of the actual roadway as a proof-of-concept (Finley et al., 2016).

In-vehicle Systems and Sign Identification Systems

With the increasing advancement of vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) technologies, the potential for onboard vehicle systems to alert wrong-way drivers is continuing to increase. Firstly, this system will give audible and visual alerts to the driver when the vehicle moves in the wrong-way. In addition, with CV technology, the right-way drivers will be alerted as they approach a WWD vehicle.

Several automobile manufacturers are developing similar systems. For instance, Ford is planning to equip its vehicles with on-board traffic sign recognition technology. In addition, the vehicles will use GPS data to check if the vehicle is on the right path. On-board cameras that are installed on the windshield will recognize the posted speed limit, "Wrong-way," and "Do Not Enter" signs. When a vehicle enters a road with a "No Entry" sign, the vehicle will start its alarm to warn the driver. Ford tested this technology on its test track with "No Entry" signs in Belgium (Harman, 2018). In previous years, some similar technology was being considered for adoption in the Mercedes-Benz S-Class and E-Class models (Szczesny, 2013).

Directional Rumble Strips

Zhou and Luo (2018) evaluated the feasibility of using the directional rumble strips in preventing wrong-way drivers from entering a roadway. The directional rumble strips

are a series of rumble strips especially designed to alert wrong-way drivers. Regular conventional rumble strips provide the same amount of noise and vibration when vehicles move over them from either way. Road surface conditions can affect driver's driving experience, such as tire-pavement noise, road friction, etc. (Nafis & Wasiuddin, 2021). Therefore, when drivers drive over the directional rumble strips the wrong-way, they will experience elevated noise and vibration compared to the regular conventional rumble strips. However, when a right-way driver drives over directional rumble strips, they experience similar noise and vibration as the regular conventional rumble strips (Zhou & Luo, 2018).

Several studies have been conducted to determine the performance baseline of directional rumble strips. Different designs of directional rumble strips have been identified by the national survey of transport practitioners, vendors and through literature review and field tests. Researchers are conducting a series of experiments to determine and recommend the most suitable configuration of directional rumble strips, which provide minimum noise and vibration to the right-way drivers, but alert the wrong-way drivers with elevated noise and vibration (Roadway Safety Institute, 2016).

2.4.3 Locations for Deploying Technology-Based WWD Countermeasures

In addition to the existing traditional WWD countermeasures (e.g., signage and pavement marking), strategies that adopt advanced technologies can be used to identify and mitigate WWD incidents. Deployment of these technology-based countermeasures can be costly, especially if the entire roadway corridors need to be covered. Optimization algorithms could be used to identify certain sections to deploy these technology-based

countermeasures. Sandt et al. (2017) used an algorithm with a previous segment model, done by Rogers et al. (2016), to maximize the WWD crash risk reduction on any limited-access roadway network for a given investment level. The segment model was developed considering the geometric design of interchanges, WWD events, and traffic volumes to determine the WWD crash risk on overlapping multi-interchange segments of limited-access facilities (Rogers et al., 2016). To analyze the effectiveness of the algorithm, the researchers used a two-phase installation strategy to find the optimal locations to install "Wrong-way'' signs with Rapid Flashing Beacons on the Central Florida Expressway Authority (CFX) toll road system in Florida. These proposed optimization algorithms assist agencies to strategically deploy countermeasures at high-risk locations. These algorithms could help detect locations where countermeasure implementation will result in the highest reduction of WWD crashes. In addition, constraints consisting of existing countermeasures and roadway coverage can be added to the optimization algorithm to replicate real-world scenarios. The algorithm can be modified based on the agency's specific requirements by incorporating other constraints as well. In addition to prioritizing the locations for installing WWD countermeasures, this algorithm can also work considering budget allocation. For instance, agencies with limited resources can utilize this algorithm to install advanced technology-based WWD countermeasures at a few interchanges instead of having to install them on entire corridors (Sandt et al., 2017; Kayes et al., 2018; Arafat et al. 2020).

2.5 Studies on Nonparametric Models for Crash Analysis

Descriptive statistics and linear regression models were often used to analyze WWD crashes and identify influential factors. One of the major limitations of the linear

regression models is that they use a linear relationship between WWD crash severity and the influential variables, leading to inaccurate estimations of injury severity (Mussone et al., 1999). Researchers have recently been using data mining techniques to understand the factors contributing to crash severity (Kuhnert et al., 2000; Sohn & Shin, 2001; Chang and Wang, 2006; Kashani et al., 2011; Pakgohar et al., 2011; Das et al., 2018; Rahman et al., 2019; Nafis et al. 2021). Data mining procedures use artificial intelligence and statistical analyses to extract processable knowledge from databases. Unlike the regular regression models, data mining techniques have the ability to identify and explain the complex patterns associated with crash risk without needing to use a functional form (Kashani et al., 2011). The WWD crash dataset is heterogeneous due to the variations in the roadway geometry and less homogeneous road sections. Moreover, missing values for some of the factors can make WWD crash data analysis cumbersome. Machine learning (ML) or data mining techniques can address these issues. Since the crash data are heterogeneous by nature (Karlaftis & Tarko, 1998), certain essential relationships useful for influencing the cause of crashes may remain hidden if they are not sectioned in subsets (Depaire et al., 2008). Acuna and Rodrigues (2004) concluded that missing data samples affect statistics-based algorithms, and nonparametric classifiers perform better than parametric classifiers for datasets with missing values. Moreover, if not addressed well, the presence of heterogeneity in the dataset may lead to biased results (Karlaftis & Tarko, 1998). Karlaftis and Tarko (1998) found that models based on clustered heterogeneous datasets yield improved and accurate results compared to the models based on the pooled heterogeneous dataset. Some other studies also found similar improved results with clustering analysis (Rajib et al. 2019). Decision trees (DT) is a data mining technique that is suitable for

analyzing crashes because they do not presume any relationship between the dependent (i.e., crash severity or crash frequency) and the independent variables (i.e., crash contributing factors), and do not require any preset probabilistic knowledge on the study because of their nonparametric approach.

Another aspect that DTs tackle well is multi-collinearity. Unlike other linear regression methods, the structure of DTs is such that the relation between contributing factors and crash severity could be explained by an 'if-then' relationship in the crash data set (Kashani et al., 2011). Furthermore, essential decision rules (DR) could be derived from DTs. For instance, Pande and Abdel-Aty (2009) developed data mining rules from closely related crash characteristics. De Oña et al. (2013) developed DTs using crash data from Spain and derived specific decision rules. In Italy, Montella (2012) used classification trees along with association rules to analyze pedestrian crashes. All these studies indicate the usefulness of data mining techniques in transportation safety studies (Pande & Abdel-Aty, 2009; Montella, 2012; De Oña et al., 2013; Khan et al., 2017; Tanvir et al., 2019; Morshed et al., 2021).

Among the many algorithms used to build DTs, Classification and Regression Trees (CART) developed by Breiman et al. (1984) is the most popular approach to investigate crash severity. Pakgohar et al. (2011) used the CART and multinomial logistic regression to investigate drivers' role in crashes. The CART models are more dependable because they are simple, and their data representation is easily understandable (Pakgohar et al., 2011). Kashani and Mohaymany (2011) found that the results are easy to understand, and the correlation problem from traffic crash data is not of great concern while using

CART models. Beshah and Hill (2010) compared different classification models and concluded that the CART models provide both theoretical and applied advantages over parametric models. Despite having many benefits, similar to every other method, the CART method has some disadvantages as well. Since tree models are formed based on their random seed number, it is often unstable, and outcomes may vary. Although data mining has been used in many studies, their application on WWD crashes is less to none.

2.6 Summary

WWD crashes have been an area of concern for over five decades. Researchers in the United States and across the world have been analyzing WWD crashes. The existing studies on WWD crashes have primarily focused on freeways. The studies examined different risk factors, causes, patterns, and contributing factors associated with WWD crashes on limited-access facilities. However, studies on WWD crashes on non-limited access facilities are close to none. Moreover, most of the previous studies have used parametric techniques to analyze WWD crashes on limited-access facilities only. This research will explore the usability of nonparametric data mining techniques in analyzing WWD crashes on both limited and non-limited access facilities.

CHAPTER 3
METHODOLOGY

This chapter presents the methods in detail that are adopted to achieve the two research objectives mentioned in Chapter 1, Introduction. Section 3.1 discusses the following five parametric models: logistic regression, Ridge, Lasso, Linear Discriminant Analysis (LDA), and Gaussian Naïve Bayes (GNB). Section 3.2 discusses the following three nonparametric models: Decision Trees (DT), Random Forests (RF), and Support Vector Machine (SVM). Finally, Section 3.3 discusses the approach used to identify factors influencing the severity of WWD crashes on arterials.

3.1 Parametric Models

The response variable was binomial, and datasets were divided into training sets and test sets for each parametric model. For each parametric model, K = 5 fold cross-validation was performed on the dataset to remove potential model bias toward a particular training set. The following parametric models were used in this research.

3.1.1 Logistic Regression

Logistic regression is a popular technique for classification and has been used in many studies. Binary logistic regression is used in this research because the output variables are two classes. Therefore, for the binary outcome variable, logistic regression can be expressed in the following form (Al-Ghamdi, 2002):

$$logit(p_i) = log\left(\frac{p_i}{1-p_i}\right) = \alpha + \beta' X_i \tag{3-1}$$

where,

p_i = prob(y_i = 1) is the response probability;

$1 - p_i$ = prob(y_i = 0);

α = intercept parameter;

β' = vector of estimation coefficients; and

X_i = vector of predictor variables.

3.1.2 Lasso Regression

The least absolute shrinkage and selection operator (LASSO) is a widely used fast algorithm. LASSO has the ability to shrink variables using ℓ1 penalty. The method shrinks unimportant variables by shrinking them to zero and selects important features (Friedman et al., 2010). The lasso estimator uses the ℓ1 penalize, and the equation can be explained as (Friedman et al., 2010):

$$\hat{\beta}(lasso) = argmin_{\beta} \parallel y - X\beta \parallel_2^2 + \lambda \parallel \beta \parallel_1 \quad (3\text{-}2)$$

where,

$\|\beta\|_1$ = $\sum p_j$ or, the ℓ1 -norm penalty on β; and

λ = the tuning parameter ($\lambda \geq 0$).

For some suitably chosen λ, the ℓ1 penalty enables the Lasso to simultaneously choose important variables and shrink some components of $\hat{\beta}(lasso)$ to zero (Friedman et al., 2010).

3.1.3 Ridge Regression

Ridge regression (RR) (Hoerl & Kennard, 1970) is used when a dataset is drawn from a normal distribution and ideal when there are too many predictor variables (Friedman et al., 2010). Unlike LASSO, RR does not force coefficients to shrink to zero and cannot

select a model with selected variables. It uses ℓ2 norm, and the equations can be explained as:

$$\hat{\beta}(ridge) = argmin_{\beta} \parallel y - X\beta \parallel_2^2 + \lambda \parallel \beta \parallel_2^2 \tag{3-3}$$

$$\parallel y - X\beta \parallel_2^2 = \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 \tag{3-4}$$

where,

$\parallel \beta \parallel_2^2$ = $\sum p_j$ is the ℓ2 -norm penalty on β;

$\parallel y - X\beta \parallel_2^2$ = ℓ2 -norm or, loss function;

x^T = i-th row of X; and

λ = tuning parameter, controls the power of the penalty term ($\lambda \geq 0$).

The value of λ is dependent on the dataset. The larger the value it has, the larger the shrinkage. The value of λ can be estimated with different methods, such as cross-validation. Aside from this method, there are many other methods for estimating the shrinkage parameter lambda available in the literature (e.g., Hoerl & Kennard, 1970; Lawless & Wang, 1976; Kibria, 2003; Kibria & Lukman, 2020, etc.).

3.1.4 Linear Discriminant Analysis

The Linear Discriminant Analysis (LDA) is a parametric model that can separate two or more classes. It assumes the dataset variables have a gaussian distribution with different means but a common covariance matrix (Ripley et al., 2013; Venables & Ripley, 2013; Worth & Cronin, 2003). The LDA function passes through the centroids of the two groups to discriminate between the groups.

3.1.5 Gaussian Naïve Bayes

A Naïve Bayes (NB) classifier classifies calculating the most votes on a certain class. While calculating, the conditional probability is expressed as $P(y|X)$. It is the product of simpler probabilities, utilizing the naïve independence assumption (Mitchell, 1997):

$$P(y|X) = \frac{P(y)P(X|y)}{P(X)} = \frac{P(y)\prod_{i=1}^{n} P(X|y)}{P(X)} \tag{3-5}$$

The Gaussian Naïve Bayes (GNB) implements the classification by presuming the likelihood of the features to be Gaussian:

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i-\mu_y)^2}{2\pi\sigma_y^2}\right) \tag{3-6}$$

where, the parameters σ_y and μ_y are calculated by maximum likelihood (Cao et al., 2003; Murakami et al., 2010).

3.2 Nonparametric Models

Similar to the parametric model analysis, the response variable was binomial, and datasets were divided into training sets and test sets for each nonparametric model. For each nonparametric model, K = 5 fold cross-validation was performed on the dataset to remove potential model bias toward a particular training set. The following nonparametric models were used in this research.

3.2.1 Random Forests

Random Forests (RF) are an ensemble learning method composed of a collection of unpruned randomized DTs. In this technique, a multitude of modified weak (fewer

features) DT classifiers are built in parallel. Each tree in the RF returns a vote (serious injury or not serious injury) and the corresponding misclassification rate. The RF then predicts by the unweighted majority of class votes and misclassification rates from these DT classifiers. As the number of trees in RF grows, the misclassification rate converges to a limit and reduces the over-fitting problem (Breiman, 2001). The **pseudocode** of the algorithm is presented below:

- First, a bootstrap sample is selected from the original data for each tree in the forest.
- The bootstrapped sample is obtained by randomly choosing instances from the original data with replacement. The number of observations in the bootstrapped sample is of the same size as the original data set.
- From each bootstrapped sample, an unpruned DT is then grown to the maximum extent possible using a modified DT learning algorithm.
- The modification in the tree-learning algorithm is applied as follows: at each node, a subset of features, a certain number of variables (Mtry), are randomly selected rather than the complete feature set to compete for the best split. Number of trees (ntree) and Mtry are tuned by increasing or decreasing from an initial value until a minimum error rate is obtained.
- Once all the modified trees are constructed, the final predictions are achieved by averaging individual predictions of the trees.

Figure 3-1 provides the framework of a RF model to predict crash severity. In a RF model, two measures are commonly used for evaluating variable importance: Out-of-bag (OOB) error rate and Gini index. For each of the bootstrap sample in a particular tree, about

two-thirds of the data points are used for training, while the remaining data points are used for testing, known as OOB samples. The OOB error rate is computed by the number of times that the voted class label is not the same as the true class. The average OOB rate error indicates the model accuracy.

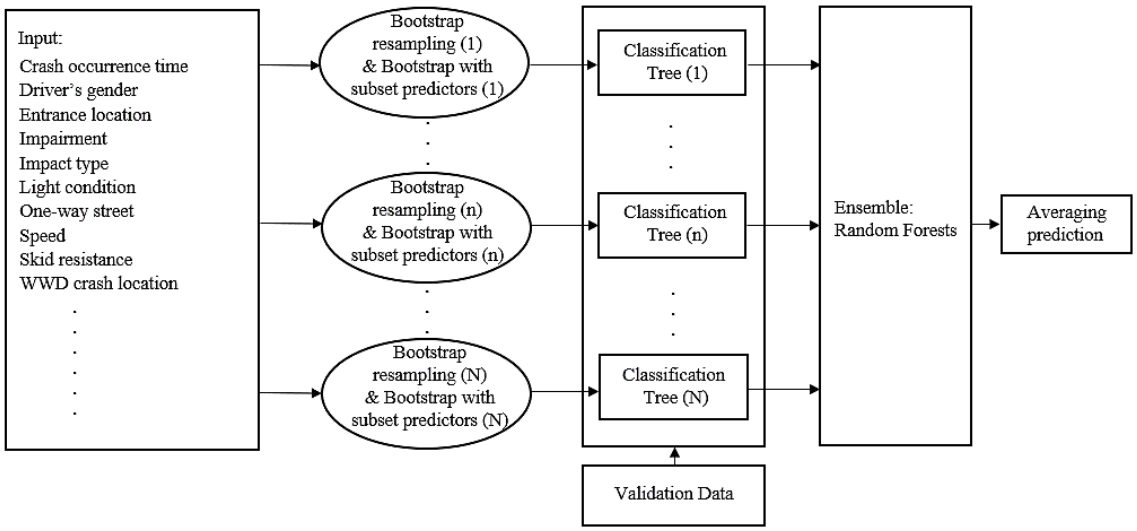


Figure 3-1. Random Forests Ensemble Technique Framework

In this research, the Gini index was used in RF to measure the importance of significant variables contributing to the WWD crash severity. The Gini index evaluates the impurity and quality of the split of each node in a particular tree. More important variables have nodes with higher purity and result in a higher decrease in Gini.

The RF algorithm results in an unbiased estimate of the generalization error and achieves lower variance. By randomizing and modifying the tree learning algorithm to choose features from random subsets, the correlation between the trees comprising the ensemble is reduced. This way, RF tends to attain superior performance (Zhang & Haghani, 2015; Saha et al., 2015).

3.2.2 Decision Trees

Classification is the process of finding a model that explains and differentiates data classes (Mining et al., 2006). A Decision Tree is called a "classification tree" when the target variable is nominal. Since the crash injury is the target variable in this research, and it is nominal, the CART analysis will be used in this research. In the CART model, the topmost node, "root node," is divided into two nodes based on the independent variable (i.e., splitter) such that each child node data is homogeneous. Each node is continually divided into child nodes with more homogeneous data until a node reaches the highest level of homogeneity. The last child node where the data cannot be divided any further is called a terminal node or "leaf" with no branches (Breiman et al., 1984). Of the many well-known methods that are available for splitting root nodes into child nodes, the Gini index is the most commonly used method and is discussed below (Kashani & Mohaymany, 2011).

$$P(k|m) = \frac{p(k,m)}{p(m)},\ P(k,m) = \frac{\pi(k)N_k(m)}{N_k},\ p(m) = \sum_{k=1}^{k} p(k,m) \tag{3-7}$$

$$\text{Gini (m)} = 1\text{-} \sum_{k=1}^{k} p^2\,(\text{k|m}) \tag{3-8}$$

where,

k	=	selected variable or class;
$P(k\|m)$	=	the conditional probability of *k* in node *m*;
p(m)	=	the proportion of total observations in node *m*;
N_k	=	the number of observations of class *k* in the root node;
N_k(m)	=	the number of observations of class *k* in node *m*;
$\pi(k)$	=	the prior probability for class *k*; and

Gini (m) = Gini index, an indication of the proportion of variable impurity.

If the Gini index is zero, it means impurity is the lowest in the node. If the Gini index is one, this means the node is very impure. If the observation ratio in a node becomes the same as the root node, then the maximum Gini index value is obtained. For each tree formed, the misclassification error rate can be calculated as:

$$Error = \sum_{m=1}^{M} p(m)\,[1- \sum_{k=1}^{k} p^2\,(k/m)] \qquad (3\text{-}9)$$

In Equation 3-9, m is the number of terminal nodes. In this method, the tree branch is pruned when the increase in the misclassification cost is considerably lower than the decrease in the complexity cost. More details of CART analysis can be found in Breiman et al. (1984).

3.2.3 Support Vector Machine

The Support Vector Machine (SVM) uses the statistical risk minimization principle. Figure 3-2(a) shows a two-category problem solved by the SVM, separating these two groups. The figure shows SVM can map input vector X into high dimensional feature space. SVM makes a hyperplane separating several groups by choosing an optimal non-linear mapping priori. The hyperplane helps to divide while maximizing boundaries between the classes [Figure 3-2 (a) and Figure 3-2 (b)].

The SVM model uses a training set to develop a model and uses a test set for validation. The equation of the hyperplane for separating outcomes can be written as follows:

$$W.X - b=0 \qquad (3\text{-}10)$$

In Equation 3-10, W is a normal vector, perpendicular to the hyperplane, and "." denotes the dot product. Maxwins voting strategy is used in this method for conducting the classification (Lingras & Butz, 2007; Li et al., 2012).

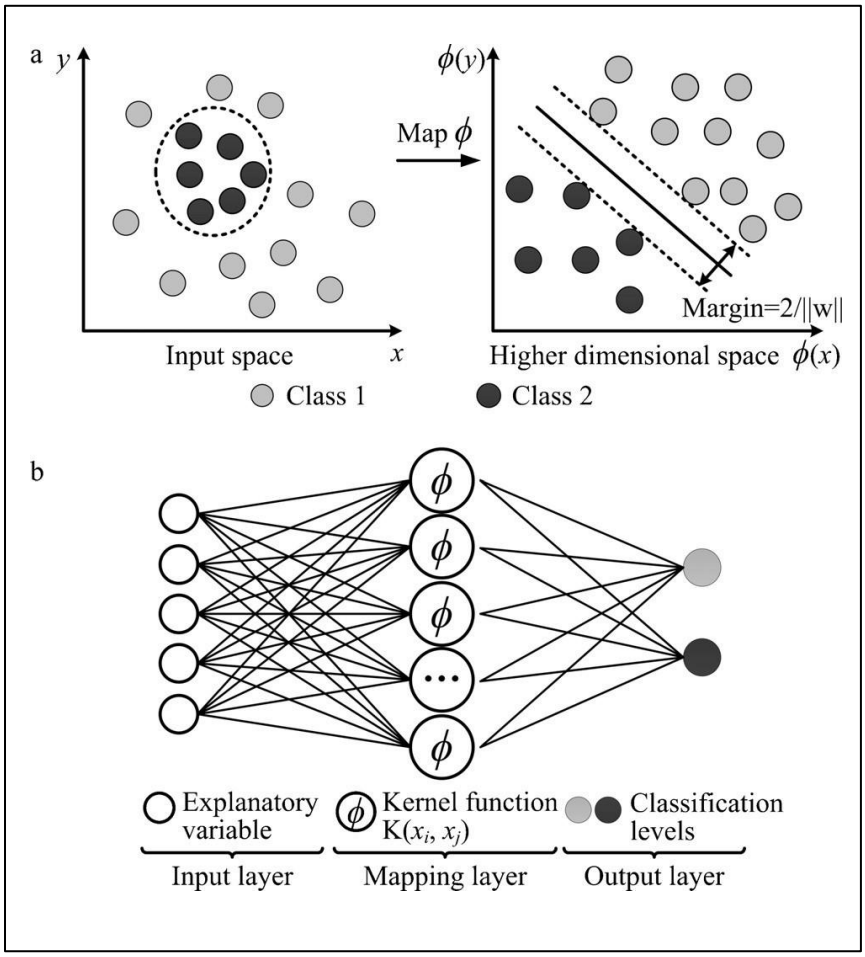


Figure 3-2: SVM Methodology (Li et al., 2012)

3.3 Combination of Data Mining Techniques

This research used the combination of AHC, RF, DT, and DR methods to identify and perform the pattern analysis of the contributing factors that affect WWD crash severity on non-limited access facilities. Specific steps adopted to accomplish the research goal include:

- Employ Agglomerative Hierarchical Clustering (AHC) to identify the homogeneous clusters and form natural subsets.
- Use Random Forests (RF) to identify and prioritize significant variables.
- Utilize the CART to explore WWD crash injury patterns on non-limited access facilities.
- Construct Decision Rules (DRs) to explain crash severity patterns.

Figure 3-3 shows the framework of the combination of the tree-based data mining processes, and the following subsections discuss these data mining techniques.

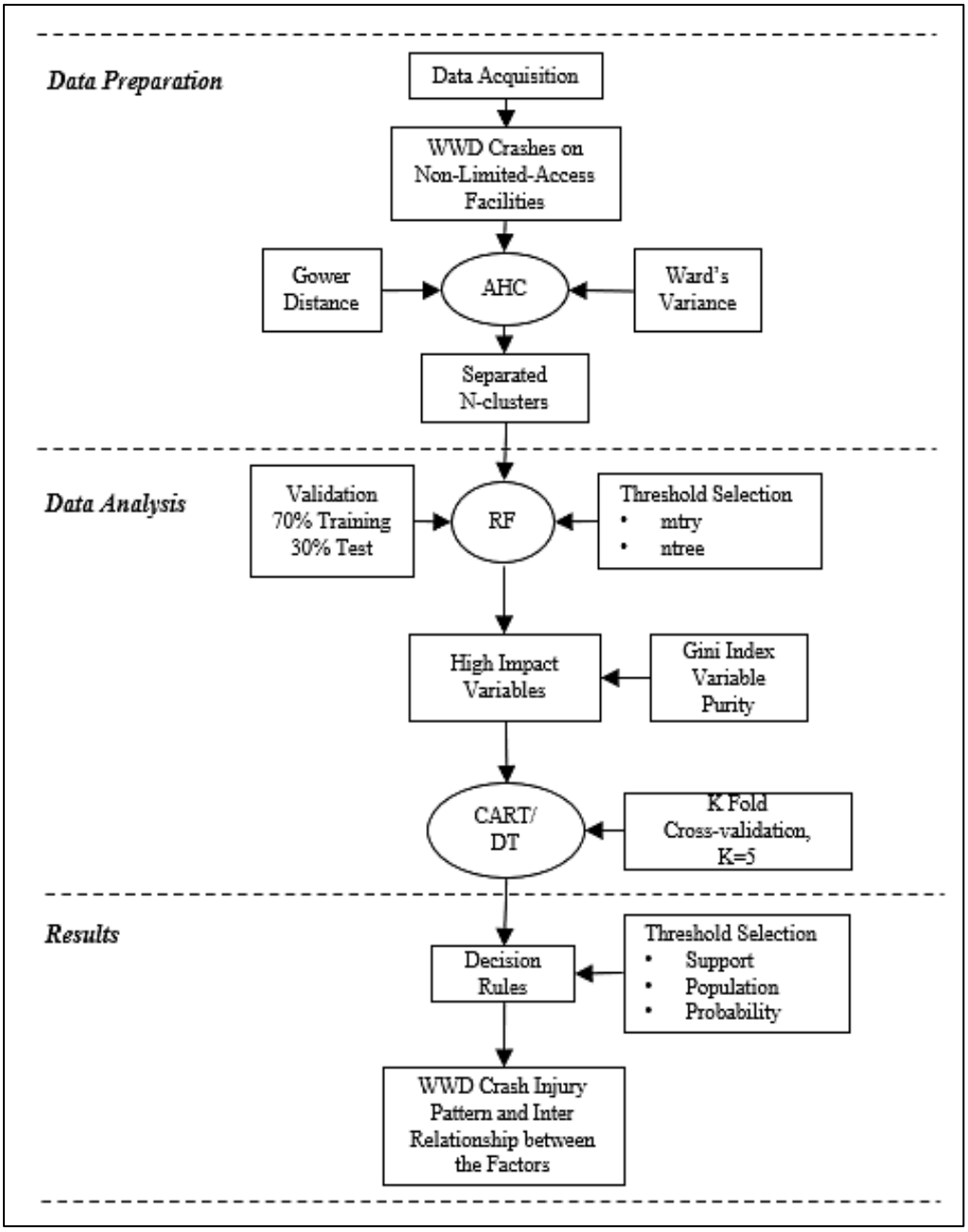


Figure 3-3: Framework of WWD Severity Analysis Using Nonparametric Models

3.3.1 *Agglomerative Hierarchical Clustering*

In this research, the AHC approach was used to segment the dataset into a few homogenous subsets to handle the heterogeneity within the crash dataset (Johnson, 1967). The general scheme of the AHC algorithm is as follows: Initially, each object (n) is

assigned to its own cluster. Afterward, the algorithm continues iteratively, at each stage aggregating the two most similar clusters, progressing until there is only a single cluster (Müllner, 2018). There are two types of distance that need to be measured to generate an AHC dendrogram and group the clusters. One of them is the distance between the records or observations, and the other type is for dissimilarity measurement among the clusters. In this research, the Gower distance was used to determine the distance between observations and manage the mixed type data (Gower, 1971). For dissimilarity measurement between the clusters and grouping the clusters, Ward's minimum variance linkage was used, as it reduces the total within-cluster variance (Murtagh & Legendre, 2014).

