MOUNTAIN-PLAINS CONSORTIUM

MPC 19-381 | P. Lu, B. Lantz, D. Tolliver, and Z. Zheng

Analysis of the Relationship of Roadside Inspections on Large Truck Crashes

MOUNTAIN-PLAINS CONSORTIUM REGION 8

A University Transportation Center sponsored by the U.S. Department of Transportation serving the Mountain-Plains Region. Consortium members:

Colorado State University
North Dakota State University
South Dakota State University
University of Colorado Denver
University of Denver
University of Utah
Utah State University
University of Wyoming

Analysis of the Relationship of Roadside Inspections on Large Truck Crashes

Dr. Pan Lu
Associate Professor/ Associate Research Fellow
Pan.lu@ndsu.edu
Tel: 701-212-3795

Dr. Brenda Lantz
Associate Director
Brenda.Lantz@ndsu.edu
Tel: 303-871-7773

Dr. Denver Tolliver
Director
Denver.Tolliver@ndsu.edu
Tel: 701-231-7190

Zijian Zheng
Graduate Assistant
Zijian.Zheng@ndsu.edu

Department of Transportation, Logistics, and Finance/College of Business
Mountain-Plains Consortium/Upper Great Plains Transportation Institute
North Dakota State University

April 2019

Acknowledgements

The present study was funded by the U.S. Department of Transportation (USDOT) through the Mountain-Plains Consortium (MPC), University Transportation Center (UTC), the Research and Innovative Technology Center, the North Dakota State University. The authors deeply appreciate their support. The authors also would like to thank the three peer reviewers, Tom Jirik, and Michael Stein for their professional technical review and editorial review of the final report.

ABSTRACT

About 3,341 large truck related fatal crashes happened in United States in 2011, and more than 3,700 people died, which accounted for more than 11% of all motor vehicle fatal crashes and fatalities, even though large trucks only accounted for 4% of registered vehicles and 9% of vehicle-miles traveled (VMT) (USDOT, 2013). Large truck VMT decreased 6.7% from 2010 to 2011; however, large truck related fatal crashes still increased 2% from 2010 to 2011 (USDOT, 2013).

To reduce serious accidents involving these vehicles, the Federal Motor Carrier Safety Administration (FMCSA [formerly the Federal Highway Administration Office of Motor Carriers]) provides support for states to perform roadside inspections of commercial vehicles (large trucks, commercial buses, and hazardous materials vehicles) and drivers, compliance reviews, and other safety programs (GAO, 1997).

On-site reviews of motor carriers' compliance with federal safety regulations are known as compliance reviews, which can be used to determine a safety fitness rating. The safety rating is used to determine how well each carrier is fit to operate safely on the nation's highway.

Roadside inspections occur on a particular driver or vehicle, most often when the drivers/vehicles are en route to their destination. Violations found during the inspection can be divided into two groups: 1) minor and 2) out-of-service (OOS). Minor means those violations that do not pose any immediate danger, and the driver/vehicle can return to the road even before the violations are addressed. Out-of-service violations are those that require the vehicle/driver violations to be fixed immediately before the vehicle can return to service. The purpose of the OOS is to ensure that a vehicle and/or its driver are not allowed to proceed back on the road until the conditions are safe. In return, it can decrease accident rates caused by mechanical defects. (Randhawa, Miller, Bell, and Montagne, 1998). Even though some researchers found there are some problems, such as an officer being unable to remain at the site to make sure the violation is corrected (Patten, 1989), great numbers of researchers still found that the roadside inspection is a useful tool to remove some potential unsafe vehicles from the highway and reduce the commercial vehicle accident rate. (Patten, 1989; GAO, 1997; Mitchell, Friswell, and Mooren, 2012; Schoor, Niekerk, and Grobbelaar, 2001; Randhawa, Miller, Bell and Montagne, 1998; Hall, and Intihar, 1997).

Previous researchers have shed light toward better understanding the relationships among commercial vehicle safety performance, roadside inspection data, and motor carrier's position (Lantz, B.M. 1993; Britto, R.A., Corsi, T.M., and Grimm, C.M., 2010); however, much of this research is outdated. Moreover, very few previous researchers focused on small motor carriers. There is need to revisit investigating the relationships among a motor carrier's position (such as size and financial status), safety performance, and roadside inspection data. This research will provide additional evidence on such relationships, especially for various sized motor carriers. In addition, the trend in the relationships will be provided by comparing the most up-to-date analysis with previous ones.

This research seeks to investigate crash severity predicting models and contributing factor explorations through the application of data mining models. There are 21 variables found to be associated with commercial truck injury severities. The importance analysis indicates the variable relative important levels for contribution. The top 11 variables account for more than 80% of injury forecasting. For property damage only, the most important variable is “Carrier State,” which indicates that the variable of Carrier State makes the most contributions, as compared with the other variables in explaining property damage only crashes. Variables contribute differently when explaining different crash severities. A variable showing significant importance for a certain severity level may be less crucial for another. For instance, “Cargo Body Type” is the second most important factor for predicting fatality crashes, but is much less important for predicting property damage only crashes (severity=0). However, it is clear that Carrier State is the most influential factor for all severity levels. Marginal effects of important variables are conducted and summarized in the research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1 Background

In 2011 there were about 3,341 fatal crashes involving large trucks in the United States. Of those crashes, more than 3,700 people died, which accounted for more than 11% of all motor vehicle fatal crashes and fatalities, even though large trucks accounted for only 4% of registered vehicles and 9% of vehicle-miles traveled (VMT) (USDOT, 2013). Large truck VMT decreased 6.7% from 2010 to 2011; however, fatal crashes involving large trucks still increased 2% from 2010 to 2011 (USDOT, 2013).

To reduce serious accidents involving these vehicles, the Federal Motor Carrier Safety Administration (FMCSA [formerly the Federal Highway Administration Office of Motor Carriers]) provides support for states to perform roadside inspections of commercial vehicles (large trucks, commercial buses, and hazardous materials vehicles) and drivers, compliance reviews, and other safety programs (GAO, 1997).

On-site reviews of motor carriers' compliance with federal safety regulations are known as compliance reviews, which can be used to determine a safety fitness rating. The safety rating is used to determine how well each carrier is fit to operate safely on the nation's highways.

Roadside inspections occur on a particular driver or vehicle, most often when the drivers/vehicles are en route to their destinations. Violations found during the inspection can be divided into two groups: 1) minor and 2) out-of-service (OOS). Minor violations do not pose any immediate danger, and the driver/vehicle can return to the road even before the violations are fixed. OOS violations require the vehicle/driver violations to be fixed immediately before returning to service. The OOS is intended to ensure that a vehicle and/or its driver are not allowed to return to the road until conditions are safe. In return, it can decrease accident rates caused by mechanical defects. (Randhawa, Miller, Bell, and Montagne, 1998). Even though some researchers found problems, such as an officer cannot always remain at the site to ensure the violation is corrected (Patten, 1989), several researchers still found that the roadside inspection is a useful tool to remove some potential unsafe vehicles from the highway and reduce commercial vehicle accident rates. (Patten, 1989; GAO, 1997; Mitchell, Friswell, and Mooren, 2012; Schoor, Niekerk, and Grobbelaar, 2001; Randhawa, Miller, Bell, and Montagne, 1998; Hall and Intihar, 1997).

Previous research shed light to help us better understand the relationships among the commercial vehicle safety performance, roadside inspection data, and motor carrier's position (Lantz, B.M. 1993; Britto, R.A., Corsi, T.M., and Grimm, C.M., 2010); however, much of such research is outdated. Moreover, very few previous researchers focused on small motor carriers. There is a need to reinvestigate the relationships among a motor carrier's position, such as size and financial status, safety performance, and roadside inspection data. This research will provide additional evidence regarding such relationships, especially for various sized motor carriers. In addition, the trend in the relationships will be provided by comparing the most up-to-date analyses with previous ones.

1.2 Research Objectives

The primary objective of this project is to conduct a statistical analysis to better understand the relationships among roadside inspection data and safety performance data by using data from the Motor Carrier Management Information System (MCMIS).

The following major tasks have been included in the scope of the study:

1) Literature Review: A national and state literature review pertaining to commercial motor carriers' safety performance and roadside inspection will be performed to identify analysis techniques and data requirements. The review will cover journal articles and government reports.

2) Data Collection and Preparation: MCMIS is the main data resource. The primary data documentations from MCMIS will be explored to summarize required inputs to perform the analysis and they are:

- Crash file documentation
- Census file documentation
- Inspection file documentation

The crash file contains data from state police crash reports involving drivers and motor carrier vehicles in the United States. Each report contains about 80 data elements, such as motor carrier, driver, and vehicle information. The crash file may contain multiple records for a crash, which can be distinguished by the crash report number field. Separate report records exist for each commercial motor vehicle involved in the same crash.

