Transportation Consortium of South-Central States

Solving Emerging Transportation Resiliency, Sustainability, and Economic Challenges through the Use of Innovative Materials and Construction Methods: From Research to Implementation

Investigating Problem of Distracted Drivers on Louisiana Roadways

Project No. 17SALSU10

Lead University: University of Louisiana at Lafayette

Addressing Region 6 Transportation Needs

Final Report

December 2018

Disclaimer

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

Acknowledgments

The assistance received from all the members of Transportation Consortium of South-Central States (Tran-SET) who were involved in the project is gratefully acknowledged. The research team especially acknowledges the assistance of 'Cecil J. Picard Center for Child Development and Lifelong Learning' at the University of Louisiana at Lafayette for their support for providing their software facilities.

TECHNICAL DOCUMENTATION PAGE

1. Project No.	2. Government Accession No.	3. Recipient's Catalog No.
17SALSU10		

4. Title and Subtitle Investigating Problem of Distracted Drivers on Louisiana Roadways	**5. Report Date** Dec. 2018 **6. Performing Organization Code**
7. Author(s) PI: Xiaoduan Sun, Ph.D., P.E. https://orcid.org/0000-0001-7282-1340 GRA: M. Ashifur Rahman https://orcid.org/0000-0001-6940-1599	**8. Performing Organization Report No.**
9. Performing Organization Name and Address Transportation Consortium of South-Central States (Tran-SET) University Transportation Center for Region 6 3319 Patrick F. Taylor Hall, Louisiana State University, Baton Rouge, LA 70803	**10. Work Unit No. (TRAIS)** **11. Contract or Grant No.** 69A3551747106
12. Sponsoring Agency Name and Address United States of America Department of Transportation Research and Innovative Technology Administration	**13. Type of Report and Period Covered** Final Research Report May 2017 – May 2018 **14. Sponsoring Agency Code**

15. Supplementary Notes

Report uploaded and accessible at: Tran-SET's website (http://transet.lsu.edu/)

16. Abstract

Due to the escalating usage of cellphone and social networking, distracted driving is and will remain as one of the most serious problems faced by Departments of Transportation (DOTs) and law enforcement agencies. Under the aim of in-depth investigation of distracted driving crashes in Louisiana, the specific objectives of this study are: (1) reviewing the crash reports for the quality of distracted driving crash reporting, (2) analyzing distracted driving-related crashes through regression model and data mining algorithm to link the severity of distracted driving crashes with the contributing factors collected in crash data, (3) investigating the observable characteristics of distracted driving roadside and video survey, and (4) recommending the countermeasures utilizing the analysis results and reviews.

About 60,000 crashes from ten-year crash data, three types of distracted driving related crashes are modeled: Fatal (K) and Severe (A) Injury; Moderate (B) and Complaint (C) Injury; and Property-damage only (PDO). One statistical method was used for prediction, multinomial logistic regression, and one data mining algorithms was used, random forest. Higher speed limit, curved road, head-on crashes were identified among the key factors. Data mining algorithms performed better in prediction compared to the multinomial logistic regression when sensitivity and specificity were used to compare the predicted results. Fisher's exact tests of roadside manual observation data shows that gender has no significant influence in cellphone distraction (regardless of distraction type), however age can be influential and associated with driver distraction. Association rule mining of observation data shows that the most predominant type of cellphone use is manipulating mainly occurs at intersections, whereas talking is more associated with segments. In-vehicle video data were coded by the software FaceReader, which captures facial expressions of drivers while driving. Initial results do suggest valence in emotion can be attributed to timing before, during, and after cellphone calls and texting. Physical countermeasure development towards reducing the distraction-related crash severity should be targeted at preventing lane departure crashes. Physical countermeasure development towards reducing the distraction-related crash severity should be targeted at preventing lane departure crashes. Strict enforcement of texting ban with awareness campaign are also expected to prevent distracted driving.

17. Key Words	18. Distribution Statement
Distracted Driving; Louisiana; Driving Behavior, Cellphone Distraction	No restrictions.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	46	

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

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

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

APPROXIMATE CONVERSIONS FROM SI UNITS

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

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS, ABBREVIATIONS, AND SYMBOLS

AAA	American Automobile Association
ARM	Association Rule Mining
AADT	Annual Average Daily Traffic
DOT	Department of Transportation
FARS	Fatality Analysis Reporting System
GPS	Geographic Positioning System
LADOTD	Louisiana Department of Transportation and Development
LSU	Louisiana State University
MLR	Multinomial Logistic Regression
NCSA	National Center for Statistics and Analysis
NSC	National Safety Council
NHTSA	National Highway Traffic Safety Administration
NOPUS	National Occupant Protection Use Survey
PDO	Property Damage Only
SHRP2	Second Strategic Highway Research Program
SHSP	Strategic Highway Safety Plan

EXECUTIVE SUMMARY

While ongoing developments of autonomous vehicles show a great promise to reduce fatalities and injuries, full implementation will take years to become a reality. Due to the escalating usage of cellphone and social networking, distracted driving is and will remain one of the most serious problems faced by Departments of Transportation (DOTs) and law enforcement agencies. From the review of state-level, national-level, and other existing guidelines in reporting distraction-related information in a crash, it is expected that a wider range of classification in driver distraction type would be helpful in collecting more accurate distraction information and understanding the relationship between distraction type and severity of crashes. Crash data is still an important resource for identification of factors related to distracted driving. Louisiana is one of the worst states in road safety performance in the United States while distracted driving remains a key source of road crashes in the state.

Under the aim of in-depth investigation of distracted driving crashes in Louisiana, the specific objectives of this study are:

- Reviewing the crash reports for the quality of distracted driving crash reporting.
- Analyzing distracted driving-related crashes through regression model and data mining algorithm to link the severity of distracted driving crashes with the contributing factors collected in crash data.
- Investigating the observable characteristics of distracted driving roadside and video survey.
- Recommending the countermeasures utilizing the analysis results and reviews.

About 60,000 crashes from ten-year crash data, three types of distracted driving related crashes are modeled: Fatal (K) and Severe (A) Injury; Moderate (B) and Complaint (C) Injury; and Property-damage only (PDO). One statistical method was used for prediction, multinomial logistic regression, and one data mining algorithms was used – random forest. Higher speed limit, curved road, head-on crashes were identified among the key factors. Data mining algorithms performed better in prediction compared to the multinomial logistic regression when sensitivity and specificity were used to compare the predicted results. Fisher's exact tests of roadside manual observation data shows that gender has no significant influence in cellphone distraction (regardless of distraction type), however age can be influential and associated with driver distraction. Association rule mining of observation data shows that the most predominant type of cellphone use is manipulating mainly occurs at intersections, whereas talking is more associated with segments. In-vehicle video data were coded by the software FaceReader, which captures facial expressions of drivers while driving. Initial results do suggest valence in emotion can be attributed to timing before, during, and after cellphone calls and texting. Physical countermeasure development towards reducing the distraction-related crash severity should be targeted at preventing lane departure crashes. Physical countermeasure development towards reducing the distraction-related crash severity should be targeted at preventing lane departure crashes. Strict enforcement of texting ban with awareness campaign are also expected to prevent distracted driving.

IMPLEMENTATION STATEMENT

Distracted driving in Louisiana is a serious concern to transportation researchers and the DOT. This study presents findings from a severity prediction analysis with underlying contributing factors, roadside distracted driving observation, in-vehicle video data coded by facial expression capturing software, FaceReader. The prediction of severity models provides insight to researchers and enforcement agencies to identify underlying factors behind distracted driving crashes. The roadside manual observation data does show some interesting relationships between distraction with driver and roadway characteristics. The resulting relationships of cellphone distraction with gender, roadway type, age group, vehicle type can be strongly conclusive with larger sample size. The application of FaceReader has remarkable potential in detecting various types of driver distraction.

The results of this study will be disseminated though graduate and undergraduate class lectures and presentations. The project team will also lecture in high schools to educate young students on the seriousness of distracted driving.

1. INTRODUCTION

Distracted driving is engaging in any activity that diverts attention from the primary task of driving *(1)*. There was substantial concern of safety in the past due to distracted driving activities such as rubbernecking, talking to other passenger(s) in the vehicle, eating, drinking, smoking, and reading, among others. With technological innovations of new gadgets, drivers have been distracted by fiddling with both vehicular and non-vehicular objects – such as stereo, entertainment systems, navigation system, etc. Due to the escalating usage of cellphone and social media in the last two decades, distracted driving has been and will probably remain as one of the most serious problems faced by Departments of Transportation (DOTs) and law enforcement agencies.

The evolution of cellphones from just a source of communication to a mode of various daily activities has added a significant number of distracting elements and has put its user in higher risk of involvement in a crash while driving. People are distracted by cellphones (specifically smartphones) while driving in a number of ways – hand-held or hands-free talking, manipulating via texting, or using other apps – not limited to navigational purposes. According to the latest study by the AAA (American Automobile Association) Foundation for Traffic Safety, an estimated 60.5% of drivers talk on hands-free cellphones, 49.1% talk on hand-held cellphones, 44.9% of drivers read a text message or email while driving, and 34.6% of drivers type or send a text message or email while driving *(2)*. One cellphone service provider shows that 40% of its subscribers with smartphones use social media while driving *(3)*. Moreover, new vehicle models continue to add more features like in-vehicle informant systems, which require more visual and cognitive demands resulting in more distraction from higher interaction time *(4)*.

Distracted driving is likely to significantly affect road safety in upcoming years, even though the ongoing developments of autonomous vehicles show a great promise to reduce fatalities and injuries. Through the gradual progression of automation level in vehicles, incremental reduction of crashes is expected *(5)*. Full automation, removal of human drivers, is expected to reduce crashes from 65% to up to an ambitious 90%. However, the complicated phase with a mixture of automated and manually driven vehicles is yet to come. Furthermore, full implementation will take years to become a reality.

Driver distraction and inattention has been identified as a major influence in traffic crashes. In crash data reporting and analysis, inattention is considered to be one of the distraction modes, which is categorized by activities like drowsiness, daydreaming, etc. Some reports or news articles use the terms inattention and distraction synonymously *(6)*. However, theoretically, driver inattention means insufficient or no attention to activities critical for safe driving, and driver distraction is just one form of driver inattention *(7)*.

Louisiana is one of the worst road safety performers in the United States. In line with the ambitious goal of “Destination Zero Deaths”, the state has addressed “Distracted Driving” as a key emphasis area in its Strategic Highway Safety Plan (SHSP) *(8)*. According to the Crash 1 database, distracted/inattentive driving is considered as a serious contributor and enormous challenge to Louisiana’s highway safety, as distracted driving fatalities represented 20.6% of all fatalities and 36.2% of severe injuries. When distracted/inattentive driving fatalities and severe injury percentages from 2006 to 2010 were compared to the percentages from 2011 to 2015, it was observed that there was no reduction in fatalities (20% in each five-year group) and a 2% increase in severe injuries (35% to 37%) *(8)*. The need to address this issue and implement effective countermeasures is crucial for roadway safety improvement.

The magnitude of Louisiana's distracted driving problem has also been observed from the data collected through driving-related smartphone applications. A smartphone application named 'EverDrive', available both in iPhones and android devices, identified Louisiana as the least safe state in the U.S. regarding distracted driving in its 2016-17 safe driving report. The application typically records cellphone uses in addition to abnormal vehicle movements during driving (e.g., speeding, acceleration, braking, turning *etc.*). It found 43% of drivers participated in at least one distracted driving event in Louisiana *(9)*. The Louisiana Department of Transportation and Development (LADOTD) and other organizations have been reiterating the necessity of preventing crashes resulting from in-vehicle distraction of cellphone use.

With the significant rise in cellphone and online social media usage, substantial research efforts have been placed towards understanding distracted driving related issues in recent years. While distracted-driving-affected crash data analysis provides useful insight to the distraction related contributing factors leading to crashes *(10 – 12)*, the large-scale underreporting of these crashes is a serious concern *(13, 14)*. For example, the Fatality Analysis Reporting System (FARS) documents nationwide fatal crashes documented with detailed information of distraction. However, the FARS crash database suffers from severe underreporting in various states, as identified by the National Safety Council (NSC) *(13)*, even though it doesn't include injury and property damage only (PDO) crashes. Extensive crash report review shows that Louisiana crash database *(15)* lacks details of distraction-related information, but analysis using Louisiana crash data could provide powerful understanding into injury and PDO crashes.

The literature on comprehensive analysis with distracted driving crash data is limited. One study with national sample of 449,049 teenage driver involved crashes in 2003 developed a multinomial logit model to predict the likelihood that a driver will be involved in one of three common crash types: an angular collision with a moving vehicle, a rear-end collision with a moving lead vehicle, and a collision with a fixed object from four driver distraction categories: cognitive, cell phone related, in-vehicle, and passenger-related distractions. The study found a clear influence of distractions on the likelihood of each crash type. Cognitive distractions and passenger-related distractions were found to have increased the likelihood of rear-end collisions even at intersections. Cell phone related distractions increase the likelihood of rear-end collisions when compared to fixed object collisions *(12)*. Another study in Canada also found over-involvement of cellphone distraction with rear-end collisions compared to non-distracted crashes using logistic regression method *(16)*. Another study identified 'distraction' as one of the key contributing factors leading towards novice teenage driver-involved crashes in Connecticut. A total of 260 crash-involved teenage drivers were interviewed in the study. It was found that 23% of at-fault drivers reported being distracted prior to the crash, compared with 3% of not-at-fault drivers, which was a significant difference ($p = 0.002$ from chi-squared test) *(11)*.

The risk of different distracted driving behaviors is widely studied using driving simulators typically through quantification and assessment of driving performance measures such as – lane position, perception-reaction time, speed *etc*. The results obtained, however, may vary substantially in terms of characteristics of the simulators especially the level of realism *(17)*. One Louisiana study used two variables – lane position variability and mean velocity as performance measures during handheld phone conversation, texting, and passenger conversation to respectively represent lateral and longitudinal control of a driving simulator developed in Louisiana State University (LSU). From F-tests, participants demonstrated significantly reduced control for the texting task but not for the handheld phone and passenger conversations. For lateral control, participants demonstrated significantly reduced control for the texting task as well as the passenger conversation task but not for the handheld phone conversation task *(18)*.

Observational studies are more direct investigation of distraction in real life which often include characterizing and drawing inferences on the types, occurrences, and associated characteristics of secondary activities based on a sample of observed distracted drivers. Generally, two types of observational studies are practiced: 1) Naturalistic driving study (NDS) (a research method that involves equipping volunteer participants' vehicles with unobtrusive cameras and instrumentation to record real-world driver behavior and performance), 2) Fixed-site observations (can be performed by using camera installed at roadside *(19) (20)*, or by manual observations *(21)*.

The 100-Car Naturalistic Driving Study was the first large-scale NDS study which was a great resource for transportation research and policies including distraction-related components in crash or near-crash incidents *(22)*. The SHRP2 study used a large and expensive data acquisition system including multiple cameras, accelerometers, vehicle network information, Geographic Positioning System (GPS); onboard computer vision lane tracking, and data storage capability, etc. *(23).* The study indicated that distraction-related activities occurred more frequently in near-crash events *(24)*. Clearly, carrying out a comprehensive NDS study requires a large amount of advanced technological resources.