The main feature of Gower distance, also known as Gower's coefficient, is its ability to handle nominal, ordinal, and (a)symmetric binary data even when different types occur in the same data set. At first, each feature (column) is standardized by dividing each entry by the standard deviation of the corresponding feature after subtracting from the mean value. In Gower distance, the dissimilarity between two rows is the weighted average of the contributions of each variable (Müllner, 2018). It can be described as:

$$d_{ij} = d(i,j) = \frac{\sum_{k=1}^{p} w_k \, \delta_{ij}^{(k)} d_{ij}^{(k)}}{\sum_{k=1}^{p} w_k \, \delta_{ij}^{(k)}} \tag{3-11}$$

where,

d_{ij} = Gower distance between observation i and j;

It is the weighted mean of $\boldsymbol{d_{ij}^{(k)}}$ with weights;

w_k = weights[k];

$d_{ij}^{(k)}$ = distance between x[i,k] and x[j,k]. It is the distance for k-th variable contributing to the total distance;

δ = 0 or 1. It becomes zero when the variable x[,k] is asymmetric binary and values are zero in both rows (i and j) or when the variable is absent in either or both rows. In all other conditions, it is 1.

All the dissimilarity measures are then formed into a matrix to further agglomerate into clusters using linkage. Ward's linkage method, also known as Ward's minimum variance method, explores and locates partitions with a small sum of squares while creating clusters. The process runs as follows:

- Generating clusters for each point, where every point is in its own cluster and the sum of squares = 0
- Then merging two clusters that result in the smallest increase in merging cost or sum of squares.
- Iteratively merging until all clusters aggregates into one single cluster.

The merging cost among two clusters (A and B) or, the sum of squares of the clusters increases when aggregating the clusters:

$$\Delta(A,B) = \sum_{i\epsilon A\cup B} \|x_i - m_{A\cup B}\|^2 - \sum_{i\epsilon A} \|x_i - m_A\|^2 - \sum_{i\epsilon B} \|x_i - m_B\|^2$$

$$= \frac{n_A n_B}{n_A + n_B} \|m_A - m_B\|^2 \qquad (3\text{-}12)$$

where,

m_j = center of cluster j;

n_j = number of points in cluster j; and

$\Delta(A,B)$ = the merging cost of aggregating the clusters A and B.

The sum of squares starts out at zero in AHC, and then grows as the clusters are merged. Ward's linkage method keeps this growth as small as possible (Murtagh, & Legendre, 2014).

3.3.2 RF and DT

After the clustering analysis, RF and DT analysis need to be performed, as discussed in Section 3.2.1 and 3.2.2, respectively. First, RF is performed in each cluster to rank the variables. The Decision Trees are then created for discovering the hidden patterns. After the DT analysis, decision rules are created from each tree, as discussed in the following section.

3.3.3 Decision Rules

To better understand the model and interpret the results, the CART model was transformed into decision rules. The structure of the DR is X →Y, where the part X is called the antecedent, and Y is called the consequent. The X→Y is expressed in an IF-THEN relationship. For instance: IF (collision type = Head-on & driver's age ≥ 50 years) THEN (WWD crash severity = serious injury).

In the CART models, the rules initiate from the root node, where the IF condition begins. Each time a variable intervenes in a tree division, an IF condition is added to the

rule. This ends at the end child node (i.e., terminal node) with a THEN condition. The following three parameter thresholds were used to identify essential rules:

- The support (S), which supports the total single rule and indicates that the frequency of the combination of antecedents and consequents appearing in the database;
- The population (Po), which is the percentage of the population in a particular node compared to the root node; and
- The probability (P), which indicates the percentage of cases in which the rules are accurate.

The relationship between these three measures is in the following equation:

$$P=S/Po \tag{3-13}$$

Note that Equation 3-13 is in percentages. The concepts of support (S) and probability (P) are core in association rules and DRs, which have been used in many previous studies (Pande & Abdel-Aty, 2009; Montella, 2012; De Oña et al.,2013; Khan et al., 2017). Figure 3-3 shows the structure of the combination of data mining processes used in the study for factor identification and pattern recognition.

3.4 Summary

This chapter discussed the framework of the methodology that was adopted to achieve the research goal and objectives. The analysis was based on WWD crashes that occurred on arterial facilities during the years 2012-2016. The WWD crash severity analysis was conducted using the Agglomerative Hierarchical Clustering technique. Decision Trees and Decision Rules were used to identify specific factors contributing to WWD crash severity.

CHAPTER 4

DATA PREPARATION

This chapter discusses the data required to achieve the research objectives. The first section provides a detailed discussion of the crash data used in this research. The second section describes the police report review process undertaken to collect additional information about the WWD crash data. The third section focuses on the roadway characteristics data. Finally, the fourth section summarizes the data needs.

4.1 Crash Analysis Reporting (CAR) System Data

Crash data for the years 2012-2016 were obtained from the FDOT's CAR system. The CARS database includes three files:

- crash level data file,
- vehicle-driver-passenger level data file, and
- non-motorist level data file.

The crash level file comprises crash-related information such as roadway ID, crash number, milepost of the crash location, crash severity, where the crash occurred, etc. The vehicle-driver-passenger file contains the road user-related data for each crash recording; therefore, it has information on crash numbers, all vehicles affected in the crash, all drivers, occupants, and road users involved in the crash, etc. The non-motorist level file contains information about all non-motorists involved in a crash, such as type of non-motorist, crash number, non-motorist injury severity, non-motorist location, etc. The following variables were extracted from the FDOT's CARS database:

- Date and Time of Crash

- Day of Week of the Crash
- Crash Severity
- First Harmful Event
- Collision Type
- Alcohol Involvement
- Max Speed
- Driver's Age
- Driver's Gender
- Light Condition
- Weather Condition
- Road Surface Friction
- Road Surface Condition
- Type of Junction
- Vehicle Body Type
- Vehicle point of Impact

4.2 Police Report Review

The analysis was based on five years of crash data from 2012-2016. The crash data shapefiles for the years 2012-2016 were downloaded from the FDOT Unified Basemap Repository (UBR). Since the scope of this research project is limited to state-maintained non-limited access facilities, the data were downloaded only for the state roads in Florida. The variable FL_WRNGWAY, a yes/no flag that indicates WWD involvement, was used to identify WWD crashes. At the time of this research, the FDOT State Safety Office has

not yet finalized the 2016 crash data shapefiles. WWD crashes for the year 2016 were identified using the following code in the vehicle-driver-passenger extract file: Driver Action at Time of Crash = “21” (wrong side or wrong-way). From 2012-2016, a total of 2,879 crashes were identified as potential WWD crashes. Signal Four Analytics is a web-based geospatial crash analytical tool. Police reports for the 2,879 arterial WWD crashes. Police reports of these 2,879 crashes were downloaded and reviewed in detail. Each of the police reports was manually reviewed, and the following data were collected:

1. *Is it a WWD crash?*
   - *Yes*
   - *No – passed over the median*
   - *No – reason* _____
   - *Unknown*
2. *Where did the WWD crash occur?*
   - *Middle of intersection*
   - *In very close proximity to an intersection*
   - *On major approach*
   - *On minor approach*
   - *On divided roadway*
   - *On undivided roadway*
   - *On Two-Way Left-Turn Lane*
   - *Other* _____
   - *Not sure*
3. *Did WWD crash occur on a one-way street?*
   - *Yes*
   - *No*
   - *Not Sure*
4. *If divided, type of median*
   - *Paved*
   - *Raised Traffic Separator*
   - *Curb*
   - *Vegetation*
   - *Curb and vegetation*
   - *Other*

5. *Is there roadside lighting?*

- *Yes*
- *Not Sure*
- *No*

6. *Is there a WWD Sign on the leg where the crash occurred?*

- *Yes*
- *Unsure*
- *No*

7. *Where did the WWD possibly enter the wrong-way?*

- *At signalized intersection*
- *At a four-way Stop sign*
- *At a two-way Stop sign*
- *From a driveway*
- *Make a U-turn*
- *Not Sure*

8. *Did the police report state where the WWD possibly entered the wrong-way?*

- *Yes*
- *No*
- *Unsure*

9. *The coordinates where the WWD possibly entered the wrong-way?*

- *Lat:* _____
- *Lon:* _____

10. *The blood alcohol concentration level of wrong-way driver:* _____

Of the 2,879 potential WWD crashes, only 1,890 crashes (i.e., 65.6%) were categorized as actual WWD crashes resulting from vehicles traveling the wrong-way. A total of 945 crashes (i.e., 32.8%) were not considered to be WWD crashes. Of these 945 non-WWD crashes, a sizable number (353 crashes) were head-on crashes. A majority of these head-on crashes were due to a driver crossing over the centerline, especially on undivided roadways.

4.3 Roadway Characteristics Data

This section discusses the data preparation efforts undertaken to extract roadway information for one-way streets and divided and undivided facilities in Florida.

4.2.1 One-Way Streets

FDOT maintains an *All Road Base Map (ARBM)* that was built on the NAVSTREETSTM base map from HERE (formerly known as NAVTEQ). To link this ARBM with its linear-referenced roadway and crash databases, FDOT added the linear references (i.e., roadway IDs and mileposts) to all road segments in the map. With the linear references, the map's state road portion can be populated with roadway data from FDOT's RCI database. While the RCI is a comprehensive and well-maintained database, it is available for only state roads and a small portion of local roads. This leaves a majority of the local roads in the map, numbering over one million segments and growing, without the same data. FDOT has since added some major variables to the map, including functional class and roadside information, which are needed for safety analysis, among its other applications. The following steps were undertaken to extract one-way streets from the FDOT's ARBM.

Step 1: Generate a one-way street layer based on the All Road Base Map.

The attribute "DIR_TRAVEL" of the ARBM shows travel direction, which is the legal travel direction for a navigable link. The definitions of the direction of travel are as follows:

- The direction of travel 'F' is applied when the direction of travel is one way from the reference node to the non-reference node.
- The direction of travel 'T' is applied when the direction of travel is one way to the reference node from the non-reference node.
- The direction of travel 'B' is applied when travel is allowed in both directions between the reference and the non-reference nodes.
- The direction of travel 'Not Applicable' is applied to non-navigable links.

  The one-way street layer only includes the streets when the direction of travel is one way from the reference node to the non-reference node.

Step 2: Generate a one-way street layer without specific types of roads.

The attribute "FUNC_CLASS" in the ARBM defines a hierarchical network used to determine a traveler's logical and efficient route. The definitions of functional class (NAVTEQ definition) are shown as follows:

- Functional Class '1' roads allow for high volume, maximum speed traffic movement between and through major metropolitan areas.
- Functional Class '2' roads are used to channel traffic to Functional Class '1' roads for travel between and through cities in the shortest amount of time.

- Functional Class ‘3’ is applied to roads that interconnect Functional Class ‘2’ roads and provide a high volume of traffic movement at a lower level of mobility than Functional Class ‘2’ roads.
- Functional Class ‘4’ is applied to roads that provide for a high volume of traffic movement at moderate speeds between neighborhoods. These roads connect with higher functional class roads to collect and distribute traffic between neighborhoods.
- Functional Class ‘5’ is applied to roads whose volume and traffic movement are below the level of any functional class. In addition, walkways, truck-only roads, bus-only roads, and emergency vehicles-only roads receive Functional Class ‘5’.

Since arterials and collectors are the focus of this research, the next step was to exclude freeways, major connectors, and local roads from the “one-way” database. Hence, all the one-way streets with Functional Class “1” and “2” were excluded from the dataset. The following specific types of roads were also excluded from the one-way street layer: ramps, tollways, bridges, tunnels, and private roads.

Step 3: Generate the final one-way streets layer.

Even when multiple rules are applied to extract the one-way streets, there may still be some small road segments (e.g., exclusive left-turn bays) that are not necessarily one-way streets. This issue was addressed by dissolving all road segments within the selected one-way street layer based on roadway ID. Finally, only the one-way segments that are longer than 0.25 miles are included in the dataset. Figure 4-1 shows the final one-way streets layer that was included in the analysis.

4.2.2 Divided Roads and Undivided Roads

Similar to the approach used to extract one-way streets, the selection of divided roads and undivided roads were also based on the ARBM. The RCI attribute "ROADSIDE" in the ARBM identifies the road segments as undivided (C) or divided (L or R). Obviously, the undivided roads layer includes records with "ROADSIDE" coded as "C". Figure 4-2 shows the final undivided roads layer. For extracting divided roadway sections, the following two rules were applied: (a) include only the records with "ROADSIDE" coded as "L" or "R"; and (b) exclude the records with "FUNCLASS" equal to "01", "02", "11" and "12". Figure 4-3 shows the final divided roads layer that was included in the analysis.

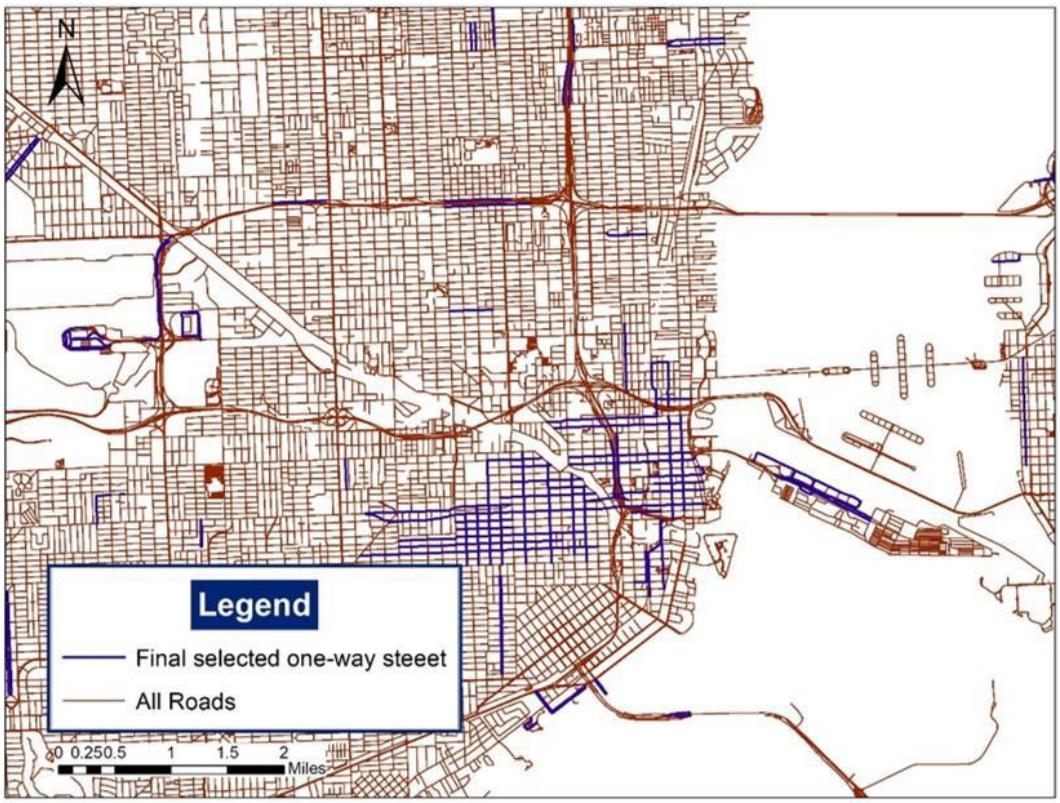


Figure 4-1: One-Way Streets Layer

Figure 4-2: Undivided Roads Layer

Figure 4-3: Divided Roads Layer

4.4 Summary

The WWD crash data analysis was based on five years of crash data from 2012-2016. During the analysis period, a total of 2,879 crashes were identified as potential WWD crashes. Police reports of these 2,879 crashes were obtained and reviewed in detail. Each police report was manually reviewed, and the following information was collected:

- The location where the wrong-way driver potentially turned the wrong-way, if available.
- The exact location of the WWD crash.
- Any cues pertaining to the Wrong-Way incident, if present.
- Other roadway characteristics that may have contributed to WWD crashes (e.g., street lighting, pavement markings, one-way streets, etc.)
- Information related to the crash, such as alcohol involvement, age of the wrong-way driver, the familiarity of the wrong-way driver with the roadway network, etc.

Of the total of 2,879 potential WWD crashes on arterial statewide from 2012-2016, only 1,890 crashes (i.e., 65.6%) were categorized as actual WWD crashes resulting from vehicles traveling the wrong-way. In addition to reviewing the police reports, the RCI and CARS datasets were also processed to identify certain characteristics, such as one-way streets, divided and undivided facilities, etc. RCI and CARS datasets were merged with police report extracted data to prepare the final dataset for analysis.

CHAPTER 5
RESULTS AND DISCUSSION

This chapter is divided into three main sections. The first section presents the descriptive statistics of WWD crashes based on the information extracted from crash summary records. The second section focuses on the comparison of prediction performance of the parametric and the nonparametric statistical methods. Finally, the third section shows identification and pattern recognition of the factors influencing arterial WWD crash severity using a combination of tree-based machine learning techniques.

5.1 Descriptive Statistics

This section provides descriptive statistics of WWD crashes based on the information extracted from crash summary records. Information analyzed included: crash severity, temporal characteristics, first harmful event, WWD driver vehicle speed, lighting conditions, and driver characteristics. Table 5-1 provides a summary of the WWD crash frequency that occurred on arterials during the years 2012 through 2016. As can be inferred from the table, of the total of 2,879 potential WWD crashes, only 1,890 crashes (i.e., 65.6%) were categorized as WWD crashes which occurred as a result of vehicles traveling the wrong-way. A total of 945 crashes (i.e., 32.8%) were not considered to be WWD crashes. Of these 945 non-WWD crashes, a sizable number, i.e., 353 crashes, were head-on crashes that occurred as a result of a driver crossing over the median. From these statistics, it can be inferred that head-on crashes on arterials are frequently incorrectly flagged as WWD crashes.

Table 5-1: Descriptive Statistics by Year

Year	WWD Crash	Not a WWD Crash	Not Sure	Total
2012	206	271	7	484
2013	309	255	6	570
2014	452	166	15	633
2015	491	121	6	618
2016	432	132	10	574
Total	1,890	945	44	2,879

5.1.1 Crash Severity

Table 5-2 provides the WWD crash statistics by year and crash severity. On average, about 6.9% of all WWD crashes resulted in a fatality, while 52.5% of all WWD crashes resulted in an injury. Figure 5-1 provides the distribution of the 1,890 WWD crashes by crash severity.

Table 5-2: WWD Crash Statistics by Year and Crash Severity

Year	Fatal		Injury		PDO		Total
	No.	%	No.	%	No.	%	
2012	14	6.8	121	58.7	71	34.5	206
2013	18	5.8	157	50.8	134	43.4	309
2014	37	8.2	237	52.4	178	39.4	452
2015	32	6.5	246	50.1	213	43.4	491
2016	29	6.7	231	53.5	172	39.8	432
Total	130	6.9	992	52.5	768	40.6	1,890

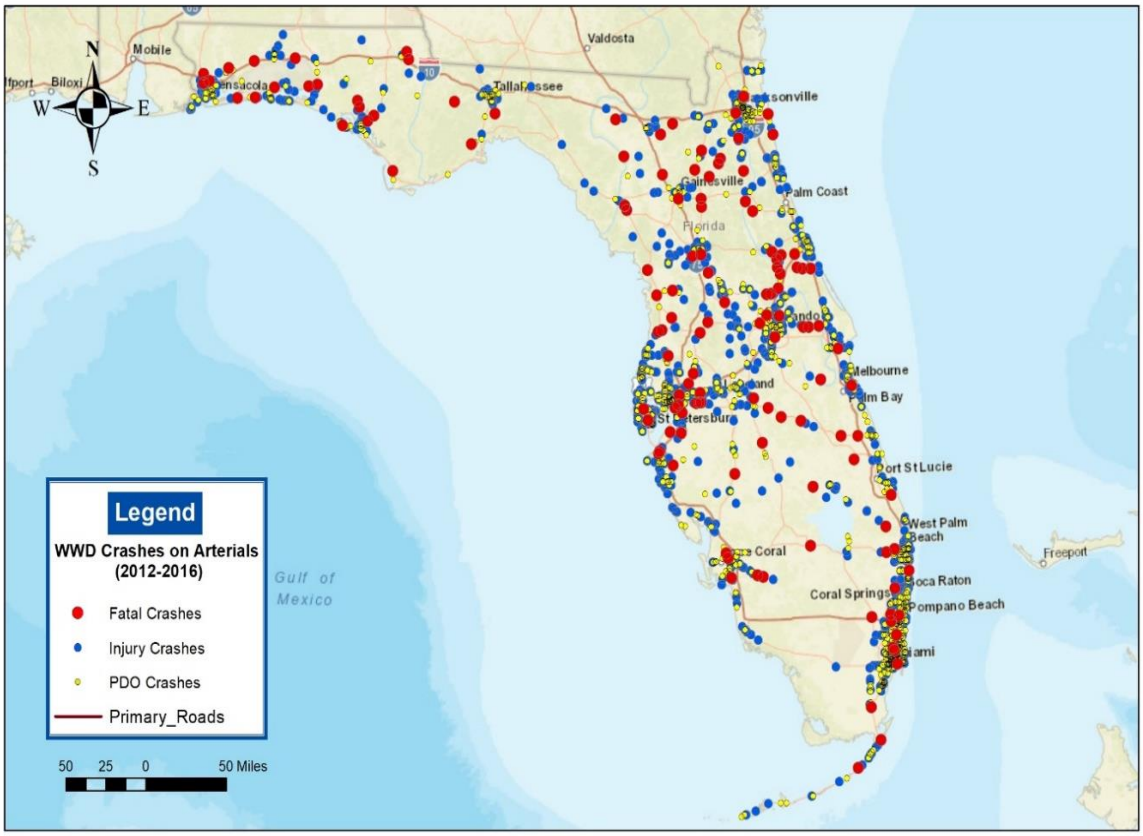


Figure 5-1: WWD Crashes on Arterials in Florida (2012-2016) (Alluri et al., 2019)

5.1.2 Temporal Characteristics

Table 5-3 provides WWD crashes by day of the week. The proportion of fatal WWD crashes (8.3%) on Fridays was higher than the proportion of fatal WWD crashes on typical weekdays (Monday to Thursday) and on weekends (Saturday and Sunday).

Table 5-3: WWD Crash Statistics by Day of Week and Crash Severity

Day of Week	Fatal		Injury		PDO		Total
	No.	%	No.	%	No.	%	
Monday-Thursday	70	7.0	501	50.5	422	42.5	993
Friday	22	8.3	135	51.1	107	40.5	264
Weekend	38	6.0	356	56.2	239	37.8	633
Total	130	6.9	992	52.5	768	40.6	1,890

Table 5-4 gives the distribution of WWD crashes by crash severity and crash time. About 8.4% of the WWD crashes that occurred between 6 am and noon resulted in fatalities. Similarly, 8.1% of the WWD crashes that occurred between midnight, and 6 am were fatal. In terms of both WWD crash frequency and crash severity, the most critical time was found to be from midnight to 6 am, and 6 am to noon. Figure 5-2 presents the hourly distribution of WWD crashes on arterials. WWD crashes were found to be more frequent from 6 pm till about 3 am.

Table 5-4: WWD Crash Statistics by Crash Time and Crash Severity

Time	Fatal		Injury		PDO		Total
	No.	%	No.	%	No.	%	
6 am – Noon	47	8.4	290	51.7	224	39.9	561
Noon – 6 pm	25	5.3	253	53.2	198	41.6	476
6 pm – Midnight	15	4.7	171	53.6	133	41.7	319
Midnight – 6 am	43	8.1	278	52.1	213	39.9	534
Total	130	6.9	992	52.5	768	40.6	1,890

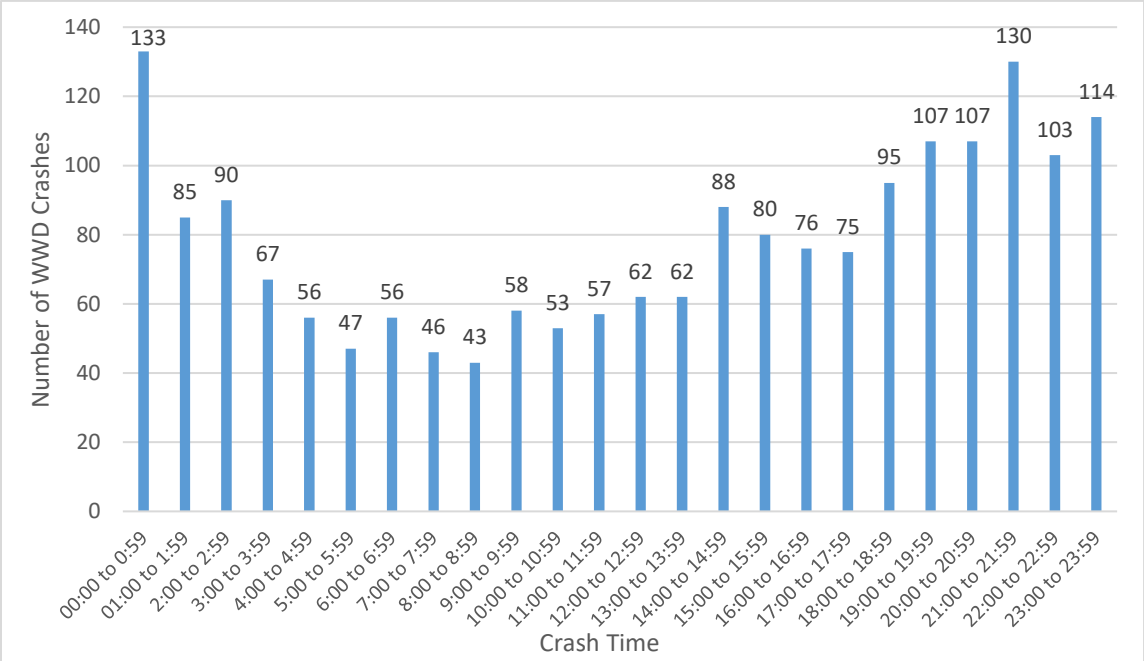


Figure 5-2: Hourly Distribution of WWD Crashes

5.1.3 First Harmful Event

Table 5-5 provides the WWD crash statistics by first harmful event and crash severity. As expected, the proportion of fatalities in WWD crashes that involved collision with other motor vehicles (7.3%), which often result in head-on crashes, was highest compared to other categories. The crashes with first harmful event as non-collision were those that involved mostly single vehicle overturns, rollover, ran into water or canal, etc. Other crashes involved collisions with objects or other road users.