The census file contains data for a steadily growing number of active Carriers. Each census record contains: 1) census information, 2) business/operation data, 3) cargo classification, 4) hazardous materials, 5) equipment and driver data, and 6) carrier review data.

The inspection file contains data from state and federal inspection actions involving motor carriers, shippers, and US transporters of hazardous materials. Most of the inspections were conducted at the roadside by state personnel under the Motor Carrier Safety Assistance Program (MCSAP); however, the file also includes mandatory periodic inspections data.

3) Analysis and Evaluation: In this task, a statistical analysis with collected input data is performed. The results from the analysis are summarized and compared with previous research conclusions. Trends of the relationship are provided and analyzed.

4) Train one Ph.D. student on the various theoretical and applicable methods employed.

5) Develop publications and associated research reports.

These research objectives will further the overall goals of promoting economic development, safety, interdisciplinary education, workforce development, and technology transfer that serves the critical needs of the Mountain-Plains Region.

1.3 Report Organization

This section introduces the organization of the report:

Section 2 conducts a complete literature review on the commercial truck crash severity analysis and forecasting models.

Section 3 introduces the data used in this research.

Section 4 summarizes the application findings with data mining analysis.

Section 5 summarizes the conclusions and recommendations from the study.

2. LITERATURE REVIEW

2.1 Background Introduction

In the U.S., there are more than 268 million registered vehicles, and 218 million people holding a valid driver's license (Statista, 2017). In 2015, there were approximately 6.3 million traffic accidents in the U.S. (Statista, 2017). Traffic accidents create an annual average loss to the economy of $871 billion (PBS, 2014), which is more than double national public spending on transportation and water infrastructure at $416 billion. In the United States, 30,000 people die in traffic accidents each year. Every year, traffic accidents cause more than a million deaths around the world. Truck crashes not only interrupt traffic flow, but also cause economic loss. Moreover, truck crashes contribute to a large number of injuries and fatalities due to additional risks, such as larger vehicle size, heavier weight, and possible hazardous material release. Truck crashes are overall more likely to result in more severe outcomes, such as a fatality. In 2014, there were 14 fatalities in large truck crashes per 100 million large truck vehicle miles traveled, while there were only 10.5 fatalities in passenger vehicle crashes per 100 million passenger vehicles miles traveld.. Additionally, there were 29.4 injury crashes involving large trucks per 100 million VMT by large trucks, compared with 58.5 for passenger vehicles (Federal Motor Carrier Safety Administration, 2014).

The need to improve commercial trucking companies' safety performance has been a major social concern in the United States for decades. Transportation agencies and other stakeholders must first identify the complete picture of factors that contribute to the severity levels of commercial truck collisions, then provide directions for commercial truck operation policies to reduce the severe crash rates of commercial trucks.

Previous studies on modeling truck crash severities provide great insights and findings (Lemp, J. et al., 2011; Zhu and Srinivasan, 2011). However, some factors are overlooked and not considered in those studies. Intuitively thinking, management characteristics, organization, culture, strategies, and financial situations in a trucking company should be closely associated with the company's safety performance. For example, safety culture shapes the attitude and behavior of employees. Building a strong safety culture has a great effect on incident reduction (U.S. Department of Labor, 2016). Furthermore, a strong safety culture will result in better trained employees, who will react better when they encounter a potential crash situation, which may result in a less severe crash outcome. Moreover, sufficient capital and profit promote truck maintenance and technology development so equipment performs well, which will minimize risk of equipment failure. In return, incident likelihood and crash severity level would be reduced. Although several studies have been carried out to investigate contributing variables to truck crash severity outcomes, the literature review revealed it is still not clear how some commercial trucking company and driver characteristics impact crash severity levels. The research intends to investigate commercial truck crash severity and contributing factors, especially trucking company related factors, through the application of a data mining model.

Generalized linear models (GLMs) are the most popular statistical models favored by researchers in transportation safety studies (Lu and Tolliver, 2016). GLMs are able to construct an easy-to-interpret quantitative relationship between a dependent variable and its contributors with a mathematical equation. However, GLMs have several limitations (Lu and Tolliver, 2016): 1) They can only handle structured data. When a dataset is complex, especially when it includes a mixture of interval, nominal, ordinal, and numerical variables, and there are a large number of redundant and irrelevant variables (Brusilovkey, 2016), GLMs can have low performance. 2) GLM performance can also be affected when the data are highly heterogeneous and have a high percentage of missing values and outliers. Safety data are usually noisy, and consist of various types of variables. In addition, the high percentage of missing values and large number of redundant variables can all affect GLM performance. 3) GLMs rely heavily on predefined assumptions. When handling big data, all the predefined assumptions can hardly be satisfied at the same time.

Compared with GLMs, data mining models are usually suited to handle more data. In a GLM, a hypothesis has to be proposed before testing the model. Therefore, the noisy data type and redundant variables greatly affect the model and variable selection. However, data mining models require no such hypothesis. In fact, they can reveal the underlying pattern among variables. GLMs are mainly limited by the required predefined assumptions. Moreover, data mining models assume no specific linear relationships, so they are able to discover nonlinear complicated relationships between dependent variables and associated factors.

Statistics and data mining share the same goal: discover and identify structure of the data and turn the data into valuable information. Even though data mining greatly relies on statistics theories, it utilizes knowledge from other fields as well, including machine learning, computer science, and database technology (Priyadharshini, 2017). In a statistics study, a hypothesis needs to be proposed and mathematic functions and models are built up to test the hypothesis. In data mining, no hypothesis is pre-required. The links between the target variable and its associated factors are automatically established.

2.2 Literature

Vehicle types, such as passenger cars or commercial trucks, should have a different impact on crash severity outcomes. Numerous studies focused on single contributor effects on truck crash frequency (Young and Liesman, 2007; Bai, Yang, Li, 2015; Curnow, 2002; Braver, et al., 1997; Anderson, et al., 2012); however, few studies focused on understanding commercial truck crash severity contributors (Khattak, Schneider, and Targa, 2003; Naik et al., 2016; Campbell, 1991; Uddin and Huynh, 2017; Zou, Wang, and Zhang, 2017). Uddin and Huynh (2017) studied influential factors of crash severity involving hazardous materials trucks. Contributors identified in the study are number of vehicle occupant, crash, vehicle, roadway, environmental, and temporal characteristics. Pahukula, Hernandez, and Unnikrishnan (2015) studied the effect of contributor variables on truck crash injury severity in large populated urban areas. Key contributors were identified as traffic flow, light conditions, surface conditions, time of year, traffic flow patterns, speeding and changing lane patterns, and percentage of trucks on the road. Naik et al. (2016) investigated the impact of weather conditions on single-vehicle truck crash injury severity. Their results indicated that wind speed, rain, air temperature, humidity, and icy or snowy road surfaces are associated with crash severities. Campbell (1991) collected survey data about trucks involved in fatal crashes, and analyzed the effect of drivers' ages on fatal crash likelihood. Khattak, Schneider, and Targa (2003) investigated the effect of associated factors with truck-involved single-vehicle crash severity levels. The study found that crashes with greater severity levels were associated with curves and, especially, dangerous driving behaviors, including reckless driving, speeding, passing violations and alcohol/drug use. Zou, Wang, and Zhang (2017) link truck crash severity with spatial location and time of day. Their results revealed that individual truck crashes are spatially dependent events for single- and multi-vehicle crashes.

Among all previous research, understanding how trucking company size attributes and driver's licenses influence crash injury severity is still unclear. Several studies discussed that the little research on how trucking company characteristics impact crash severity is due to the lack of available company data (Chen, 2008). Moreover, research in previous studies was conducted with limited contributors, and few did the study with a large dataset containing comprehensive potential contributors. This research focuses on risk factors for commercial truck crash severity, particularly how company related characteristics affect crash severity, and with a more comprehensive truck crash dataset from the Federal Motor Carrier

Safety Administration (FMCSA). The detailed information regarding this database is described later in the data description section.