Manual roadside observation of drivers is perhaps the most conventional yet pragmatic approach which enables exploring distracted driving behaviors in real-world situations. Previous studies have performed statistical assessments of selected distracted driving behaviors through categorization of the distracted drivers and other observable distraction-related traits aiming at identifying the prevalent groups. The National Highway Traffic Safety Administration (NHTSA) performs nationwide roadside surveys of drivers' electronic device use annually with the "probability sample" data on about 50,000 vehicles at about 1,500 intersections. Using complex multistage probability sample, the NHTSA analyzes the percentages of different groups in the driver attributes and compares those attributes between the last two years of survey for three types cellphone/electronic device use (i.e. holding phones to their ears, speaking with visible headsets on, and visibly manipulating hand-held devices). For example, the analysis in 2016 shows that there has been a significant increase in drivers aged 16 to 24 years old speaking with visible headsets on between 2015 and 2016 *(25).* The NHTSA suggests against producing the results state-by-state as they use probability sample and indicates that their results rather provide the best tracking of the extent to which people in the whole United States use cell phones and other electronic devices while driving.

Majority of the roadside distraction observation studies around the world are based on a city or multiple cities. The recent studies attempted to answer different research questions – most commonly association of a multitude of human factor, roadway, and vehicle related variables with cellphone use at intersections. The study by Huemer et al. *(26)* can be referred for a detailed systematic review of observational studies on secondary task engagement while driving. A good number of studies produced contradictory results especially when finding whether gender can be among influential factors for distraction. Logically, the results may vary among locations, which also warrants independent studies in local level.

Apart from the national survey by NHTSA, roadside distraction observation studies are not uncommon at the local level and in other countries. Most of the studies involve cellphone use at intersections to answer specific research questions. For example, in one small-scale French study conducted in traffic signal-controlled intersections, the researchers wanted to observe the tendency of drivers to use cellphones at red lights. The study found that drivers who use cellphones at red lights tend to continue cellphone conversations significantly longer than other visual-manual intersections like texting (chi-squared test, $p < 0.001$, phi = 0.6) *(27)*. Another study in UK found males are more likely to be distracted than females in almost all types of distraction during driving *(28)*. A roadside observation study in Alabama found that the proportion of drivers talking on cellphones was not statistically different across vehicle speeds, however a comparative large portion of vehicles traveling at higher speed (>50 mph) were observed with drivers texting ($p = 0.07$) *(29)*.

Application of laws aimed at reducing cellphone distraction-related crashes varies state by state. Laws restricting the use of cellphones while driving are becoming stricter over time. Bans can be categorized in two types: complete bans of any use of cellphones while driving, and bans particularly focusing on texting while driving. Several states employ stricter bans for newly-licensed drivers or young drivers. Some states have banned cellphone use specifically for school bus drivers. Increasing monetary fines are common for multiple violation offenses. As shown in Figure 1, a ban on cellphone texting for all drivers is the most common law, with an exception in several states *(30)*.

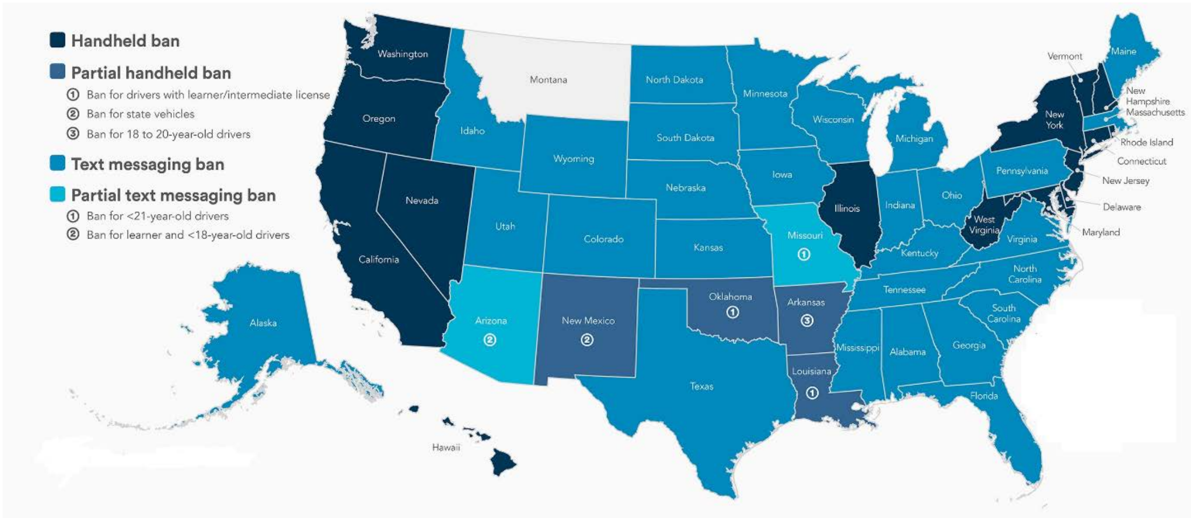

Figure 1. Breakdown of laws against using cellphones while driving (by state) ***(30)*****.**

Several studies estimated the effectiveness of driver cellphone hand-held and texting bans. A study modeled state monthly insurance collision claims per insured vehicle year before/after hand-held cellphone bans in California, Connecticut, New York, and District of Columbia, with 2 or more neighboring control states with the data of 18–33 months before and 12–29 months after bans were effective. The results indicate non-significant small declines in claim rates in California and the District of Columbia relative to control states, and significant small increases in Connecticut and New York *(31)*. A 2013 study used 2000-2010 state-level annual rates of crash deaths per miles traveled, number of drivers in fatal crashes per capita, and number of drivers in fatal crashes for 8 different age groups to model crash measures with all-driver hand-held cellphone ban with primary enforcement status and set of control variables across 50 states and Washington D.C. The control group in the study included states with secondary enforcement cellphone laws. It was found that hand-held bans with primary enforcement are not significantly associated with fatality rates per miles traveled or per capita in the full models but significantly associated with reductions in total number of drivers in fatal crashes and number of drivers in fatal crashes for age groups under 55 *(32)*.

One key research questions in a 49-state (excluding Alaska) study was whether of single-vehicle, single-occupant fatal crash frequency has an association with varying level of texting bans. Strong texting ban status was assumed where texting ban was primary enforcement and for all-drivers, while weak texting ban status was considered where texting ban was a secondary enforcement for all-driver or covering only young drivers. Using 2007-2010 crash data, the study found that the number of single-vehicle, single-occupant fatal crashes was lower (statistically insignificant) in states with strong texting bans vs. states without bans. However, single-vehicle, single-occupant fatal crash counts was significantly higher in states with weak bans vs. states without bans *(33)*

2. OBJECTIVES

This project aims to improve public safety by conducting an in-depth investigation on the scope of the distracted driving problem and providing recommendations to address distracted driving. The investigation studies the scale of the problem in Louisiana, analyzes characteristics of distracted drivers, and how their behaviors affect roadway safety. Under this aim, the specific objectives of the research are:

- Reviewing the crash reports for the quality of distracted driving crash reporting.
- Analyzing distracted driving-related crashes through regression model and data mining algorithm to link the severity of distracted driving crashes with the contributing factors collected in crash data.
- Investigating the observable characteristics of distracted driving roadside and video survey.
- Recommending the countermeasures utilizing the analysis results and reviews.

3. SCOPE

In distracted driving-related crash analysis, the focus was on inattentiveness due to in-vehicle distractions – cellphone, other electronic device, and other in-vehicle sources. In the application of statistical analysis and data mining, the research team decided to exclude the crashes where drivers were supposedly distracted by outside sources due to the ambiguity and insufficient information with regard to exact external sources. Information on about 60,000 crashes with driver-at-fault distracted by in-vehicle distraction sources (cellphone, other electronic device, other source inside the vehicle) with was used for the analysis.

In the roadside observational part, the variables for which information were collected were distraction type, vehicle type, driver's gender, driver's age, presence of passenger(s). A total of 827 distracted drivers from a sample of 3,727 observed drivers were found from the manual data collection of 10 one-hour sessions both at intersections and on segments in both rural and urban area.

To analyze the distracted driving behavior in younger drivers, about 230 minutes of video clips of facial expressions, while voluntary UL Lafayette students were driving, were collected for facial expression capturing software 'FaceReader'. The trips of those students were usually short trips in and around Lafayette city, which were usually limited to and from home, work and school.

4. METHODOLOGY

Methodology has been reported in four major steps. First a review of reporting distracted driving crashes is presented. Then, a Louisiana crash data analysis with statistical and data mining approach will follow. Roadside observation data analysis will be presented next. Lastly in-vehicle driver observation analysis using facial expression through a face capturing software will be presented.

4.1. Reporting Distracted Driving Crashes

Analyzing distraction-related crash reports is the most direct way to measure the impact of distracted driving. Collection of distraction-related information in a crash report is, therefore, particularly significant in assessing the safety impact of different distraction modes. The query of distraction-related crash data collection is twofold:

- Which data is collected? Since, distraction can be generated from multiple sources, how the distraction sources are grouped for the purpose of reporting is significant.
- How accurately is the data collected? The limitations which might cause underreporting need to be identified.

To answer those queries, the research team first investigated the format of reporting distraction-related information from three different sources: (1) Louisiana crash database, (2) FARS database, (3) existing standard (i.e., the Model Minimum Uniform Crash Criteria (MMUCC) Guideline). The comparison of these three formats indicates whether there are any deficiencies in collecting distraction information in a crash and whether it is necessary to collect any particular attributes in crash report.

Secondly, inadequacy in the quantity of distraction-related crashes is also investigated. It is understandable that distraction-related data would be underreported, since in most cases distraction related information can be collected only from driver's statement. It is difficult to identify the scale of underreporting without a thorough investigation. However, from the experience of reading crash reports, the limitations which might cause underreporting can be identified. Inaccuracy in reporting by police can be identified through cross-checks between variables.

4.2. Crash Data Analysis

The research team initially conducted a simple crash analysis of 10 years (2006 – 2015) of Louisiana crash data collected from "Crash 1 Database" *(15)*. Since crashes on non-state roadways lack a significant number of attribute information, crashes on state-controlled highways were selected first. The crash data used for preliminary crash analysis with distracted driver at fault which also included all types of distraction: both inside and outside distraction. The purpose was to identify trends and use it as a basis for further statistical analysis. Although distracted driving-related crashes are presumably underreported, the research team found the available number of crashes is large enough to identify key contributing factors. Data for statistical analysis and mining didn't include external distractions. The basic steps of crash data analysis have been presented in Figure 2.

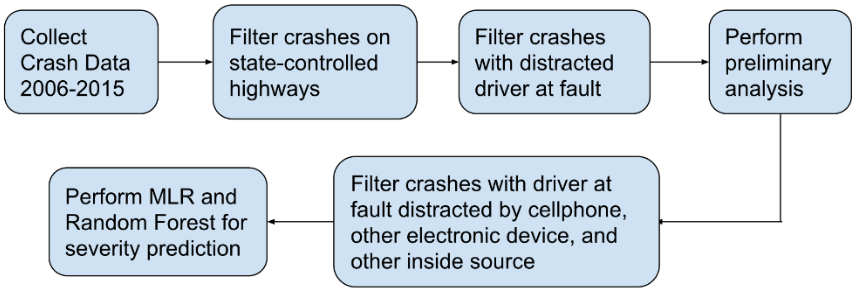

Figure 2. Crash data analysis procedure.

4.2.1. Data for Statistical Analysis and Data Mining

Steps of crash data analysis is presented in Figure 3. Statistical analysis was conducted with a purpose to predict the injury severity from the crash characteristics. The dependent variable (Injury Severity) is nominal with more than two levels: (1) Fatal (K) and Severe (A) Injury, (2) Moderate (B) and Complaint (C) Injury, and (3) Property damage only (PDO). The research team identified that the "Driver Condition - Distracted" does not clearly indicate whether the driver at-fault was distracted or not. For example, 350 drivers in 2015 in Louisiana have been identified as "not distracted", although later they were found to be distracted in corresponding reports. Therefore, the research team relied on the variable 'Driver distracted by'. Three types of distraction which resulted in crashes were considered: cellphone, other electronic device, and other inside (source). The crash description of crashes with external distraction contained unclear information about the source of distraction. Therefore, the research team decided to exclude external distraction-related crashes. Considering the deficiency in recording distraction-related crashes, only the variable "Driver Distracted by" was filtered for those three distraction types for the drivers at fault. Nevertheless, Variable information of a total of 59,919 crashes was obtained, whereas with only the "Driver Condition - Distracted" criteria could only provide 50,878 crashes.

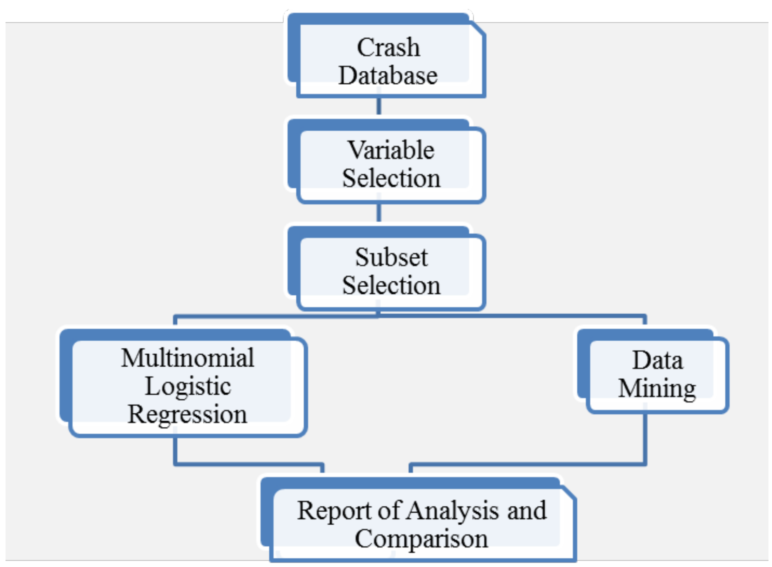

Figure 3. Analysis procedure for in-vehicle distraction-affected crash data.

The variables initially selected for severity prediction are presented in Table 1. The research team selected these variables based on reporting accuracy, possible influence in distractive driving or the crash severity. For example, estimated operating speed is expected to be significantly influential in distracted driving crash severity. However, only 9% (according to 2015 crash data) of the cases, operating speeds are reported accurately. In 91% of the cases, operating speed is either unknown or inaccurately reported. The research team then considered higher posted speed limit as a probable influential variable in absence of accurate and adequate data of operating speed limit. It would also be interesting to see which type of roadway can be dominant in case of distracted driving crash severity. Table 1 also presents the frequency and percentage distribution of variables selected.

The crash data analysis was performed using statistical software R version 3.5.1. For severity prediction by multinomial logistic regression and data mining, the selection of suitable variables was necessary – which was done using the R-package “leaps”. The “leaps” package performs an exhaustive search for the best subsets of the variables, using an efficient branch-and-bound algorithm *(34)*. Out of 16 total initially selected predictor variables, eight were selected for the final subset. The “Injury Severity” was the outcome variable. The annual average daily traffic (AADT) was the only continuous variable, whereas the rest are nominal variables. Figure 4 illustrates the boxplot of the AADT and indicates that the majority of crash locations had an AADT of less than 50,000 per day. All the variables initially considered and then finally selected are presented in Table 1 with percentage of each item under the variables. The AADT was also one of the selected variables in the best subset.