Table 5-5: WWD Crash Statistics by First Harmful Event and Crash Severity

First Harmful Event	Fatal		Injury		PDO		Total
	No.	%	No.	%	No.	%	
Non-collision	2	5.0	23	57.5	15	37.5	40
Collision with Non Motorists	1	6.7	13	86.7	1	6.7	15
Collision with Motor Vehicle	123	7.3	900	53.4	664	39.4	1,687
Collision with Other Non-fixed Objects	0	0.0	3	50.0	3	50.0	6
Collision with Fixed Objects	4	2.8	57	40.1	81	57.0	142
Total	130	6.9	992	52.5	768	40.6	1,890

5.1.4 Vehicle Speed

Table 5-6 gives the WWD crash statistics by the speed of the WWD vehicle and crash severity. As expected, a high proportion of WWD crashes involving vehicles traveling over 45 mph resulted in fatalities. Vehicle speed, as expected, was found to play a significant role in crash severity.

Table 5-6: WWD Crash Statistics by Vehicle Speed and Crash Severity

Speed	Fatal		Injury		PDO		Total
	No.	%	No.	%	No.	%	
15 – 30 mph	1	0.6	68	42.8	90	56.6	159
35 – 45 mph	31	3.3	483	51.3	427	45.4	941
> 45 mph	96	21.0	246	53.8	115	25.2	457
Unknown	2	0.6	195	58.6	136	40.8	333
Total	130	6.9	992	52.5	768	40.6	1,890

5.1.5 Lighting Condition

Table 5-7 provides the WWD crash statistics by lighting condition and crash severity. Over 55% of all WWD crashes occurred during dark conditions; over 64% of all fatal WWD crashes occurred during dark conditions. Moreover, 19% of all WWD crashes that occurred in the dark with no lighting resulted in fatalities. On the other hand, only 4.9% of all WWD crashes that occurred during the daytime resulted in fatalities.

Table 5-7: WWD Crash Statistics by Lighting Condition and Crash Severity

Lighting Condition	Fatal		Injury		PDO		Total
	No.	%	No.	%	No.	%	
Daylight	38	4.9	414	53.6	320	41.5	772
Dusk	4	9.8	23	56.1	14	34.1	41
Dawn	4	13.8	16	55.2	9	31.0	29
Dark Lighted	21	3.0	355	50.3	330	46.7	706
Dark Not lighted	63	19.0	181	54.5	88	26.5	332
Dark Unknown	0	0.0	3	37.5	5	62.5	8
Other	0	0.0	0	0.0	2	100.0	2
Total	130	6.9	992	52.5	768	40.6	1,890

5.1.6 Driver Characteristics

5.1.6.1 Alcohol and Drug Related Crashes

Table 5-8 provides the WWD crash statistics by alcohol and drug involvement and crash severity. Of the 1,890 WWD crashes, 680 crashes (36%) involved intoxicated drivers. It can be inferred from the table that drugs were found to result in more fatalities compared to just alcohol. About 30.5% of WWD crashes that were drugs- and alcohol-related resulted in fatalities, 24.6% of WWD crashes that involved only drugs led to fatalities, while 6.2% of WWD crashes that were alcohol-related resulted in fatalities. When the driver is not impaired, only 4.0% of crashes were fatal.

Table 5-8: WWD Crash Statistics by Driver Impairment and Crash Severity

Alcohol & Drug Involvement	Fatal		Injury		PDO		Total
	No.	%	No.	%	No.	%	
None	48	4.0	638	52.7	524	43.3	1,210
Alcohol	31	6.2	271	54.1	199	39.7	501
Drugs	15	24.6	24	39.3	22	36.1	61
Alcohol & Drugs	36	30.5	59	50.0	23	19.5	118
Total	130	6.9	992	52.5	768	40.6	1,890

5.1.6.2 Driver's Age

Table 5-9 provides the WWD crash statistics by driver's age and crash severity. Approximately 8.8% of WWD crashes involving a driver aged between 35 and 64 years were fatal; 7.6% of WWD crashes involving a driver aged 65 years and older resulted in a fatality.

Table 5-9: WWD Crash Statistics by Driver's Age and Crash Severity

Age	Fatal		Injury		PDO		Total
	No.	%	No.	%	No.	%	
≤ 20 years	6	2.8	107	49.8	102	47.4	215
21 - 34 years	37	5.8	355	55.7	245	38.5	637
35 - 64 years	64	8.8	361	49.5	305	41.8	730
≥ 65 years	23	7.6	166	55.0	113	37.4	302
Unknown	0	0.0	3	50.0	3	50.0	6
Total	130	6.9	992	52.5	768	40.6	1,890

5.1.7 Crash Location

The location where the WWD crashes occurred were divided into the following six categories:

- In close proximity to an intersection,
- Middle of an intersection,
- On a divided roadway,
- On an undivided roadway,

- On a two-way left-turn lane (TWLTL), and
- Other

Table 5-10 provides the WWD crash statistics by crash location and crash severity. Over 50% of WWD crashes (i.e., 986 out of 1,890) occurred at or near intersections. While most of these crashes did not result in fatalities, over 11% of WWD crashes that occurred on divided facilities were fatal. Divided facilities also experienced a high number of WWD crashes; 764 out of 1,890 WWD crashes (40.4%) occurred on divided roadways. WWD crashes on undivided facilities, although relatively rare, resulted in a high proportion of fatalities.

Table 5-10: WWD Crash Statistics by WWD Crash Location and Crash Severity

Location of WWD	**Fatal**		**Injury**		**PDO**		**Total**
	No.	**%**	**No.**	**%**	**No.**	**%**	
In Close Proximity to an Intersection	23	4.0%	286	49.1%	273	46.9%	582
Middle of an Intersection	5	1.2%	216	53.5%	183	45.3%	404
On Divided Roadway	86	11.3%	413	54.1%	265	34.7%	764
Two-way Left-turn Lane	1	4.8%	15	71.4%	5	23.8%	21
On Undivided Roadway	13	12.0%	57	52.8%	38	35.2%	108
Other	2	18.2%	5	45.5%	4	36.4%	11
Total	**130**	**6.9%**	**992**	**52.5%**	**768**	**40.6%**	**1,890**

The police report of each WWD crash was carefully reviewed to determine the precise location where the wrong-way driver might have entered the wrong-way. Table 5-11 provides the WWD crash statistics by the location where the wrong-way driver had potentially entered the wrong-way. As can be observed from the table, the largest proportion of the wrong-way drivers (718 out of 1,890; 38%) turned the wrong-way at a signalized intersection, while 154 drivers turned the wrong-way at a stop sign. About 17.8% of wrong-way drivers were found to enter the wrong-way from a driveway.

Table 5-11: WWD Crash Statistics by Entering Location of the Wrong Way Driver

Entering Location of the Wrong-Way Driver	Number of WWD Crashes
At a Stop Sign	154
At a Signalized Intersection	718
From a Driveway	338
Made a U-turn	26
Other	654
Total	1,890

The location where the WWD might have potentially entered the wrong-way was identified for about 85% of the WWD crashes. This information was used to estimate the distance between the points where the wrong-way drivers possibly entered the wrong-way and the points where the WWD crashes occurred. The actual path that the driver traveled could not be determined based on the information available in the police reports; therefore, the shortest distance between the two points was calculated. Over 95% of the time, WWD crashes were found to occur within 400 ft from where the drivers potentially entered the wrong-way. Figure 5-3 presents the cumulative probability curve of the distance between the WWD entrance location and the WWD crash location. It can thus be inferred that wrong-way drivers, especially on arterials, do not travel far prior to getting involved in a crash. It could also be possible that the drivers traveling the wrong-way on an arterial may quickly recognize their error and turn around. Thus, there would be less exposure for longer distances. Based on the curve in the figure, the probability reaches 1 at around 450 feet; implying that the wrong-way driver has either crashed or turned around beyond 450 feet.

Of the 1,890 WWD crashes, the entering location of the wrong-way driver was mentioned in the police report for 1,068 crashes (i.e., 56.5%). This information was deduced from the illustrations and crash diagram for 737 WWD crashes (39.0%). This information was unavailable for 85 WWD crashes.

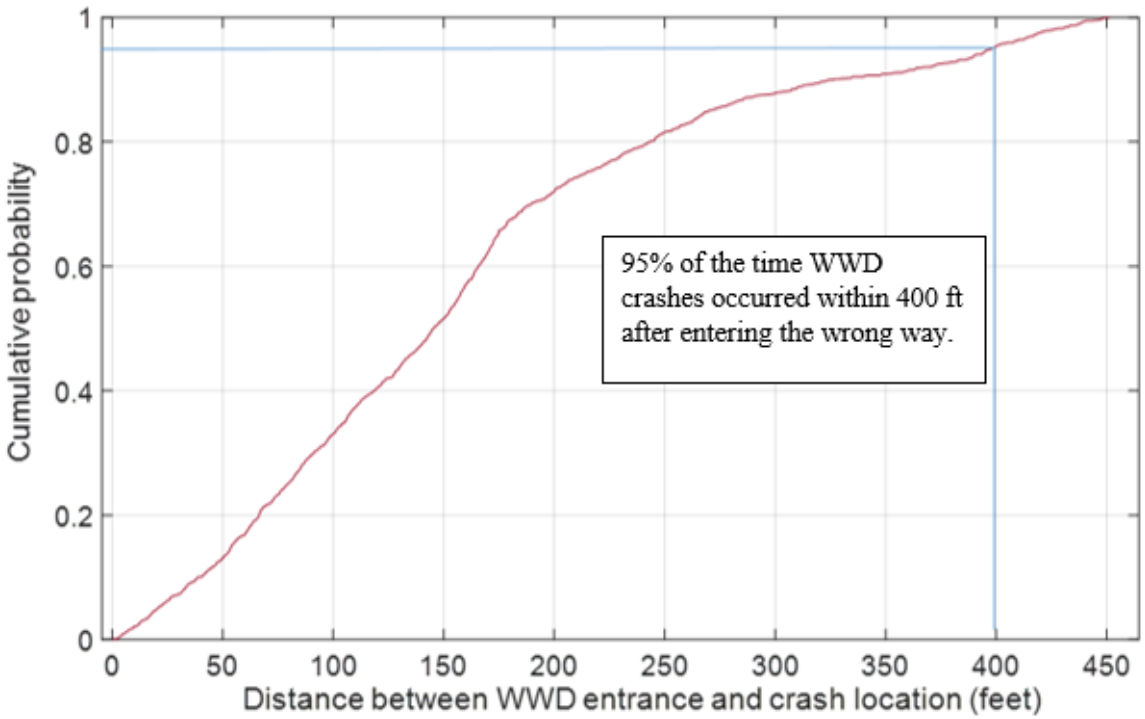


Figure 5-3: Cumulative Probability Curve of the Distance between WWD Entrance Location and WWD Crash Location

5.1.8 Roadway Cross-Section

Table 5-12 provides the WWD crash statistics by roadway cross-section and crash severity. About 13.5% of all WWD crashes were found to occur on one-way streets, and also, these crashes were relatively more severe. About 7.5% of crashes that occurred on one-way streets resulted in a fatality, while the proportion was only 2.4% for crashes that occurred on two-way streets.

Table 5-12: WWD Crash Statistics by Roadway Cross-Section and Crash Severity

Cross-Section	Fatal		Injury		PDO		Total
	No.	%	No.	%	No.	%	
Two-way Street	6	2.4	863	53.8	621	38.7	1,604
One-way Street	120	7.5	118	46.3	131	51.4	255
Unknown	4	12.9	11	35.5	16	51.6	31
Total	130	6.9	992	52.5%	768	40.6%	1,890

5.1.9 Blood Alcohol Concentration

The CARS database has a column to identify if the driver was intoxicated at the time of the crash. This information was extracted from the police reports, where the law enforcement officials enter the Blood Alcohol Concentration (BAC) level of suspected intoxicated drivers. However, the BAC level may not be complete as oftentimes the police reports need to be updated after the crash summary records are populated. For example, in a fatal crash, the BAC level might not be available until after a few days. Also, when the driver refuses to take the BAC test in the field, it may take a few days to receive the test results from the laboratory.

Since the actual BAC level of the driver is not available in the crash summary records, this information is collected from the police reports. Table 5-13 presents the WWD crash statistics by the wrong-way driver's BAC level and crash severity. This information was only available for 60.7% (i.e., 1,148 of 1,890) of all WWD crashes. It can be inferred from the table that a relatively lower 6.1% of WWD crashes involving sober drivers were fatal. When WWD crashes involving intoxicated drivers are considered, almost 60% of all WWD crashes involving drivers with BAC < 0.08 were fatal, while about 14.5% of WWD crashes involving drivers with BAC > 0.08 were fatal. These statistics are counterintuitive. Although there is no empirical evidence to prove, here are a few thoughts and insights:

- If a drunk driver is under 0.08 BAC, the driver may not think that he/she is as impaired as he/she really is, so the driver may not drive too cautiously. On the other hand, if the impaired driver is over 0.08 BAC, the driver probably is feeling a little drunk, and therefore drives slower and more cautiously (so he/she does not get caught for DUI), which significantly reduces the risk of a fatal crash. Impaired

drivers over the 0.08 BAC limit may get involved in more crashes, but since they are going slower, may result in more injuries and fewer fatalities.

- The most likely explanation of this apparent anomaly is that there is a very small sample size of WWD crashes involving impaired drivers with BAC < 0.08, only 22; versus the WWD crashes involving impaired drivers with BAC > 0.08, which has 235. Therefore, the comparison of the percentages is biased.

Table 5-13: WWD Crash Statistics by Driver's BAC Level and Crash Severity

Blood Alcohol Concentration	Fatal		Injury		PDO		Total
	No.	%	No.	%	No.	%	
None	54	6.1	480	53.9	357	40.1	891
Below 0.08	13	59.1	7	31.8	2	9.1	22
Over 0.08	34	14.5	110	46.8	91	38.7	235
Unknown	29	3.9	395	53.2	318	42.9	742
Total	130	6.9%	992	52.5	768	40.6	1,890

5.1.10 WWD Warning Signs

Table 5-14 presents the WWD crash statistics by the presence of WWD warning signs and crash severity. Some of the sample WWD warning signs include DO NOT ENTER, KEEP RIGHT/LEFT, and ONE WAY. It can be inferred from the table that only 309 of 1,890 (16.3%) WWD crashes occurred at locations where there are WWD warning signs. WWD crashes at these corridors were found to be relatively less severe compared to the WWD crashes at corridors where there are no WWD warning signs.

Table 5-14: WWD Crash Statistics by WWD Warning Signs and Crash Severity

Presence of WWD Warning Signs	Fatal		Injury		PDO		Total
	No.	%	No.	%	No.	%	
Yes	10	3.2	146	47.2	153	49.5	309
No	104	8.0	685	52.8	508	39.2	1,297
Unknown	16	5.6	161	56.7	107	37.7	284
Total	130	6.9	992	52.5	768	40.6	1,890

Yes – WWD signs present were DO NOT ENTER, KEEP RIGHT/LEFT, and/or ONE WAY

Table 5-15 shows that over 50% of crashes that occurred on roadways with WWD warning signs occurred on one-way streets, while 47.6% occurred on two-way roadways. For crashes involving roadways with no WWD signs, approximately 94.9% of crashes occurred on two-way roadways, while 4.1% of crashes occurred on one-way roadways. As expected, it appears the WWD warning signs such as DO NOT ENTER, ONE WAY, KEEP RIGHT/LEFT, etc., are prominent along one-way corridors.

Table 5-15: WWD Crash Statistics by WWD Warning Signs and Cross-Section

Presence of WWD Warning Signs	Roadway Cross-Section						
	Two-way		One-way		Unknown		Total
	No.	%	No.	%	No.	%	
Yes	147	47.6	157	50.8	5	1.6	309
No	1231	94.9	53	4.1	13	1.0	1,297
Unknown	226	79.6	45	15.8	13	4.6	284
Total	1,604	84.9	255	13.5	31	1.6	1,890

5.1.11 Roadside Lighting

Corridors with adequate street lighting may experience fewer (and/or less severe) crashes, especially at night. Table 5-16 presents the WWD crash statistics by the presence of roadside lighting and crash severity. WWD crashes that occurred along the corridors with no street lighting were found to be relatively more severe compared to the crashes that occurred along the corridors with street lighting.

Table 5-16: WWD Crash Statistics by Roadside Lighting and Crash Severity

Presence of Roadside lighting	Fatal		Injury		PDO		Total
	No.	%	No.	%	No.	%	
No	89	15.1	327	55.4	174	29.5	590
Yes	39	3.1	647	51.2	578	45.7	1,264
Unknown	2	5.6	18	50.0	16	44.4	36
Total	130	6.9	992	52.5	768	40.6	1,890

5.2 Comparison of Parametric and Nonparametric Models

This research compared the prediction performance of the following five parametric and three nonparametric statistical methods.

- Parametric methods:
  - Logistic Regression (Logit)
  - Least Absolute Shrinkage and Selection Operator (LASSO)
  - Ridge Regression
  - Linear Discriminant Analysis (LDA)
  - Gaussian Naive Bayes (NB)
- Nonparametric methods:
  - Decision Trees (DT)
  - Random Forests (RF)
  - Support Vector Machine (SVM)

Comparisons were based on predicted accuracies of WWD crash severity, where the dependent variable classes include serious injury (20%) and not serious injury (80%). The underlying factors responsible for crash severity consisted of the 21 independent variables, as shown in table 5-17.

Of the 1,890 WWD crashes, crashes with missing information were excluded from the analysis. The final dataset used for analysis contained 1,475 WWD crashes. Five levels of injury severity were included in the crash dataset: fatal (K), incapacitating injury (A), non-incapacitating injury (B), possible injury (C), and property damage only (O). Since the dataset was small, with a data imbalance for these five injury classifications (Table 5-17), the data mining model outcome variable was grouped into two severity categories, Serious

Injury and Moderate to No Injury/Not serious. A serious injury was defined as (KA) = fatal (K) + incapacitating injury (A), and a moderate to no injury was defined as (BCO) = non-incapacitating injury (B) + possible injury (C) + property damage only (O). Note that several researchers recommended using two-class target variables when the dataset has multiclass target variables (Kashani et al., 2011; Allwein et al., 2000).

Table 5-17 provides the summary statistics of the independent variables included in this research. Note that Average Annual Daily Traffic (AADT) was found to range from approximately 17,000 veh/day to 138,500 veh/day and was transformed using a ‘log-based 10' scale for analysis. Other variables were categorized based on data distribution and previous studies.

Table 5-17: Summary Statistics of Variables for Crash Severity Analysis

Variable Type	Variable Name	Variable Category and Description	Frequency	Percent (%)
Response Variable	Original Data Crash Severity	Fatal (K)	118	8
		Incapacitating (A)	181	12
		None- Incapacitating (B)	306	21
		Possible (C)	275	19
		None (O)	595	40
	Binomial Crash Severity	Serious Injury (KA) = Fatal (K) + Incapacitating injury (A)	299	20
		Not serious injury (BCO) = Non-incapacitating injury (B) + Possible injury (C) +No Injury (O)	1176	80
Traffic	Speed Limit	30mph = less than 35mph	363	25
		40 mph = 40mph to 45mph	689	47
		50mph = 50mph to 55mph	311	21
		60mph or above = 60mph and over	112	7
	Traffic Sign	No WWD Related sign	1224	83
		One Way or Do not enter	251	17
Driver-Related	Driver's Age	29 years or younger	463	31.4
		30 to 49 years	504	34.4
		50 years or older	508	34.2
	Driver Impairment	Sober	940	63.6
		Impaired	537	36.4
	Driver Gender	Female	219	15
		Male	1256	85
Temporal	Day of Week (Day)	Weekday = Monday/ Tuesday/ Wednesday/ Thursday	786	53.2
		Weekend = Friday/ Saturday/ Sunday	691	46.8
	Crash Occurrence Time (Time)	00:00-03:59	302	20.5
		04:00-07:59	157	10.6
		08:00-11:59	167	11.3
		12:00-15:59	233	15.8
		16:00-19:59	272	18.4
		20:00-23:59	346	23.4
Vehicle and Collision	Collision/ Impact Type	Head-on	637	43.1
		Angle	357	24.2
		Sideswipe	226	15.3
		Single Vehicle	257	17.4
	Vehicle Point of Impact	Front	1107	75
		Left	126	9
		Other	87	6
		Rear	26	2
		Right	129	9
	Vehicle Body Type	Passenger Car	940	64
		Pickup Utility/Truck / Other	535	36

Table 5-17 (Cont'd): Summary Statistics of Variables for Crash Severity Analysis

Variable Type	Variable Name	Variable Category and Description	Frequency	Percent
Geometric	Road Condition	Wet	184	12.5
		Dry	1,291	87.5
	Median Type	Curb only	401	27.1
		Vegetation only	126	8.5
		Paved only	492	33.3
		Combination of Curb and Vegetation	280	19
		No median (includes one-way street)	178	12.1
	Shoulder Type	Curb	644	44
		Paved	520	35
		Unpaved	311	21
	Crash Location	In very close proximity to an intersection	463	31
		Middle of Intersection	322	22
		On Roadway = Two way left turn lane/ Major approach/Minor approach	690	47
	Possible Entrance Location	At a two way four-way Stop sign	141	141
		At signalized intersection	596	40
		From a driveway	275	19
		Other	463	31
	Number of Lanes	one lane	1219	83
		two lanes	179	12
		three lanes	77	5
	One-way Street	Yes	204	14
		No	1268	1268
	Related to Junction	Junction = Any type of junction related;	369	369
		Non-junction	1106	75
	Skid Resistance / Road Surface Friction Number	30	43	3
		40	929	63
		50	475	32
		60	28	2
	First Harmful Event	Moving Object = Vehicle/ Bicycle/ Pedestrian	1336	91
		Other = Fixed Object / Non- collision	139	9
Environment - Related	Light Condition	Dark-Lighted	652	44
		Daylight	493	33
		Dark-not lighted	330	22
	Weather Condition	Clear	1134	77
		Cloudy	225	15
		Rain	16	8

The final models were based on 1,475 WWD crashes that occurred on arterial roads in Florida from 2012-2016. To make the prediction performance comparison more reasonable, five-fold cross-validation was performed for each parametric and nonparametric model, and the average prediction was selected. Of the 1,475 WWD crashes, 80% (i.e., 1,180 crashes) of the data was identified as a training set, while the remaining 20% (i.e., 295 crashes) was identified as a testing set. Training sets were used to develop each model, and the model was used to predict crash severity in the testing sets.

5.2.1 Prediction Accuracy Comparison

The prediction accuracy was determined by calculating the difference between the predicted crash severity values for the 295 observations in the testing set with the original crash severity values for the same 295 observations. Class accuracy, also known as Sensitivity and Specificity of a categorical prediction analysis, was calculated and summarized as a classification table (Table 5-18), also known as a confusion matrix. The original conditions are actual observed events, and the predicted conditions are conditions obtained from prediction models. In this research, 'not serious' injuries were assigned a positive class, and 'serious injuries' were assigned a negative class. The total number of true positive (TP) and true negative (TN) observations indicate the number of correct predictions when the predictions match the original condition. The number of false positive (FP) and false negative (FN) observations indicate the number of incorrect predictions against observed conditions.

Table 5-18: Classification Table

		Original Condition	
		Positive (i.e., Not Serious)	**Negative (i.e., Serious)**
Predicted Condition	Positive (i.e., Not Serious)	True Positive (TP)	False Positive (FP)
	Negative (i.e., Serious)	False Negative (FN)	True Negative (TN)

Each prediction model was assessed through the following prediction performance criteria: sensitivity, specificity, and accuracy, as shown in Equations 5-1 through 5-3. As not serious is a positive class, here, sensitivity and specificity indicate the accuracy of correctly predicting not serious injury and serious injury, respectively.