The literature search also reveals that most prior studies are based on logit, probit, and their extension statistical models (Lemp, J., et al., 2011; Zhu and Srinivasan, 2011; Wu et al., 2016; Charbotel et al., 2003). However, these statistical models all required certain assumptions. One of the common assumptions is that the effects of contributing factors are assumed identical across different severity levels. These assumptions are inappropriate and do not hold true in most circumstances. Once violated, numerous errors will be generated. In addition, truck crashes are affected by a set of heterogeneous variables (Kumar and Toshniwai, 2015). A goal of the study is to be able to extract hidden, valuable information from large, complex datasets. Thus, instead of applying statistical models, the non-parametric gradient boosting (GB) model, a data mining technique, is selected in this study to overcome the shortcomings and achieve more convincing conclusions. The GB model does not have any predefined data assumptions. Moreover, the GB model inherits most of the tree-based data mining models' advantages. It is also superior to most of the tree-based data mining models with its missing data handling techniques, robustness with data noise, and resistance to over-fitting (Friedman and Meulman, 2003; Zheng, Lu, and Tolliver, 2016). The GB model has proven its success in crash prediction analysis (Chung, 2013; Saha, Alluri, Gan, 2015); however, it has been rarely seen in a truck crash injury severity explanatory study. Therefore, in this study, the authors adopted a GB model to comprehensively analyze influential factors on truck crash injury severity.

2.3 Gradient Boosting Model

The gradient boosting method is also known as multiple additive trees (MAT), and is a machine-learning data-mining technique for regression and classification problems proposed by Friedman (2001, 2003) at Stanford University. A GB model can be viewed as a series expansion approximating the true functional relationship (Salford-Systems). Therefore, the GB model inherits all the advantages of tree-based models while improving other aspects, such as forecasting accuracy (Friedman and Meulman, 2003). In general, a GB model starts by fitting the data with a simple decision tree model, which has a certain level of error in terms of fitness with the data. The simple DT model is referred as a weak learner. Considering the errors having the same correlation with outcome value, the GB model then develops another decision tree model on the errors or residuals of the previous tree.

The detailed algorithm of GB is described as follows (De'ath, 2007; Hastie, Tibshirani, and Friedman, 2009):

$$f(x) = \sum_n f_n(x) = \sum_n \beta_n g(x, \gamma_n) \tag{25}$$

Where x is a set of predictors, and $f(x)$ is the approximation of the response variable. $g(x,\gamma_n)$ are single decision trees with the parameter γ_n indicating the split variables. β_n (n=1,2,…,n) are the coefficients, and determine how each single tree is to be combined. Loss function measures prediction performance, such as deviance. Friedman (2001) proposed a numerical optimization method called functional gradient descent, which is an iterative tree-building process. The process keeps adding trees until all observations are perfectly fitted. To avoid over-fitting, the model is also validated with a testing dataset. Iterative training will stop when the performance of the model reaches a point where the model predicts well for both the training and testing datasets.

3. DATA

3.1 Data Sources

In this study, truck crash data were obtained from the FMCSA. Crash data files, census file, and inspection files from the MCMIS are selected for the research. The MCMIS datasets contain 1) records from state police crash reports, including information on drivers, crash conditions, environmental factors when the crash occurred, and crash-involved truck conditions; 2) motor carrier corporation variables and operational factors; and 3) motor carrier safety inspection records. This study examines truck crash related data for crashes that occurred in North Dakota and Colorado from 2010 to 2016. The selection of the two states is due to the availability of data, research interest, and data size limitation; however, the research can be extended to a national level or include additional states if this is of interest.

3.2 Data Analyzed

The authors excluded irrelevant, privacy variables and four redundancy variables from the raw data before performing mathematical analysis. As summarized in Table 3.1, 38 variables are removed from analysis.

Table 3.1 Summary of Unanalyzed Variables

Variable	Rationale for removal
Carrier related variables	
Address	Irrelevant variable
Zip code	Irrelevant variable
Country	Irrelevant variable
Phone	Irrelevant variable
Identification number	Irrelevant variable
Last updated date	Irrelevant variable
May have undeliverable physical address	Irrelevant variable
May have undeliverable mailing address.	Irrelevant variable
Carrier name	Irrelevant variable
USDOT number	Irrelevant variable
City	Irrelevant variable
Crash related variables	
Crash ID	Irrelevant variable
Crash year	Redundant variable with variable “Year”
Crash quarter	Irrelevant variable
Federal recordable	Irrelevant variable
Officer badge	Irrelevant variable
Crash number	Irrelevant variable
Crash date	Irrelevant variable
Crash time	Redundant variable with variable "Time of Day"
Officer badge	Irrelevant variable
Record status	Irrelevant variable

(Table 3.1 continued)

Matched status	Irrelevant variable
SAFETYNET input date	Irrelevant variable
MCMIS upload date	Irrelevant variable
Number days to SAFETYNET	Irrelevant variable
Number days to MCMIS	Irrelevant variable
Counter	Irrelevant variable
Vehicle configuration desc	Redundant variable with variable "Vehicle configuration"
GVW rating Desc	Redundant variable with variable "GVW"
City	Irrelevant variable
City code	Irrelevant variable
County	Irrelevant variable
County code	Irrelevant variable
Number assigned to motor carriers engaging in interstate or foreign operations	Irrelevant variable
Registered as a common carrier: A- Active registration, I-Inactive registration, N- no registration	Irrelevant variable
Driver related variables	
First name	Irrelevant variable
Last name	Irrelevant variable
Mid name	Irrelevant variable

Detailed information of the data analyzed in this research is shown in Table 3.1. In general the data variables can be grouped into the following five categories: trucking company characteristics, crash characteristics, environmental characteristics, driver characteristics, and truck characteristics.

The authors selected 24 variables to be investigated and tested. Of those, 21 are categorical variables (labeled with "$" in Table 3.2) and two of them are numeric variables. In this study, the target variable (injury severity) is classified as: 0=property damage only; 1=injury only (no fatalities); 2=only one fatality; 3=two or more fatalities. There were 16,389 recorded truck-involved crashes in North Dakota and Colorado from 2010 to 2016. Of those crashes, 72.13% (11,822) resulted in property damage only (severity=0); 24.22% (3,969) were injury only (severity=1); 1.97% (323) caused one fatality (severity=2); and 1.68% (275) caused two or more fatalities (severity=3). Trucking companies are divided into five categories based on company size: 1=single truck companies; 2=small truck companies; 3=medium-size truck companies; 4=large truck companies; and 5=very large truck companies.

Table 3.2 Variable Description

Variable	Total	Missing Records	Missing Percentage	Data Description
Trucking Company Characteristics				
Carrier State$	16,389	0	0	State/District/Province of the principal place of business of the carrier registered
Company Size$	13,221	3,168	19.33	1, 2, 3, 4, 5
Indicator$	13,269	3,120	19.04	'S' = Safety; 'I' = Insufficient Data; 'N' = Intrastate Safety; 'R' = Random
Inspection Value	13,269	3,120	19.04	Inspection value. Ranging from 0 to 100 with 100 indicate the worst performance
Interstate Carrier$	15,376	1,013	6.18	Is carrier an interstate carrier? Yes/No
New Entrant$	16,389	0	0	Is carrier a new registered carrier? Yes/No
Crash Characteristics				
Day of Week$	16,389	0	0	Sun.; Mon.; Tue.; Wed.; Thu.; Fri.; Sat.
First Harmful Event$	16,114	275	1.68	The first injury or damage-producing event
Time of Day$	16,250	139	0.85	12:00 - 2:59 AM; 3:00 - 5:59 AM; 6:00 - 8:59 AM; 9:00 - 11:59 AM; 12:00 - 2:59 PM; 3:00 - 5:59 PM; 6:00 - 8:59 PM; 9:00 - 11:59 PM
Tow Away$	16,389	0	0	Is accident vehicle towed away? Yes/No
Number of Vehicles	16,388	1	0.01	The total number of vehicles or vehicle combinations involved in the crash.
Environmental Characteristics				
Light Condition$	16,371	18	0.11	Dark – Lighted; Dark - Not Lighted; Dark - Unknown Roadway Lighting, Dawn; Daylight; Dusk; Other; Unknown
Road Surface Condition$	16,382	7	0.04	Dry; Ice; Other; Sand, Mud, Dirt, Oil Or Gravel; Slush; Snow; Water(Standing, Moving); Wet; Unknown;
Traffic Way Type$	16,388	1	0.01	Not Reported; One-Way Trafficway, Not Divided; Two-Way Trafficway, Divided, Positive Barrier; Two-Way Trafficway, Divided, Unprotected Median; Two-Way Trafficway, Not Divided
Weather Condition$	16,378	11	0.07	Blowing Sand, Soil, Dirt, Or Snow; Fog; No Adverse Conditions; Other; Rain; Severe Crosswinds; Sleet, Hail; Snow; Unknown
Driver Characteristics				
Driver's Age$	16,389	0	0	<26; 26 – 35; 36 – 45; 46 – 55; 56 – 65; 66 – 75; >75
Driver's License Class$	15,816	573	3.5	A, B, C, D
Driver's License State$	16,020	369	2.25	The license state/district/province of the driver.
Valid Driver's License$	16,148	241	1.47	If driver's license is valid or not. Yes/No
Truck Characteristics				
GVWR$	16,382	7	0.04	Gross Vehicle Weight Rating in pounds: < 10,000; 10,001-26,000;>26,000
Cargo Body Type$	16,333	56	0.34	Auto Transporter; Bus Seats For 9-15 People, Including Driver; Bus Seats For > 15 People, Including Driver; Cargo Tank; Concrete Mixer; Dump; Flatbed; Garbage/Refuse; Grain, Chips, Gravel; Intermodal; Logging; Not Applicable/No Cargo Body; Other; Pole; Van/Enclosed Box; Vehicle Towing Another Vehicle
Vehicle Configuration$	16,370	19	0.12	Light Truck(Only If Vehicle Displays Hm Placa; Single-Unit Truck (2-Axle, 6 Tire); Single-Unit Truck (3 Or More Axles); Tractor/Double; Tractor/Semi-Trailer; Tractor/Triples; Truck Tractor (Bobtail); Truck/Trailer; Unknown
Vehicle License State$	16,356	33	0.2	The license state/district/province of the truck.
Target Variable				
Severity	16,389	0	0	0=no injuries and no fatalities; 1=injuries and no fatalities; 2=one fatality and no injuries; 3=one fatality and injuries, or two or more fatalities