Table 1. Frequency and percentage of each item under the variables.

Variable	Frequency	Percentage
Driver Age (year)[2]		
<25	22,467	37.50
25-34	15,437	25.76
35-44	9,157	15.28
45-54	6,601	11.02
55-64	3,839	6.41
65-74	1,608	2.68
>=75	735	1.23
Unknown	75	0.13
Driver Distracted By		
Cellphone	12,480	20.83
Other Electronic Device	3,102	5.18
Other Inside	44,337	73.99
Driver Gender		
Female	27,208	45.41
Male	32,577	54.37
Unknown	134	0.22
Number of Occupants[1]		
Single	17,113	28.56
Multiple	42,806	71.44
Posted Speed Limit (mph)		
<=30	4,324	7.22
>30 to <=40	15,916	26.56
>40 to <=50	20,626	34.42
>50 to <=60	13,980	23.33
>60 to <=70	3,471	5.79
>70 to <=80	84	0.14
Unknown	1,518	2.53
Vehicle type		
Light Truck	15,735	26.26
Passenger Car	29,687	49.55
SUV	10,526	17.57
Truck/Tractor/Trailer/Bus/Others	1,680	2.80
Van	2,291	3.82
Crash Time		
12am - 6am	3,559	5.94
12pm - 6pm	26,170	43.68
6am - 12pm	14,207	23.71
6pm - 12am	15,983	26.67
Alignment		
Curve-Level	3,718	6.21
Curve-Level-Elevated	535	0.89
Dip, Hump-Curve	16	0.03
Dip, Hump-Straight	41	0.07
Hillcrest-Curve	78	0.13
Hillcrest-Straight	438	0.73
On Grade-Curve	551	0.92
On Grade-Straight	1,217	2.03
Straight-Level	51,605	86.12
Straight-Level-Elevated	1,589	2.65
Unknown other	131	0.22
Lighting Condition		
Continuous Street Light	6,213	10.37
Dark No Street Lights	4,733	7.90
Dawn	532	0.89
Daylight	45,919	76.64
Dusk	976	1.63
Street Light At Intersection	1,317	2.20
Unknown other	229	0.38

Variable	Frequency	Percentage
Collision Type		
Head On	586	0.98
Left-turn Angle	536	0.89
Left-turn Opposite Direction	684	1.14
Left-turn Same Direction	343	0.57
Single Vehicle	8,407	14.03
Rear End	38,877	64.88
Right-turn Opposite Direction	103	0.17
Right-turn Same Direction	237	0.40
Right Angle	3,477	5.80
Sideswipe Opposite Direction	839	1.40
Sideswipe Same Direction	2,576	4.30
Other	3,254	5.43
Intersection		
No	41,834	69.82
Yes	18,085	30.18
Surface Condition		
Contaminant	12	0.02
Dry	53,871	89.91
Ice	38	0.06
Snow/Slush	16	0.03
Unknown other	57	0.10
Wet	5,925	9.89
Weather Condition		
Clear	44,414	74.12
Cloudy	11,219	18.72
Fog/Smoke	321	0.54
Rain	3,842	6.41
Snow Sleet Hail	31	0.05
Unknown other	92	0.15
Highway Type		
Ramp/Exit/Interstate Exit	87	0.15
Rural two-Lane	15,612	26.06
Rural Interstate	1,307	2.18
Rural Multilane	1,111	1.85
Service/Frontage Road	58	0.10
Urban two-Lane	13,956	23.29
Urban Interstate and Freeways	5,425	9.05
Urban Multilane	22,363	37.32
Occupant Protection System		
Child Safety Seat Improperly Used	26	0.04
Child Safety Seat Used	41	0.07
Helmets Used	49	0.08
Lap Belt Only Used	144	0.24
None Used - Vehicle Occupant	1,619	2.70
Restraint Use Unknown	4,606	7.69
Shoulder and Lap Belt Used	52,707	87.96
Shoulder Belt Only Used	727	1.21
Injury Severity		
Fatal and Severe Injury	466	0.78
Moderate and Complaint Injury	21,009	35.06
Property Damage Only	38,444	64.16

[1]The selected variables are depicted in both bold and italics. [2]The variables discarded were only named in italics.

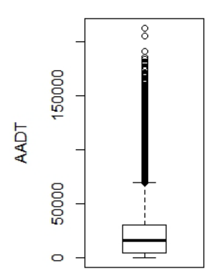

Figure 4. Boxplot of AADT.

4.2.2. Multinomial Logistic Regression

Multinomial logistic regression (MLR) is a regression analysis which has been used in the crash data analysis to describe data and to explain the relationship between one dependent nominal variable with more than two levels or categories (severity of distracted driving crashes – FSI, MCI, PDO) and one or more independent variables (e.g. vehicle type, crash time).

In our analysis, the dependent variable has three categories. The severity prediction will be presented for k = 1 or Fatal and Serious Injury (FSI), and for k = 2 or Moderate and Complaint Injury (MCI) with a reference to k = 3 or Property Damage Only (PDO). The MLR estimates the k-1 log odds for each category (k) with the last category as reference. The regression functions are estimated as:

$$logit\ (y = FSI) = \log(\frac{p(y = FSI)}{1 - p(y = FSI)})$$

$$= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{in}\ for\ i = 1,2, \ldots\ldots, n \qquad [1]$$

$$logit\ (y = MCI) = \log(\frac{p(y = MCI)}{1 - p(y = MCI)})$$

$$= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{in}\ for\ i = 1,2, \ldots\ldots, n \qquad [2]$$

4.2.3. Random Forest Algorithm

Random forest algorithm is a supervised algorithm which can be used both for classification and regression. The random forest algorithm starts with a standard machine learning technique called a “decision tree” which, corresponds to a weak learner. In a decision tree, an input is entered at the top and as it traverses down the tree the data gets bucketed into smaller and smaller sets. The random forest (illustrated *(13)* in Figure 5) takes this notion to the next level by combining trees. This algorithm utilizes bagging (i.e. bootstrap aggregation) to reduce the variance in the model.

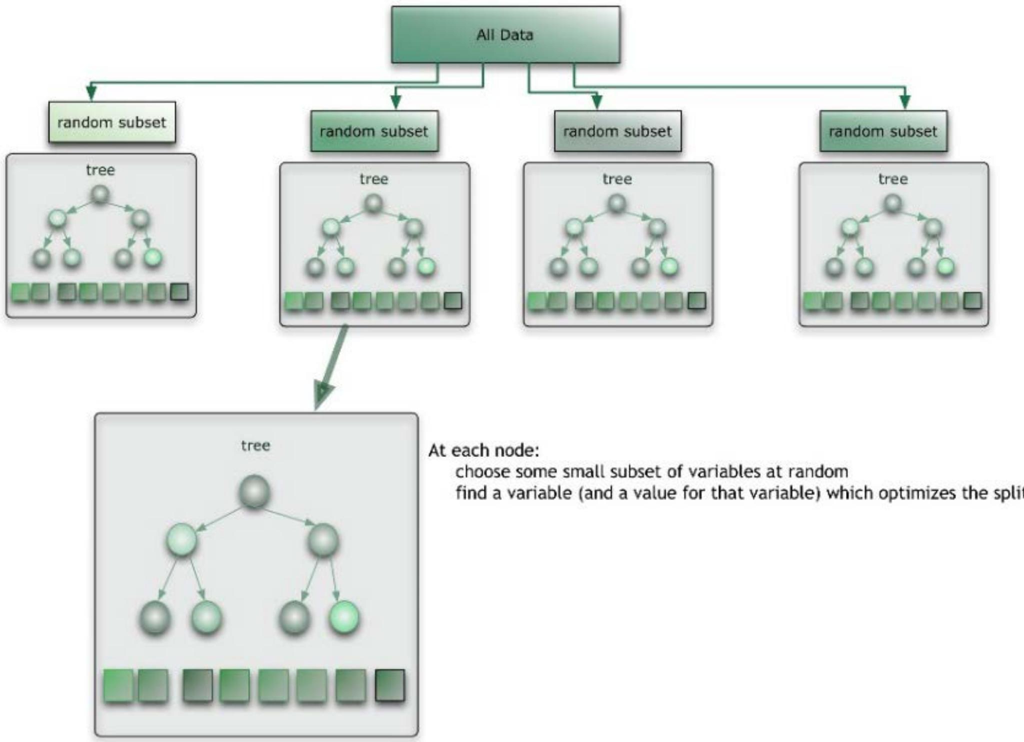

Figure 5. Depiction of random forest algorithm *(13)*.

4.3. Roadside Observation of Driver Electronic Device Use

4.3.1. Observation Sites

Initially, the research team proposed setting up high resolution cameras at selected locations to identify distracted driving behaviors. However, it was discovered high definition cameras fail to identify distracted driving activities through shaded glasses, especially in moving vehicles. Therefore, a decision was made for roadside observations at randomly selected sites in order to achieve a representative population sample and to fulfill the assumptions of Fisher's Exact tests to be used for identifying associations.

The roadside observations were performed in Lafayette Parish in Louisiana. Four observations were conducted at intersections controlled by stop signs or signals and six observations were conducted on segments away from intersections. Signalized intersections were selected with a purpose to capture driving behaviors while stopped in traffic, as drivers are known to be distracted by cellphones at intersection stops. Observations on segments were also done to assess whether drivers take risks to use cellphones while moving. Data was collected both at intersections and on segments in both rural and urban areas. The breakdown of number of sites and observations based on area setting (urban or rural) and road section type (intersection or segment) is:

- three observations at an urban intersection;
- five observations at two urban segments;
- one observation at a rural segment; and

- one observation at a rural intersection.

The urban intersections were four-legged signalized intersections, where the major roads were continuous five-lane highways with a left-turn lane in the middle. The minor roads were two-lane with an additional left-turn lane at the intersections. The rural intersection was three-legged signalized intersection. The major road was two-lane highway, which had a left-turn lane only at the intersection. The minor road leg with two-lanes did not have any additional turning lane at the intersection. However, both the left-turn and right-turn movements were allowed from the minor road.

4.3.2. Observation Procedure

Observation data were collected between October 2017 and March 2018 mainly during weekdays. One observation data collection was made during one weekend. The observations were conducted mainly in the morning and afternoon peak hours. The roadside observation and data collection were performed typically for one hour in 10 one-hour sessions.

Three graduate students and six undergraduate students participated in the observation and collection of the data. The observers positioned themselves in unobtrusive locations to the drivers and collected data from the vehicles for identification whether the drivers were distracted with cellphones. The driver and vehicle information were collected for non-distracted drivers as well.

The observation sites on segments were chosen away from intersections where traffic flow is continuous and is not affected by red lights at nearby intersections. At intersections, the data collection related to distraction, driver, and vehicle from as many vehicles as possible began from the first vehicle stopped at the red light and continued up to the start of the green light. On segments, not all of the flowing vehicles were targeted for observation; rather, information was collected depending on the observers' ease, because of the difficulty involved.

4.3.3. Measures

The following variables were collected during roadside observation:

- Typically, cellphone uses are categorized into these two basic types and hence were recorded during observation. There are many ways a cellphone can distract its user while driving, including working a navigation system or talking or texting on a cell phone. However, considering the "identification time limitation" of roadside observation, two of the most common recognized distractions were observed with following benchmark.
    - *Talking* (either talking on a cell phone by holding the phone up to the user's ear or by holding it between their ear and shoulder, or using headphone/earbuds or phone loudspeaker).
    - *Manipulating* by looking at the screen (manually dialing or manipulating buttons on a cellphone or virtual keypads for texting, initiating or ending a call, using apps for navigation, entertainment, or other purposes, etc.).
- Vehicle Type: The observed vehicles were coded as of four categories: passenger car, SUV/van, light truck, and other. Motorcycles (not bicycles), tractor-trailer trucks, buses, and any vehicles besides those in the first three categories were listed as "Other".

- Driver’s Gender: If the observer could not detect the driver as “Male” or “Female”, the observation will be coded as “Unknown”.
- Driver’s Age: Driver’s age was grouped as “<30 years”, “30-60 years”, and “>60 years”. If the age could not have been detected by the observer, the age was coded as “Unknown”.
- Presence of Passenger(s): The research team also wanted to assess whether the presence of any passenger(s) played a role in the driver’s cellphone distraction. The variable was classified as “Yes” or “No”.

4.3.4. Data and Analysis Methods

This study applied Fisher’s exact tests, logistic regression, and association rule mining algorithms to explore associations of distraction types with observed driver, vehicle, and roadway characteristics. Fisher’s exact tests describes the qualitative association between observed variables and two types of observed cellphone use. Association rule mining is a data mining algorithm which is used to find frequent co-occurring associations among a collection of items. This algorithm was utilized to find associations rules involving three distracted cellphone uses (talking and manipulating) with multiple associations of traits, like area type (urban/rural), road section type (segment/intersection), driver gender, driver age, number of occupants, and vehicle type. Finally, Logistic Regression was used to identify significance of observed characteristics with regard to cellphone use type.

Initially, 827 observed drivers were found to be distracted by cellphone, 22.2% of the total observed drivers (3,727). Table 2 presents the percentages of observable or known characteristics/items of each variable by cellphone use type with regard to total observed drivers. For simplicity in analysis, a dataset of 825 observations was used excluding only two observations with unknown gender and unknown age group.

Figure 6 illustrates the Relative frequency of each item within the variable. The obtained sample size of observed distracted-driving is relatively large in urban setting than in rural setting, although collected data at intersection and on segment are relatively similar. More drivers were involved in manipulating than talking on the phone. Majority of the distracted drivers had no passengers in their vehicles. Relative frequency of older driver (>60 years) in the variety of age groups is small compared to younger (<30 years) and middle-aged drivers (30-60 years).

Table 2. Percentages of each items of observed variables by cellphone use type.