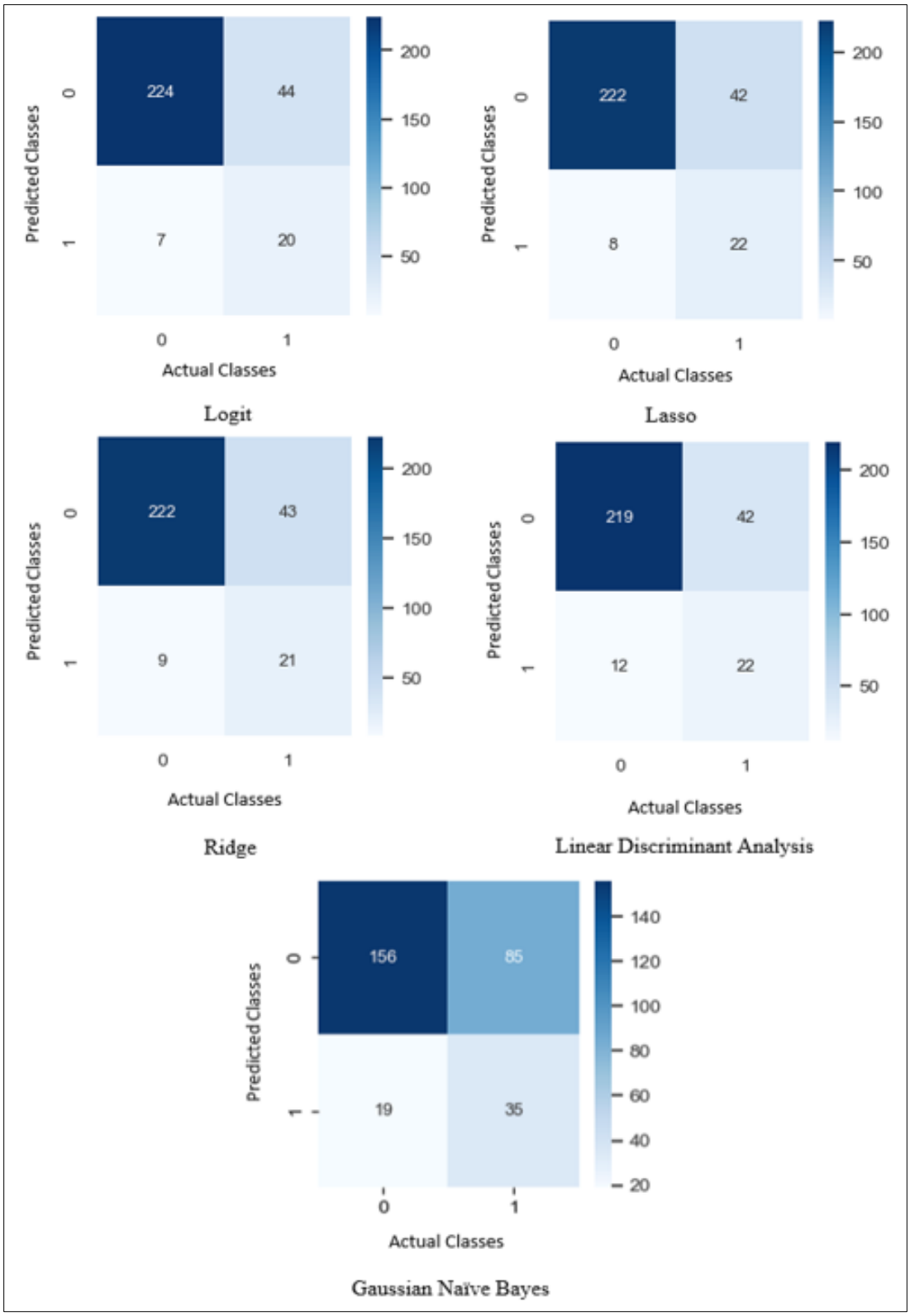
$$Sensitivity = \frac{TP}{TP + FN} \times 100\% \tag{5-1}$$

$$Specificity = \frac{TN}{TN + FP} \times 100\% \tag{5-2}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{5-3}$$

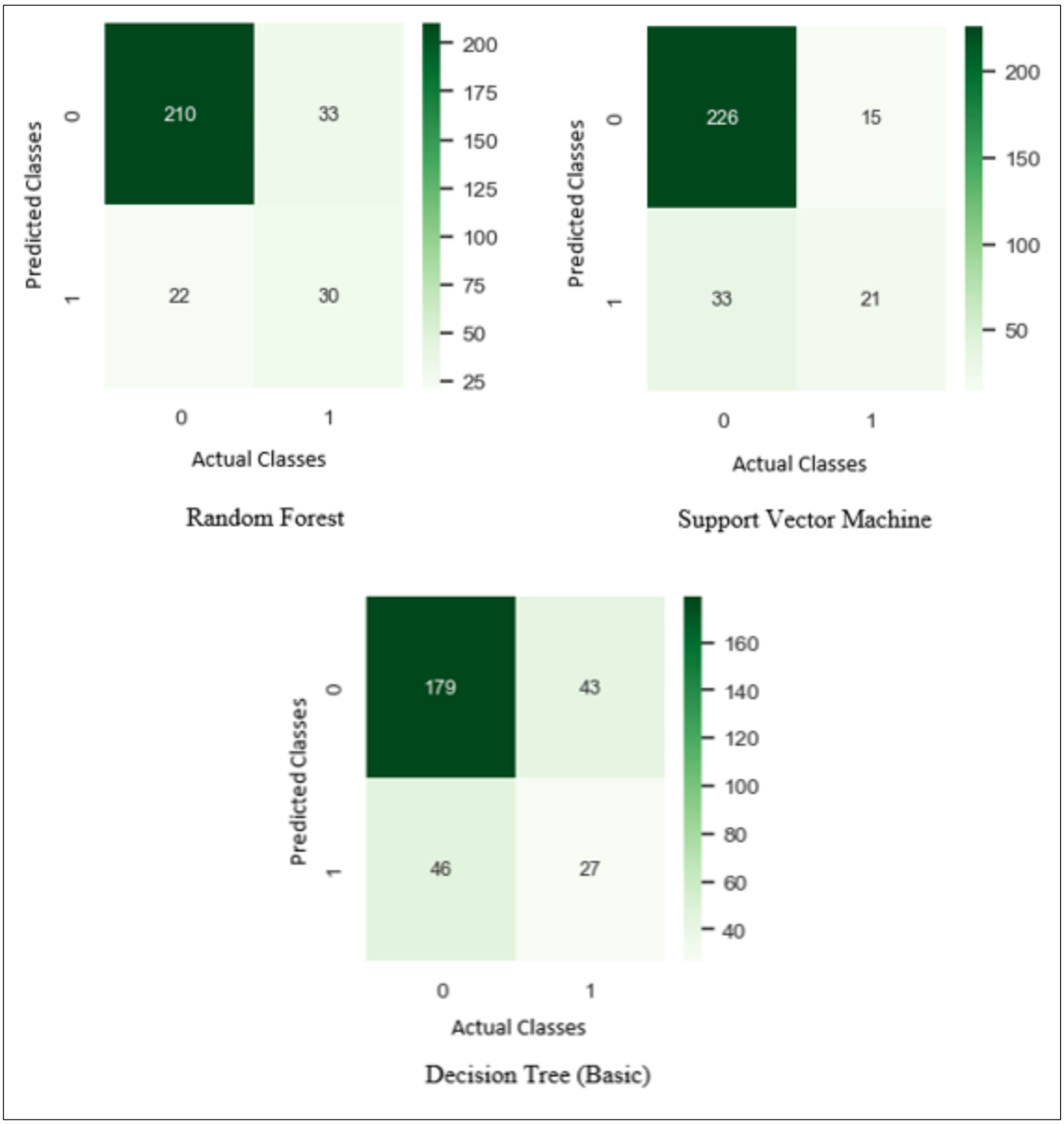
Figures 5-4 and 5-5 provide confusion matrix heatmaps for the parametric and nonparametric methods, respectively. The heatmaps show the distribution of the predicted classes in contrast to the originally observed classes for 295 crashes from the model test sets. As the training set and the testing sets were randomly selected each time, the total number of serious injuries might vary among the testing sets. The left side of the heat maps shows predicted classes, where 0 is a positive class (i.e., not serious injury), and 1 is a negative class (i.e., serious injury). The heatmaps correspond to Table 5-18, where a higher value of 0,0 (true positive) and 1,1 (false negative) indicates higher accuracy. The sensitivity, specificity, and accuracy of each prediction method were obtained from these

values. Table 5-19 provides the average sensitivity, specificity, and prediction accuracy of the parametric and nonparametric methods applied to the 1,475 WWD crashes. Note that the measures are based on testing sets.



[Note: Positive class (not serious injury) = 0; Negative class (serious injury) = 1]

Figure 5-4: Confusion Matrix Heatmap for Parametric Methods



[Note: Positive class (not serious injury) = 0; Negative class (serious injury) = 1]

Figure 5-5: Confusion Matrix Heatmap for Nonparametric Methods

As the data set is imbalanced (i.e., serious injuries are fewer compared to not serious injuries), the specificity is lower than the sensitivity. Note that all the parametric and nonparametric methods gave more weightage to not serious injury crashes. Among the parametric techniques, lasso has the highest accuracy. None of the parametric models performed well in terms of specificity. Among the nonparametric models, RF and support

vector machine (SVM) performed the best with very good specificity. Among the eight models, Gaussian Naïve Bayes performed the worst. It should be noted that accuracy itself does not prove if a model is good or not. Since the goal of this research is to identify factors that affect serious injuries, specificity is also very important. It can be found from the table that despite having an imbalanced dependent variable DT, RF, and SVM, all three data mining approaches not only provided accurate predictions but also better specificity.

Table 5-19: Parametric and Nonparametric Model Prediction Accuracies Summary

Method		Sensitivity % (Not Serious)	Specificity % (Serious)	Accuracy %
Parametric Models	Logit	96.86	31.12	82.74
	Lasso	96.35	34.52	83.01
	Ridge	96.03	33.33	82.38
	Linear Discriminant Analysis	94.62	34.42	81.69
	Gaussian Naïve Bayes	89.14	29.16	64.74
Nonparametric Models	Decision Tree (Raw)	79.55	38.57	69.83
	Random Forests (Raw)	90.48	47.50	81.28
	Support Vector Machine	87.25	58.33	83.72

Since LASSO was found to outperform the remaining three parametric models, the LASSO model results are provided in Table 5-20. This method is suitable for variable selection and regularization. However, due to the WWD severity classification's poor specificity, this model is not the most suitable approach and was only used as a reference in further analyses.

Table 5-20: LASSO Model Summary

Variables	Coefficient for LASSO model
Intercept	-2.591
Max Speed (60 or above)	0.022
Day of Week (Weekend)	0.219
Time Interval (12 to 15:59)	0.307
Time Interval (0 to 3:59)	0.285
Time Interval (4 to 7:59)	-0.314
Light Condition (Dark Not Lighted)	1.673
Shoulder Type (Paved)	0.001
Median Type (Paved)	0.023
Median Type (Vegetation)	0.161
Median Type (Curb and Vegetation)	-0.070
Number of Lanes (2 Lane)	-0.079
One Way Street (Yes)	-0.409
Traffic Sign (One way/ Do Not Enter)	-0.109
Entrance Location (At Signalized Intersection)	-0.162
Crash Location (Middle of Intersection)	-0.010
Crash Location (On Roadway/ Not Near Intersection)	0.695
Impact Type (Head-on Collision)	0.343
Vehicle Point of Impact (Left Side)	0.160
Driver's Age (30 years to 49 years)	-0.351
Driver's Age (50 Plus)	0.301
Impairment (Impaired)	0.811

Among the nonparametric models, the SVM and RF performed relatively better. However, SVM cannot directly produce any variable importance ranking. Overall, the RF approach performed better than the DT method because RF are a bootstrap of multiple smaller DTs. On the other hand, the DT can produce a visible pattern between independent and dependent variables using a tree structure that RF cannot directly show. Although DTs are simple tree models but coupled with clustering and variable selection RF prior to DT analysis and pruning the tree, DT can achieve higher accuracy levels. The next section discusses a more in-depth analysis of the applicability of nonparametric techniques in

WWD crash severity analysis. It discusses the marginal effect of data mining techniques followed by the application of a combination of data mining techniques on WWD crashes.

5.2.2 Marginal Effect of Data Mining Technique

Generalized linear models (GLMs) are the most popular models favored by analysts and many researchers, as they establish a quantitative relationship between response variables and explanatory variables. However, GLMs are based on several assumptions, and the results could become questionable when these assumptions are violated. On the contrary, data mining models have a strong capability in handling complex databases and do not require assumptions. However, one of the limitations of the newer predictive models, such as machine learning (ML) techniques, is the lack of interpretation. For instance, when using black-box ML algorithms such as the SVM or RF, it is hard to understand the relations between predictors and the model outcome. In such models, it is necessary to evaluate the magnitude of the relationship and determine the direction of the relationship between the dependent and independent variables. For that reason, the marginal effect analysis is conducted in this research by showing the partial dependent plots. Marginal effect analysis is a process to assess the effect of explanatory variables on the response variable. It describes the relationship between the response variable and the explanatory variables through its range while holding the other variables constant (Goldstein et al., 2015; Fish & Blodgett, 2003). The partial dependence plot (PDP) shows the marginal effect of each explanatory variable on the predicted result of a data mining model (Friedman, 2001). The PDP can be viewed as a graphical representation of contributor coefficients for each independent variable. PDPs are not directly drawn from the raw data; instead, they are constructed from predictions based on the model. The

marginal effect analysis has not been implemented with RF in transportation studies yet. But were used in some other fields of research. For instance, Berk and Bleich (2013) used RF and the associated PDPs to accurately forecast criminal behaviors and demonstrated the advantage of this technique. The study showed an accurate predictor-response relationship, even when the response variable categories were imbalanced. In this research, PDPs were constructed from the RF model to analyze the marginal effect of the contributing factors on WWD crashes. First, the variable importance plot was generated (Figure 5-6). Details of variable importance are explained in Section 5.2 (Combination of Data mining, after AHC). Figures 5-7 through Figure 5-21 illustrate the PDPs of RF on WWD crash severity analysis.

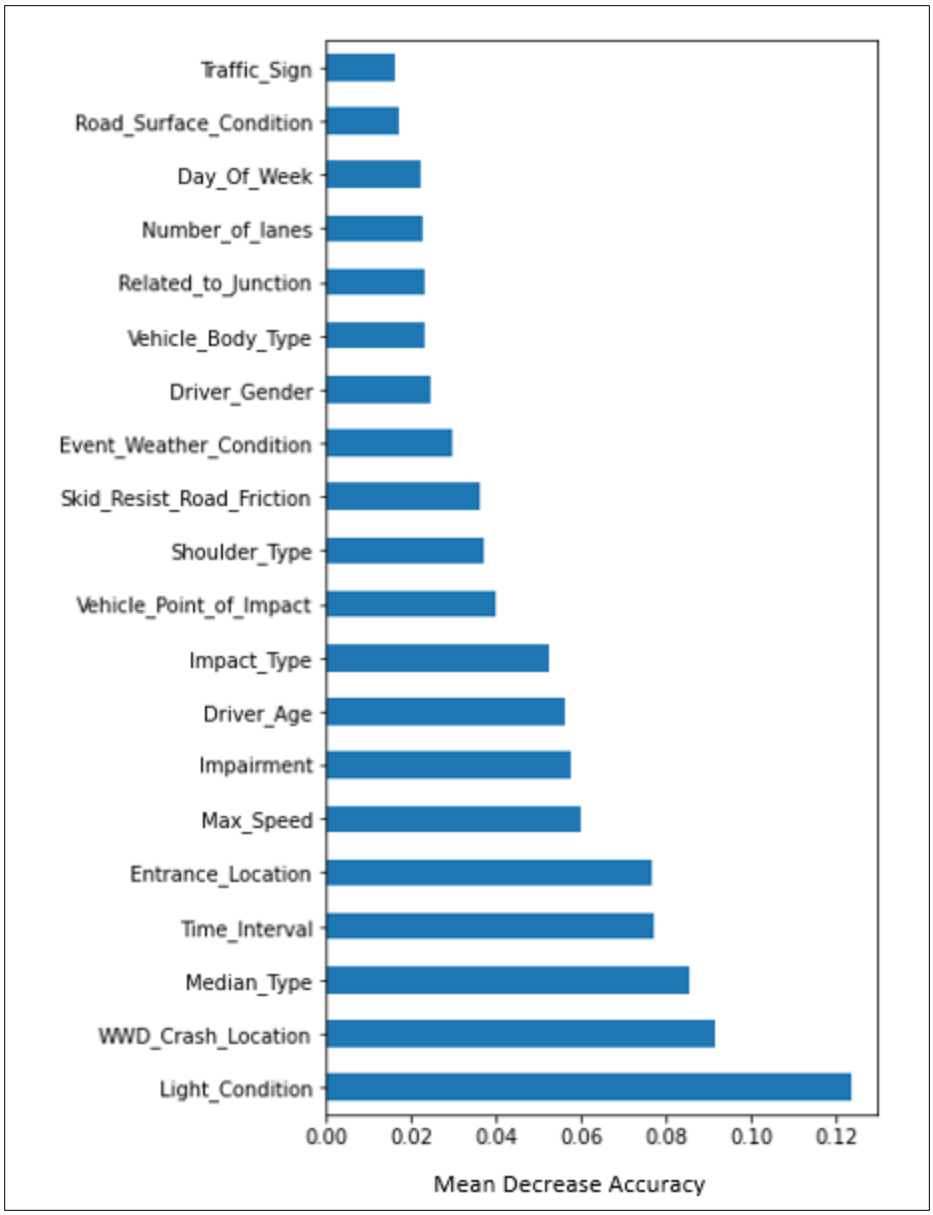


Figure 5-6: Random Forests Variable Importance

From Figures 5-7 through Figure 5-21, the y-axis represents the dependent variable. A positive y value implies that the contributing variable positively affects the classification in the model or predicts higher likelihood of "serious injury". On the contrary, a negative value indicates the opposite. The blue shaded area shows the level of confidence. The x-axis has the assigned feature. A feature value '0' on the x-axis means when a condition is

false or not true or otherwise of that feature class. In contrast, ‘1’ on the x-axis means true for the feature. For instance, Figure 5-7 has the marginal effect of ‘light condition – dark not lighted’ on crash severity is illustrated. When the ‘light condition – dark not lighted’ value increases from 0 to 1, from left to right of the x-axis, the value of y increases. This implies that when light condition = dark not lighted becomes true, the serious injury is more likely to occur. On the contrary, in Figure 5-8, when light condition = dark not lighted’ increases and becomes 1 (true), then the y value becomes negative, indicating the probability of serious injury decreases. Similarly, in Figure 5-9, when light condition = daylight, the probability of serious injury decreases. Therefore, from Figure 5-7, 5-8, 5-9, we can infer that light conditions with daylight or dark-lighted condition a negative correlation with serious injury, and dark-not lighted conditions will have a positive correlation with serious injury. Analyzing Figures 5-7 through 5-21, Table 5-21 is constructed for marginal effect on severity for all the influential variables.

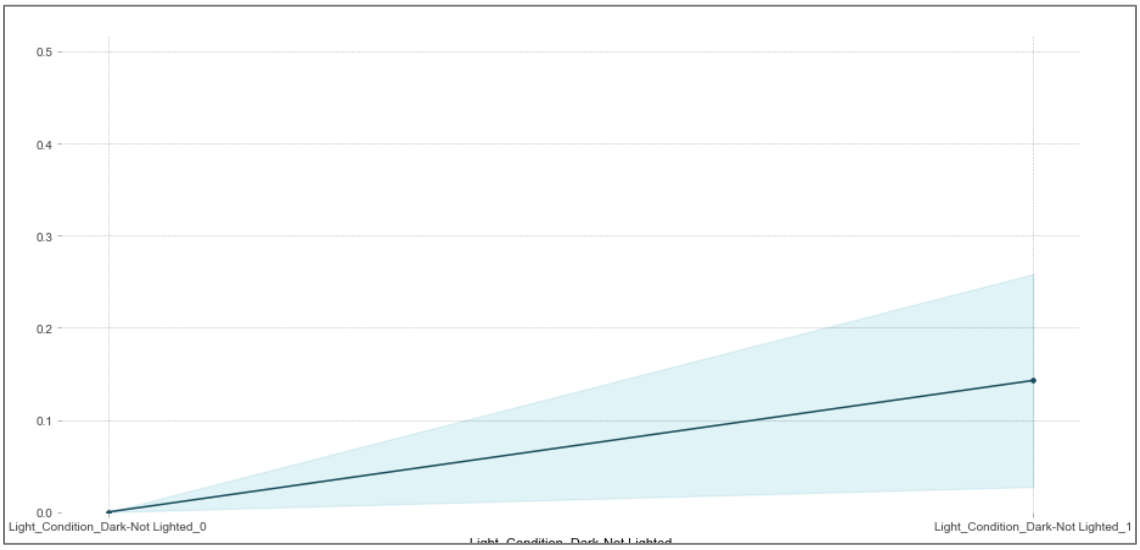


Figure 5-7: PDP (RF) for Severity vs. Light Condition Dark-Not Lighted

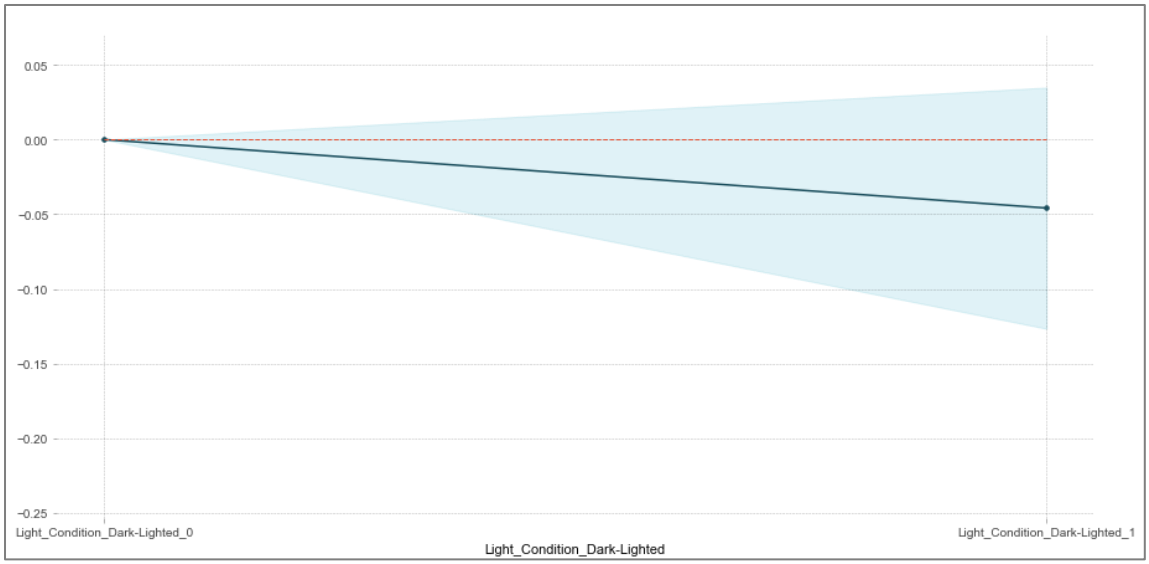


Figure 5-8: PDP (RF) for Severity vs. Light Condition Dark-Lighted

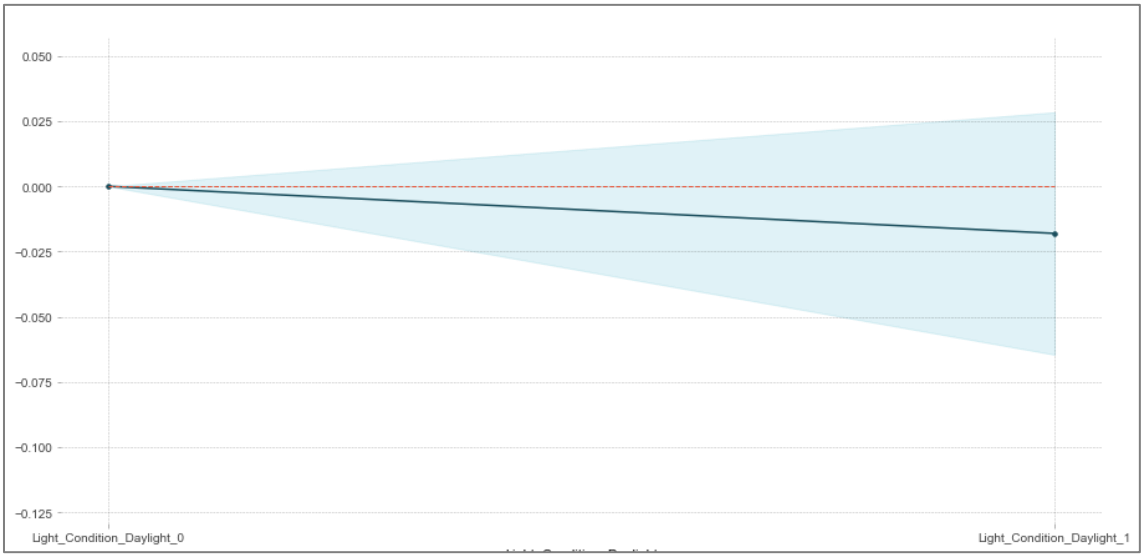


Figure 5-9: PDP (RF) for Severity vs. Light Condition-Day Light

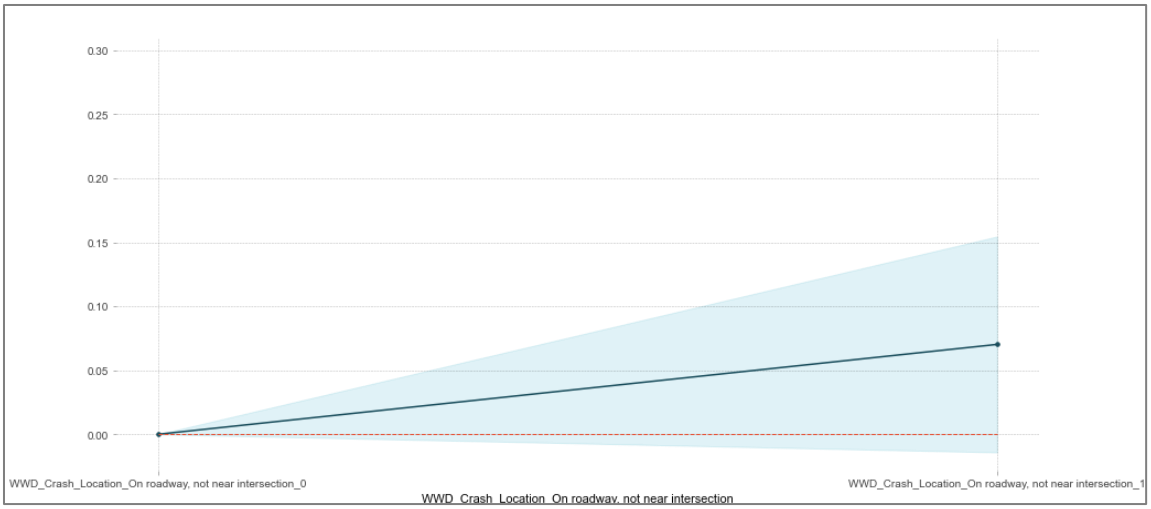


Figure 5-10: PDP (RF) for Severity vs. Crash Location-On Roadway/Not Near Intersection

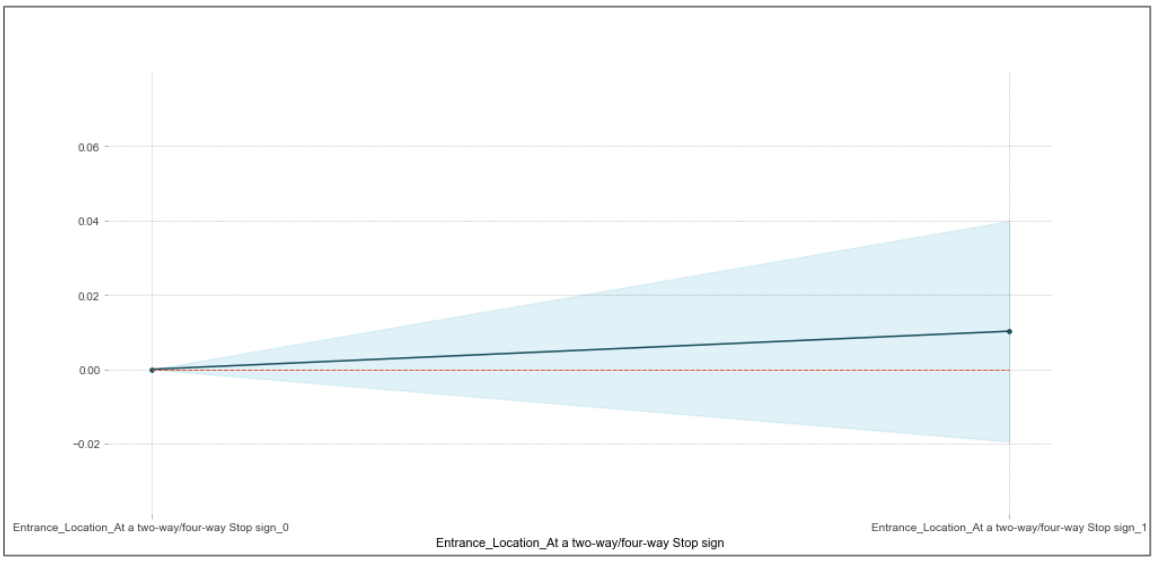


Figure 5-11: PDP (RF) for Severity vs. Entrance Location-Two-way/ Four-way Stop Sign