4. ANALYSIS RESULTS AND DISCUSSIONS

4.1 Result Analysis

The raw crash data from North Dakota and Colorado are fit into the GB model, and 25 contributor variables are tested as predictors of injury severity. Of these variables, 21 are found to be associated with injury severities. To further understand the importance of the 21 contributors, the relative variable importance analysis for causal importance of inputs is also conducted. Table 4.1 presents variable importance under four injury severity levels. The importance of a variable in a simple single tree is measured by the number of times the variable is used as a splitter, and the improvement on mean squared error attributed to the tree due to the splits by the variable. After summing the importance score computed in a simple single tree over the ensemble of trees and the average value of the summation is scaled to the most important variable, which scores 100, the scaled average value is then regarded as the variable's importance in the model. A high value of variable importance indicates a high contribution a variable makes to the prediction (Friedman and Meulman, 2003). As noted in Table 4.1, the top 11 variables account for more than 80% of injury forecasting. For Injury, the most important variable is Carrier State, which indicates that this variable makes the most contributions as compared with the other variables in explaining injury crashes. First Harmful is the second most important contributor, and it accounts for 85% of the importance that Carrier State contributes to. The column Cum % in Table 4.1 indicates the absolute cumulative contribution of the variables. For Injury, one can say that two contributors, Carrier State and First Harmful, account for 30% of contributions among all the contributors to explain crash severity.

Table 4.1 Variable Importance under Each Level of Severity

Damage only (Severity=0)			Injury (Severity=1)			One Fatality (Severity=2)			Multiple Fatalities (Severity=3)		
Variable	**Score**	**Cum %**	**Variable**	**Score**	**Cum %**	**Variable**	**Score**	**Cum %**	**Variable**	**Score**	**Cum %**
Carrier State	100	16%	Carrier State	100	16%	Carrier State	100	17%	Carrier State	100	16%
Tow away	71	28%	First Harmful Event	85	30%	First Harmful Event	69	28%	Number of Vehicles	66	27%
First Harmful Event	53	36%	Tow away	50	39%	Cargo Body Type	55	38%	First Harmful Event	50	35%
Cargo Body Type	50	44%	Vehicle Configuration	44	46%	Time of Day	45	45%	Cargo Body Type	45	43%
Time of Day	46	52%	Traffic Way Type	44	53%	Day of Week	39	52%	Time of Day	43	50%
Day of Week	40	58%	Cargo Body Type	43	60%	Vehicle Configuration	37	58%	Weather Condition	41	57%
Driver Age	33	64%	Road Surface Condition	40	67%	Driver Age	35	64%	Day of Week	33	62%
Weather Condition	31	69%	Light Condition	32	72%	Road Surface Condition	33	69%	Light Condition	30	67%
Number of Vehicles	26	73%	Number of Vehicles	30	77%	Weather Condition	24	73%	Tow away	25	72%
Vehicle Configuration	26	77%	Weather Condition	25	81%	Company Size	22	77%	Vehicle Configuration	25	76%
Company Size	22	81%	Driver Age	22	85%	Driver's License Class	21	81%	Inspection Value	25	80%
Light Condition	19	84%	Time of Day	19	88%	Inspection Value	17	84%	Driver Age	24	84%
Road Surface Condition	17	87%	Day of Week	16	91%	Number of Vehicles	16	86%	Traffic Way Type	22	88%
GVWR	15	89%	Company Size	15	93%	Indicator	15	89%	Company Size	20	91%
Indicator	14	92%	GVWR	9	95%	Light Condition	15	92%	Road Surface Condition	17	94%
Driver's License Class	13	94%	Indicator	8	96%	Tow away	14	94%	Indicator	10	96%
Inspection Value	13	96%	Interstate Carrier	7	98%	Traffic Way Type	14	96%	New Entrant	8	97%
Traffic Way Type	10	98%	Inspection Value	6	99%	Interstate Carrier	7	98%	GVWR	5	98%
Interstate Carrier	9	99%	Driver's License Class	6	100%	GVWR	5	99%	Interstate Carrier	5	99%
New Entrant	2	100%	Valid Driver's License	1	100%	Valid Driver's License	5	100%	Driver's License Class	4	100%
Valid Driver's License	1	100%	New Entrant	0	100%	New Entrant	2	100%	Valid Driver's License	1	100%

As illustrated in Table 4.1, variables contribute differently when explaining different crash severities. A variable showing significant importance for a certain severity level may be less crucial for another. For instance, Cargo Body Type is the second most important factor for predicting fatality crashes, but is much less important for predicting property damage only crashes (severity=0). However, it is clear that Carrier State is the most influential factor for all severity levels. First Harmful Event also plays an important role in predicting all severity levels. Figure 4.1 better indicates the relative importance levels.

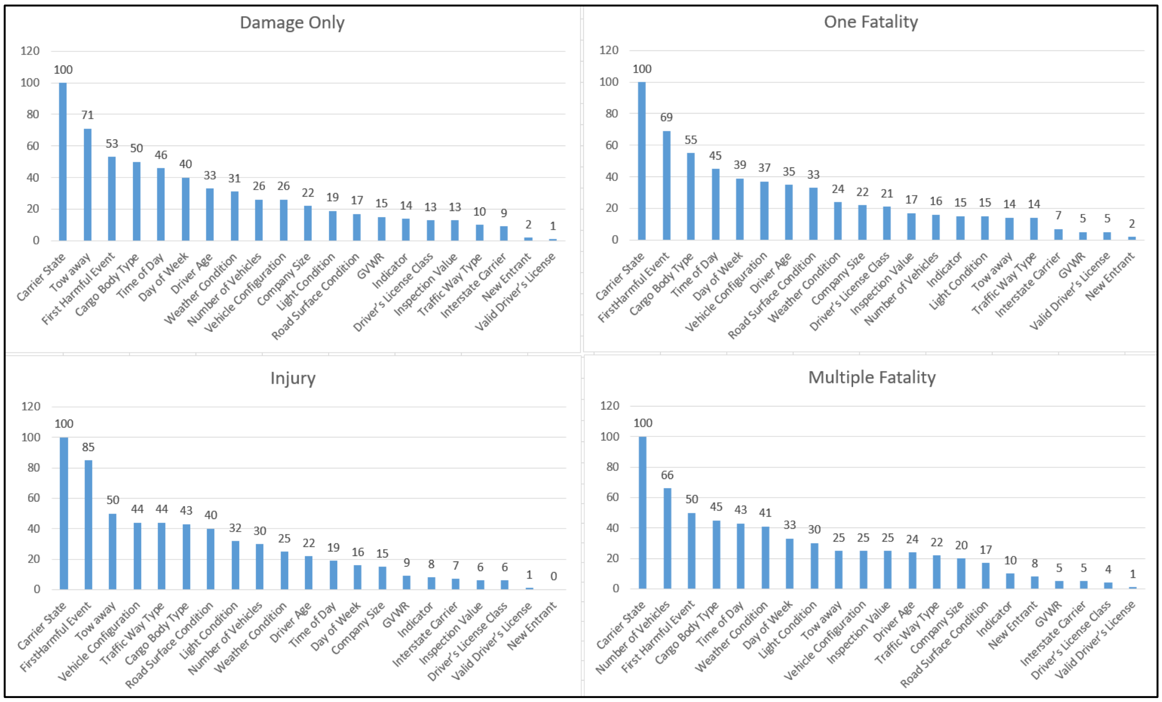

Figure 4.1 Importance Bar-Charts for Four Severity Levels

From Figure 4.1 and Table 4.1, one can tell that for Damage Only, the top contributors are Carrier State, Tow Away, First Harmful Event, Cargo Body Type, Time of Day, and Day of Week; but for very severe crash results, Multiple Fatalities, the top contributors are Carrier State, Number of Vehicles, First Harmful Event, Cargo Body Type, Time of Day, and Weather Condition.