Variable	Talking	Manipulating	Percentage Distracted (frequency)
Setting			
Urban	8.2	12.0	20.3[a] (755)
Rural	1.1	0.8	1.9 (72)
Cross-section			
Intersection	4.0	6.2	10.2 (379)
Segment	5.3	6.7	12 (448)
Driver gender			
Male	4.3	6.1	10.4 (388)
Female	5.0	6.7	11.7 (437)
Age group			
<30y	4.0	5.6	9.6 (358)
30-60y	5.1	6.8	11.9 (445)
>60y	0.2	0.4	0.6 (23)
Vehicle type			
Car	4.2	6.4	10.6 (395)
Truck	1.7	2.1	3.8 (142)
SUV/Van	3.2	4.3	7.5 (281)
Other	0.007	0.004	0.011 (9)
Passenger presence			
Yes	0.8	1.9	2.7 (101)
No	8.6	10.9	19.5 (726)
Total	9.3[b]	12.9[b]	22.2 (827)

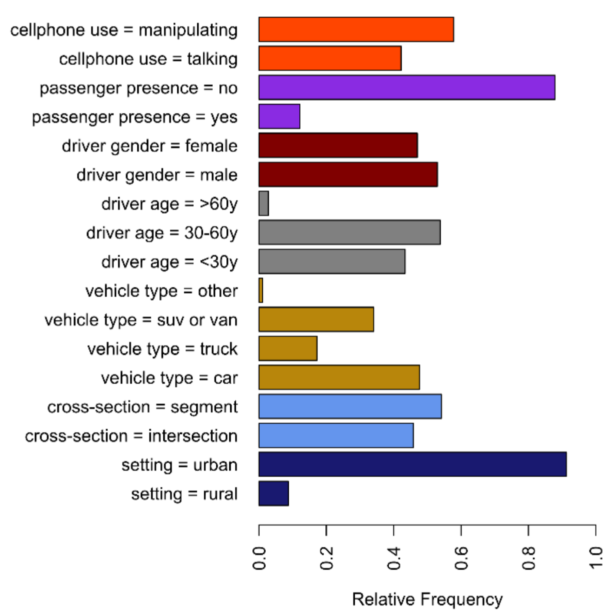

Figure 6. Relative frequency of the items observed.

Fisher's Exact Tests: Fisher's exact test, proposed by Ronald Fisher *(35)*, assesses the null hypothesis of independence applying hypergeometric distribution of the numbers in the cells of contingency tables formed from the observed data to determine nonrandom associations between two categorical variables – in this study, the frequency of cellphone distraction type (talking and manipulating) and variables (like driver's age group, driver's gender, vehicle type, and number of occupants, etc.). The chi-squared test, most commonly used for finding associations of categorical variables, relies on an approximation of sample distribution which works well for large sample sizes. According to the most popular thumb rule, the approximation in Chi-squared test becomes inadequate when more than 20% of cells have expected frequencies < 5. Fisher's exact test was chosen over the chi-squared test to overcome the inadequacy of applying approximation in very small frequencies in some data. Fisher's exact test is popularly used for small samples in 2 × 2 contingency tables, but also works well for contingency tables of larger sizes *(36, 37)*. Fisher's exact test was discouraged due to its large computational demand in earlier years, however multiple studies have argued that feasibility of Fisher's exact test with large sample size isn't a case of computing power in modern age *(38, 39)*.

If two categorical variables X and Y have m and n observed states, respectively Now form an $m \times n$ matrix in which the entries a_{ij} represent the number of observations in which $x = i$ and $y = j$. The row and column sums are R_i and C_j, respectively, and the total sum is:

$$\mathrm{N} = \sum_{\mathrm{i}} \mathrm{R_i} = \sum_{\mathrm{j}} \mathrm{C_j} \qquad [3]$$

of the matrix. The conditional two-tailed probability of getting the actual matric given the particular row and column sums, given by:

$$\mathrm{P_{Cutoff}} = \frac{(\sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{R_i}!)(\sum_{\mathrm{j=1}}^{\mathrm{n}} \mathrm{C_j}!)}{\mathrm{N}! \prod_{\mathrm{i,j}} \mathrm{a_{ij}}!} = \frac{(\mathrm{R_1}!\mathrm{R_2}!\ldots\mathrm{R_m}!)(\mathrm{C_1}!\mathrm{C_2}!\ldots\mathrm{C_n}!)}{\mathrm{N}! \prod_{\mathrm{i,j}} \mathrm{a_{ij}}!} \qquad [3]$$

which is a multivariate generalization of the hypergeometric probability function. Now all possible matrices of nonnegative integers consistent with the row and column sums R_i and C_j can be found. For each one, the associated conditional probability can be calculated using equation (2), where the sum of these probabilities must be 1. In line with previous studies, a cutoff p-value of less than 0.05 was considered statistically significant for the inference of any association results. The 'rcompanion' package *(40)* of 'R' (version 3.5.1) statistical software *(41)* was used for estimating p-values in Fisher's exact test. Specific argument in R software allows estimating non-simulated p-values in Fisher's exact test for contingency tables larger than 2 × 2 even with large frequency.

Association Rule Mining: Three measurements are commonly used to quantify the association rules:

Support: Support is an indication of frequency of a combination of items in the dataset. If X is a combination of variable items (area type, road section type, driver gender, driver age, number of occupants, vehicle type) and Y is the targeted item (in our case cellphone use type), X→Y an association rule and V is a complete observation in a dataset – the support of X i.e. S(X), with regard to observation V is defined by the proportion of observations 'v' the dataset which contains the combination of items X.

$$Supp(X) = \frac{|\{v \in V; X \subseteq v\}|}{|V|} \qquad [4]$$

Confidence: Confidence is a measure of how often the rule, X→Y, is true in the dataset, i.e. how often each item in Y appears in observations that contain X.

$$Conf(\mathrm{X} \rightarrow \mathrm{Y}) = \frac{S(X \cup Y)}{S(X)} \qquad [5]$$

Lift: The lift of a rule X→Y is the confidence of the rule divided by the expected confidence, assuming X and Y are independent.

$$Lift(\mathrm{X} \rightarrow \mathrm{Y}) = \frac{S(X \cup Y)}{S(X) * S(Y)} \qquad [6]$$

A lift value greater than 1 is an indication that X and Y appear more often together than expected. This can be restated as – the occurrence of X has a positive effect on the occurrence of Y or that X is positively correlated with Y. A lift value smaller than 1 indicates that X and Y appear less often together than expected, and therefore, X is negatively correlated with Y. A lift value near 1 indicates that X and Y appear almost as often together as expected; this means that the occurrence of X has almost no effect on the occurrence of Y or that X and Y have zero correlation.

The Apriori algorithm of Association rules mining (ARM) follows a breadth-first search and was used to find out the key antecedents (X) for the different types of cellphone use as consequent (Y). Optimization of support and confidence is a key issue in generating unique rules. Using combination of optimized values of support and confidence enable the researchers to avoid generating either less frequent or replicated rules. Insignificant rules (either less frequent or replicated) with respect to more general rules, which exist only in the presence of high confidence can be pruned using specific R software code.

Logistic Regression: Logistic regression is a method for modeling when the outcome usually expressed by a binary response variable. Predictor variables can be numerical or categorical (including binary). In order to identify significant characteristics between talking on cellphone and manipulating, 825 observed distracted driving data were used. The binary response variable in the model, Y can be denoted as 1 for cellphone manipulating; it can be denoted as 0 for talking. Typically, 1 denotes “yes” or “true”, and 0 denotes “no” or “false” in dichotomous response.

If Y is the binary response variable, it is assumed that P(Y=1) is possibly dependent on X, where X is a vector of predictor values. In logistic regression, the purpose is to model:

$$p(X) \equiv p(Y = 1|X) \qquad [7]$$

$p(X)$ is modeled as a linear function of predictor variables:

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \qquad [8]$$

Then the fitted model can result in estimated probabilities outside of [0,1]. Therefore, it is better is to assume that:

$$p(X) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n)} \qquad [9]$$

where, $x_{1,} x_{2,} \dots \dots, x_n$ may be the original set of explanatory or contributing variables.

Therefore:

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \qquad [10]$$

$\log\left(\frac{p(X)}{1-p(X)}\right)$ is called the logit. The estimate of $p(X)$ is between 0 and 1, irrespective of the value of $\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_n x_n$. The unknown parameters (the coefficients, $\beta_0, \beta_1, \beta_2 \dots, \beta_n$) are typically estimated by maximizing likelihood estimation.

4.4. In-Vehicle Observation of Electronic Device Use

4.4.1. Use of FaceReader Software

A small-scale naturalistic driving study was conducted in this study where a face capture software was used for analyzing simple unobtrusive video recorded data. Face capture is a facial recognition technology which represents the process of “netting” a person’s facial expression. It converts expressions into a digital form and recognizes gaze direction, head orientation, mouth, and head orientation (open or closed) and measures valences of several emotions. It usually presents distribution basic emotions derived facial expression graphically in the form of a pie or bar chart.

FaceReader software, a specific type of face capture software, allows the user to apply stimuli at any time and records the occurrence (applying stimuli) for any set duration. FaceReader is utilized in various areas of research, e.g. consumer behavior, psychology, human computer interaction, etc. *(42)*. The FaceReader software evaluates frequency and duration of facial expression in response to stimuli. Details and efficient facial expression analysis is enabled by FaceReader which are coded to determine characteristics relevant to a research model. Seven types of facial expressions are captured by FaceReader: neutral, happy, sad, angry, surprised, scared, and disgusted. Valence of each emotion are estimated in every tenth of a second through detection of facial expression. Figure 7 shows the distribution of all seven emotions of a driver, whereas blue circles and thick lines are the time when the driver was observed to be distracted by cellphone usage for a very shot and relatively longer period of time respectively. Figure 8 shows an example of FaceReader interface when it is used for estimating the valence of emotions along with the playback of the videos.

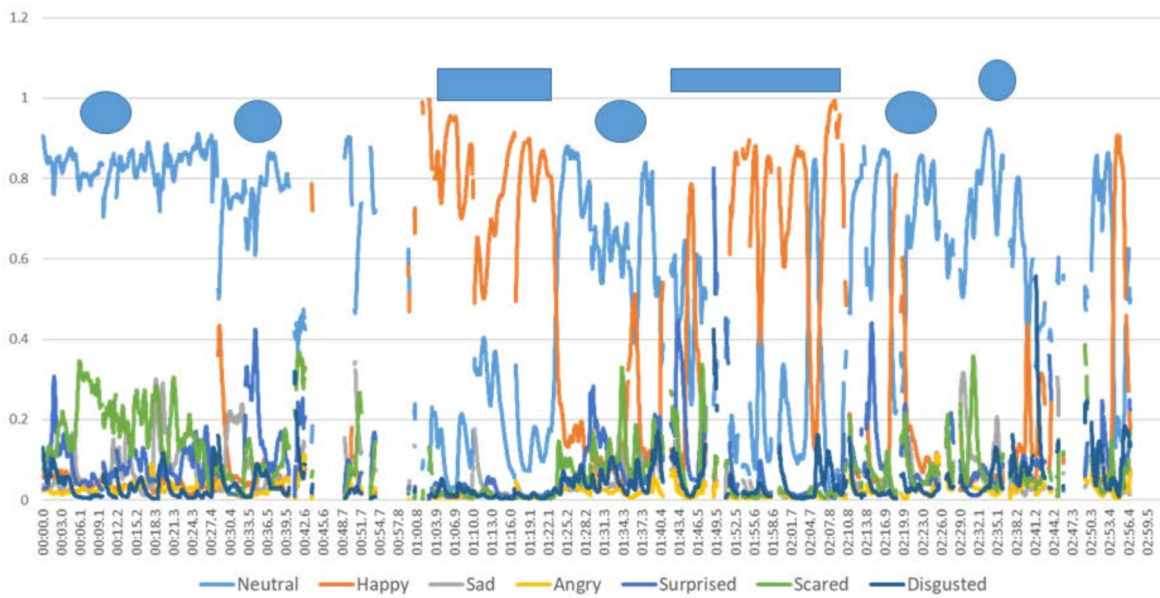

Figure 7. Distribution of all seven emotions of a driver while using cellphone.

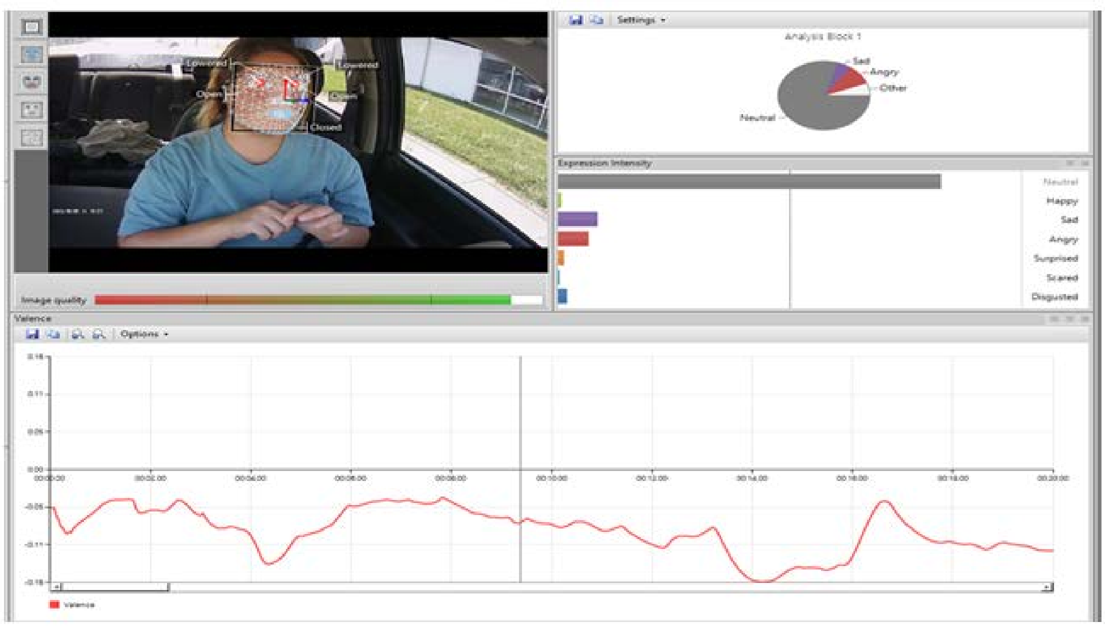

Figure 8. FaceReader interface while encoding a video.

4.4.2. Study Design

A total of 40 college-aged undergraduate students from the Department of Civil Engineering and Psychology at the University of Louisiana at Lafayette voluntarily participated in this study. The trips of those students were usually short trips in and around Lafayette city, which were usually limited to and from home, work and school. Dash cameras were placed in each students' cars for a 24 -72 hour period, and snapshots were obtained of driving-related behavior in three-minute blocks of time. Average duration of clips per participant were 9 minutes. To analyze the distracted driving behavior in younger drivers, about 230 minutes of video clips of facial expressions, while voluntary UL Lafayette students were driving, were found ideal for facial expression capturing software ‘FaceReader’. Running those captured videos, the research team obtained and recorded FaceReader measures (i.e. valence of emotions).

Two research questions were asked:

- Is there a difference in human emotions before, during, and after a cellphone call received while driving?
- Can cellphone distractions be recognized using the observed emotions.

5. FINDINGS

5.1. Review of Reporting Distracted Driving Crashes

5.1.1. Distraction-Related Information in Crash Report

The Model Minimum Uniform Crash Criteria (MMUCC) Guideline, an NHTSA initiative, suggests a minimum, model set of variables (data elements) to describe a motor vehicle crash. Their aim is to provide uniformity in generating the information necessary to describe motor vehicle crashes nationally. According to the 4^{th} and 5^{th} edition of MMUCC, the distraction-related attribute in a crash report "Driver Distracted By" should be collected in two subfields – distractive action taken by the driver and the source of distraction *(43, 44)*. The rationale is to mitigate the effects of distracting activities through identification of specific distracted driving behavior and the source of distraction during a crash. In the prior editions of MMUCC, distracted driving attribute was limited to the type of electronic device source, and whether the source was external or internal *(45)*.