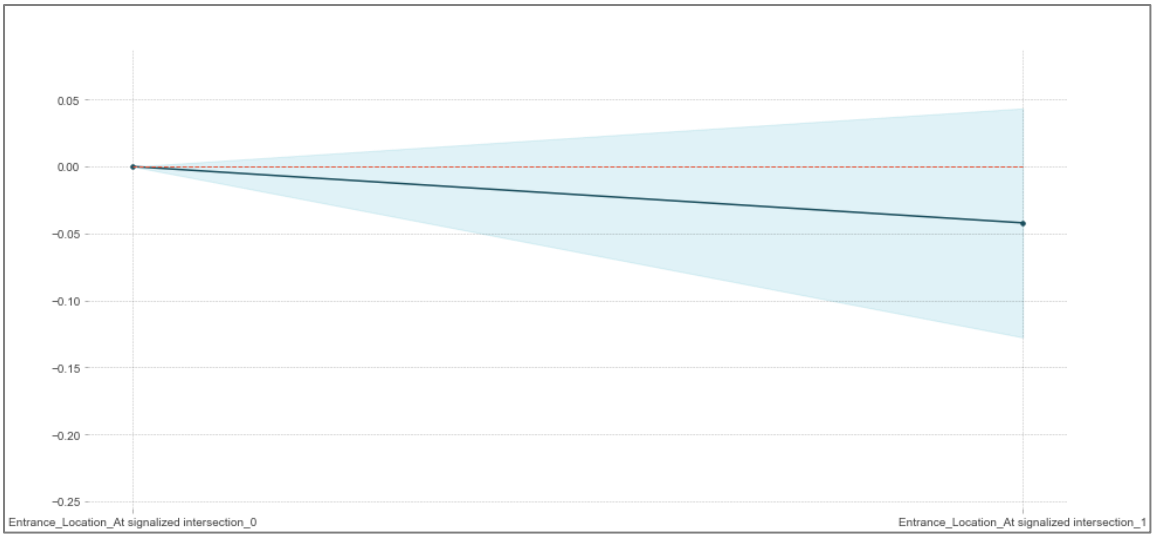


Figure 5-12: PDP (RF) for Severity vs. Entrance Location-At Signalized Intersection

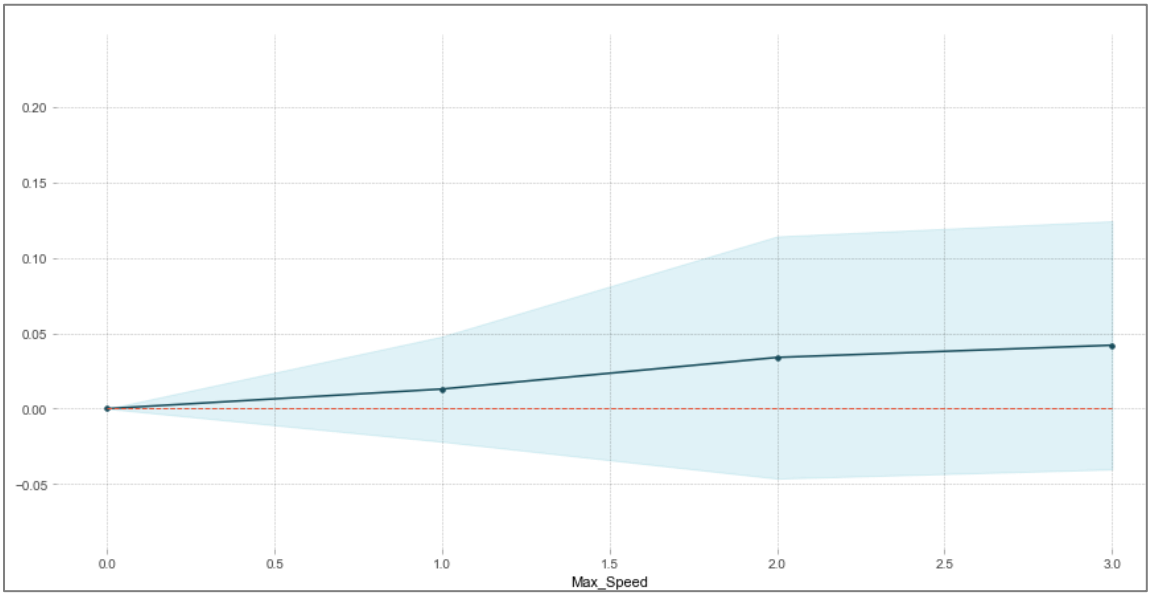


Figure 5-13: PDP (RF) for Severity vs. Max Speed

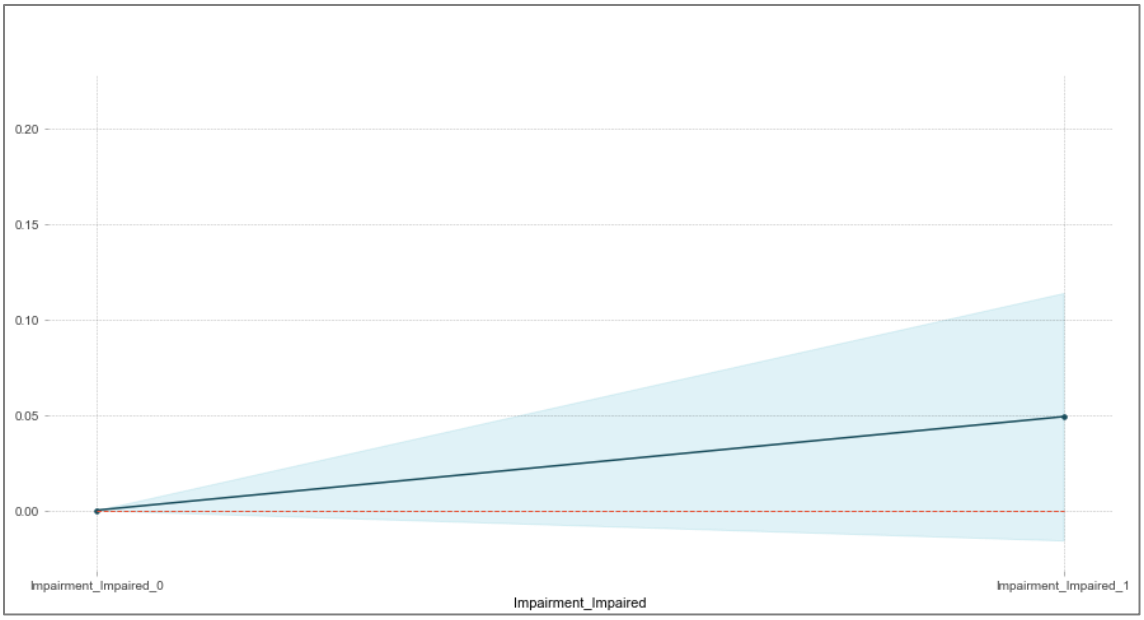


Figure 5-14: PDP (RF) for Severity vs. Impairment-Impaired

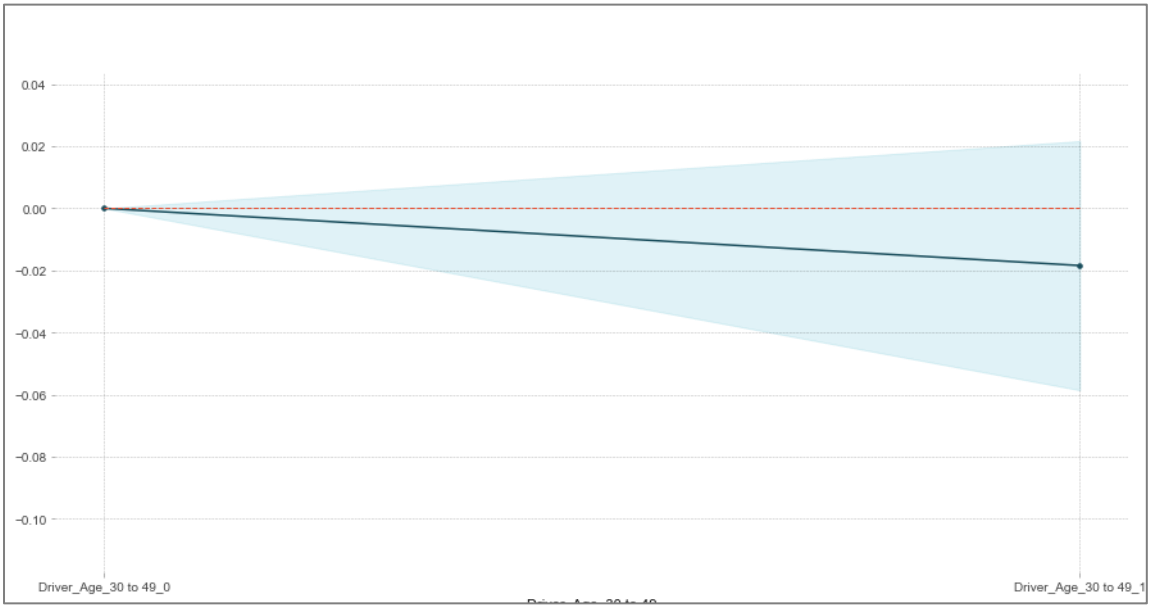


Figure 5-15: PDP (RF) for Severity vs. Driver Age-(30 to 49)

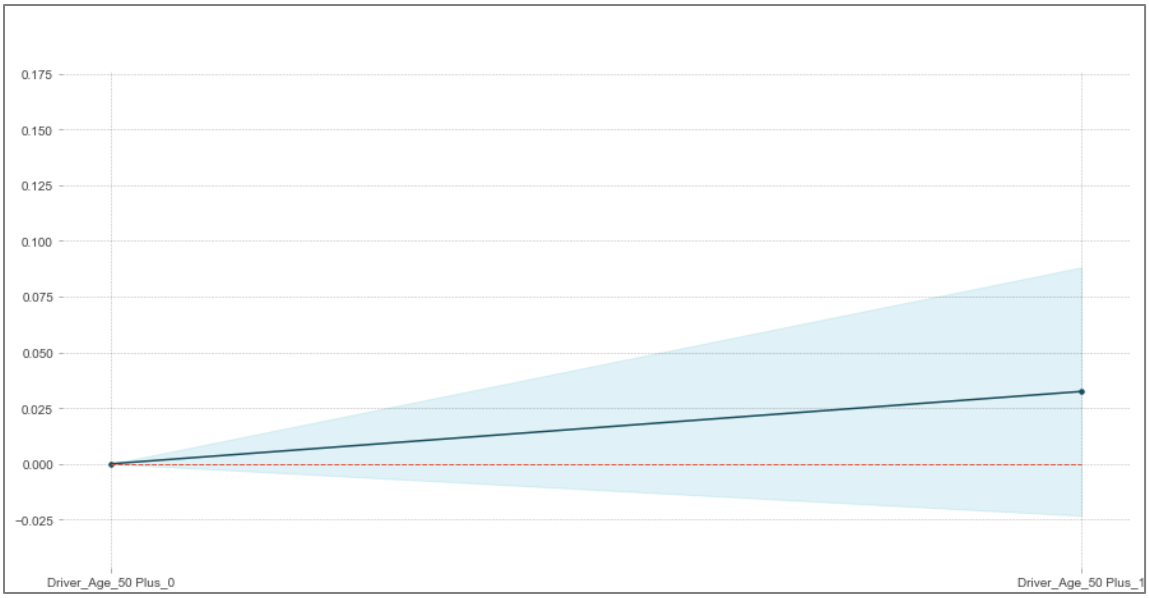


Figure 5-16: PDP (RF) for Severity vs. Driver Age-(50 and up)

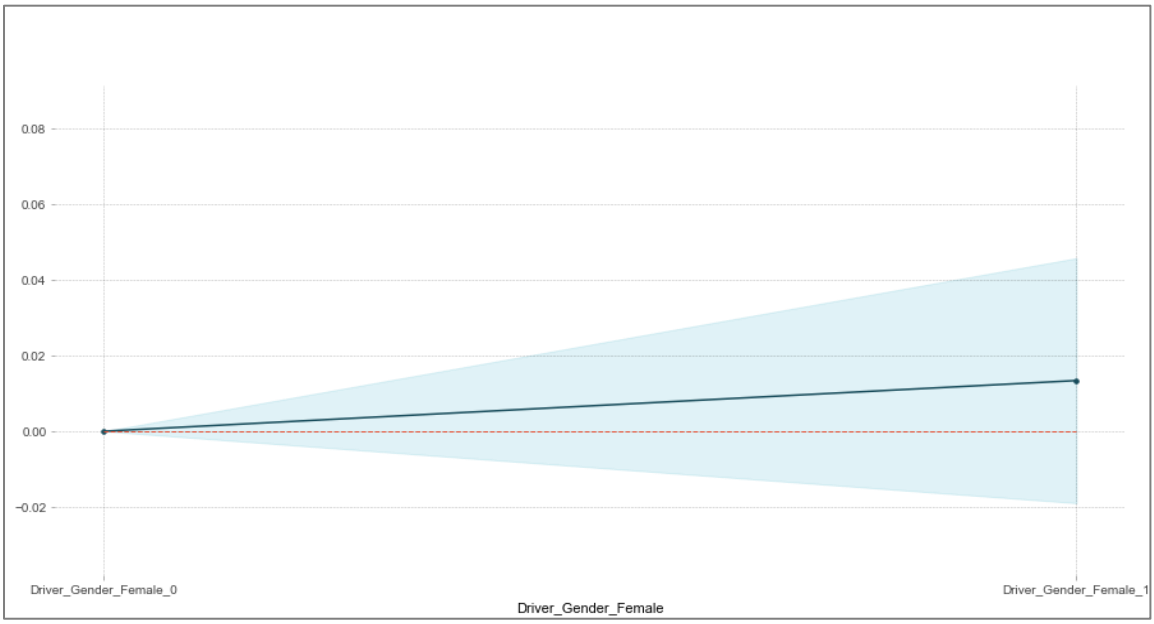


Figure 5-17: PDP (RF) for Severity vs. Driver Gender-Female

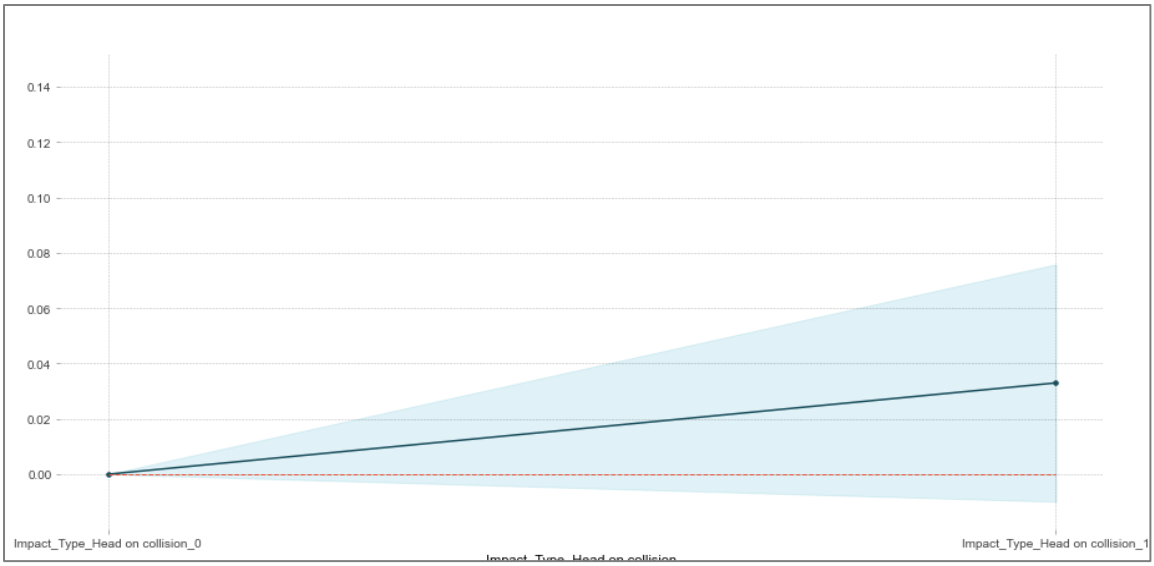


Figure 5-18: PDP (RF) for Severity vs. Impact Type Head-on Collision

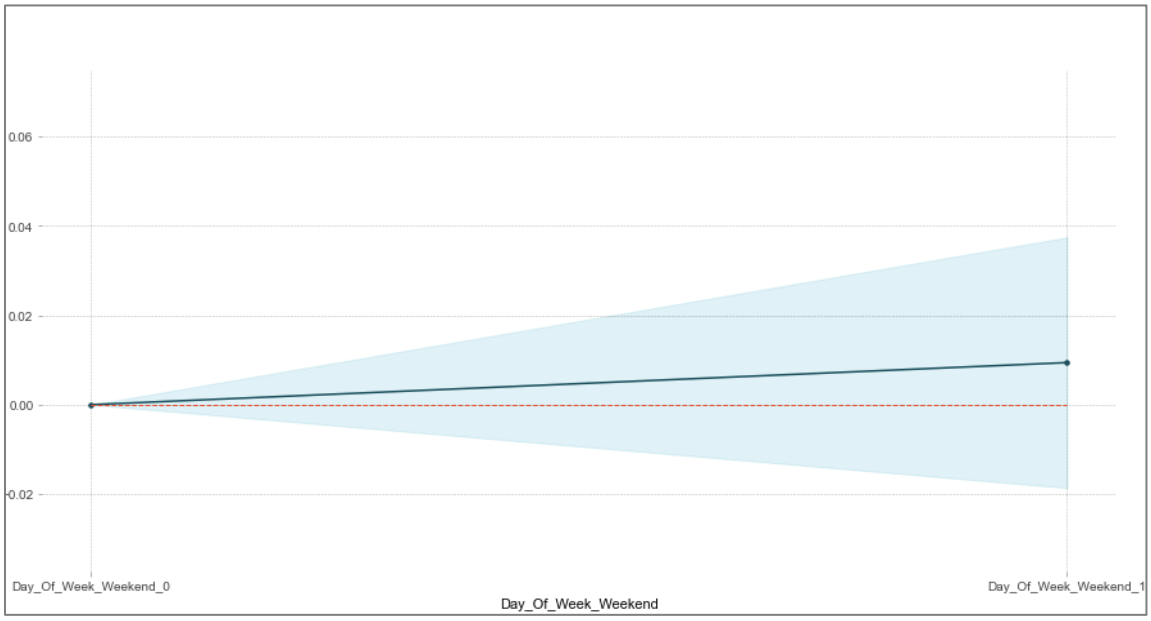


Figure 5-19: PDP (RF) for Severity vs. Day of Week Weekend

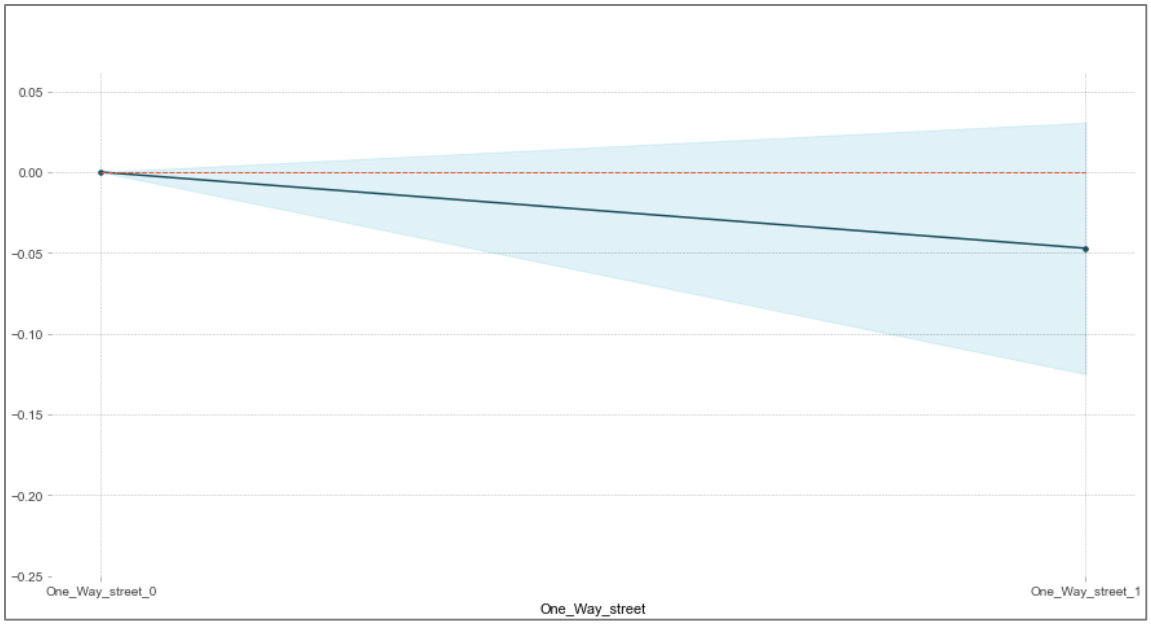


Figure 5-20: PDP (RF) for Severity vs. One-way Street

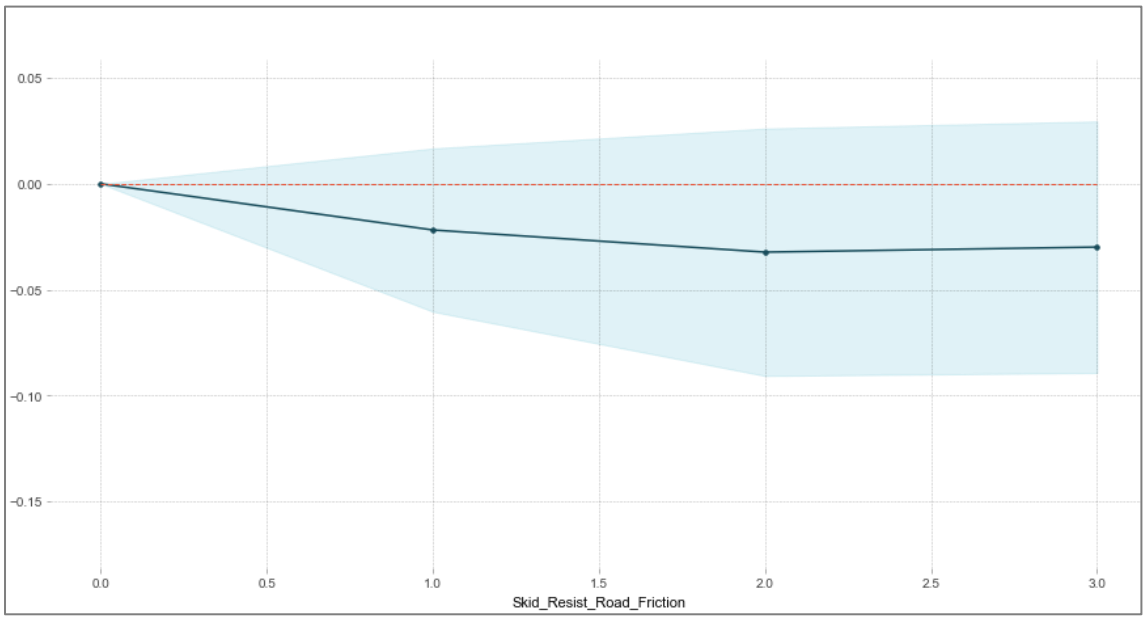


Figure 5-21: PDP (RF) for Severity vs. Skid Resistance- Road Friction

Table 5-21: Marginal Effect of Influential Variables from Random Forests

Influential Variables	Figure No.	Correlation	Probability of Serious Injury
Light Condition: Dark Not Lighted	5-4	Positive	Increase
Crash Location: On Roadway/Not Near Intersection	5-7	Positive	Increase
Entrance Location: Two-way/ Four-way Stop Sign	5-8	Positive	Increase
Increase of Max Speed	5-10	Positive	Increase
Impairment: Impaired	5-11	Positive	Increase
Driver Age: 50 and up	5-13	Positive	Increase
Gender: Female	5-14	Positive	Increase
Impact Type: Head-on Collision	5-15	Positive	Increase
Day of Week: Weekend	5-16	Positive	Increase
light condition dark-lighted	5-5	Negative	Decrease
Light Condition-Day Light	5-6	Negative	Decrease
Entrance Location-At Signalized Intersection	5-9	Negative	Decrease
Driver Age-(30 to 49)	5-12	Negative	Decrease
One-way Street	5-17	Negative	Decrease
Increase in Skid Resistance- Road Friction	5-18	Negative	Decrease

The results from Table 5-21 show the marginal effect of different independent variables in the dependent variable. From the RF marginal effect analysis of partial dependence plot ‘Light Condition: Dark Not Lighted’, ‘Crash Location: On Roadway/Not Near Intersection’, ‘Entrance Location: Two-way/ Four-way Stop Sign’, ‘Increase of Max Speed’, ‘Impairment: Impaired’, ‘Driver Age: 50 and up’, ‘Gender: Female’, ‘Impact Type: Head-on Collision’, and ‘Day of Week: Weekend’ are found to have a positive correlation with the response variable ‘crash severity’, meaning increase or presence of these variables increase the likely hood of serious injuries. On the other hand, ‘Light condition dark-lighted’, ‘Light Condition-Day Light’, ‘Entrance Location-At Signalized Intersection’, ‘Driver Age-(30 to 49)’, ‘One-way Street’, ‘Increase in Skid Resistance- Road Friction’ are negatively correlated with the response variable ‘crash severity, meaning increase or presence of these variables decrease likelihood of serious injuries, and increase the likelihood of not serious injuries.

For comparison purposes, this research compared the nonparametric RF model's marginal effects with the parametric model's coefficients. The comparison shows that marginal effect correlation results are consistent and aligns with the predicted correlation of the LASSO model's selected variables. Some variables were dropped by LASSO.

As the RF model's variable marginal effect results are consistent with the prediction of the parametric model (Table 5-22), the study demonstrates that data mining techniques such as RF are valid and can serve as an alternative tool in transportation safety studies. For more detailed analysis, the next section uses a combination of data mining techniques on WWD crashes on arterials to explore the underlying pattern of the factors affecting the WWD crashes.

Table 5-22: RF Marginal Effects Contrast to LASSO Coefficients

Variables	Coefficient for LASSO Model	RF
Intercept	-2.591	NA
Max Speed (60 or above)	0.022	Positive
Day of Week (Weekend)	0.219	Positive
Time Interval (12 to 15:59)	0.307	Positive
Time Interval (0 to 3:59)	0.285	Positive
Time Interval (4 to 7:59)	-0.314	Negative
Light Condition (Dark Not Lighted)	1.673	Positive
Shoulder Type (Paved)	0.001	Positive
Median Type (Paved)	0.023	Negative
Median Type (Vegetation)	0.161	Positive
Median Type (Curb and Vegetation)	-0.070	Negative
Number of Lanes (2 Lane)	-0.079	Negative
One Way Street (Yes)	-0.409	Negative
Traffic Sign (One way/ Do Not Enter)	-0.109	Negative
Entrance Location (At Signalized Intersection)	-0.162	Negative
Crash Location (Middle of Intersection)	-0.010	Negative
Crash Location (On Roadway/ Not Near Intersection)	0.695	Positive
Impact Type (Head-on Collision)	0.343	Positive
Vehicle Point of Impact (Left Side)	0.160	Positive
Driver's Age (30 years to 49 years)	-0.351	Negative
Driver's Age (50 Plus)	0.301	Positive
Impairment (Impaired)	0.811	Positive
Road Friction	Null (Shrinkage)	Negative
Gender (Female)	Null (Shrinkage)	Positive

5.3 Crash Severity Prediction Using Combination of Tree-based Models

In this research, the WWD dataset output variable was class imbalanced, where the serious injury (KA) class had 1,176 observations, and the moderate injury to no injury (BCO) class had 299. To remedy this class imbalance, the library ROSE in RStudio (Menardi & Torelli, 2014), along with the ovun.sample function's 'both' resampling method, with an embedded mixed resampling technique, was used. This method balances the dataset by randomly resampling from the dataset and giving equal weight to both classes of the output binomial variable (Lunardon et al., 2014; Menardi & Torelli, 2014). This method was performed after clustering analysis and before the RF and DT analysis.