Some other interesting findings are observed in the analysis: 1) three variables are all identified as important contributors for all four severity levels: Carrier State, Cargo Body Type, and First Harmful Event; 2) Time of Day and Day of Week play more important roles in explaining damage only crash and fatality crash, but less importance in explaining injury; 3) Driver Age plays a more important role for damage only and one fatality, but less importance for injury and multiple fatalities; 4) Vehicle Configuration plays a more important role in injury and one fatality than in damage only and more fatalities; 5) Number of Vehicles plays a very important role in multiple crashes, and a relatively important role in property damage and injury crashes, but relatively low importance in one fatality crash; 6) Inspection Value plays a relatively important role in fatality crashes, but plays a relatively less important role in damage only crashes and injury crashes; 7) Road Surface Condition plays relatively important role in injury and one fatality, but plays a relatively less important role in property damage and multiple fatalities; and 8) trucking company size, road surface condition, safety inspection value, valid driver's license, and driver's license class all significantly impact crash severities at different levels. The GB successfully identified the contribution variables to crash severities and prioritized their importance. A marginal effect analysis of each influential variable is also analyzed to provide a further detailed understanding of how they contribute to various crash severities.

4.2 Marginal Effects Analysis Result Discussions

Marginal effects of important variables are summarized in Table 4.2. For categorical variables with various levels, due to the space limitations, Table 4.2 only shows selected levels for a significant contributor categorical variable with the most significant impacts. Moreover, levels with more outstandingly significant impacts are bolded. For example, from Table 4.1, one can tell that weather condition is the 8th significant contributor variable for a damage only crash. In Table 4.2, only the positive effect of the weather condition of snow is listed, which has a much more significant impact on damage only crashes than any other weather conditions. And the level of snow is bolded. The first column, Variable, lists influential variables whose impact on severity prediction is valuable. Positive effect (P) means that the corresponding categories for the influential variable will increase the probability of a certain severity level (column severity=0, 1, 2, 3), while negative effect (N) means it will decrease that likelihood.

Examining Carrier State as one example, if a carrier is registered in Massachusetts, Mississippi, Ohio, or Wisconsin, this has a significantly positive effect on damage only, and if a carrier is registered in North Dakota or Texas, this has a significantly negative effect on damage. All other unlisted carrier states have no significantly different contribution for damage only crashes. And there is no bolded state among all listed positive or negative impact carrier states, which indicates their positive/negative impact effects are not significantly different within their corresponding category.

Table 4.2 Marginal Effect of Influential Variables

Variable	Effect	Damage only (Severity=0)	Injury (Severity=1)	One Fatality (Severity=2)	Two or More Fatality (Severity=3)
Trucking Company Characteristics					
Carrier State$	P	MA, MS, OH, WI	AL, OR, WI, MS	MO, KS	MI, MB, NC, ND, PA
	N	ND, TX	KS, MO	GA, NY, PA	AL, MA, MS, FL, OH, OR
Inspection Value	P	<30	<25	<45	30-70, >90
	N	>30	>25	>50	80-90
Company Size$	P	1, 5 (very small or very large)	1, 4 (very small or large size)	1, 3 (very small, medium)	4, 5 (large, very large)
	N	2, 4 (small or large)	2, 3, 5 (small, medium, or very large)	2, 4, 5 (small, large, very large)	1, 2, 3 (very small, small, medium)
Interstate Carrier$	P	N	N	Y	Y
	N	Y	Y	N	N
New Entrant$	P	N/A	N	Y	Y
	N	N/A	Y	N	N
Indicator$	P	N, R	I, N, R	S	N, S
	N	I, S	S	I, N, R	I, R

(Table 4.2 continued)

Crash Characteristics					
First Harmful Event$	P	**Involving Animal;** Involving Fixed Object; Involving Other Movable Object; Involving Train; Involving Unknown Movable Object; Work Zone Maintenance Equipment; **Eqp Failure; Cargo Loss or Shift;** Cross Median/Centerline; Downhill Runaway; **Explosion Or Fire; Jackknife;** Other; Overturn (Rollover); **Separation Of Unit;** Unknown	Involving Fixed Object; Involving Pedalcycle; Involving Unknown Movable Object; Work Zone Maintenance Equipment; Eqp Failure; **Cargo Loss or Shift;** Downhill Runaway; Explosion Or Fire; Jackknife; Other; **Overturn (Rollover);** Ran Off Road; Separation Of Unit; Unknown	Involving Animal; Involving Fixed Object; Involving Other Movable Object; **Involving Pedalcycle; Involving Pedestrian; Involving Train; Involving Unknown Movable Object;** Work Zone Maintenance Equipment; Downhill Runaway; Ran Off Road; Other	Involving Motor Vehicle In Transport; **Involving Parked Motor Vehicle;** Involving Pedestrian; **Involving Train;** Work Zone Maintenance Equipment; **Cross Median/Centerline;**
	N	Involving Motor Vehicle In Transport; Involving Parked Motor Vehicle; Involving Pedalcycle; Involving Pedestrian; Involving Train; Ran Off Road;	Involving Animal; Involving Motor Vehicle In Transport; Involving Other Movable Object; Involving Parked Motor Vehicle; Involving Pedestrian; Involving Train; Cross Median/Centerline;	Involving Motor Vehicle In Transport; Involving Parked Motor Vehicle; Eqp Failure; Cargo Loss Or Shift; Cross Median/Centerline; Explosion Or Fire; Jackknife; Overturn (Rollover); Separation Of Unit;	Involving Animal; Involving Fixed Object; Involving Other Movable Object; Involving Pedalcycle; Involving Unknown Movable Object; Eqp Failure; Cargo Loss Or Shift; Downhill Runaway; Explosion Or Fire; Jackknife; Other; Overturn (Rollover); Ran Off Road; Separation Of Unit;
Number of Vehicle in Crash	P	<2	<2	>2	>4
	N	>2	>2	<2	<4
Time of Day$	P	**9-12AM;** 12PM-3PM; **3-6 PM**	12 AM - 3 AM; 6AM - 9 AM; 12 PM - 3 PM; 3PM - 6 PM;	3AM - 6 AM; **9:00 PM - 12PM**	0 AM - 3 AM; 3 AM - 6 AM; 12PM - 3 PM; 6 PM - 9 PM;
	N	The rest	3 AM - 6 AM; 9 AM - 12 AM; 6 PM - 9 PM; 9 PM - 12 PM	0AM - 3 AM; 6 AM - 9 AM; 9 AM - 12 AM; 12 PM - 3 PM; 3:00 PM - 6 PM; 6 PM - 9 PM;	6 AM - 9 AM; 9AM - 12 AM; 3 PM - 6 PM; 9 PM - 12 PM
Day of Week$	P	Mon. Wed. **Thu**. Fri.	**Mon**. Wed. Fri.	**Tue. Sat.** Sun.	**Mon.** Wed. Thu. Sat. Sun.
	N	Tue. Sat. Sun.	Tue. Thu. Sat. Sun.	Mon. Wed. Thu. Fri.	Tue. Fri.
Tow Away$	P	Y	N	N	Y
	N	N	Y	Y	N
Environment Characteristics					
Weather Condition$	P	Rain; Severe Crosswinds; Sleet, Hail; **Snow**; Unknown	**Blowing Sand, Soil, Dirt, Or Snow;** Fog; **Rain**; Severe Crosswinds; Sleet, Hail; Snow; **Unknown**	No Adverse Conditions;	**Fog**; Other; **Severe Crosswinds**; Sleet, Hail; Unknown
	N	Blowing Sand, Soil, Dirt, Or Snow; Fog; No Adverse Conditions; Other;	No Adverse Conditions; Other;	Blowing Sand, Soil, Dirt, Or Snow; Fog; Other; Rain; Severe Crosswinds; Sleet, Hail; Snow; Unknown	Blowing Sand, Soil, Dirt, Or Snow; No Adverse Conditions; Rain; Snow
Road Surface Condition$	P	Ice; Slush	Ice; Sand, Mud, Dirt, Oil Or Gravel; Slush; Snow; Unknown; Water(Standing, Moving);	Dry; Ice; Other; Sand, Mud, Dirt, Oil Or Gravel; Unknown; Water(Standing, Moving); Wet	Dry; Wet