The Fatality Analysis Reporting System (FARS), the nationwide fatal crash census, collects fatal crash information at the national level using a designated form. The FARS system collects a wide array of data for distracted driving attributes, including inattention and carelessness of the driver. It expands the internal and external distractions into several more types for a clearer description of the crash. It strictly follows NHTSA guidelines to identify whether driver's behavior should be counted as distracted or not. For example, driving while daydreaming or lost in thought is identified as distracted driving. However, physical conditions/impairments (fatigue, alcohol, medical condition, etc.) or psychological states (anger, emotional, depressed, etc.) are not identified as distractions *(46)*.

The state of Louisiana Uniform Motor Vehicle Traffic Crash Report was last issued in 2005. The standard form lists values related to the attribute "Driver Distracted By" which are similar to the MMUCC (prior to 4^{th} edition) guideline. Table 3 compares the "Driver Distracted By" attribute and its classifications for the abovementioned three guidelines. The expansion of the distraction-related variable in a crash report through addition of items over time, guidelines for categorizing the distraction by action and source, the disparity in the range of the components of the same information – all indicate that deficiency exists in distraction-related crash data collection.

Table 3. Distraction-related attributes in various crash databases and reports.

Louisiana Crash Report (2005)	MMUCC (3rd edition, 2008)	MMUCC (4th and 5th edition, 2012-17)	FARS Database (2017)
• Cellphone • Other Electronic Device (pager, palm pilot, navigation device, etc.) • Other Inside the Vehicle • Other Outside the vehicle • Not Distracted • Unknown	• Not Distracted • Electronic Communication Device • Other Electronic Device (navigation device, DVD player, etc.) • Other Inside the Vehicle • External Distraction (outside the vehicle) • Unknown	*Action* • Not Distracted • Talking/listening • Manually Operating (texting, dialing, playing game, etc.) Other Inside the Vehicle • Other Action (looking away from task, etc.) • Unknown *Source* • Hands-Free Mobile Phone • Hand-Held Mobile Phone • Other Electronic Device • Vehicle-Integrated Device • Passenger/Other Non-Motorist • External (to vehicle/non-motorist area) • Other Distraction (animal, food, grooming) • Not Applicable (Not Distracted) • Unknown	• Not Distracted • Looked But Did Not See • No Driver Present / Unknown if Driver Present • Not Reported • By Other Occupant(s) • By a Moving Object in Vehicle • While Talking or Listening to Cellular Phone • While Manipulating Cellular Phone • Adjusting Audio or Climate Controls • While Using Other Component/Controls Integral to Vehicle • While Using or Reaching For Device/Object Brought Into Vehicle • Distracted by Outside Person, Object or Event • Eating or Drinking • Smoking Related • Other Cellular Phone Related • Distraction/Inattention • Distraction/Careless • Careless/Inattentive • Distraction (Distracted), Details Unknown • Inattention (Inattentive), Details Unknown • Lost in Thought / Day Dreaming • Other Distraction • Unknown if Distracted

5.1.2. Quality of Reporting

It is well-known that distracted driving is under-reported, although the scale of underreporting is unascertained. The National Safety Council (NSC) indicates substantial underreporting in cellphone-affected fatal crashes according to their review of the national fatal crash data (FARS) of three years. The assumed large underreporting is often attributed to the driver's acknowledgement being the primary source of recoding distraction information in a crash report. The NSC presents a hypothetical depiction (Figure 9) and claims cell-phone crash underreporting is unavoidable *(13)*.

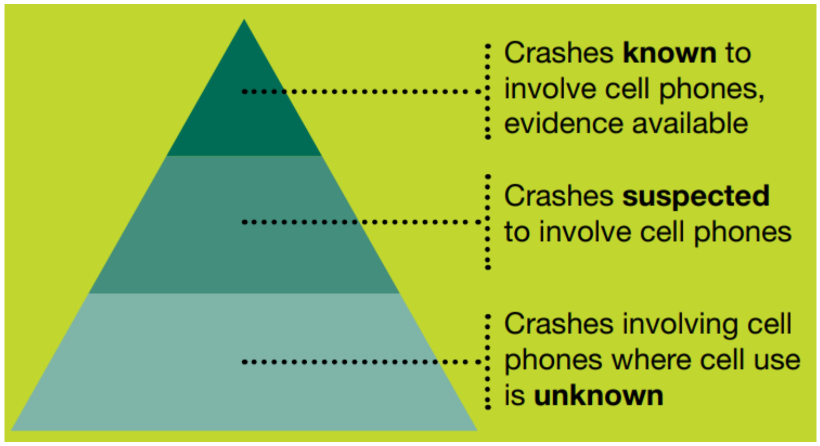

Figure 9. Hypothetical depiction of underreporting of distracted driving crashes *(13)*.

Figure 9 shows the distraction and cellphone-affected fatal crashes at the national level during 2010-15 from NHTSA data, in which cellphone-affected crashes are claimed to be greatly under-reported by NSC. The difficulties in obtaining distracted driving crash data has also been supported in the Louisiana SHSP *(8)*.

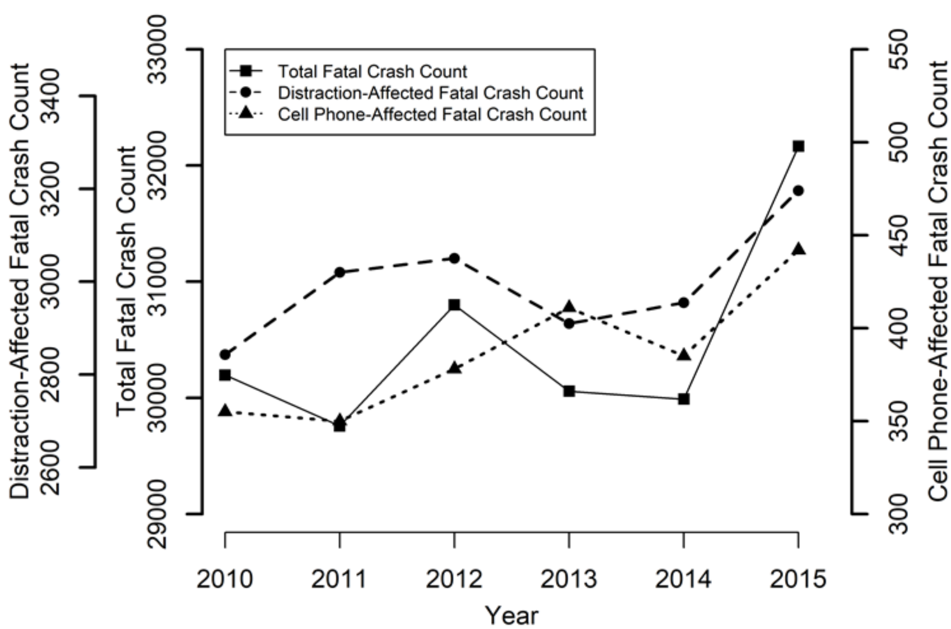

Figure 10. Distraction and Cellphone-affected fatal crashes in comparison with total fatal crashes in last six years in the US *(25)*.

Several aspects can be mentioned from the extensive review of distraction related crash report:

- Cellphone records are thoroughly checked in the cases of fatal and severe injury crashes for possibility of cellphone distraction.
- In the cases of crashes resulting in moderate to no injury, drivers' and witnesses' statement may often be considered as main source by the assigned police officer.

- The driver at fault often claims to be distracted by necessary smartphone apps, specifically navigational apps, which might require additional verification.
- Coding error is another issue which should be addressed. For example, 350 drivers in 2015 in Louisiana have been identified as "not distracted", although later they were found to be distracted in corresponding reports.

5.2. Crash Analysis

Preliminary analysis from the crashes recorded by the police over the last decade (2006-2015 including external crashes) show a general increasing pattern, and an overall increase of 36.4%. Distraction-affected crashes (including external crashes) were recorded about 2.5 times more at non-intersection segments than at intersections. Both intersection and non-intersection distracted driving crashes have increasing trends, which can be seen in Figure 11.

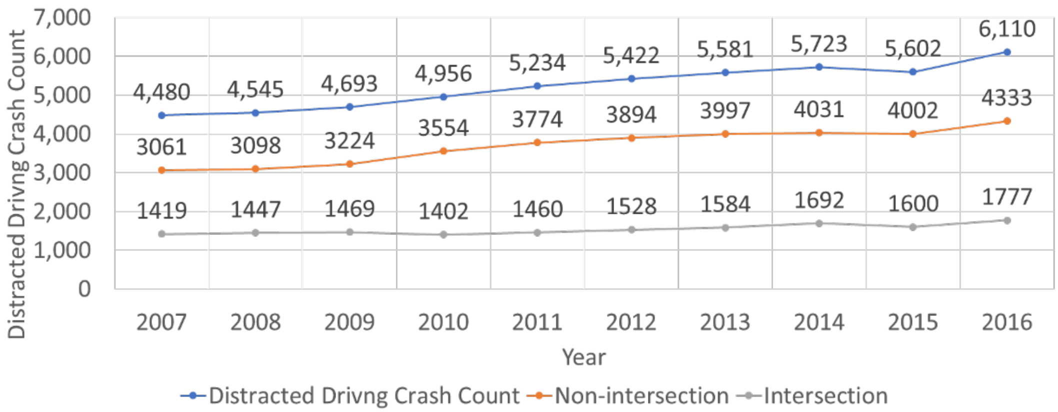

Figure 11. Louisiana crashes where driver condition was recorded as "Distracted".

Figure 12 shows distraction-related crashes by severity. Both injury and PDO crashes due to driver distraction are increasing over the last 10 years, by 28.5% and 41%, respectively. Distraction-related fatal crashes are random events.

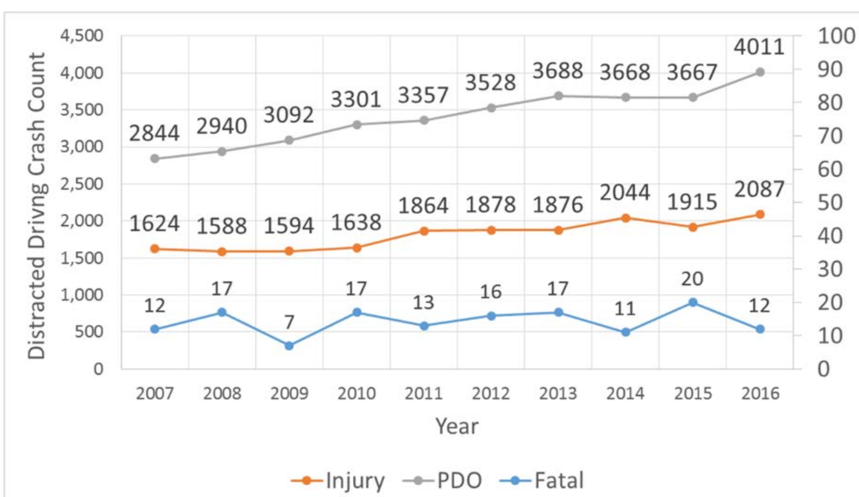

Figure 12. Crashes by severity where driver condition was recorded as "Distracted".

Rear-end crashes were the vast majority among all police-recorded distraction-affected crashes. 27% of those crashes occurred at intersections, whereas 73% occurred at non-intersection segments. Figure 13 illustrates the share of distraction-related crashes by collision type.

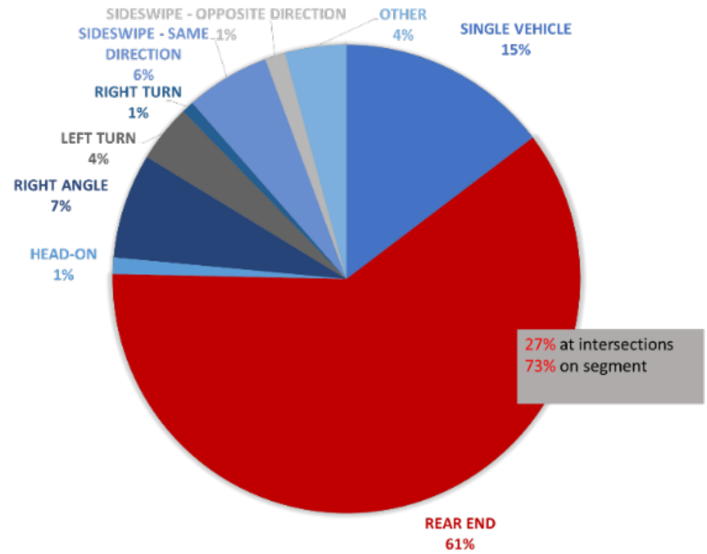

Figure 13. Manner of collision distribution of Louisiana crashes during 2007-2016 where driver condition was recorded as "Distracted".

Figure 14 depicts the crashes by hour. It is interesting to note that big share of distraction-related crashes occurred during the afternoon period.

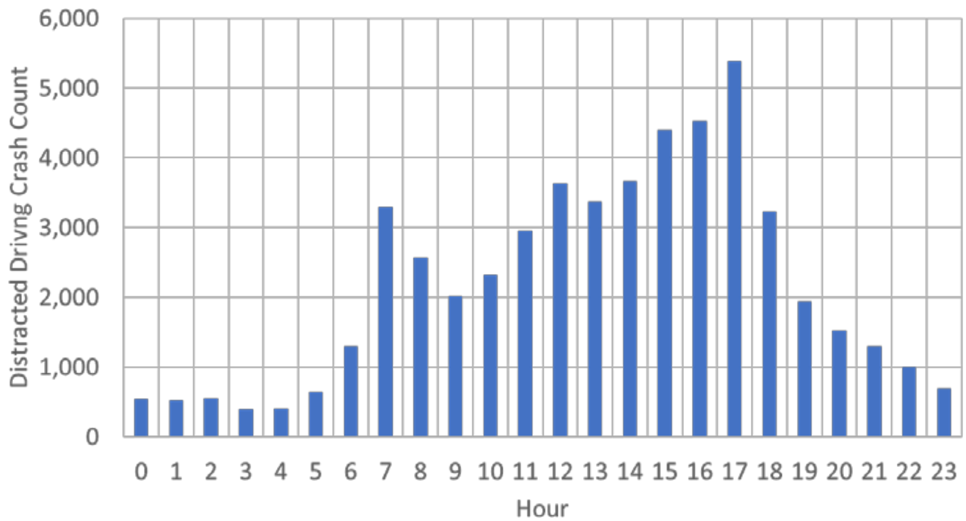

Figure 14. Hourly distribution of Louisiana crashes during 2007-2016 where driver condition was recorded as "Distracted".

5.2.1. Multinomial Logistic Regression Analysis Results

The results of multinomial logistic regression model are presented in Table 4. The coefficient values can be interpreted as increase or decrease of one unit compared to the base, property damage only crashes. For example, the multinomial logit estimate comparing single occupants to multiple occupants is estimated for fatal and serious injury (FSI) crashes or moderate and complaint injury (MCI) crashes relative to property damage only (PDO) crashes given the other variables in the model are held constant. The multinomial logit for single occupants relative to multiple occupants is 0.551 unit lower for being in fatal and serious injury crash to property damage only crash given all other predictor variables in the model are held constant. Similarly, the multinomial logit for single occupants relative to multiple occupants is 0.231 unit lower for being in MCI crash to PDO crash given all other predictor variables in the model are held constant.