For clustering analysis, the AHC algorithm was set to produce four dendrograms for the dataset, separated based on roadway geometric features. The Hclust package in RStudio software was used in this analysis (Müllner, 2018). This function performs the AHC analysis with the use of a set of dissimilarities for the *n* objects being clustered. All the geometric features from Table 5-17, along with AADT were used as input variables to cluster the data. Figure 5-22 illustrates the dendrograms, where the y-axis shows the height of the dendrogram, and the x-axis shows the dissimilarity measured in the distance, d. For distance measurement in the AHC, Ward's minimum variance linkage method performed better than other linkage methods. This is because Ward's method performs well even when there is some noise in the dataset. Gower distance was used to calculate the dissimilarity measure because it can handle mixed-type data or categorical data. Finally, data within each cluster were extracted for further analysis. Table 5-23 provides the sample size in each cluster.

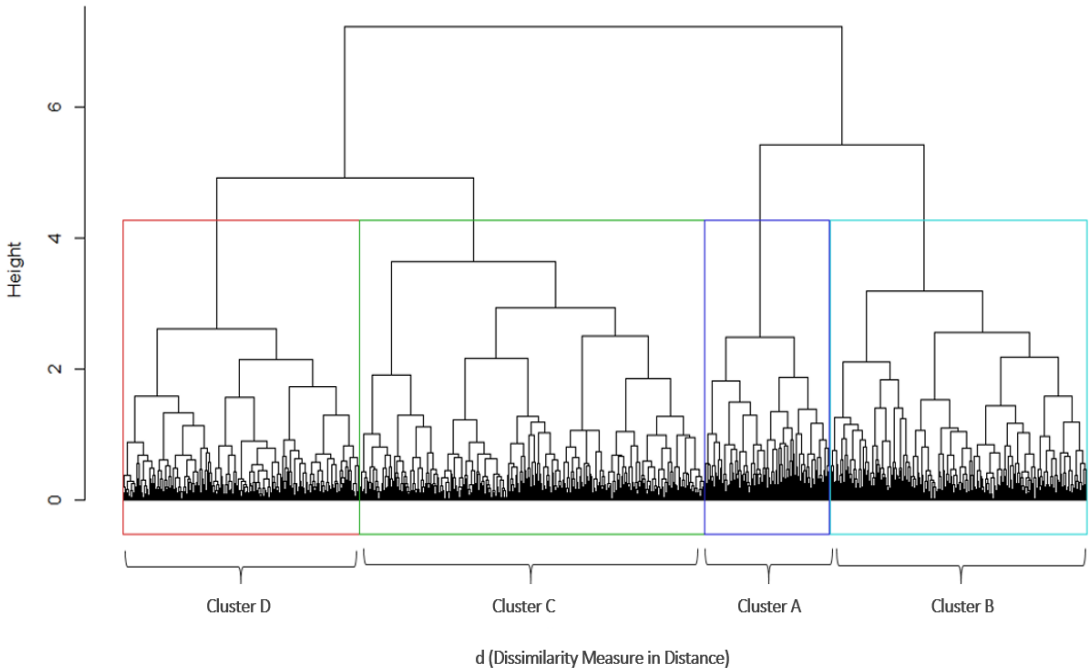


Figure 5-22: Data Divided into Four Clusters for Analysis

5.3.1 Variable Importance

After the application of AHC, each cluster sample was analyzed using the "randomForest" library in the RStudio software (Liaw & Wiener, 2002). In RF, a forest was grown using 500 trees (ntree) and randomly selecting three features (Mtry) at a time. Classification trees and association rules can suffer from an extreme risk of type I error due to the large number of patterns considered (Webb, 2007). Therefore, to reduce the risk of type I error, validation was performed, and the dataset was randomly split into two parts: a training sample (70% of the total dataset) and a test sample (30% of the total dataset). To indicate variable purity, the variables were arranged in decreasing order, from most important to least important, based on the Gini index measure. Figure 5-23 provides the results from the RF algorithm performed on the four clusters. WWD crash location, time

of the crash, light condition, impairment, driver's age, impact type, median type, speed limit, skid resistance (i.e., surface friction), vehicle body type, vehicle point of impact, weather condition, WWD entrance location are among the high impact variables in the clusters. The Gini index cut-off value was set in such a way that only the top 15 variables were selected for further analysis of each cluster in the CART model (Table 5-23).

Table 5-23: Cluster Sample Distribution and Summary of RF

Cluster	Serious Injury Crashes (K & A)	Moderate to No Injury Crashes (B, C & O)	Total Sample Size	High Impact Variables from RF (Top 15)	RF Accuracy (%) (1-OOB)%
A	35	155	190	WCL, COT, LC, EL, MT, DA, VPI, SR, IT, SP, IM, WC, VBT, ST, RC	76.32
B	84	310	394	EL, WCL, COT, LC, SP, IT, DA, MT, IM, VBT, VPI, SR, WC, DG, D	79.31
C	98	431	529	LC, MT, EL, COT, IT, SP, IM, SR, DA, VPI, WCL, NL, D, VBT, WC	87.56
D	82	280	362	WCL, IT, COT, LC, MT, SP, IM, DA, WC, EL, VPI, OWS, SR, VBT, NL,	80.69
Total	299	1176	1,475	Not Applicable	

COT = Crash Occurrence Time/ Time Interval, D = Day of Week, DA = Driver's Age, DG = Driver's Gender, EL = Entrance Location, IM = Impairment, IT = Impact Type, LC = Light Condition, MT = Median Type, NL = Number of Lanes, OWS = One-way Street, RC = Road Surface Condition, SP = Speed, SR = Skid Resistance, ST = Shoulder Type, VBT = Vehicle Body Type, VPI = Vehicle Point of Impact, WC = Weather Condition, WCL = WWD Crash Location.

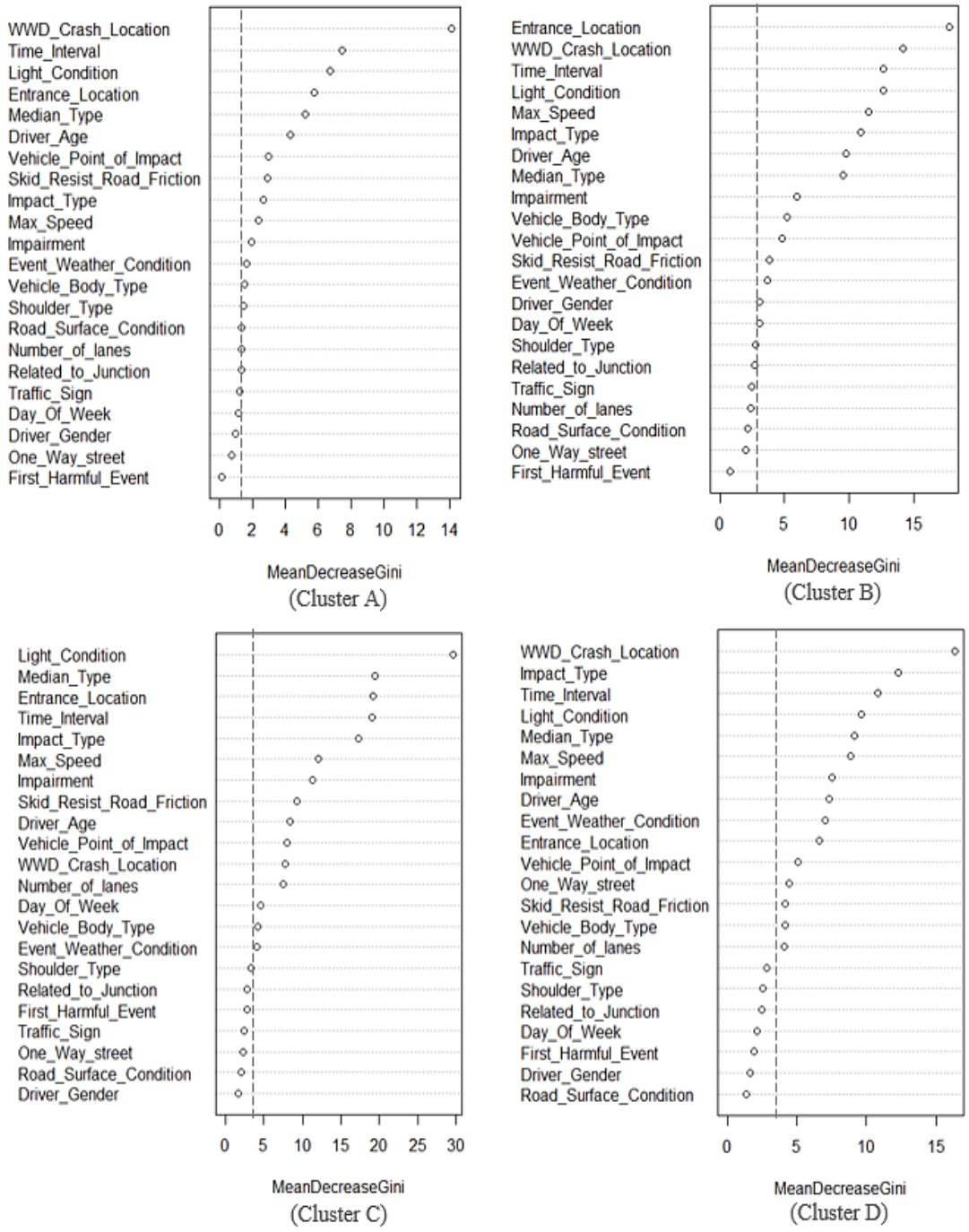
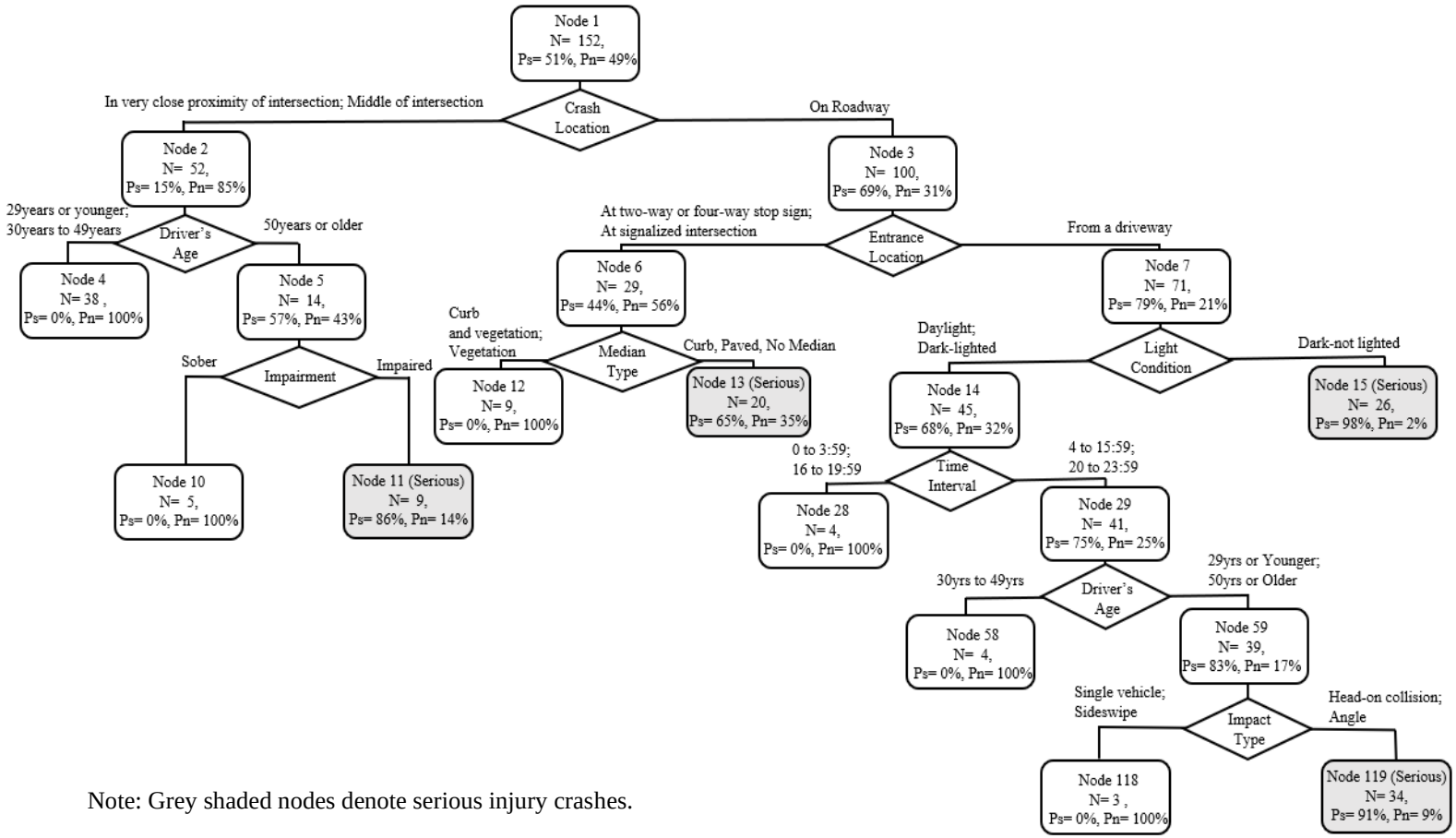


Figure 5-23: Variable Importance Ranking Using Random Forests Algorithm

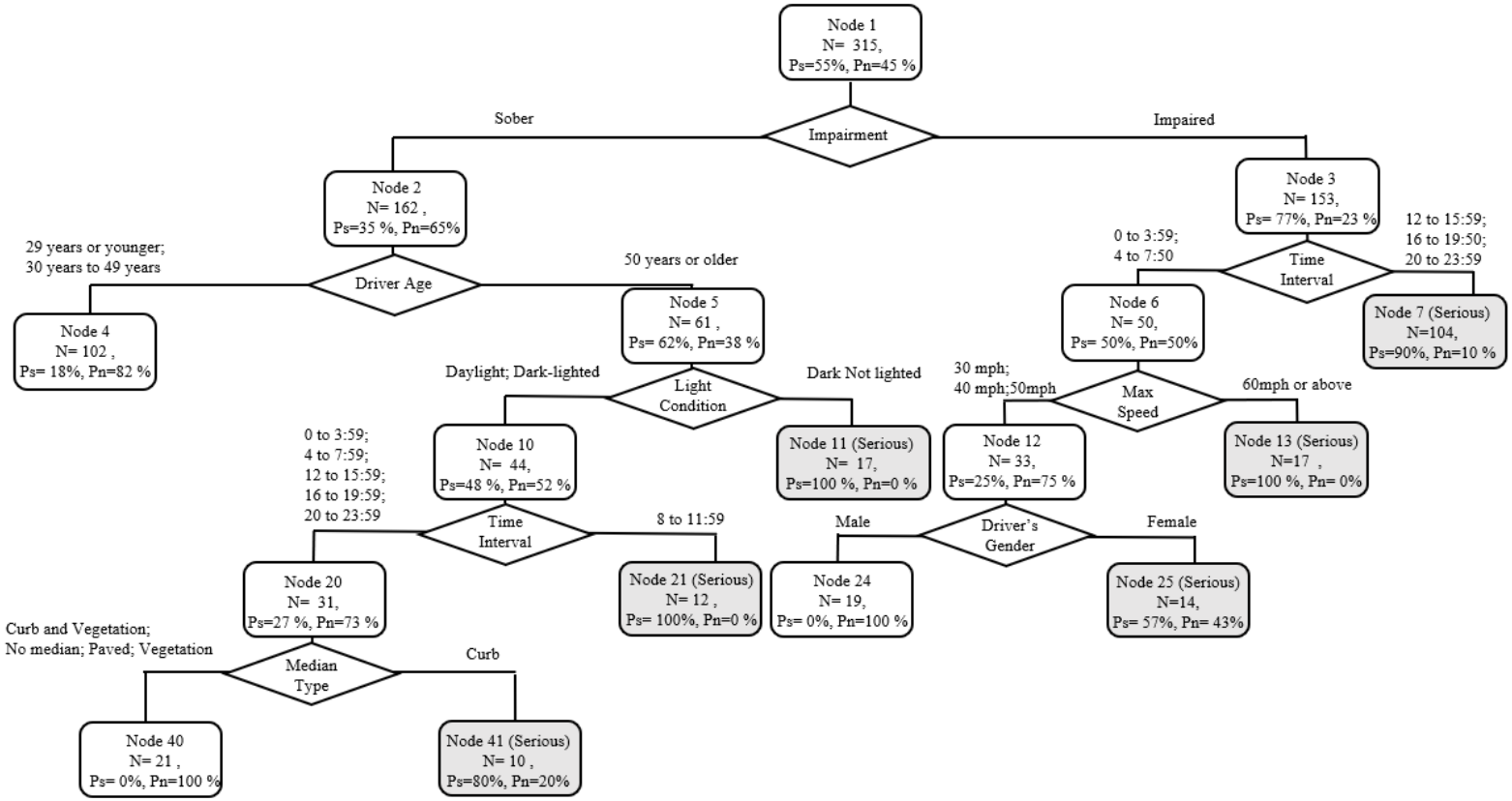
5.3.2 Tree Models

CART analysis was performed using the "*tree*" library in RStudio after obtaining the important variables from RF (Ripley, 2016). Here, a K = 5 fold cross-validation of data for each cluster was performed to remove potential model bias toward a particular training set. The output trees of each cluster were pruned, based on the minimum misclassification error rate, to obtain the best performing tree models free of overfitting the training data. Figures 5-24 through 5-27 show the output DTs from the CART analysis for A through D, respectively. In these figures, the node numbers are at the top of each node. If the parent node is '*n*', then the left and right child nodes are numbered as '*2n*' and '*2n+1*', respectively. '*Serious Injury*' crashes are in grey-shaded nodes. Each node in the DT diagram has its node number, the number of observations *N*, and the probabilities *(Ps = probability of serious injury; Pn = probability of not serious injury/ moderate to no injury)*.



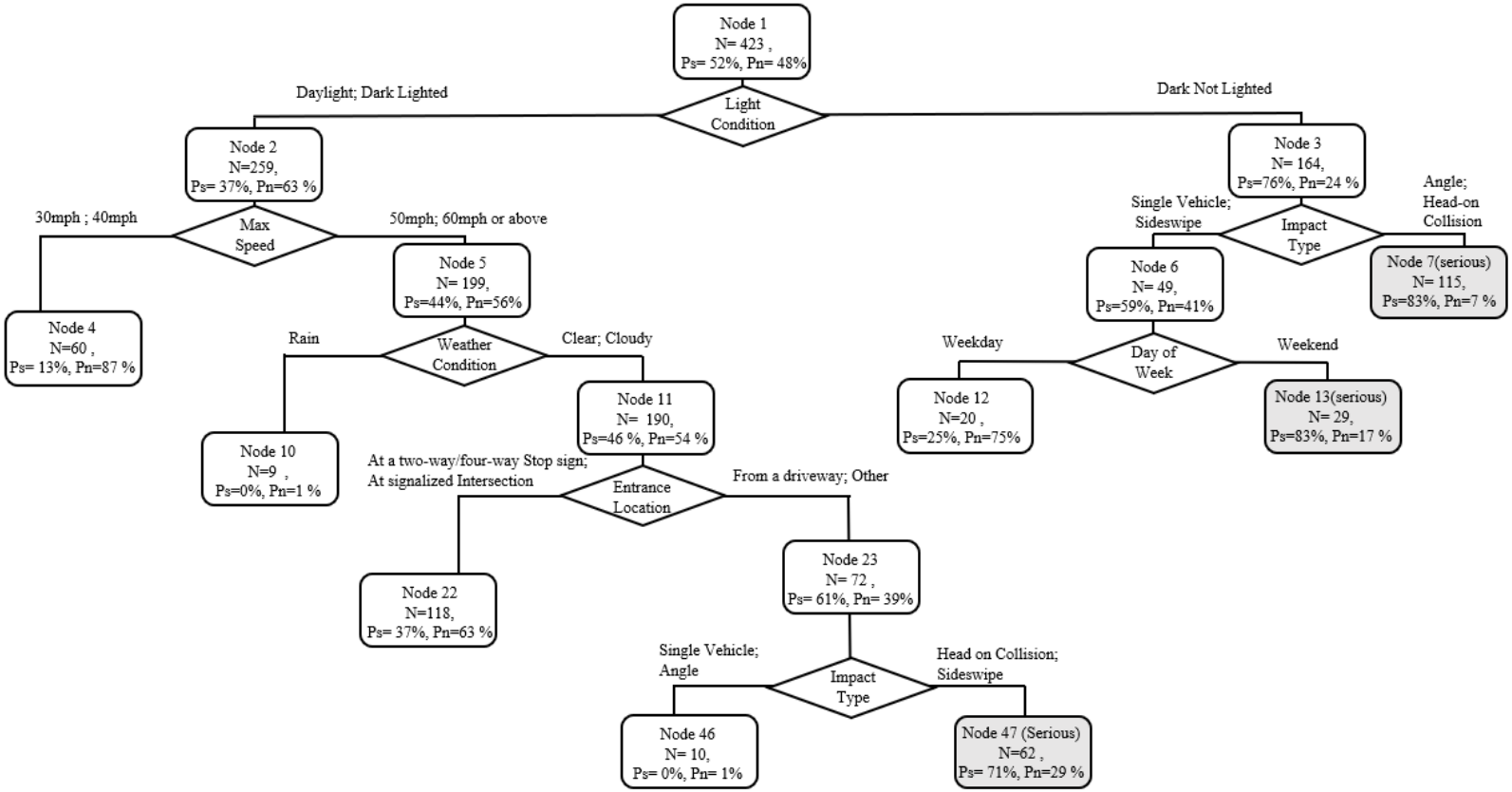
Note: Grey shaded nodes denote serious injury crashes.

Figure 5-24: Pruned DT from CART Model for Cluster A



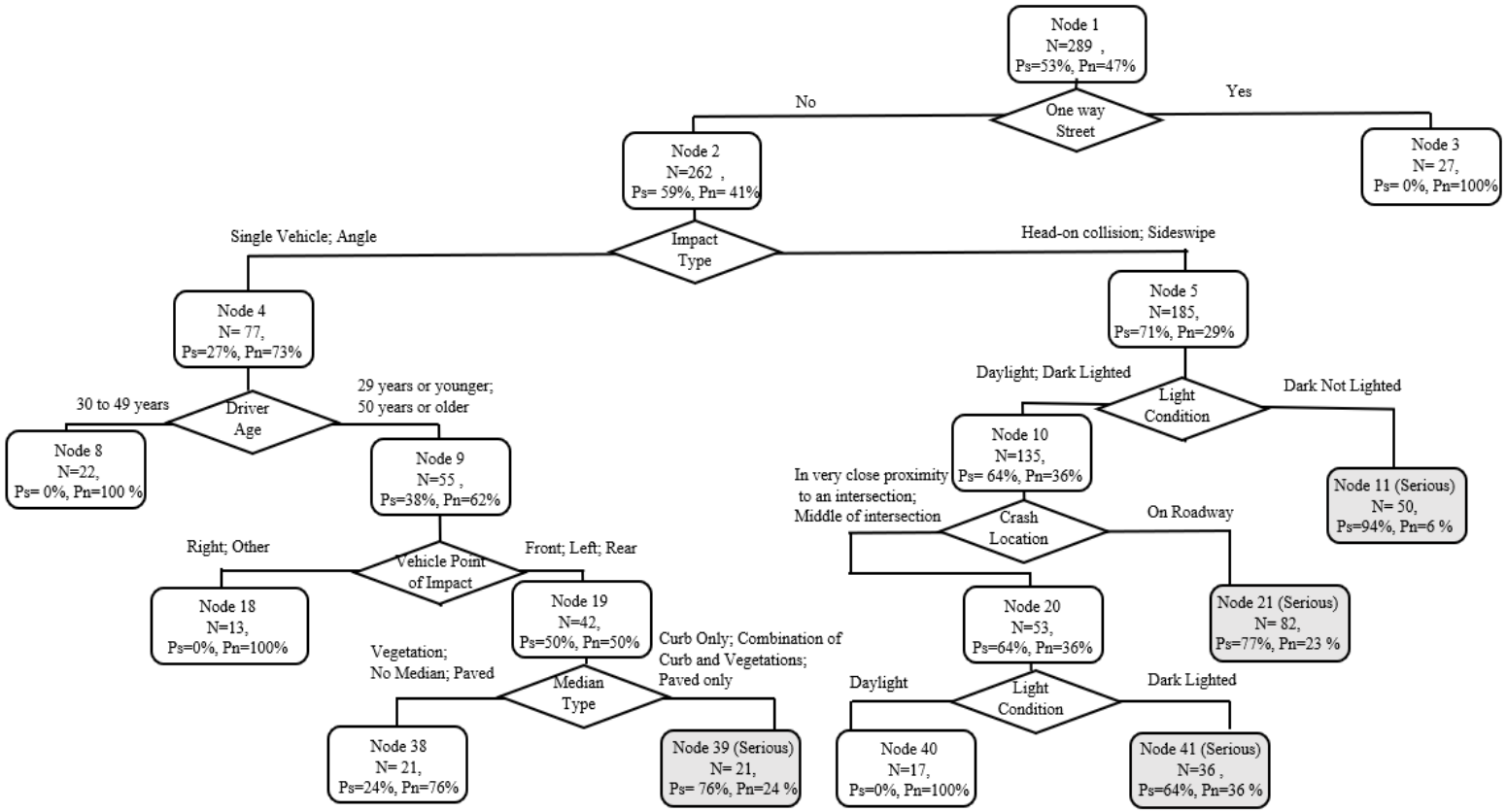
Note: Grey shaded nodes denote serious injury crashes.

Figure 5-25: Pruned DT from CART Model for Cluster B



Note: Grey shaded nodes denote serious injury crashes.

Figure 5-26: Pruned DT from CART Model for Cluster C



Note: Grey shaded nodes denote serious injury crashes.