(Table 4.2 continued)

	N	Dry; Other; Sand, Mud, Dirt, Oil Or Gravel; Snow; Unknown; Water(Standing, Moving); Wet	Dry; Other; Wet	Slush; Snow;	Ice; Other; Sand, Mud, Dirt, Oil Or Gravel; Slush; Snow; Unknown; Water(Standing, Moving);
Light Condition$	P	Dark – Lighted; Dark - Unknown Roadway Lighting, Dawn; Daylight; Other; Unknown	Dark – Lighted; Dark - Unknown Roadway Lighting, Dawn; Daylight; Other; Unknown	Dark – Lighted; Dark - Not Lighted; Dusk	Dark - Not Lighted;
	N	Dark - Not Lighted; Dusk	Dark - Not Lighted; Dusk;	Dark - Unknown Roadway Lighting, Dawn; Daylight; Other; Unknown	Dark – Lighted; Dark - Unknown Roadway Lighting, Dawn; Daylight; Dusk; Other; Unknown
Trafficway Type$	P	Not Reported; One-Way Trafficway, Not Divided; Two-Way Trafficway, Divided, Positive Barrier;	One-Way Trafficway, Not Divided; Two-Way Trafficway, Divided, Unprotected Median;	Two-Way Trafficway, Divided, Unprotected Median;	One-Way Trafficway, Not Divided; Two-Way Trafficway, Divided, Unprotected Median; Two-Way Trafficway, Not Divided
	N	Two-Way Trafficway, Divided, Unprotected Median; Two-Way Trafficway, Not Divided	Not Reported; Two-Way Trafficway, Divided, Positive Barrier; Two-Way Trafficway, Not Divided	Not Reported; One-Way Trafficway, Not Divided; Two-Way Trafficway, Divided, Positive Barrier; Two-Way Trafficway, Not Divided	Not Reported; Two-Way Trafficway, Divided, Positive Barrier;
Driver Characteristics					
Driver Age$	P	26-45	26-45, 66+	<25, 45-65, 75+	75+, 25-
	N	the rest	the rest	the rest	26-45
Driver's License Class$	P	**B**, C	**B, C**	A, D	B, C, D
	N	A, D	A, D	B, C	**A**
Valid Driver's License$	P	N	Y	N	N
	N	Y	N	Y	Y
Truck Characteristics					
Cargo Body Type$	P	**Auto Transporter; Bus Seats For 9-15 People, Including Driver;** Bus Seats For > 15 People, Including DriverDump; Intermodal; Logging; Pole; Van/Enclosed Box;	Auto Transporter; Bus Seats For 9-15 People, Including Driver; Bus Seats For > 15 People, Including Driver; Concrete Mixer;	Auto Transporter; **Cargo Tank; Concrete Mixer;** Dump; Flatbed; Garbage/Refuse; Grain, Chips, Gravel; Intermodal; Logging; Other; Pole; **Vehicle Towing Another Vehicle**	Cargo Tank; Dump; Flatbed; Garbage/Refuse; Grain, Chips, Gravel; Logging; Not Applicable/No Cargo Body; Other; Van/Enclosed Box; Vehicle Towing Another Vehicle
	N	; Cargo Tank; Concrete Mixer; Flatbed; Garbage/Refuse; Grain, Chips, Gravel; Not Applicable/No Cargo Body; Other; Vehicle Towing Another Vehicle	Cargo Tank; Dump; Flatbed; Garbage/Refuse; Grain, Chips, Gravel; Intermodal; Logging; Not Applicable/No Cargo Body; Other; Pole; Van/Enclosed Box; Vehicle Towing Another Vehicle	Bus Seats For 9-15 People, Including Driver; Bus Seats For > 15 People, Including Driver; Not Applicable/No Cargo Body; Van/Enclosed Box	Auto Transporter; Bus Seats For 9-15 People, Including Driver; Bus Seats For > 15 People, Including Driver; Concrete Mixer; Intermodal; Pole
Vehicle Configuration$	P	Light Truck(Only If Vehicle Displays Hm Placa; Single-Unit Truck (2-Axle, 6 Tire); Tractor/Triples; Truck/Trailer; Unknown	Light Truck(Only If Vehicle Displays Hm Placa; Single-Unit Truck (3 Or More Axles); Tractor/Triples; Truck/Trailer; Unknown	Tractor/Double; Tractor/Semi-Trailer; Truck Tractor (Bobtail); Truck/Trailer; Unknown	Single-Unit Truck (2-Axle, 6 Tire); Tractor/Double; Tractor/Semi-Trailer

(Table 4.2 continued)

	N	Single-Unit Truck (3 Or More Axles); Tractor/Double; Tractor/Semi-Trailer; Truck Tractor (Bobtail);	Single-Unit Truck (2-Axle, 6 Tire); Tractor/Double; Tractor/Semi-Trailer; Truck Tractor (Bobtail);	Light Truck(Only If Vehicle Displays Hm Placa; Single-Unit Truck (2-Axle, 6 Tire); Single-Unit Truck (3 Or More Axles); Tractor/Triples;	Light Truck(Only If Vehicle Displays Hm Placa; Single-Unit Truck (3 Or More Axles); Tractor/Triples; Truck Tractor (Bobtail); Truck/Trailer; Unknown
GVWR$	P	10,001-26,000;	10,001-26,000; >26,000	>26,000	>26,000
	N	>26,000	>26,000	10,001-26,000;	10,001-26,000;

From Table 4.2, one can tell trucks registered in different states perform differently in terms of crash severity. For single fatality analysis, crashes with trucks for carriers registered in Missouri and Kansas are significantly more prone to have a single fatality crash, while trucks from carriers registered in Georgia, New York, and Pennsylvania are prone to not have a single fatality crash. Other states not listed have no significant difference with the likelihood of a single fatality crash.

Trucking companies own various numbers of trucks, which have different influences on different crash severity levels. For damage only, trucks from very large or single-truck companies are more likely to be involved in damage only crashes, while trucks from small or large companies are less likely to be involved in damage only crashes. Trucks from medium sized companies have no significant difference in involvement in damage only crashes. In examining single-truck companies, one can see they have a high risk to be involved in damage only, injury, or single fatality crashes, but have a low risk to be involved in multiple fatality crashes. Very large companies, those that own more than 100 trucks, tend to have a high risk to be involved in either damage only or multiple fatality crashes, but a low risk to be involved in injury or single fatality crashes. It is also notable that small truck companies, those with two to five trucks, are found to be the best safety performance companies in terms of crash severity, given that they are estimated to have a negative impact on all levels of crash severity. In other words, small sized companies are significantly different than any other size company in their involvement in any crash severity level. The underlying reasons for this observation are unclear, and further investigation is needed. It is inferred that in a fatal crash, trucks from large and very large companies are more likely to cause multiple deaths. A potential rationale could be larger truck companies own heavier and larger trucks, and those trucks are hard to maneuver and need more time to perform brake operations in an emergency situation.

As expected, the company inspection value has a significant impact on their safety conditions. FMCSA has defined three general categories based on the inspection value: higher risk carriers have inspection values of 75 or greater, medium risk carriers have inspection values between 50 and 74, and low risk carriers have inspection values less than 50. The results indicate that low inspection values, less than 30 or 25, respectively, are positively associated with the likelihood of a less severe crash result, such as damage only or injury. In other words, low inspection values generally indicate better performing companies; however, the benchmark value is around 25 to 30 rather than 50 in terms of crash severity. For fatality analysis, the benchmark value is 45, which is close to the 50 used in FMCSA inspection categories. The most interesting finding is for multiple-fatality crashes. A truck with a company inspection value between 30 and 70 or greater than 90 is significantly more prone to multiple-fatality crashes. For a company with an inspection value greater than 90, this finding was not a surprise. The inspection value is based on the prior safety record of the company. Those companies with higher values have had more crashes and more violations in past inspections, and thus are more likely to continue to have safety issues. Another rationale for this observation could be that those companies have much larger trucks, but further study is suggested to verify the hypothesis. However, for a company with a value between 30 and 70, the finding is very surprising. Further single factor investigation is needed to understand how the inspection value is associated with crash severity.