For FSI crashes to PDO crashes, the z test statistic for the predictor science (-0.551/1.66e-11) = -3.31e+10 is with an associated p-value of <2e-16. With $\alpha = 0.05$, we reject the null hypothesis and conclude that the difference between single occupant and multiple occupant has been found to be statistically significant for FSI relative to PDO crashes given that the rest of the variable are in the model. Similar conclusions can be made for MCI crashes with regard to number of occupants.

Table 4. Multinomial logistic regression results.

	Fatal and	Severe	Injury	Crashes	Moderate and	Complaint	Injury	Crashes
	Coefficient	Std. Error	z-value	P-value	Coefficient	Std. Error	z-value	P-value
(Intercept)	-20.056	2.62e-11	-7.67e+11	<2e-16	-0.083	4.81e-12	-1.73e+10	<2e-16
Number of Occupants (Ref: Multiple)								
Single	-0.551	1.66e-11	-3.31e+10	<2e-16	-0.231	3.65e-12	-6.33e+10	<2e-16
Posted Speed Limit (Ref: <=30)								
>30 to <=40	-0.005	4.21e-12	-1.15e+09	<2e-16	0.193	1.39e-12	1.39e+11	<2e-16
>40 to <=50	0.297	7.07e-12	4.20e+10	<2e-16	0.326	1.81e-12	1.80e+11	<2e-16
>50 to <=60	0.447	7.24e-12	6.17e+10	<2e-16	0.600	1.32e-12	4.56e+11	<2e-16
>60 to <=70	1.164	5.49e-12	2.12e+11	<2e-16	0.658	1.63e-12	4.04e+11	<2e-16
>70 to <=80	0.381	5.52e-14	6.91e+12	<2e-16	0.471	1.41e-14	3.34e+13	<2e-16
Unknown	0.416	7.42e-13	5.60e+11	<2e-16	0.197	1.06e-13	1.86e+12	<2e-16
Vehicle Type (Ref: Pickup Truck)								
Passenger Car	-0.029	1.26e-11	-2.31e+09	<2e-16	-0.054	2.49e-12	-2.18e+10	<2e-16
SUV	0.132	4.80e-12	2.75e+10	<2e-16	-0.025	8.97e-13	-2.82e+10	<2e-16
Van	0.445	1.37e-12	3.26e+11	<2e-16	0.017	2.30e-13	7.39e+10	<2e-16
Truck/Tractor/Trailer/Bus/Others	0.825	1.89e-12	4.37e+11	<2e-16	0.157	6.64e-13	2.37e+11	<2e-16
Crash Time (Ref: 12am – 6am)								
6am - 12pm	-0.835	5.48e-12	-1.52e+11	<2e-16	-0.113	1.26e-12	-8.98e+10	<2e-16
12pm - 6pm	-0.818	8.88e-12	-9.21e+10	<2e-16	-0.046	2.21e-12	-2.09e+10	<2e-16
6pm - 12am	-0.572	7.78e-12	-7.35e+10	<2e-16	-0.011	1.36e-12	-8.30e+09	<2e-16
Alignment (Ref: Curve-Level)								
Straight-Level	-0.088	2.01e-11	-4.38e+09	<2e-16	-0.056	4.10e-12	-1.36e+10	<2e-16
Straight-Level-Elevated	0.619	2.63e-12	2.35e+11	<2e-16	0.147	5.81e-13	2.53e+11	<2e-16
Curve-Level-Elevated	-0.368	3.02e-13	-1.22e+12	<2e-16	-0.139	5.19e-14	-2.68e+12	<2e-16
On Grade-Straight	-0.548	4.51e-13	-1.22e+12	<2e-16	0.208	2.22e-13	9.39e+11	<2e-16
On Grade-Curve	-0.664	1.28e-13	-5.18e+12	<2e-16	0.204	5.11e-14	4.00e+12	<2e-16
Hillcrest-Straight	1.634	1.16e-12	1.40e+12	<2e-16	0.320	4.12e-13	7.76e+11	<2e-16
Hillcrest-Curve	-18.486	3.22e-22	-5.74e+22	<2e-16	0.091	1.24e-14	7.30e+12	<2e-16
Dip, Hump-Straight	-11.464	3.27e-19	-3.51e+19	<2e-16	-0.571	6.13e-15	-9.32e+13	<2e-16
Dip, Hump-Curve	-5.647	1.47e-17	-3.85e+17	<2e-16	0.712	2.73e-15	2.61e+14	<2e-16
Unknown/Other	-28.881	4.83e-26	-5.98e+26	<2e-16	-0.072	3.32e-14	-2.18e+12	<2e-16
Collision Type (Ref: Head On)								
Single Vehicle	-1.853	4.62e-12	-4.01e+11	<2e-16	-0.829	1.36e-12	-6.11e+11	<2e-16
Rear End	-2.270	1.42e-11	-1.60e+11	<2e-16	-1.003	4.32e-12	-2.32e+11	<2e-16
Right Angle	-1.062	2.05e-12	-5.18e+11	<2e-16	-0.354	5.45e-13	-6.49e+11	<2e-16
Left-turn Angle	-1.002	5.07e-13	-1.98e+12	<2e-16	-0.626	1.15e-13	-5.44e+12	<2e-16
Left-turn Opposite Direction	-1.375	2.22e-13	-6.21e+12	<2e-16	-1.691	2.10e-14	-8.04e+13	<2e-16
Left-turn Same Direction	-1.575	2.31e-13	-6.81e+12	<2e-16	-0.923	3.92e-14	-2.36e+13	<2e-16
Right-turn Same Direction	-0.014	6.97e-13	-2.00e+10	<2e-16	-1.563	1.31e-13	-1.20e+13	<2e-16
Right-turn Opposite Direction	-34.713	1.21e-28	-2.87e+29	<2e-16	-0.751	1.08e-14	-6.94e+13	<2e-16
Sideswipe Same Direction	-3.008	1.01e-12	-2.98e+12	<2e-16	-1.914	2.57e-13	-7.44e+12	<2e-16
Sideswipe Opposite Direction	-2.818	6.25e-14	-4.51e+13	<2e-16	-1.058	4.71e-14	-2.25e+13	<2e-16
Other	-1.687	1.90e-12	-8.90e+11	<2e-16	-1.046	3.97e-13	-2.64e+12	<2e-16
Surface Condition (Ref: Contaminant)								
Dry	3.930	2.44e-11	1.61e+11	<2e-16	0.805	4.39e-12	1.83e+11	<2e-16
Wet	3.519	1.71e-12	2.05e+12	<2e-16	0.710	4.92e-13	1.44e+12	<2e-16
Ice	-12.156	1.70e-21	-7.15e+21	<2e-16	0.767	4.59e-15	1.67e+14	<2e-16
Snow/Slush	5.525	4.54e-14	1.22e+14	<2e-16	-0.594	5.28e-15	-1.12e+14	<2e-16
Unknown/Other	-14.176	2.95e-22	-4.81e+22	<2e-16	0.023	6.47e-15	3.48e+12	<2e-16
AADT	-2.57e-06	1.43e-06	-1.792	0.0731	-3.37e-06	2.56e-07	-13.168	<2e-16
Highway Class (Ref: Ramp/Exit/Interstate Exit)								
Rural Two-Lane	5.271	1.01e-12	5.24e+12	<2e-16	-0.052	2.95e-13	-1.75e+11	<2e-16
Rural Multilane	5.236	6.75e-13	7.75e+12	<2e-16	0.075	2.25e-13	3.35e+11	<2e-16
Rural Interstate	5.299	2.33e-12	2.28e+12	<2e-16	0.058	7.84e-13	7.43e+10	<2e-16
Urban Two-Lane	4.568	2.34e-12	1.95e+12	<2e-16	-0.109	7.83e-13	-1.40e+11	<2e-16
Urban Multilane	4.739	9.45e-12	5.01e+11	<2e-16	-0.093	2.55e-12	-3.63e+10	<2e-16
Urban Interstate and Freeways	5.027	1.04e-11	4.85e+11	<2e-16	-0.092	1.65e-12	-5.57e+10	<2e-16
Service/Frontage Road	-13.430	8.25e-22	-1.63e+22	<2e-16	-0.125	3.30e-14	-3.78e+12	<2e-16
Occupant Protection System (Ref: Child Safety Seat Improperly Used)								
Child Safety Seat Used	10.930	4.91e-14	2.23e+14	<2e-16	-0.829	6.26e-15	-1.32e+14	<2e-16
Helmets Used	13.207	3.08e-13	4.29e+13	<2e-16	1.904	2.53e-13	7.53e+12	<2e-16
Lap Belt Only Used	10.393	1.23e-13	8.44e+13	<2e-16	-0.296	2.24e-14	-1.33e+13	<2e-16
Shoulder Belt Only Used	10.365	6.99e-13	1.48e+13	<2e-16	-0.110	1.43e-13	-7.64e+11	<2e-16
Shoulder and Lap Belt Used	9.318	1.95e-11	4.77e+11	<2e-16	-0.318	4.86e-12	-6.54e+10	<2e-16
None Used - Vehicle Occupant	11.512	2.01e-12	5.73e+12	<2e-16	1.024	1.25e-12	8.18e+11	<2e-16
Restraint Use Unknown	10.249	3.45e-12	2.97e+12	<2e-16	-0.275	6.56e-13	-4.20e+11	<2e-16

The multinomial logit for all posted speed limit groups (except >30 to <=40 mph for FSI crashes) to speed limit <=30 mph are higher for being both in FSI crash and MCI crash to PDO crash. In all these cases, p-value is less than 0.05, which indicates higher posted speed limits are significantly different in more severe distraction-related crashes. The severity risk is specifically higher in case of posted speed limit between 60 to 70 mph.

Compared to pickup truck, the multinomial logit for passenger car is 0.029 unit lower for being in fatal and serious injury crash and is 0.054 unit lower for being in moderate and complaint injury crash to property damage only crash. With p-value less than 0.05, it can be said that the difference is significant. Same can be said for van and other vehicles compared to pickup truck.

The impact of a distraction-affected crash on curve-level road can be more severe compared to straight-level road. Although rear-end crashes are most frequent when it comes to in-vehicle distraction, head on crashes can turn out to be more severe. Both dry and wet roads can result in severe crashes, as can be seen from the comparison of coefficients.

The multinomial logit for rural two-lanes relative to Ramp/Exit/Intestate Exit is 5.271 unit higher. Drivers on rural two-lane highways are more prone to distracted driving fatal crashes compared to PDO crashes, followed by rural multilane roads, rural interstate, urban freeways and interstate, urban multilane and urban two-lanes. In rural multilane and interstate highways, drivers at fault are likely to be involved in moderate and complaint injury crashes compared to PDO crashes. In case of FSI crashes relative to PDO crashes, using both lap and shoulder belts produces the lowest multinomial log-odds compared with only lap belt or only shoulder belt. This indicates drivers using both lap and shoulder belts might have lowest risk of being involved in an FSI crash.

5.2.2. Random Forest Results

The Random Forest algorithm does not provide an equation to predict severity directly as it is a supervised algorithm, rather it presents a variable importance plot. Visualization of random forest prediction results with all categorical type of predictor data is complicated. However, Random Forest often predicts more accurately than statistical regression model.

A variable importance plot (Figure 15) indicates what variables had the greatest impact in the classification model through the estimation of mean decrease of accuracy. The more the accuracy of the random forest decreases due to the exclusion (or permutation) of a single variable, the more important that variable is deemed, and therefore variables with a large mean decrease in accuracy are more important for classification of the data.

These importance values can be used to perform additional analysis, like principal component analysis or to make simpler models with fewer important variables. Collision type is the most important variable and surface condition is the least important variable in the prediction of severity according to random forest algorithm.

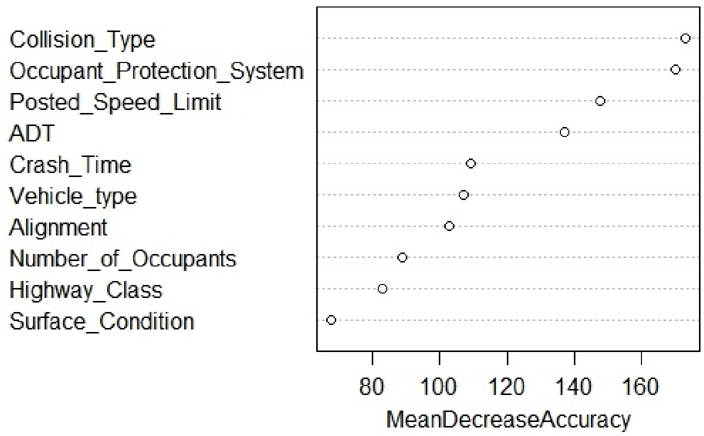

Figure 15. Importance of random forest variables.

5.2.3. Comparison of MLR and Random Forest

In order to compare the accuracy of these two approaches – multinomial logistic regression and Random Forest data mining, the database was randomly split into 70% and 30%, training set and testing set respectively. The training set was used to generate the prediction models of regression and random forest both. Both models were then used to predict severity with the predictor variables in the testing dataset. Predicted severity (by both models) and actual severity were then compared to estimate the accuracy of both models. Sensitivity and specificity are two quantified measures for estimating prediction accuracy. Test sensitivity is the ability of a test to correctly identify those with the actual result in the testing set (true positive rate), whereas test specificity is the ability of the test to correctly identify severity without the disease (true negative rate). The Receiver operating characteristic (ROC) Curve as illustrated in Figure 16, which includes both the sensitivity and specificity shows more accuracy in prediction, a 54.91% area under curve compared to 52.62% for multinomial logistic regression.

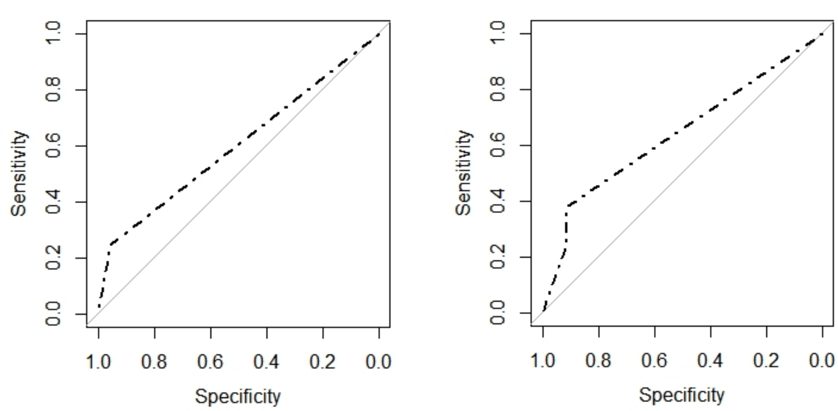

Figure 16. ROC curve for multinomial logistic regression (left) and random forest (right).

5.3. Roadside Observation Results

5.3.1. Fisher's Exact Test Results

Fisher's exact tests assess the null hypothesis that there are no relationships between two variables with a p-value of 0.05 was used as a cutoff point. A p-value <0.05 indicates we reject the null hypothesis that there is no association between two classifications. A Fisher's exact p-value >0.05 shows we cannot reject the null hypothesis that two classifications have no association. If p-value of driver's gender is greater than 0.05 – meaning it cannot be rejected that driver's gender have no association with distraction type.