Figure 5-27: Pruned DT from CART Model for Cluster D

5.3.2.1 Tree Model #1: Cluster A

Tree model #1 is based on Cluster A, which contains 190 crashes (Figure 5-24). Root Node 1 includes a total of 152 (N) observations. Within this cluster, 51% (Ps) of the crashes are '*Serious Injury (KA)',* and 49% (Pn) are '*Moderate to No Injury (BCO)*'. Below root node '*Crash Location'* is the first splitting variable for this tree. The names on top of Node 2 suggest that this node is the outcome of crash locations '*in close proximity of an intersection*' or '*Middle of Intersection'*. Node 11 has nine crashes, where 86% of the crashes are '*Serious Injury (KA)*'. As Node 11 has a higher probability of '*KA*' crashes, it is categorized as '*Serious*'. Node 11 shows that when the crash location is '*in close proximity of an intersection*' or '*middle of intersection'*, drivers age is 50 years and older, and impaired drivers have a *Serious Injury (KA)* rate of 86%. However, if the driver is sober (i.e., not impaired), the crash injury is not serious (Node 10). The tree also indicates that the probability of serious injury is high when a wrong-way vehicle enters from a driveway, and the light condition is dark-not lighted (Node 15). Node 119 shows that younger and older drivers are more prone to serious injury when the impact type is a head-on collision.

5.3.2.2 Tree Model #2: Cluster B

Tree model #2 is based on 394 crashes in Cluster B (Figure 5-25), where driver impairment emerges as the splitting predictor of the root node. Node 11 indicates drivers age 50 and older are at high risk (Ps= 100%) of serious injury when driving the wrong-way in dark conditions without any road lighting. Node 13 identifies that WWD crashes involving impaired drivers, driving between the hours of midnight to 8:00 AM, will result in serious injury when driving at higher speeds, 60 mph or higher. Similarly, Node 7

denotes that the probability of serious injury is high from 12:00 PM to midnight if the driver is impaired. This node shows a high number of observations (104), implying that a large number of serious WWD crash injuries are associated with impaired driving.

5.3.2.3 Tree Model #3: Cluster C

Tree model #3 is based on 529 crashes in Cluster C (Figure 5-26), where Node 47 shows that on high-speed roadways, head-on or sideswipe collisions will result in serious injury when a wrong-way vehicle enters the facility from a driveway. Similarly, Node 7 (115 observations) is associated with head-on crashes, which reveals serious injury will occur when the collision type is head-on or angle on a dark-not lighted roadway. This indicates that head-on collision is a principal cause of WWD crash injury. Unfortunately, the head-on collision is the most probable collision type for WWD incidents, as the WWD vehicles travel opposite to the legal flow direction. Node 13 indicates crashes occurring on dark (nighttime) roadways and weekends are more prone to serious injuries.

5.3.2.4 Tree Model #4: Cluster D

Figure 5-27 presents the DT (tree model #4) for Cluster D, based on 362 crashes. Here, the variable 'one-way street' has segregated the root node. Node 3 shows the outcome for 'not serious injury' and does not further divide. This implies that if the roadway is a one-way facility, WWD crashes will not result in serious injury for crashes in Cluster D. For this cluster, Node 11 and Node 41 indicate that WWD crashes on two-way roadways will result in serious injury when the collision type is head-on, or sideswipe and the lighting condition is nighttime, dark-lighted or dark-not lighted.

5.3.3 Decision Rules from CART

Table 5-24 lists the decision rules that result in serious injury identified from the DTs. The DR IDs are based on cluster and node numbers. For instance, ID A_11 refers to Node 11 of Cluster A. Higher values of support (S), population (Po), and probability (P) indicate stronger rules. Stronger rules with consequent '*KA'(serious injury)* need more attention when identifying and prioritizing countermeasures for implementation. The analysis generated 17 important decision rules that describe the connection between the factors that lead to serious (KA) injury crashes. The number of rules generated from each tree is as follows: Tree #1: four rules, Tree #2: six rules, Tree #3: three rules, and Tree #4: four rules. The following subsections provide a detailed discussion on the predominant factors.

Table 5-24: Decision Rules from Decision Trees

ID	Antecedent*	Consequent (Severity)	N	Po%	P%	S
A_11	Crash Location= CpI/ MoI and Driver's Age = 50≤ and Impairment = Imp	KA	9	6	0.9	5.1
A_13	Crash Location= OnRdwy and Entrance location = TwFwSS and Median Type = Cb/ Nm/ Pv	KA	20	13	0.7	8.5
A_15	Crash Location= OnRdwy and Entrance location = Dw and Light Condition = Dnl	KA	26	17	1	17.0
A_119	Crash Location= OnRdwy and Entrance location = Dw and Light Condition = Dy/Dl and Time interval = 4 to 15:59/ 20-23:59 and Driver's age = ≤29/ 50≥ and Impact Type = Ho, An, Ot	KA	34	22	0.9	20.1
B_7	Impairment = Imp and time interval 12 to 15.59/16 to 19.50/20 to 23.59	KA	104	33	0.9	29.6
B_11	Impairment = Sb and Driver Age = ≥50 and Light condition = Dnl	KA	17	5	1	5.3
B_13	Impairment = Imp and time interval 0 to 3.59 or 4 to 7.59 and max speed = ≥60	KA	17	5	1	5.3
B_21	Impairment = Sb and Driver Age = ≥ 50 and Light condition = Dy /Dl and time interval = 8 to 11.59	KA	12	4	1	3.9
B_25	Impairment = Imp and time interval 0 to 3.59 or 4 to 7.59 and max speed = ≥60 and Driver's gender = Fm	KA	14	4	0.6	2.5
B_41	Impairment = Sober and Driver Age = ≥ 50 and Light condition = Dy / DL and time interval = 0 to 3.59/ 4 to 7.59/ 12 to 15.59/16 to 19.59/ 20 to 23.59 and Median Type = Cb	KA	10	3	0.8	2.6
C_7	Light Condition = Dnl and Impact type= HO / An	KA	115	27	0.8	22.6
C_13	Light Condition=Dnl and Impact type= Siv/Ot/SS and day of week = wnd	KA	29	7	0.8	5.7
C_47	Light Condition = Dy / Dl and Max speed= ≥ 50-60 7 Weather Condition= clr/cldyand Entrance location = Dw and impact type=Ho/Ss	KA	62	15	0.7	10.4
D_11	One way street= no and impact type = Ho / Ss and light cond =Dnl	KA	50	17	0.9	16.3
D_21	One way street= no and impact type = Ho / Ss and light cond = Dy / Dl and crash location = OnRdwy and light cond=Dnl / Dl	KA	82	28	0.8	21.8
D_39	One way street= no and impact type = Siv / An / ot and driver age = 29≤ / ≥50 and vehicle point of impact = fr /lf/ rr and cb only /Vg and median type = Pv	KA	21	7	0.8	5.5
D_41	One way street= no and impact type = Ho / Ss and light condition = Dy / Dl and crash location = CpI /MoI andlight condition = Dnl/	KA	36	12	0.6	8.0

* / means OR; and means AND. Column Consequent: KA = Serious Injury (Fatal Injury + Incapacitating Injury); Column Antecedents: Median Type = (Pv **OR** Cb **OR** Ot) **AND** Collision Type = HO **AND** Driver's Age = 50+ years. An = Angle; Cb = Curb only; CbVg = Combination of Curb and Vegetation; Clr = clear; Cldy = clowdy; CpI = In very close proximity to an intersection; Dnl = dark-not lighted; Dl= dark-lighted; Dw = From driveway; Dy = daylight; Fm = Female; Fr = fron; Ho = Head-on collision; Imp = Impaired; Lf = left; MoI = Middle of Intersection; Nm = No median; OnRdwy = On Roadway; Ot = Other; Pv = Paved only; Rr = rear; Sb = sober; Siv = Single vehicle; Ss = Sideswipe; TwFwSS = two way or four way intersection; Vg = Vegetation only; Wdy = weekday; Wnd = weekend.

5.3.3.1 Collision Type

The vehicle collision type was found to play a vital role in arterial WWD crash injury. Among the 17 decision rules, seven have an association with collision type. Insights can be drawn from the following significant rules that are associated with collision types:

- Rules A_119 and C_7 (Figures 5-24 and 5-26) identify WWD crashes that involve head-on or angle collision/impact types. The probability of serious injury from head-on or angle crashes is considerably high, at 91% (Cluster A) and 83% (Cluster C), with the support and population very high as well. These results are consistent with earlier studies (Aty and Keller, 2005 and Jalayer et al., 2018b), and was expected since these types of crashes often result in fatalities.
- Rule C_13 (Figure 5-26) shows the relationship between collision/impact type and day of week. This rule identifies that on a weekend day, a single-vehicle crash or sideswipe will likely result in a serious injury (83%).
- Rule C_47 (Figure 5-26) identifies head-on or sideswipe crashes will most probably result in serious injury (71%) when a driver enters a high-speed highway from a driveway.

5.3.3.2 Speed Limit

The following significant rules show the association between arterial WWD crash severity and speed limit:

- Rule B_13 and Rule C_47 (Figures 5-25 and 5-26) identify that higher speed limits are associated with a high rate of serious injury. Both rules show a high probability of serious injury.

This result is consistent with previous studies, indicating that higher speed limits are associated with greater severity (Renski et al., 1999; Rifaat et al., 2015; Aty & Keller, 2005).

5.3.3.3 Median Type

The following significant rules show the association between arterial WWD crash severity and median type:

- Rule B_41 (Figures 5-25) identifies curbed medians were associated with serious WWD crash injury (80%).

This association is also consistent with previous studies. Das and Aty (2010) found that curb and paved or curb and lawn medians increase crash severity and explained that the presence of a curb makes it difficult for the driver to avoid crashes.

5.3.3.4 Temporal Factors and light condition

The following significant rules show the association between arterial WWD crash severity, temporal factors, and light condition:

- Rule A_15, C_7, Rule D_11, and Rule D_41 (Figures 5-24, 5-26, and 5-27) identify the interaction between collision type and light condition. Findings indicate that head-on, sideswipe, or angle crashes have a high probability of serious injury under dark-not-lighted conditions.

- Rule C_13 (Figure 5-26) shows crashes are prone to serious injuries on dark-not lighted roadways on weekends.

The identified crash injury time and day of the week are consistent with several previous WWD studies on both freeways and arterials (Copelan, 1989; Cooner et al., 2004b; Braam, 2006; Lathrop et al., 2010; Alluri et al., 2018a; Kitali et al., 2020).

5.3.3.5 Driver-related Factors

The following significant rules show the association between arterial WWD crash severity and driver-related factors:

- Rule A_11 (Figure 5-24) shows the relationship between two driver-related factors, driver's age and driver impairment (86%).
- Rule B_7 and Rule B_13 (Figure 5-25) indicate that serious injuries (KA) are associated with impairment, where Rule B_7 has the highest support value (29.6) among all the DRs. This implies that driver impairment causes a high probability of serious injuries.
- Rule B_25 identifies that female impaired drivers are prone to serious injury from 12 am to 8 am, even at lower speed limit roadways.
- Rule B_11 identifies that older drivers age 50 years and older will encounter serious injury compared to drivers younger than 50 years. Rule B_11, Rule B_21, and Rule B_41 are related to serious injuries associated with older drivers.

This is consistent with several previous studies which suggest impaired drivers are highly associated with WWD crashes (Hossain et al., 2021); female and older drivers are more prone to serious injuries. Previous studies explain that older drivers' frail physical

condition, as well as their increased perception reaction time, may explain the higher fatality risk (Meuser et al., 2009; Kim et al., 2013; Kadeha et al., 2020). As driver-related factors contribute to a large number of crashes, public awareness and educating the road users are equally important in mitigating WWD incidents (Alluri et al., 2018b; Nafis et al., 2018).

5.3.3.6 Crash Location and Entering Location

The WWD vehicle entering location and crash locations were found to affect the crash severity significantly. The following are the critical observations from the DRs relating to the vehicle entering location:

- Rule A_13 (Figure 5-24) identifies that serious injury (65%) will result when a WWD vehicle enters from a two-way or four-way stop sign when the median type is curb or paved, or there is no median.
- Rule A_119 (Figure 5-24) identifies younger (≤ 29 years) and older (age 50 years ≥) drivers are more susceptible to serious injury (91%) from a head-on collision when the WWD driver enters the wrong-way from a driveway. Similarly, Rule A_15 indicates that WWD crash will result in serious injury in dark not-lighted areas when a WWD vehicle enters from a driveway.

These scenarios may arise when a vehicle deviates from its legal path while turning and accidentally ends up going the wrong-way. The sudden appearance of a WWD vehicle does not give enough time for vehicles already on the road to react, resulting in injury crashes. Results from this research are consistent with crash severity analysis concepts and several previous studies. For instance, Eccles et al. (2017) found that the expected number

of target total crashes and target fatal and injury crashes are influenced by the available intersection sight distance.

The following significant rules show the association between arterial WWD crash severity and crash location:

- Rule A_11, A_13, A_15, and A_119 (Figure 5-24) are all associated with the crash location. Rule A_11 identifies older impaired drivers as prone to serious injury when a WWD crash occurs in very close proximity to an intersection or middle of an intersection. Rule A_13 identifies that a WWD vehicle entering at a signalized intersection or a two-way or four-way stop sign, will result in a serious crash when the median type is curb or paved, and the crash location is on a roadway segment.
- Rule D_21 (Figure 5-27) identifies that the probability of injury is high (77%) when a head-on crash occurs on a roadway segment (i.e., major approach, minor approach, or two-way left turn lane (TWLTL)).

The rules suggest that entering location of the wrong-way driver and WWD crash location have a substantial impact on crash severity. These locations have critical conflict points, affecting the severity of WWD crashes. For example, more angle and left-turn crashes occur at intersections, while undivided roadway segments, and those with TWLTLs, experience head-on crashes, which are all generally more severe (Aty & Keller, 2005; Wang & Aty, 2008; Haule et al., 2018).

5.3.4 Accuracy of the Tree Models

The tree models were further processed to calculate their accuracy. For each tree model, after five-fold cross-validation, the output tree's misclassification rate was

generated for the best tree size, which can be found in Table 5-25. Among the cluster tree models, Cluster C performed best and has the lowest misclassification rate of 15.79% (i.e., 84.21% accuracy). Note that the DT analysis on the crash data prior to clustering and use of RF shows a higher misclassification rate of 30.79%.

The modified output DT result from the combination of data mining techniques performed better in all clusters compared to the raw DT model. In addition to that, the cluster 3 DT model performed better than all the parametric and nonparametric models discussed in Table 5-19. In Table 5-19, the highest accuracy of 83.72% was from the SVM model, 83.72 %, and serious injury (specificity) varied from 29.16% to the highest 58.33%. In contrast to that, the updated DT model serious injury prediction has boosted to 85% for some clusters.

The results show that the application of clustering, resampling, and variable selection with RF prior to using DT increased the accuracy of the predicted trees. The models are sound and can predict the predictor-response relationship well. Therefore, the results demonstrate the applicability of nonparametric data mining techniques in crash severity analysis. Figure 5-28 illustrates the Receiver Operator Characteristic (ROC) curve generated for the tree models. The ROC curves show that all of the tree models from the clusters performed quite well and can correctly classify more than 50% of crash injuries with a far less false alarm rate.

Table 5-25: Decision Tree Model Performance

Evaluation Parameter	Tree Model				
	Without Cluster	#1 Cluster A	#2 Cluster B	#3 Cluster C	#4 Cluster D
DT Misclassification Rate (%)	30.79	28.13	21.52	15.79	24.17
DT Accuracy Rate (%)	69.21	71.87	78.48	84.21	75.83
DT Sensitivity (%)	82	80	80	84	78
DT Specificity (%)	46	30	69	85	68

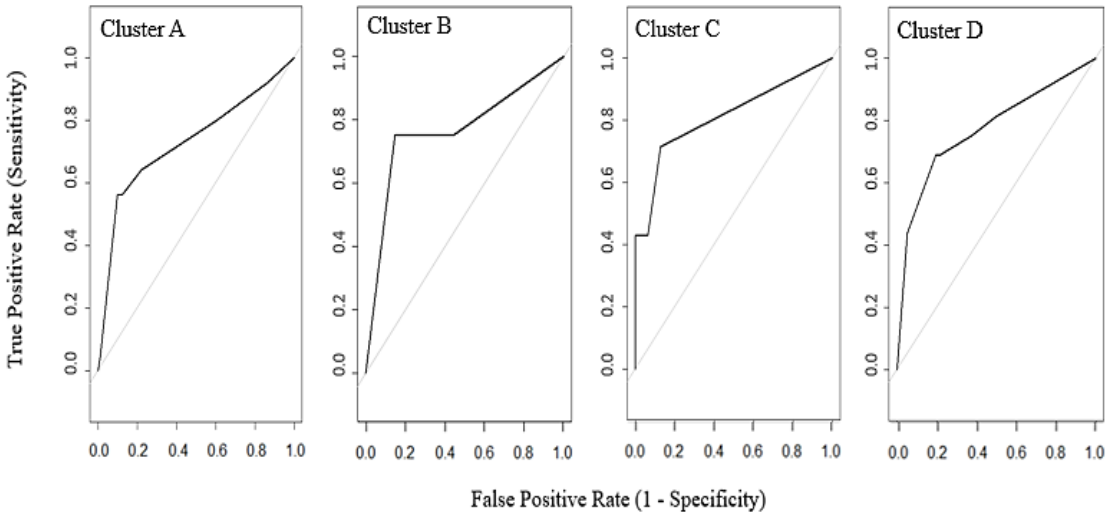


Figure 5-28: Receiver Operator Characteristic (ROC) Curve for Tree Models

5.4 Summary

Three nonparametric data mining models, DT, RF, SVM, were found to be reliable in predicting the WWD crash severity on arterials. These models' predictions were as good as the parametric models and sometimes performed better in predicting rare events, the serious injury crashes (i.e., better specificity). The RF model was selected to represent the nonparametric model for analyzing the correlation between predictor-response variables. Note that the RF models produced better overall accuracy and were also generating variable importance directly. By conducting marginal effect analysis, the outputs of the RF model

were more interpretable, not acting as a black-box. This marginal effect can be used with any of the nonparametric models used in this research and generate a response-predictor variable relationship. These steps demonstrated that the nonparametric models are robust for predicting crash severity classes.

Next, the combination of AHC, RF, and DT models was used to identify the influential contributing factors of the arterial WWD crash injury severity analysis. Based on WWD crash data collected on non-limited access facilities in Florida, the AHC model was developed to cluster the crash environment into four groups. The RF model was then developed for each cluster to select important variables that contribute to crash severity. The top important influential variables identified by the RF models were WWD crash location, WWD vehicle entrance location, light condition of the roadway, time of the crash, median type, speed limit, and collision type.

Finally, DTs were created using the important variables to investigate crash severity patterns. The tree pattern from the DTs demonstrated that the independent variables are not completely independent from each other. The DT results show that head-on collisions, weekend days, high-speed facilities, road surface friction, crashes involving vehicles entering from a driveway, dark-not lighted roadways, older drivers, and driver impairment are some of the significant factors that play a crucial role in resulting injuries on arterial facilities.

These results are consistent with the existing literature, reflecting the robustness of the study approach. In addition to identifying contributing factors, the DRs also recognize how the factors are connected and which combinations of factors influence the crash

severity of WWD crashes on arterial roadway networks. By conducting prediction accuracy comparison, contributor variables' marginal effect analysis, variable importance evaluation, and crash severity pattern recognition analysis, the nonparametric models have been demonstrated to be valid and proved to serve as an alternative tool in transportation safety studies.

CHAPTER 6
CONCLUSIONS

The goal of this research was to investigate the applicability of nonparametric data mining techniques to identify crash severity patterns of arterial wrong-way driving crashes. This goal was achieved using the following two components: (a) by comparing the prediction performance of parametric and nonparametric statistical models, and (b) identify factors that affect serious WWD crash injuries on arterial roadways using nonparametric data mining models. This chapter provides a summary of this effort, research contributions, and potential future research.

6.1 Summary and Conclusions

WWD crashes have been an area of concern for transportation agencies and the traveling public for many years and are still an ongoing phenomenon. In addition to this, WWD crashes on non-limited access facilities have not been studied as much as WWD crashes on limited-access facilities. Therefore, additional research, with new techniques, is needed to address this gap. The current rapid progression of computer technologies enhanced the availability of quality data, enabling a foundation for data-driven decision-making in transportation safety. Parametric models such as generalized linear models (GLMs) are most popular among transportation safety researchers and decision-makers as they produce easily interpretable functional forms by establishing a quantitative relationship between the response variable and the explanatory features. However, parametric techniques depend on predefined assumptions and require to follow a certain distribution for the dependent and independent variables, which may not always be correct,

especially when handling safety data with complex patterns and structure. On the contrary, nonparametric data mining techniques can overcome these issues. Data mining techniques do not use predetermined assumptions as GLMs do, require minimum data processing, define a nonlinear pattern in the data structure, address the correlation problem among explanatory variables, display results graphically, and simplify the potentially complex relationship between the variables. However, the use of nonparametric techniques is limited in transportation safety research and is criticized as a weak exploratory tool like a black-box model, even though they have higher prediction capability.

The main objective of this research was to demonstrate the applicability of data mining techniques and identify factors that affect severe WWD crash injuries on arterial roadways. First, the prediction accuracy of three nonparametric methods (DT, RF, SVM), and five parametric models (logit, Ridge, Lasso, LDA, GNB) was computed. All the models were then compared based on classification sensitivity, specificity, and overall prediction accuracy. The results showed that nonparametric models provided better predictive accuracy on predicting serious injury (i.e., specificity), compared to parametric models. By conducting prediction accuracy comparison, contributor variables' marginal effect analysis, variable importance evaluation, and crash severity pattern recognition analysis, the nonparametric models have been demonstrated to be valid and proved to serve as an alternative tool in the transportation safety analysis.

Once the enhanced prediction accuracy of nonparametric methods was established, the combination of the AHC, RF, and DT model was used to identify the factors influencing the severity of WWD crashes on arterials. Based on WWD crash data collected from non-

limited access facilities in Florida, the AHC model was developed to cluster the crash environment into four groups. The RF model was then developed for each cluster to select important variables that contribute to crash severity. Some of the influential variables identified by the RF models were WWD crash location, WWD vehicle entrance location, light condition of the roadway, time of the crash, median type, speed limit, and collision type. Finally, DTs were created using the important variables to investigate crash severity patterns. The tree pattern from the DTs demonstrated that the independent variables are not completely independent from each other. The DT results showed that head-on collisions, weekends, high-speed facilities, road surface friction, crashes involving vehicles entering from a driveway, dark-not lighted roadways, older drivers, and driver impairment are some of the significant factors that play a crucial role in resulting injuries on arterial facilities. These results are consistent with the existing literature, reflecting the robustness of the study approach. In addition to identifying contributing factors, the DRs also recognize how the factors are connected and which combinations of factors influence the crash severity of WWD crashes on the arterial roadway network. This research also highlighted the need for an in-depth study on the locations of WWD crashes on the arterial road network because the entering location and crash location have a substantial impact on crash severity. Information presented in this dissertation is a step toward the mitigation of WWD crashes on non-limited access facilities. The decision rules produced from this research can be useful for agencies, researchers, and road safety analysts in planning risk management and identifying and deploying targeted WWD crash mitigation strategies and countermeasures. This research also demonstrates that the nonparametric data mining models are robust tools

to predict and explain crash severity, indicating the applicability of data mining techniques in transportation safety analyses.

6.2 Research Contributions

The State Departments of Transportation and local transportation agencies have been investing a substantial amount of resources in developing strategies to mitigate WWD crashes. Agencies have been struggling since WWD crash mitigation on arterials is not a straightforward process. Mitigating WWD crashes is challenging, especially on the arterial network, because of their characteristics and crash data heterogeneity. Parametric models such as generalized linear models (GLMs) are the most popular models; however, GLMs are based on several assumptions, and the results could become questionable when these assumptions get violated. Unlike the regular regression models, data mining techniques have the ability to identify patterns associated with crash risk without needing assumptions about the dataset. However, in spite of being accurate prediction methods, data mining techniques are often overlooked by transportation researchers due to difficulty in interpreting their results.

While there are many existing methods in crash severity prediction, and much research has been conducted in this area, this dissertation offers new contributions by reviewing nonparametric models in the context of transportation safety in general and WWD crash severity in particular. This research discussed the shortcomings of the existing GLM methods used to predict crash severity and proposed nonparametric methods with better interpretation techniques.

For the first time, this research extended the previous efforts on WWD crash severity prediction on arterials by implementing nonparametric data-mining techniques. The research showed an accurate predictor-response relationship using nonparametric methods, even when the response variable categories were imbalanced, which can be improved further by resampling.

Furthermore, for the first time, this research used marginal effect analysis implemented with RF in transportation studies to better understand the complex dependence among relationships between explanatory variables and the response variable. The use of the partial dependence plot for the nonparametric method's marginal effect is also new in transportation safety.

The main contribution of this research is that it found the effectiveness in the application of data mining techniques such as tree-based methods like random forest and decision tree in accurately predicting crash severity and finding the relationship between response predictor variables like parametric models do, but without imposing any assumptions. This helps in the data analysis process as researchers do not have to abide by the assumptions. The results stated that the nonparametric models used in this research are as good as the parametric models used in this research. However, as the nonparametric models do not require any assumption, nonparametric methods may become more useful in developing crash prediction models.

The research demonstrated the full potential of nonparametric data mining in transportation crash severity prediction by ranking and selecting features, pattern recognition, generating rules, and showing the correlation between response-contributor

variables. All these findings reflect the robustness of the study approach, proving nonparametric can be used effectively in safety analyses.

6.3 Future Work

This research is not without limitations. The performance of the RF and DT models is highly dependent on the learning procedure, data cleaning, grouping of the parameter categories, number of observations in each class of outcome variable, and the total number of observations. The study model had binomial severity output due to the smaller sample size, previous study suggestions, and the importance of severe injury crashes. For future research, a larger dataset with five-level severity output can be explored with these models. While this research employed a combination of several data mining techniques, other data mining techniques to compare prediction accuracy were not considered. It would be interesting to see the application of different data mining techniques, such as neural networks and k-nearest neighbors, in future WWD research. In addition, other combinations of models, such as combining clustering analysis with k-nearest neighbors or similar methods, can be explored.

REFERENCES

Abdel-Aty, M., & Keller, J. (2005). Exploring the overall and specific crash severity levels at signalized intersections. *Accident Analysis & Prevention, 37(3),* 417-425. doi: 10.1016/j.aap.2004.11.002

Acuna, E., & Rodriguez, C. (2004). The treatment of missing values and its effect on classifier accuracy. *In Classification, Clustering, and Data Mining Applications,* 639-647. Springer, Berlin, Heidelberg.

Al-Ghamdi, A. S. (2002). Using logistic regression to estimate the influence of accident factors on accident severity. *Accident Analysis & Prevention, 34(6),* 729-741.

Alluri, P., Wu, W., Nafis, S., Kadeha, C., & Hagen, L. (2019). *Strategies to mitigate wrong-way driving incidents on arterials* (BDV29-977-50). Tallahassee, Florida: Florida Department of Transportation.

Alluri, P., Wu, W., Nafis, SR., & Hagen, L. (2018a). *A data-driven approach to implementing wrong-way driving countermeasures.* (BDV29-977-36). Tallahassee, Florida: Florida Department of Transportation.