It was found that trucks owned by interstate companies are more prone to have fatality crashes, while those owned by an intrastate company are more likely to be involved in damage or injury only crashes. Interstate company truck drivers usually have longer driving distances, which could cause drowsy driving and run-off-the-road fatal crashes. Thus, further data collection to support significant testing on how drowsy driving by interstate and intrastate company truck drivers contributes to crash severity levels between is recommended.

In addition, interstate companies may hire more drivers registered in the same state as the companies. Drivers from other states may be not familiar with local driving behavior when driving in states other than their own. For example, comparing North Dakota and Colorado, most ND highways are rural, and ND drivers are less aggressive than CO drivers. Thus, when ND truck drivers are on roads in Colorado, they may not familiar with the driving behavior, and may not react in time. Because of limited data, it is suggested that law enforcement officers record state registration of the license for each truck driver involved in a crash. If the hypothesis is verified, it is recommended that truck drivers be trained to drive in various driving behavior environments.

Newly registered companies are found to be non-significant for damage-only crashes. However, trucks owned by newly registered companies have a higher risk of fatality crashes. This meet the expectations because newly registered companies are usually less experienced in fleet management and safety practices.

Regarding crash characteristics, the first harmful event is one of the most significant explanatory variables in crash outcome prediction. A conclusion can be drawn that the huge difference in speed and weight among vehicles involved in crashes is one of the major contributors to fatality crashes. In such cases, the more vulnerable road users expose themselves to a high risk of fatal crashes. For example, when a truck hits a passenger car, the fatal outcome could be due to the huge impact at the moment of collision. As expected, the more vehicles involved in a truck crash increases the probability of a more severe outcome. Crash severity level also changes over the times in a day. It is notable that early morning (3 am to 6 am) is considered the most dangerous time, given that both single-fatality and multiple-fatality crashes are more likely to happen during this period. This may result from difficulty in a driver making appropriate responses when it is dark or suffers from lack of sleep (Pahukula, Hernandez, and Unnikrishnan, 2015). During weekends, crashes are more likely to be fatal, while on Fridays, crashes are prone to be non-fatal.

Regarding environmental characteristics, it is not surprising that weather condition is a significant factor affecting crash severity levels. Interestingly, fatal crashes are less likely to happen on a snowy or rainy day when drivers are more cautious than usual. Nevertheless, single-fatality crashes are more likely to happen with no adverse weather conditions. The reasons could be more truck traffic and/or higher travel speeds in good weather, and also drivers tend to fully focus on driving, slow down, and keep their eyes on the road under bad weather. Thus, traffic exposure data can be very helpful to better understand the relationship between weather and severity. Validation of such a hypothesis can result in warning signs to remind drivers to obey speed limits or apply speed enforcement during good weather. Fog and severe crosswinds negatively affect drivers' visualization and make large trucks hard to control. Thus, under these conditions, the probability of multiple-fatality crashes is predicted to increase. An icy road surface is a definite crash factor. Drivers usually pay more attention than usual; however, an icy road surface can make crashes inevitable. Thus, an icy road surface increases the risk of damage-only, injury, and single-fatality crashes, but decreases the likelihood of multiple-fatality crashes, most likely because drivers lower their speed under such conditions. A slushy road condition raises the risk of damage-only and injury crashes. On the other hand, the likelihood of fatal crashes increases under a dry or wet road surface condition. Night is considered a dangerous time, because it is a positively significant contributor to all fatality level crashes. It is noteworthy that the fatal crash risk increases at night with no street lighting;

thus, visualization is negatively impacted, which is supported by previous studies (Lemp, Kockelman, and Unnikrishnan, 2011; Kockelman, Murray, and Ma, 2007).

The marginal effect of traffic way type indicates that a median barrier effectively prevents fatal crashes, because those are more likely when two-way traffic is not separated.

Regarding driver characteristics, a driver's age is a significant factor for predicting crash severity levels. Younger drivers (< 25 years old) and older drivers (> 75 years old) are found to be the most vulnerable groups for multiple-fatality crashes. The underlying reason could be that young people have less driving experience and may be more prone to dangerous actions. On the other hand, older people do not react as quickly as younger persons, and their overall health condition could also impact their risk of fatalities (Chen et al., 2015; Campbell, 1991). The driver's license class is another significant variable. Class A, B, and C are significant in improving truck safety performance in regard to crash severity level. For example, drivers with class D licenses are predicted to more likely be involved in fatal crashes. An invalid driver's license is predicted to increase risk of damage-only, single-fatality, and multiple-fatality crashes. Not surprisingly, considering it is illegal, people driving with an invalid driver's license could be less responsible, more aggressive, and possibly have a bad driving record.

Regarding truck characteristics, the cargo body type is a factor impacting injury severity. Cargo tanks, flatbeds, and grain trucks, or trucks towing another vehicle, increase the probability of high injury severities (severity=2, 3). The weight of these trucks increases operation difficulty in an emergency situation. Severity level is predicted to positively relate with gross vehicle weight.

5. CONCLUSIONS

5.1 Research Results, Summary, and Conclusions

This research details a comprehensive analysis regarding the impacts of a set of heterogeneous factors (trucking company, crash, environment, truck driver, and truck characteristics) on injury severity caused by truck crashes by analyzing six recent years of Federal Motor Carrier Safety Administration data. Gradient boosting, a data mining technique, is used to study significant influential factors and their marginal effect on injury severities.

The target variable (crash severity) is classified into four categories: property damage only, injury only, one fatality, and two or more fatalities. Based on a GB model, 22 variables are proven to significantly relate with severity. For the first time, trucking company and driver characteristics are proven to have significant impact on truck crash injury severity. Some of the results in this study reinforce previous studies' conclusions. For example, wet road surface, bad visualization (dark or low light conditions, or fog/poor weather conditions), strong crosswinds, heavy gross vehicle weight (over 26,000 lbs.), and collisions with opposing traffic are estimated to increase the likelihood of more severe outcomes. Younger drivers (under 25 years old) and older drivers (over 75 years old) are predicted as the most likely groups to be involved in crashes resulting in fatalities. Also, truck crash severity levels become higher when more vehicles are involved.

One interesting finding is that fatal crashes are more likely when weather is good or the road surface has no adverse conditions, perhaps because adverse conditions make people vigilant to potential risk. Another unique contribution of this study is to demonstrate the significant effect of the trucking company and driver characteristics on injury severities. Based on ND and CO crash data, it is estimated that carriers registered in Missouri, Manitoba, North Carolina, North Dakota, and Pennsylvania increase the likelihood of the most severe outcomes. Companies owning two to five trucks are predicted to have the lowest probability of crash risk. Carriers with inspection values of 30-70, or greater than 90, increased the possibility of high injury severities. Newly registered carriers and interstate carriers are estimated to be associated with a higher probability of fatal crashes. Drivers with a regular license (Class D) only are at greater risk of being involved in fatal crashes. Special training, experience, skill, and knowledge are efficacious to improving truck safety, and are required for safely operating a truck. The analyzed factors in this study can contribute to lowering injury severity of crashes, and provide guidance for transportation agencies to improve safety.

5.2 Research Limitation and Future Studies

Instead of demonstrating that data mining models are superior to GLMs in all aspects, research tends to promote the application of data mining models in commercial truck safety studies. This study applied a GB data mining model in commercial truck severity analysis and identified the contributor variables, especially trucking company characteristics, through the application of a data mining model to commercial trucking crash data. This study did not examine all data mining models in safety studies. In addition, as different data mining models have different features, they are probably feasible in different types of research. Thus, more data mining models, such as clustering analysis, are recommended to be tested in safety research. In practice, it is feasible that combinations of a few data mining models be used.

Moreover, in this research, crash severity and relationships of several major contributors, especially trucking company characteristics, are discovered. Underlying reasons leading to the observations are assumed and described in the research report; however, they are not analyzed due to data limitation. Further studies are recommended to demonstrate the hypothetical rationales and extend this study to the following: effects of truck configuration at the corporation level, effects of traffic exposure at corporation level, and effects of driver behavior and driving environment awareness at corporation level; and national commercial truck safety analyses are recommended for further research.

REFERENCES

United States Department of Transportation (USDOT). (2013). *Commercial Motor Vehicle Facts- March 2013, Federal Motor Carrier Safety Administration, http://www.fmcsa.dot.gov/documents/facts-research/CMV-Facts.pdf, accessed: Aug. 20, 2013.*

United State General Accounting Office (GAO). (1997). *DOT is shifting to performance-based Standards to Assess Whether Carriers Operate Safely*. GAO/RCED-98-8 Commercial Motor Carriers. Washington, D.C.