P-value from Fisher's exact test for setting was obtained as 0.0039, which indicates that cellphone use could be significantly different by area type. Higher percentage of drivers were found to be manipulating than talking on the phone in urban areas, whereas the situation is opposite in rural area. It should also be considered that the sample size in rural area is smaller.

The estimated p-value to assess the difference in texting or manipulating and talking by cross-section was 0.1574. The two basic types of cellphone use were not associated with intersection or segment.

Safety-oriented driving style can be different between men and women, however risky behaviors are predominantly attributed to males *(47 – 49)* and varies according to driving conditions *(50)*. Our roadside observation study shows that there is no significant difference between gender type and cell phone distraction type in Louisiana (p-value is 0.6215). However, national statistics continue to show higher percentage of males involved in distraction related crashes *(1)*. Studies regarding different cellphone use type shows contradictory results. One self-reported opinion survey study suggests male drivers are more likely to engage in talking on a cellphone than female drivers due to their work *(51)*. Although the 2016 NHTSA survey study on national data found almost the same percentage of young males and females are engaged in texting while driving *(52)*, one study shows while driving higher cellphone dependence and higher levels of risky behaviors could be associated with young female drivers when it comes to texting *(48)*.

The p-value of Fisher's exact test for age group and distraction is 0.7626, more than 0.05. It indicates the variable 'age group' is not influential to cellphone use type. The latest NHTSA roadside observation results suggest young drivers aged less than 25 years old *(25)* are involved in using an electronic device while driving and FARS data shows teens are killed in distraction-affected crashes more than any other age group *(53)*. However, older drivers have reported to have engaged themselves in various cellphone use while driving. According to the latest AAA report from national survey of more than 2,600 conducted in 2017, among the drivers of age 25-39 years, 66.9% reported to have talked on a hands-free cellphone, 62.2% have read a text message or email, and 55.5% have typed or sent a text message or email – more than any other age group *(2)*. Another study from anonymous survey of 500 participants showed that significant causal distracted driving predictors were prevalent among drivers aged 30-64 years and their engagement in talking on the phone while driving and/or texting while driving is primarily due to overconfidence in driving abilities and obligation to take work calls *(54)*.

However, in our study age group was not observed to be associated with cellphone use type – texting and talking.

In our observations, vehicle type was categorized based on abundance in type of vehicles, although different regulations are in place for commercial vehicle drivers in terms of cellphone use. NHTSA roadside observation data suggests passenger car drivers have higher proportion of distracted drivers than SUV, van, or pickup truck drivers *(25)*. According to our observation results, particular vehicle types were not associated with cellphone distraction type.

From experience, it is highly expected that drivers in the vehicles without any other passengers will engage more in cellphone use while driving than the drivers with passengers in the vehicle. Expectedly, the roadside observation shows single occupant (no passenger presence) is highly associated with cellphone distraction type (p-value = 0.0036).

5.3.2. Association Rule Mining Results

Using the Apriori algorithm of ARM followed by pruning, 15 significant 4-itemset rules were generated with “cellphone use = manipulating” as consequent (Table 5). To avoid unnecessary rules, minimum 4-itemset were chosen for rule generation. To optimize most frequent rules, minimum support of 0.1 and minimum confidence of 0.6 were used. The rules have been listed and ordered by higher to lower lift. All the rules contain a lift value higher than 1, which indicates these co-occurring associations are more than expected.

None of the rules included “cross_section = segment”. Manipulating, including texting typically occurs at intersections. Both males and females have been involved in texting. Drivers in both the 30-60y and <30y age groups were found to engage in manipulating. Expectedly, the absence of passengers seems to have induced the drivers to text more, since drivers with passengers have not been found in any rules.

Table 5. Apriori generated 4-itemset rules for "cellphone_use = manipulating" with minimum support of 0.1 and minimum confidence of 0.6.

Antecedent	Consequent	support	confidence	lift
cross_section = intersection, setting = urban, vehicle_type = car	cellphone_use = manipulating	0.129	0.686	1.184
driver_gender = male, setting = urban, vehicle_type = car		0.119	0.676	1.167
cross_section = intersection, passenger_present = no, vehicle_type = car		0.128	0.671	1.158
driver_age = 30-60y, driver_gender = female, setting = urban		0.102	0.656	1.133
cross_section = intersection, driver_age = 30-60y, driver_gender = female		0.104	0.647	1.116
cross_section = intersection, driver_gender = female, setting = urban		0.157	0.634	1.095
driver_age = 30-60y, setting = urban, vehicle_type = car		0.117	0.634	1.095
driver_age = <30y, driver_gender = male, setting = urban		0.120	0.631	1.089
cross_section = intersection, driver_gender = female, passenger_present = no		0.154	0.629	1.085
cross_section = segment, driver_gender = male, setting = urban		0.145	0.628	1.085
driver_gender = male, passenger_present = no, vehicle_type = car		0.110	0.628	1.084
cross_section = intersection, driver_age = 30-60y, setting = urban		0.158	0.621	1.072
cross_section = intersection, passenger_present = no, setting = urban		0.252	0.615	1.062
driver_age = <30y, driver_gender = male, passenger_present = no		0.116	0.611	1.056
driver_age = <30y, setting = urban, vehicle_type = car		0.148	0.601	1.038

A total of 13 4-itemset rules with "cellphone use = talking" as consequent were generated with Apriori algorithm of ARM, which are listed in Table 6. Minimum support and confidence were used as 0.08 and 0.45 considering the low frequency of co-occurring associations in this case. The lift value for all rules generated is also greater than 1, indicating more than expected co-occurring associations.

The results are mixed, both male and females are present within 13 rules. Handheld or hands-free conversation occurs both at segment and intersection, with only one rule including "cross-section = intersection". Car and SUV/Van drivers are most frequently engaged in talking on the cellphone, whereas specifically car drivers were mainly engaged in manipulating compared to any other vehicles.

Table 6. Apriori generated 4-itemset rules for "cellphone_use = talking" with minimum support of 0.08 and minimum confidence of 0.45.

Antecedent	Consequent	support	confidence	lift
cross_section = segment, driver_gender = female, vehicle_type = car	cellphone_use = talking	0.085	0.543	1.290
cross_section = segment, driver_gender = female, passenger_present = no		0.112	0.522	1.242
cross_section = segment, driver_age = 30-60y, passenger_present = no		0.103	0.506	1.202
cross_section = segment, passenger_present = no, vehicle_type = car		0.112	0.492	1.169
driver_age = 30-60y, driver_gender = male, passenger_present = no		0.104	0.486	1.155
driver_age = <30y, driver_gender = female, passenger_present = no		0.100	0.483	1.147
cross_section = segment, driver_gender = female, setting = urban		0.120	0.481	1.142
cross_section = segment, passenger_present = no, setting = urban		0.180	0.466	1.107
cross_section = intersection, passenger_present = no, vehicle_type = suv_van		0.081	0.459	1.091
cross_section = segment, driver_age = <30y, passenger_present = no		0.105	0.455	1.082
driver_age = 30-60y, passenger_present = no, vehicle_type = suv_van		0.087	0.453	1.076
driver_gender = female, passenger_present = no, vehicle_type = car		0.110	0.453	1.076
driver_age = <30y, driver_gender = female, setting = urban		0.097	0.452	1.074

Combination of support and confidence for the generated rules has been presented in scatterplots, in Figures 17a and 17b. Cellphone manipulating is found to be more frequent than talking according to the visual comparison of two scatterplots. All 15 rules generated for "cellphone_use = manipulating" have a confidence of 0.6, whereas 10 out of 13 rules generated for "cellphone_use = talking" have a confidence of less than 0.5.

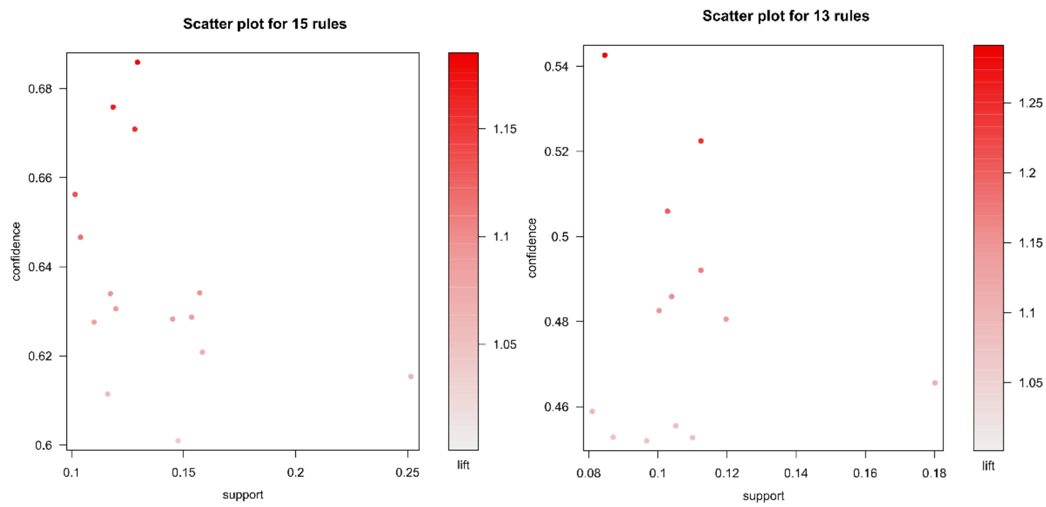

Figure 17. Support and confidence scatterplot with lift for all the rules with consequent cellphone_use = manipulating (left) and cellphone_use = talking (right).

5.3.3. Network Visualization

Figure 18 and 19 illustrated all the rules for manipulating and talking separately showing the interconnection of all the itemsets. Similar to the scatterplot, the network diagrams also display the tradeoff between support and lift. Larger circles imply higher support, while red circles imply higher lift. Most importantly, network diagrams illustrate the relationships between each antecedent item within the generated rules with the consequent. Each antecedent might have multiple connections with the consequent according the lift and support values of each co-occurring associations. From the diagrams, cross-section = intersection perhaps possess the strongest relationship the consequent "cellphone_use = manipulating" by maintaining optimized lift and support. Same can be said for the relationship between "driver_gender = female" with "cellphone_use = talking".

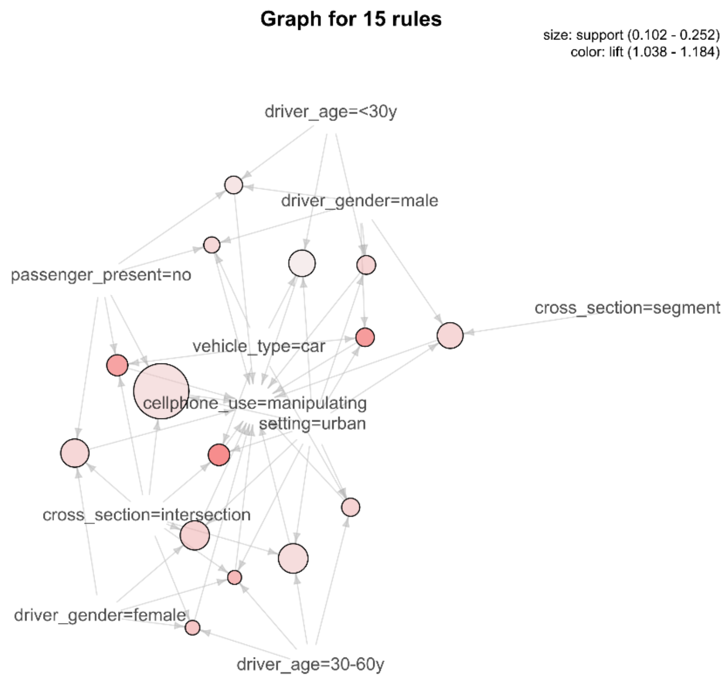

Figure 18. Network diagrams for all 15 rules generated by "cellphone_use = manipulating".

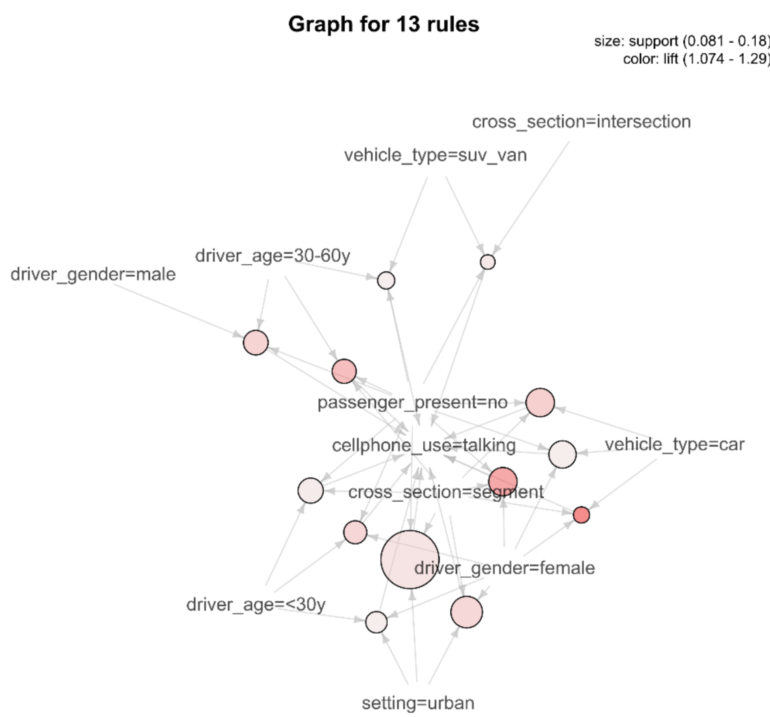

Figure 19. Network diagrams for all 13 rules generated by "cellphone_use = talking".

5.3.4. Logistic Regression Results

The Logistic regression results show that setting and passenger presence are only two variables which can be significantly associated with prediction of driver cellphone use. For example, in rural areas drivers are 1.9 times likely to engage in talking than manipulating cellphone. Without a passenger, a driver may engage in talking 2.1 times than manipulating, which could go up to 3.4 times. The rest of the variables (cross-section, gender, age group, vehicle type) are weakly associated with cellphone use.

Table 7. Results of logistic regression.

Variable	Coefficient	Standard Error	Wald	p-value	Odds Ratio	95% C.I.
Setting (ref. Urban)						
Rural	0.618	0.257	5.757	0.016	1.855	1.120, 3.072
Cross-section (ref. Segment)						
Intersection	-0.283	0.149	3.582	0.058	0.754	0.563, 1.010
Gender (ref. Male)						
Female	0.210	0.159	1.744	0.187	1.234	0.903, 1.685
Age group (ref. 30-60y)			0.411	0.814		
<30y	-0.045	0.153	0.088	0.767	0.956	0.709, 1.289
>60y	-0.273	0.452	0.366	0.545	0.761	0.314, 1.845
Vehicle type (ref. Truck)			3.708	0.295		
Car	0.265	0.220	1.443	0.230	0.767	0.498, 1.182
Other	0.893	0.740	1.458	0.227	2.444	0.573, 10.419
SUV/Van	-0.150	0.232	0.417	0.519	0.861	0.546, 1.357
Passenger presence (Ref. Yes)						
No	0.750	0.242	9.587	0.002	2.117	1.317, 3.403

5.4. In-Vehicle Observation Results

From the video clips of 40 participants' driving, only three incidents of cellphone conversation were identified. It is difficult to obtain conclusive results from only three conversations. However, average valence estimations by the FaceReader software do vary 5 seconds before, during, and 5 seconds after cellphone conversations. The result of average valence shows that during a phone call, large valence of neutral emotion counterbalances all the six emotions. When percentages are compared, large differences in valence estimates of happiness/sadness, surprise, and disgust are noticeable. Figure 20 depicts valence estimates of collective emotions before, during, and after phone call.