Alluri, P., Nafis, S., Soto, F., Gonzalez, M., & Gan, A. (2018b). *Use of communication technologies to enhance public involvement in transportation projects* (BDV29-977-32). Tallahassee, Florida: Florida Department of Transportation.

Allwein, E. L., Schapire, R. E., & Singer, Y. (2000). Reducing multiclass to binary: A unifying approach for margin classifiers. *Journal of Machine Learning Research, 1,* 113-141.

Arafat, M., Nafis, S. R., Sadeghvaziri, E., & Tousif, F. (2020). A data-driven approach to calibrate microsimulation models based on the degree of saturation at signalized intersections. *Transportation Research Interdisciplinary Perspectives, 8,* 100231.

Arizona Department of Transportation. (2017). Interstate Wrong-way Detection System. Retrieved February 3, 2018, from https://www.azdot.gov/projects/central-district-%20projects/i-17-%20wrong-way-detection-system

Athey Creek Consultants. (2016). Countermeasures for wrong-way driving on freeways. Enterprise Transportation Pooled Fund Study TPF-5 (231), West Linn, Oregon.

Beshah, T., & Hill, S. (2010). Mining road traffic accident data to improve safety: role of road-related factors on accident severity in Ethiopia. In *2010 AAAI Spring Symposium Series*.

Berk, R. A., & Bleich, J. (2013). Statistical procedures for forecasting criminal behavior: A comparative assessment. *Criminology & Pub. Pol'y, 12,* 513.

Boot, W. R., Charness, N., Mitchum, A., Roque, N., Stothart, C., & Barajas, K. (2015). Driving Simulator Studies of the Effectiveness of Countermeasures to Prevent Wrong-way Crashes. Florida State University.

Braam, A. C. (2006). *Wrong-way crashes: statewide study of wrong-way crashes on freeways in north carolina.* North Carolina Department of Transportation, Division of Highways.

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

Breiman, L., Friedman, J., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. Chapman & Hall. New York.

Campbell, B. E., & Middlebrooks, P. B. (1988). *Wrong-way movements on partial cloverleaf ramps* (FHWA-GA-88-8203). Atlanta, Georgia: Georgia Department of Transportation.

Cao J, Panetta R, Yue S, Steyaert A, Young-Bellido M, et al. (2003). A naive bayes model to predict coupling between seven transmembrane domain receptors and g-proteins. *Bioinforma Oxf Engl, 19,* 234–240.

Chang, L., & Wang, H. (2006). Analysis of traffic injury severity: An application of non-parametric classification tree techniques. *Accident Analysis & Prevention, 38(5),* 1019-1027. doi: 10.1016/j.aap.2006.04.009

Cooner, S. A., Cothron, A. S., & Ranft, S. E. (2004a). *Countermeasures for wrong-way movement on freeways: guidelines and recommended practices* (FHWA/TX-04/4128-2). Texas Transportation Institute, Texas A & M University System.

Cooner, S. A., Cothron, A. S., & Ranft, S. E. (2004b). *Countermeasures for wrong-way movement on freeways: overview of project activities and findings.* (FHWA/TX-04/4128-1). Texas Transportation Institute, Texas A & M University System.

Copelan, J. E. (1989). *Prevention of wrong-way accidents on freeways* (FHWA/CA-TE-89-2). California Department of Transportation. Publication.

Cortez, P., & Embrechts, M. J. (2011). Opening black box data mining models using sensitivity analysis. In *2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM),* 341-348. IEEE.

Das, A., & Abdel-Aty, M. (2010). A genetic programming approach to explore the crash severity on multi-lane roads. *Accident Analysis & Prevention, 42(2),* 548-557.

Das, S., Dutta, A., Jalayer, M., Bibeka, A., & Wu, L. (2018). Factors influencing the patterns of wrong-way driving crashes on freeway exit ramps and median crossovers: Exploration using 'Eclat'association rules to promote safety. *International journal of transportation science and technology, 7(2),* 114-123.

De Oña, J., López, G., & Abellán, J. (2013). Extracting decision rules from police accident reports through decision trees. *Accident Analysis & Prevention, 50,* 1151-1160. doi: 10.1016/j.aap.2012.09.006

Depaire, B., Wets, G., & Vanhoof, K. (2008). Traffic accident segmentation by means of latent class clustering. *Accident Analysis & Prevention*, *40*(4), 1257-1266. doi: 10.1016/j.aap.2008.01.007

DHSMV (Florida Department of Highway Safety and Motor Vehicles). (2016). *Wrong-way driving awareness month: stay right at night; Campaign evaluation report.* Florida Department of Highway Safety and Motor Vehicles, Tallahassee, Florida.

Doty, R. N., & Ledbetter. C. R. (1965). Full Scale Dynamic Tests on One-Way Spike Barriers.

Eccles, K., Himes, S., Peach, K., Gross, F., Porter, R. J., Gates, T. J., & Monsere, C. M. (2018). *Guidance for evaluating the safety impacts of intersection sight distance*. NCHRP research report, (875), 1-44.

Elvik, R. (2013). A re-parameterisation of the power model of the relationship between the speed of traffic and the number of accidents and accident victims. *Accident Analysis & Prevention, 50,* 854-860.

Finley, M. D., Venglar, S. P., Iragavarapu, V., Miles, J. D., Park, E. S., Cooner, S. A., & Ranft, S. E. (2014). *Assessment of the effectiveness of Wrong-way driving countermeasures and mitigation methods* (FHWA/TX-15/0-6769-1). Texas A&M Transportation Institute.

Finley, M. D., Balke, K. N., Rajbhandari, R., Chrysler, S. T., Dobrovolny, C. S., Trout, N. D., Avery, P., Vickers, D., & Mott, C. (2016). *Conceptual Design of a Connected Vehicle Wrong-Way Driving Detection and Management System* (FHWA/TX-16/0-6867-1). Texas A&M Transportation Institute.

Finley, M. D., Miles, J. D., & Park, E. S. (2017). Closed-course study to examine the effect of alcohol-impairment on a driver's ability to identify and readsigns. *Transportation Research Record: Journal of the Transportation Research Board, 2660,* 86-93.

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.

Friebele, J. D., Messer, C. J., & Dudek, C. L. (1971). *State-of-the-art of wrong-way driving on freeways and expressways.* Texas Transportation Institute, Texas A & M University.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software, 33*(1), 1.

Goldstein, A., Kapelner, A., Bleich, J., & Pitkin, E. (2015). Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *Journal of Computational and Graphical Statistics, 24*(1), 44-65.

Gower, J. (1971). A general coefficient of similarity and some of its properties. *Biometrics, 27*(4), 857. doi: 10.2307/2528823

Harman, A. (2018). Ford Technology Intercepts Wrong-Way Driving (Web Page). Retrieved February 5, 2019, from https://www.wardsauto.com/technology/ford-technology-intercepts-wrong-way-driving.

Haule, H. J., Sando, T., Kitali, A. E., Angel, M. L., & Ozguven, E. E. (2018). Influence of intersection characteristics on elderly driver crash involvement (No. 18-06625).

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, *12*(1): 55-67.

Hossain, M. M., Rahman, M. A., Sun, X., & Mitran, E. (2021). Investigating Underage Alcohol Intoxicated Driver Crash Patterns (TRBAM-21-03375).

Jalayer, M., Pour-Rouholamin, M., & Zhou, H. (2018a). Wrong-way driving crashes: a multiple correspondence approach to identify contributing factors. *Traffic Injury Prevention*, *19*(1), 35-41.

Jalayer, M., Shabanpour, R., Pour-Rouholamin, M., Golshani, N., & Zhou, H. (2018b). Wrong-way driving crashes: A random-parameters ordered probit analysis of injury severity. *Accident Analysis & Prevention*, *117*, 128-135

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, *32*(3), 241-254.

Kadeha, C., Haule, H., Ali, M. S., Alluri, P., & Ponnaluri, R. (2020). Modeling wrong-way driving (WWD) crash severity on arterials in Florida. *Accident Analysis & Prevention*, *151*, 105963.

Karlaftis, M., & Tarko, A. (1998). Heterogeneity considerations in accident modeling. *Accident Analysis & Prevention*, *30*(4), 425-433. doi: 10.1016/s0001-4575(97)00122-x

Kashani, A., & Mohaymany, A. (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. doi: 10.1016/j.ssci.2011.04.019

Kashani, A., Mohaymany, A., & Ranjbari, A. (2011). A data mining approach to identify key factors of traffic injury severity. *PROMET – Traffic & Transportation*, *23*(1), 11-17. doi: 10.7307/ptt.v23i1.144

Kayes, M. I., Al-Deek, H., Sandt, A., Rogers Jr, J. H., & Carrick, G. (2018). Analysis of performance data collected from two wrong-way driving advanced technology countermeasures and results of countermeasures stakeholder surveys. *Transportation Research Record*.

Khan, M., Ahmed, F., & Kim, K. (2017). Weldability knowledge visualization of resistance spot welded assembly design. *Procedia Manufacturing*, *11*, 1609-1616. doi: 10.1016/j.promfg.2017.07.308

Kibria, B. M. G. (2003). Performance of some new ridge regression estimators. *Communications in Statistics-Simulation and Computation*, *32*, 419-435.

Kibria, B. M. G. & Lukman, A. F. (2020). A new ridge-type estimator for the linear regression model: Simulations and applications. *Scientifica*, Article ID 9758378, 1-16.

Kim, J., Ulfarsson, G., Kim, S., & Shankar, V. (2013). Driver-injury severity in single-vehicle crashes in California: A mixed logit analysis of heterogeneity due to age and gender. *Accident Analysis & Prevention*, 50, 1073-1081.

Kittelson & Associates. (2015). *Statewide wrong-way crash study*. Final Research Report, Florida Department of Transportation, Tallahassee, FL.

Kemel, E. (2015). Wrong-way driving crashes on French divided roads. *Accident Analysis & Prevention*, *75*, 69-76.

Kuhnert, P., Do, K., & McClure, R. (2000). Combining non-parametric models with logistic regression: an application to motor vehicle injury data. *Computational Statistics & Data Analysis*, 34(3), 371-386.

Lathrop, S., Dick, T., & Nolte, K. (2010). Fatal wrong-way collisions on New Mexico's interstate highways, 1990-2004. *Journal of Forensic Sciences*, *55*(2), 432-437.

Lawless, J.F. & Wang, P. (1976). A simulation study of ridge and other regression estimators. *Communication in Statistics-Simulation and Computation*, *A5*, 307-323.

Leduc, J. L. K. (2008). Wrong-Way Driving Countermeasures [Web page]. Connecticut General Assembly, Hartford, Connecticut. Retrieved February 5, 2019, from https://www.cga.ct.gov/2008/rpt/2008-r-0491.htm.

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.

Lingras, P., Cory Butz, C. (2007). Rough set based 1-v-1 and 1-v-r approaches to support vector machine multi-classification. *Information Sciences*, *177*(18), 3782-3798.

Lin, P. S., Ozkul, S., Boot, W. R., Alluri, P., Hagen, L. T., & Guo, R. (2017). Comparing countermeasures for mitigating wrong-way entries onto limited access facilities. University of South Florida. Center for Urban Transportation Research.

Lunardon, N., Menardi, G., & Torelli, N. (2014). ROSE: A package for binary imbalanced learning. *R journal*, *6*(1).

Menardi, G., & Torelli, N. (2014). Training and assessing classification rules with imbalanced data. *Data Mining and Knowledge Discovery*, *28*(1), 92-122.

Meuser, T., Carr, D., & Ulfarsson, G. (2009). Motor-vehicle crash history and licensing outcomes for older drivers reported as medically impaired in Missouri. *Accident Analysis & Prevention*, *41*(2), 246-252

Mining, W. I. D. (2006). Data mining: Concepts and techniques. Morgan Kaufinann.

Mitchell TM (1997) Machine Learning. 1st edition. New York: McGraw-Hill.

Moler, S. (2002). Stop. You're Going the Wrong-way! *Public Roads*, *66*(2), 24-29.

Montella, A., Aria, M., D'Ambrosio, A., & Mauriello, F. (2012). Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery. *Accident Analysis & Prevention, 49,* (58)-72.

Morena, D. A., & Leix, T. J. (2012). Where these drivers went wrong. *Public Roads, 75*(6).

Morshed, S. A., Khan, S. S., Tanvir, R. B., & Nur, S. (2021). Impact of The Covid-19 pandemic on ride-hailing services based on large-scale twitter data analysis. *Journal of Urban Management.*

Müllner, D. (2018). The fastcluster package: User's manual.

Murakami Y, Mizuguchi K (2010) Applying the naïve bayes classifier with kernel density estimation to the prediction of protein-protein interaction sites. *Bioinforma Oxf Engl 26,* 1841–1848

Murtagh, F., & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion?. *Journal of classification, 31*(3), 274-295.

Mussone, L., Ferrari, A., & Oneta, M. (1999). An analysis of urban collisions using an artificial intelligence model. *Accident Analysis & Prevention,* 31(6), 705-718.

Nafis, S., Alluri, P., Wu, W., & Kibria, B. M. G. (2021). Wrong-way driving crash injury analysis on arterial road networks using non-parametric data mining techniques. *Journal of Transportation Safety and Security.*

Nafis, S., & Wasiuddin, N. M. (2021) Field performance analysis of open graded friction course: A case study in Shreveport, Louisiana. *Airfield and highway pavements. American Society of Civil Engineers.*

Nafis, S., Alluri, P., Jung, R., Ennemoser, R., and Gan, A. (2018). Use of communication technologies to enhance public involvement in transportation projects, Proceedings of the *97th Annual Meeting of the Transportation Research Board*, Washington, DC.

National Transportation Safety Board (NSTB). (2012). *Highway special investigation report: wrong-way driving* (NTSB/SIR-12/01PB2012-917003).Washington, DC.

Ozkul, S., & Lin, P. (2017). Evaluation of red RRFB implementation at freeway off-ramps and its effectiveness on alleviating wrong-way driving. *Transportation Research Procedia, 22,* 570–579.

Pakgohar, A., Tabrizi, R., Khalili, M., & Esmaeili, A. (2011). The role of human factor in incidence and severity of road crashes based on the CART and LR regression: a data mining approach. *Procedia Computer Science, 3,* 764-769.

Pande, A., & Abdel-Aty, M. (2009). Market basket analysis of crash data from large jurisdictions and its potential as a decision support tool. *Safety Science, 47*(1), 145-154. doi: 10.1016/j.ssci.2007.12.001

PennState Eberly College of Science. (2017). Introduction to Generalized Linear Models. Retrieved February 5, 2019, from https://online.stat.psu.edu/stat504/lesson/6/6.1

Ponnaluri, R. (2018). Modeling wrong-way crashes and fatalities on arterials and freeways. *IATSS Research*, *42*(1), 8-17.

Ponnaluri, R. & Heery, F., 2016. Wrong-way driving mitigation: a holistic approach in Florida, USA. *Institute of Transportation Engineers, ITE Journal*. 86(5): 43.

Ponnaluri, R. V. (2016a). Addressing wrong-way driving as a matter of policy: The Florida Experience. *Transport Policy*, *46*, 92-100.

Ponnaluri, R. V. (2016b). The odds of wrong-way crashes and resulting fatalities: a comprehensive analysis. *Accident Analysis & Prevention*, *88*, (105)-116.

Pour-Rouholamin, M., & Zhou, H. (2016). Investigating the risk factors associated with pedestrian injury severity in Illinois. *Journal of Safety Research*, *57*, 9-17.

Pour-Rouholamin, M., Zhou, H., & Shaw, J. (2014). Overview of safety countermeasures for wrong-way driving crashes. Institute of Transportation Engineers. *ITE Journal*, *84*(12), 31.

Pour-Rouholamin, M., Zhou, H., Zhang, B., & Turochy, R. E. (2016). Comprehensive analysis of wrong-way driving crashes on alabama interstates. *Transportation Research Record: Journal of the Transportation Research Board (2601)*, 50-58.

Rahman, M. S., Abdel-Aty, M., Hasan, S., & Cai, Q. (2019). Applying machine learning approaches to analyze the vulnerable road-users' crashes at statewide traffic analysis zones. *Journal of Safety Research*. *70*, 275-288.

Rifaat, S. M., Pasha, M., Shovon, M. H., Nafis, S. R., & Limon, M. K. H. (2015). Comparative speed study: A way to improve road safety condition. *IUT Journal of Engineering and Technology (JET)*, 12, 41-50.

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D., & Ripley, M. B. (2013). Package 'mass'. Cran R, 538, 113-120. Retrieved 1 January 2019, from https://mran.microsoft.com/snapshot/2017-12-11/web/packages/tree/tree.pdf

Ripley, B. (2016). Package 'tree'. Classification and Regression Trees. Version, 1-0. Retrieved 1 January 2019, from https://cran.r-project.org/web/packages/MASS/MASS.pdf

Renski, H., Khattak, A., & Council, F. (1999). Effect of speed limit increases on crash injury severity: analysis of single-vehicle crashes on North Carolina interstate highways. *Transportation Research Record: Journal of the Transportation Research Board*, 1665(1), 100-108. doi: 10.3141/1665-14

Rhode Island Department of Transportation (RIDOT). (2015). Wrong-way crash avoidance (Web page). Retrieved February 5, 2019, from http://www.dot.ri.gov/community/safety/wrong_way.php.

Rinde, E. (1978). *Off-Ramp surveillance: Wrong-way driving.* (FHWA-CA- TE-78-1). Sacramento, California.

Roadway Safety Institute. (2018). Directional Rumble Strips for Reducing Wrong-Way Driving Freeway Entries [Web page]. Retrieved February 5, 2019, from http://www.roadwaysafety.umn.edu/publications/researchreports/reportdetail.html?id=2654.

Rogers Jr, J. H., Al-Deek, H., Alomari, A. H., Gordin, E., & Carrick, G. (2016). Modeling the risk of wrong-way driving on freeways and toll roads. *Transportation Research Record*, *2554*(1), 166-176.

Saha, D., Alluri, P., & Gan, A. (2015). Prioritizing Highway Safety Manual's crash prediction variables using boosted regression trees. *Accident Analysis & Prevention*, 79, 133-144. doi: 10.1016/j.aap.2015.03.011

Saha, R., Tariq, M. T., Hadi, M., & Xiao, Y. (2019). Pattern recognition using clustering analysis to support transportation system management, operations, and modeling. *Journal of Advanced Transportation.*

Sandt, A., Al-Deek, H., & Rogers Jr, J. H. (2017). *Identifying wrong-way driving hotspots by modeling crash risk and assessing duration of wrong-way driving events. Transportation Research Record: Journal of the Transportation Research Board, 2616*, 58-68.

Scaramuzza, G., & Cavegn, M. (2007). Wrong-way drivers: extent-interventions. In the *European Transport Conference*, Leeuwenhorst Conference Centre, The Netherlands.

Shepard, F. D. (1976). *Evaluation of raised pavement markers for reducing incidences of wrong-way driving. Transportation Research Record: Journal of the Transportation Research Board, 597*, 41.

Shmueli, G. (2010). To Explain or to Predict? *Statistical Science, 25*(3), 289-310.

Simpson, S. A. (2013). *Wrong-way vehicle detection: proof of concept* (FHWA-AZ- 13-697). Phoenix, Arizona.

Simpson, S., & Bruggeman, D. (2015). *Detection and Warning Systems for Wrong-way Driving* (FHWA-AZ- 15-741). Phoenix, Arizona.

Sohn, S. Y., & Shin, H. (2001). Pattern recognition for road traffic accident severity in Korea. *Ergonomics*, *44*(1), 107-117.

SWOV. (2009). Fact sheet: wrong-way driving. Leidschendam, The Netherlands. Retrieved November 23, 2019 from https://www.swov.nl/en/facts-figures/factsheet/wrong-way-driving

Szczesny, J. (2013). Daimler debuts alert system for wrong-way drivers (Web page). The Detroit Bureau. Detroit, MI. Retrieved February 5, 2019, from

http://www.thedetroitbureau.com/2013/02/daimler-debuts-alert-system-for-wrong-way-drivers/.

Tamburri, T. (1965). *Report on wrong-way automatic sign, light and horn device*. Traffic Department, Division of Highways, Department of Public Works, State of California.

Tanvir, R. B., Aqila, T., Maharjan, M., Mamun, A. A., & Mondal, A. M. (2019). Graph theoretic and pearson correlation-based discovery of network biomarkers for cancer. *Data, 4*(2), 81.

Trimble, G. (2018). FDOT to add technology that could detect wrong-way drivers on Howard Frankland. Retrieved February 5, 2019, from

WTSP-TV:10NEWS.https://www.wtsp.com/article/news/local/fdot-to-add-technology-that-could-detect-wrong-way-drivers-on-howard-frankland/67-610495127.

U.S. News. (2019). Thermal camera system helps stop wrong-way driver in Phoenix. U.S. News & World Report L.P. Retrieved February 5, 2019, from

https://www.usnews.com/news/best-states/arizona/articles/2019-01-08/thermal-camera-system-helps-stop-wrong-way-driver-in-phoenix.

Vaswani, N. K. (1973). *Measures for preventing wrong-way entries on highways* (VHRC 72-R41). Virginia Transportation Research Council.

Venables, W. N., & Ripley, B. D. (2013). Modern applied statistics with S-PLUS. Springer Science & Business Media.

Venglar, S. P., & Fariello, B. G. (2014). Efforts to reduce wrong-way driving: a case study in San Antonio, Texas. Presented at *93rd Annual Meeting of the Transportation Research Board*, Washington, D.C.

Wang, J., Gluck, J., Ginder, A., & Ward, N. (2018). A safe system approach to reduce wrong-way driving crashes on divided highways by applying access management and traffic safety culture. (No. 18-01009)

Wang, J., Zhou, H., & Zhang, Y. (2017). Improve sight distance at signalized ramp terminals of partial-cloverleaf interchanges to deter wrong-way entries. *Journal of Transportation Engineering, Part A: Systems, 143*(6), 04017017

Wang, X., & Abdel-Aty, M. (2008). Analysis of left-turn crash injury severity by conflicting pattern using partial proportional odds models. *Accident Analysis & Prevention, 40*(5), 1674-1682. doi: 10.1016/j.aap.2008.06.001

Webb, G. (2007). Discovering significant patterns. *Machine Learning, 68*(1), 1-33. doi: 10.1007/s10994-007-5006-x

Worth, A. P., & Cronin, M. T. (2003). The use of discriminant analysis, logistic regression and classification tree analysis in the development of classification models for human health effects. *Journal of Molecular Structure: THEOCHEM, 622*(1-2), 97-111.

Xing, J. (2014). Characteristics of wrong-way driving on motorways in Japan. *IET intelligent transport systems, 9(1),* 3-11.

Zhang, B., Pour-Rouholamin, M., & Zhou, H. (2017). *Investigation of confounding factors contributing to wrong-way driving crashes on partially and uncontrolled-access divided highways.* (TRID Report Number: 17-05146).

Zhang, Y., & Haghani, A. (2015). A gradient boosting method to improve travel time prediction. *Transportation Research Part C: Emerging Technologies, 58,* 308-324.

Zheng, Z. (2018). Application of Data Mining Techniques in Transportation Safety Study. [Doctoral dissertation, North Dakota State University]

Zhou, H., Zhao, J., Fries, R., Gahrooei, M. R., Wang, L., Vaughn, B., Bahaaldin, K., & Ayyalasomayajula, B. (2012). *Investigation of contributing factors regarding wrong-way driving on freeways* (FHWA-ICT-12-010).

Zhou, H., Zhao, J., Pour-Rouholamin, M., & Tobias, P. A. (2015). Statistical characteristics of wrong-way driving crashes on Illinois freeways. *Traffic injury prevention, 16(8),* 760-767.

Zhou, H., Zhao, J., Reisi Gahrooei, M., & Tobias, P. A. (2016). Identification of contributing factors for wrong-way crashes on freeways in Illinois. *Journal of Transportation Safety & Security, 8(2),* 97-112.

Zhou, H., Xue, C., Yang, L., & Luo, A. (2018). *Directional rumble strips for reducing wrong-way-driving freeway entries*(CTS 18-04). Illinois Center for Transportation, Rantoul, Illinois.

VITA

SAJIDUR RAHMAN NAFIS

EDUCATION

2009 – 2013 B.S., Civil Engineering
Islamic University of Technology, Dhaka, Bangladesh

2016 – 2021 Graduate Research/Teaching Assistant
Department of Civil and Environmental Engineering
Florida International University, Miami, Florida

2019 – 2021 Doctoral Candidate
Department of Civil and Environmental Engineering
Florida International University, Miami, Florida

PUBLICATIONS & PRESENTATIONS

1. Nafis**,** S., Alluri, P., Wu, W., & Kibria, B. M. G. (2021). Wrong-Way Driving Crash Injury Analysis on Arterial Road Networks Using Non-Parametric Data Mining Techniques. *Journal of Transportation Safety and Security.*

2. Nafis**,** S., Wasiuddin, N. M. (2021) Field Performance Analysis of Open Graded Friction Course: A Case Study in Shreveport, Louisiana. *Airfield and highway pavements*. American Society of Civil Engineers.

3. Alluri, P., Wu, W., Nafis**,** S., Kadeha, C., and Hagen, L. (2019). *Strategies to Mitigate Wrong-way Driving Incidents on Arterials*, Final Research Report, Florida Department of Transportation, Tallahassee, FL.

4. Alluri, P., Wu, W., Nafis**,** S**.**, and Hagen, L. (2018). A Data-Driven Approach to Implementing Wrong-way Driving Countermeasures, Final Research Report, Florida Department of Transportation, Tallahassee, FL.

5. Nafis, S., Alluri, P., Jung, R., Ennemoser, R., and Gan, A. (2018). Use of Communication Technologies to Enhance Public Involvement in Transportation Projects, Proceedings of the 97th Annual Meeting of the Transportation Research Board, Washington, DC.

6. Alluri, P., Nafis, S., Soto, F., Gonzalez, M., and Gan, A. (2018). Use of Communication Technologies to Enhance Public Involvement in Transportation Projects, Final Research Report, Florida Department of Transportation, Tallahassee, FL.

7. Arafat, M., Nafis, S., Sadeghvaziri, E., & Tousif, F. (2020). A Data-Driven Approach to Calibrate Microsimulation Models Based on the Degree of Saturation at Signalized Intersections. *Transportation Research Interdisciplinary Perspectives*, *8*, 100231.

8. Nafis, S., Raihan, M. A., and Alluri, P. (2018). 'Factors Affecting Run-off-road Crashes on Rural Two-lane Undivided Roads'. Florida International University (FIU) Civil and Environmental Engineering (CEE) GSAW, Miami, FL.

9. Alluri, P., Raihan, M. A., Saha, D., Wu, W., Huq, A., Nafis, S., and Gan, A. (2017). Statewide Analysis of Bicycle Crashes, Final Research Report, FDOT, Tallahassee, FL.

10. Wasiuddin, N. M., Nafis, S., Flurry, L. (2016). HfL Demonstration Project: LA-511, Louisiana Transportation Research Council.

11. Rifaat, S. M., Mosabbir Pasha, M. H. S., Nafis, S., & Limon, M. K. H. (2015). Comparative Speed Study: A Way to Improve Road Safety Condition. IUT Journal of Engineering and Technology (JET), 12, 41-50.