Randhawa, S.U., Miller, S.G., Bell, C.A., and Montagne, P.E. (1998). "A Study of Commercial Vehicle Safety Alliance's Out-of-Service Criteria." *Accid. Anal. And Prev*., Vol. 30, No. 1, PP. 61-67.

Mitchell, R., Friswell, R., and Mooren, L. (2012). "Initial Development of a Practical Safety Audit Tool to Assess Fleet Safety Management Practices." *Accident Analysis and Prevention* 47 (2012) 102-118

Schoor, O.V., Niekerk, J.L., and Grobbelaar, B. (2001). "Mechanical Failures as a Contributing Cause to Motor Vehicle Accidents—South Africa." *Accident Analysis and Prevention* 33 (2001) 713-721

Hall, R.W., and Intihar, C. (1997). *Commercial Vehicle Operations: Government Interfaces and Intelligent Transportation Systems.* http://escholarship.org/uc/item/9p272488, *accessed: Aug. 20, 2013.*

Pattern, M.L., Carroll, J.L., and Thomchick, E.A. (1989). "The Efficiency of Roadside Inspections in Reducing Heavy Truck Accidents." *Journal of the Transportation Research Forum*, Vol. 29, No. 2, 1989, p. 269-276.

Lantz, B.M. (1993). *Analysis of Roadside Inspection Data and Its Relationship to Accident and Safety/Compliance Review Data and Reviews of Previous and Ongoing Research in These Areas.* UGPTI Publication no. 95, North Dakota State University/Upper Great Plains Transportation Institute, Fargo, North Dakota.

Britto, R.A., Corsi, T.M., and Grimm, C.M., (2010). "The Relationship between Motor Carrier Financial Performance and Safety Performance." *Transportation Journal,* Vol. 49, No. 4, 2010, pp. 42-51.

Zheng, Z., Lu, P., and Tolliver, D. "Decision Tree Approach to Accident Prediction for Highway-Rail Grade Crossings: Empirical Analysis." *Transportation Research Record*, 2545, 115-122, 2016

Lu, P., and Tolliver, D., "Accident Prediction Model for Public Highway-Rail Grade Crossings." *Accident Analysis & Prevention*, 90, 73-81, 2016

Statista, Road Accidents in the U.S. – Statistics & Facts. https://www.statista.com/topics/3708/road-accidents-in-the-us/ Accessed in June, 2017.

PBS. "Traffic Accidents in the U.S. Cost $871 Billion a Year, Federal Study Finds." Lowy, J., Associated Press. https://www.pbs.org/newshour/nation/motor-vehicle-crashes-u-s-cost-871-billion-year-federal-study-finds Accessed in Jan, 2018

Lemp, J.D., Kockelman,K.M., and Unnikrishnan, A., 2011. "Analysis of Large Truck Crash Severity Using Heteroskedastic Ordered Probit Models." *Accident Analysis and Prevention*. 43, 370-380.

Zhu, X., Srinivasan, S., 2011. "A comprehensive analysis of factors influencing the injury severity of large-truck crashes." *Accident Analysis and Prevention*. 43, 49-57.

U.S. Department of Labor. "Creating a Safety Culture. Safety and Health Program Management." https://www.osha.gov/SLTC/etools/safetyhealth/mod4_factsheets_culture.html. Accessed Jun. 15 2016

Brusilovkey, P. N.A. "Data Mining vs. Statistics." Business Intelligence Solutions. http://www.bisolutions.us/Data-Mining-vs-Statistics.php Accessed in Oct, 2016

Priyadharshini, G., S., 2017. “Data Mining vs. Statistics – How are They Different?” *Simplilearn Solutions*. https://www.simplilearn.com/data-mining-vs-statistics-article Accessed in Oct, 2017.

Young, R.K., and Liesman, J., 2007. “Estimating the relationship between measured wind speed and overturning truck crashes using a binary logit model.” *Accident Analysis and Prevention*. 39, 574-580.

Bai, Y., Yang, Y., and Li, Y., 2015. “Determining the effective location of a portable changeable message sign on reducing the risk of truck-related crashes in work zones.” *Accident Analysis and Prevention*. 83, 197-202.

Curnow, G., 2002. “Australian transport safety bureau heavy truck crash databases: what do the statistics tell us?” Natl. Heavy Veh. Saf. Semin. http://www.ntc.gov.au/filemedia/Publications/ WhatdoStatisticstellusGitaCurnow.pdf

Elisa R. Braver, Paul L. Zador, Denise Thum, Eric L. Mitter, Herbert M. Baum, and Frank J. Vilardo, 1997. “Tractor-Trailer Crashes in Indiana: A Case-control Study of the Role of Truck Configuration.” *Accident Analysis and Prevention*. 29, 79-96.

Anderson, J.E., Govada, M., Steffen, T.K., Thorne, C.P., Varvarigou, V., Kales, S.N., andBurks, S.V., 2012. “Obesity is associated with the future risk of heavy truck crashes among newly recruited commercial drivers.” *Accident Analysis and Prevention*. 49, 378-384.

Khattak, A.J., Schneider, R.J., and Targa, F., 2003. “Risk Factors in Large Truck Rollovers and Injury Severity: Analysis of Single-vehicle Collisions.” Transportation Research Record. TRB 2003 Annual Meeting CD-ROM, Transportation Research Board, National Research Council, Washington D.C. (2003)

Naik, B., Tung, L.W., Zhao, S., and Khattak, A.J., 2016. “Weather impacts on single-vehicle crash injury severity.” *Journal of Safety Research*. 58, 57-65.

Campbell, K.L., 1991. “Fatal accident involvement rates by driver age for large trucks.” *Accident Analysis and Prevention*. 23 (4), 287–295.

Uddin, M., and Huynh, N., 2017, “Truck-involved crashes injury severity analysis for different lighting conditions on rural and urban roadways.” *Accident Analysis and Prevention*. 108 PP 44-55. 2017

Wei, Z., Xiaokun, W., Dapeng, Z., 2017. “Truck crash severity in New York City: an investigation of the spatial and the time of day effects.” *Accident Analysis and Prevention*. 99, 249-261.

Pahukula, J., Hernandez, S., and Unnikrishnan, A., 2015. “A time of day analysis of crashes involving large trucks in urban areas.” *Accident Analysis and Prevention*. 75, 155-163.

Chen, G. 2008. “Impact of federal compliance reviews of trucking companies in reducing highway truck crashes.” *Accident Analysis and Prevention*. 40, 238-245.

Wu, Q., Zhang, G., Chen, C., Tarefder, R., Wang, H., and Wei, H., 2016. “Heterogeneous impacts of gender-interpreted contributing factors on driver injury severities in single-vehicle rollover crashes.” *Accident Analysis and Prevention*. 94. 28-34.

Charbotel, B., Martin, J.L., Gadegbeku, B., and Chiron, M., 2003. “Severity Factors for Truck Drivers’ Injuries.” *American Journal of Epidemiology*. 158.

Kumar, S., and Toshniwai, D., 2015. “A Data Mining Framework to Analysis Road Accident Data.” *Journal of Big Data*. 2:26.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat., 1189–1232

Friedman, J.H., Meulman, J.J., 2003. “Multiple additive regression trees with application in epidemiology.” *Stat. Med.* 22 (9), 1365–1381. http://dx.doi.org/ 10.1002/sim.1501.

Hastie, T., Tibshirani, R.J., Friedman, J.H., 2009. *The Elements of Statistical Learning*, second ed. Springer-Verlag, NY.

Yishih, C., 2013. “Factor complexity of crash occurrence: An empirical demonstration using boosted regression trees.” *Accident Analysis and Prevention*. 61, 107-118.

Saha, D., Alluri, P., and Gan, A., 2015. “Prioritizing Highway Safety Manual’s crash prediction variables using boosted regression trees.” *Accident Analysis and Prevention*. 79, 133-144.

Saha, D., Alluri, P., and Gan, A., 2015. “Prioritizing Highway Safety Manual’s crash prediction variables using boosted regression trees.” *Accident Analysis and Prevention*. 79, 133-144.

De’ath, G., 2007. “Boosted trees for ecological modeling and prediction.” *Ecology* 88 (1), 243–251.

Kockelman, K.M., Ma, J., 2007. “Freeway speeds and speed variations preceding crashes, within and across lanes.” *Journal of the Transportation Research Forum* 46 (1), 43–61.

Chen, C., Zhang, G., Tian, Z., Bogus, S.M., and Yang, Y., 2015. “Hierarchical Bayesian Random Intercept Model-based Cross-level Interaction Decomposition for Truck Driver Injury Severity Investigations.” *Accident Analysis and Prevention*. 85, 186-198.