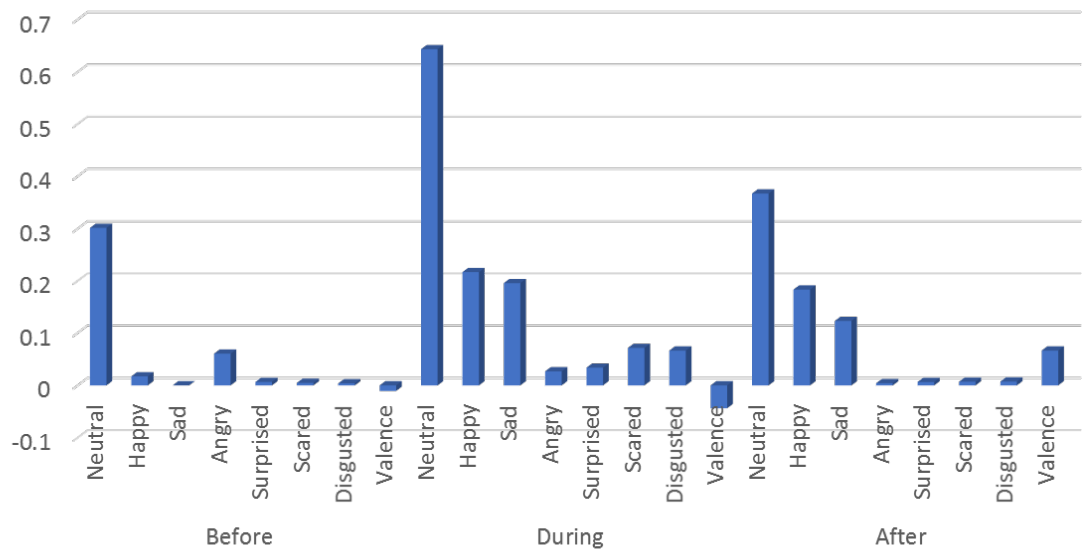

Figure 20. Estimated valence of all emotions before, during, and after phone conversation.

Figure 21 enlarges the changes (from Figure 20) of three emotions – happiness, surprise, and disgust. When it comes to emotions like surprise and disgust, the valence attained before the conversation and the valence attained after the conversation, are of almost similar quantity while the valence during the call is higher. However, happiness/sadness valence estimates are retained after the phone call.

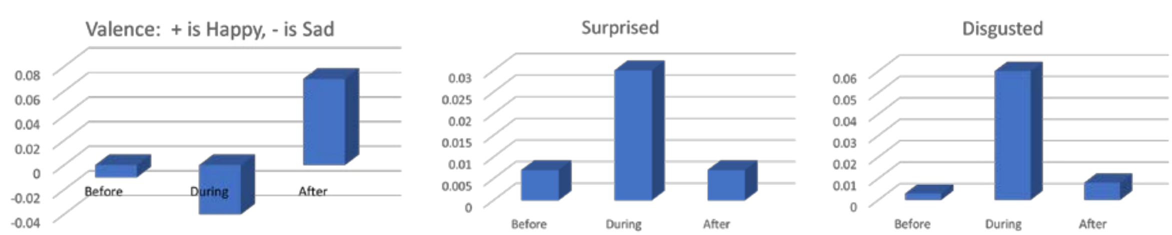

Figure 21. Estimated valence of happiness/sadness, surprise, disgust before, during, and after phone conversation.

To check whether the individual and combined emotions change during, before and after texting, an F-test was performed. Estimation of emotions 5 seconds prior to texting, during complete duration of texting and 5 seconds after texting were gathered and F-test was run by excel. The results show that individual emotions don't change before, during, and after texting while driving. However, combined valence of all emotions shows significant changes during the period of texting, 5 seconds prior to texting and 5 seconds after texting.

Table 8. F-test for emotions before, during, and after texting.

Emotion	***F***	***P-value***
Anger	1.029	0.363
Neutral	1.018	0.366
Happy	0.819	0.446
Sad	1.977	0.147
Surprised	0.901	0.411
Scared	0.384	0.683
Disgusted	0.384	0.683
Valence	**3.620**	**0.032**

Two ideal scenarios were chosen to make a distinguishable rule for cellphone conversation and manipulation from the characteristics of eye, mouth and head orientation, and estimated valence. For cellphone conversation:

- Key features: emotion (unstable), mouth movement (open), and
- Rule 1: (Mouth = Open) and [(Happy > $\mu + 2\sigma$) or (Angry > $\mu + 2\sigma$)].

For cellphone manipulation:

- Key features: look down, eyes are closed
- Rule 2: (Left eye and Right eye = Closed) and (Sad > $\mu + \sigma$)

where, μ is mean valence estimate of a particular emotion and σ is standard deviation. These rules need to be verified with multiple observations of cellphone calls and manipulation.

6. CONCLUSIONS

The crash data can be a key source to measure the impact of distracted driving. However, the deficiency in distraction-related crash reporting is clear and can further be improved for better understanding of distracted behaviors. Underreporting of distraction-related crashes should be further studied at the local level. The quality of distraction-related reporting can be improved by reviewing the categorization of “Driver Distracted By” variable and whether the category “other inside source” can be expanded into significant distraction-related behaviors, as previously learned from the MMUCC guidelines.

A wider range of classification in driver distraction might be more helpful in understanding the relationship between distraction type and severity of crashes. For simplicity of the analysis and interpretation of prediction, the five injury severity types were grouped into three injury severity types. The grouping of fatal and severe injury types was targeted at reducing the randomness of fatal crashes. However, excluding the fatal crashes, the rest of the injury types can also be separately studied.

The crash analysis results of distraction collision type and crash severity indicate several interesting remarks. Due to in-vehicle distractions, head-on crashes were found to be deadlier than any other crash types. The curve-level road was found to be more prone to fatal and serious injury crashes compared to straight-level road, as far as in-vehicle distraction is concerned. On rural roadways (two-lane, multilane, interstate), distracted drivers have higher probability of being fatally or severely injured. For reduction in severity of distraction related crashes, these particular types of roadways should be targeted for countermeasures.

The random forest works slightly better than multinomial logistic regression, although it is expected to work more effectively in severity prediction. The low accuracy is also attributed to the randomness of fatal and severe injury crashes, as those two types of crashes are only 0.78% of the total analyzed crashes. The data mining algorithms have a larger potential in the application of exploratory analysis of distraction-related crash data. This study is just a small demonstration of random forest algorithm. Algorithms like ‘support vector machine’ or ‘neural network’ can also be studied for better predicting the relationships between contributing factors and distraction-related crash severity.

The severity prediction can be helpful for DOTs and road safety organizations to improve the knowledge regarding the roles of contributing factors in severity of distraction-related crashes and to make better decision in applying appropriate countermeasures.

Roadside observation of drivers shows that both driver’s gender and age group have no significant influence in cellphone distraction type. Without a passenger, a driver may engage in talking 1.3 to 3.4 times (s)he may engage in manipulating. Association rule mining of observation data shows that the most predominant type of cellphone use is manipulating i.e. texting, followed by talking. With a larger sample, a combination of different variables can be further studied using association rules mining.

FaceReader has remarkable potential in transportation safety including identification of distracted behaviors while driving. Face Reader represents a novel approach to understanding human emotion while driving. Initial results do suggest emotional distribution before, during and after cellphone manipulating is significantly different. Further analysis is needed to examine differences attributed to cellphone conversation and manipulation across all age groups. A larger sample size is required before drawing more reliable conclusions.

7. RECOMMENDATIONS

Several recommendations are made based on the findings of the study which are as follows:

- Crash data is and will remain as an important resource for identification of factors related to distracted driving. The underreporting of distraction-affected crashes is a serious concern. It is to be reviewed whether the quality of distraction-related reporting in the incident of a crash can be improved by expanding the classification of distraction type in the current crash reporting form. Roadside observation of distracted driving can be further expanded by evaluating with the addition of more variables.
- Physical countermeasure development towards reducing the distraction-related crash severity should be targeted at preventing lane departure crashes. Rural roadways and roadways with horizontal curvature require special attention in this regard. Countermeasures such as – center line rumble strips, shoulder rumble strips, retroreflective edge line marking, and chevron signs provide visual, auditory, and vibratory guidance to drivers. These countermeasures have remarkable potential for distraction-related crash severity reduction when installed on segments that meet the installation criteria.
- From the limited observations in this study, it was found that drivers tend to engage less in distracted driving behavior, specifically using a cellphone, if a passenger is present. Further study is required to evaluate the extent of the effectiveness of carpooling with experienced drivers compared to the effect of cellphone ban during carpooling reducing distraction-related crashes.
- As shown by the previous literature, a ban on texting significantly lowers fatal crashes. Strict enforcement of the current texting ban is necessary to create a road safety culture where texting while driving is viewed as derogatory behavior. Campaigns in schools, youth organizations, and local libraries could play an important role in helping to promote safe driving habits.
- Data mining can be a very helpful in distracted driving safety data analysis and modeling. In this regard, the applicability of advanced data mining algorithms in like support vector machine, neural networks can be investigated regardless of crash data and naturalistic observation data.
- Software packages like FaceReader have significant potential in detecting various types of driver distraction. With more participants and more coded data of distracted drivers, different in-vehicle distraction types can be modeled through FaceReader.

REFERENCES

1. NHTSA. Distracted Driving, n.d. https://www.nhtsa.gov/risky-driving/distracted-driving, Accessed January 2, 2018.

2. AAA Foundation for Traffic Safety. “2017 Traffic Safety Culture Index.” AAA Foundation, Washington, D.C., 2018.

3. AT&T Inc. Smartphone Use While Driving Grows Beyond Texting to Social Media, Web Surfing, Selfies, Video Chatting, May 19, 2015. http://about.att.com/story/smartphone_use_while_driving_grows_beyond_texting.html, Accessed January 12, 2018.

4. Strayer, D. L., Cooper, J. M., Goethe, R. M., McCarty, M. M., Getty, D. and Biondi, F. “Visual and Cognitive Demands of Using In-Vehicle Infotainment Systems.” Washington, D. C.: AAA Foundation for Road Safety, 2017.

5. U.S. Energy Information Administration, “Study of the Potential Energy Consumption Impacts of Connected and Automated Vehicles.” U.S. Department of Energy.” Washington D.C., 2017.

6. NHTSA, “Overview of the National Highway Traffic Safety Administration's Driver Distraction Program.” National Highway Traffic Safety Administration, Washington D.C., 2010.

7. Regan, M. A., Hallett, C. and Gordon, C. P. “Driver distraction and driver inattention: Definition, relationship and taxonomy.” Accident Analysis and Prevention, Vol. 43, No. 5, 2011, pp. 1771–1781.

8. LaDOTD, “Louisiana Strategic Highway Safety Plan.” Louisiana Department of Transportation and Development, Baton Rouge, 2017.

9. Everdrive. EVERDRIVE Safe Driving Report 2016-2017: Louisiana. n.d. https://www.everquote.com/everdrive/safe-driving-report-2016/louisiana/, Accessed January 2, 2018.

10. Donmez, B., and Liu, Z. “Associations of distraction involvement and age with driver injury severities.” Journal of Safety Research, Vol. 52, pp 23–28, 2015. doi:10.1016/j.jsr.2014.12.001.

11. Braitman, K. A., Kirley, B. B., McCartt, A. T., and Chaudhary, N.K. “Crashes of novice teenage drivers: Characteristics and contributing factors.” Journal of Safety Research, Vol. 39, pp 47–54, 2008. doi:10.1016/j.jsr.2007.12.002.

12. Neyens, D. M., and Boyle, L. N. “The effect of distractions on the crash types of teenage drivers.” Accident. Analysis and Prevention, Vol 39, pp 206–212, 2007. doi:10.1016/j.aap.2006.07.004.

13. National Safety Council. “Crashes Involving Cell Phones: Challenges of Collecting and Reporting Reliable Crash Data”, pp1–12, 2013. http://www.nsc.org/DistractedDrivingDocuments/NSC-Under-Reporting-White-Paper.pdf.

14. Regev, S., Rolison, J. J., Feeney, A., and Moutari, S. “Driver distraction is an under-reported cause of road accidents : An examination of discrepancy between police officers’ views and road accident reports.” Fifth International Conference on Driving Distraction, DDI2017, Paris, 2017. http://www.ifsttar.fr/collections/ActesInteractifs/AII2/pdfs/DDI2017_Paper_2B-2.pdf.

15. Crash 1 Database. Louisiana Department of Transportation and Development, Baton Rouge, Louisiana.

16. Wilson, J., Fang, M., Wiggins, S. and Cooper, P., 2003. Collision and violation involvement of drivers who use cellular telephones. Traffic injury prevention, 4(1), pp.45-52.

17. Papantoniou, P., Papadimitriou, E., and Yannis, G. “Review of driving performance parameters critical for distracted driving research.” Transportation Research Procedia, Vol 25, pp1801–1810, 2017. doi:10.1016/j.trpro.2017.05.148.

18. Ishak, S., Codjoe, J., Thapa, R. and McCarter, K., 2014. *Distracted Driving and Associated Crash Risks*, Report No. FHWA/LA. 14/530, Federal Highway Administration, Baton Rouge, Louisiana.

19. Alghnam, S., Towhari, J., Alkelya, M., Binahmad, A., and Bell, T. M. The effectiveness of introducing detection cameras on compliance with mobile phone and seatbelt laws: a before-after study among drivers in Riyadh, Saudi Arabia, Injury Epidemiol. Vol 5 pp 4–11, 2018. doi:10.1186/s40621-018-0161-z.

20. Johnson, M. B., Voas, R. B., Lacey, J. H., McKnight, A. S., and Lange, J. E. “Living Dangerously: Driver Distraction at High Speed.” Traffic Injury Prevention, Vol. 5, pp1–7, 2004. doi:10.1080/15389580490269047.

21. Huemer, A. K., Schumacher, M., Mennecke, M., and Vollrath, M. “Systematic review of observational studies on secondary task engagement while driving.”, Accident. Analysis and Prevention, Vol 119, pp 225–236, 2018. doi:10.1016/j.aap.2018.07.017.

22. Neale, V.L., Klauer, S.G., Knipling, R.R., Dingus, T.A., Holbrook, G.T. and Petersen, A., 2002. The 100 car naturalistic driving study, Phase I-experimental design (No. HS-809 536,).

23. Campbell, K. The SHRP 2 Naturalistic Driving Study. n.d. https://insight.shrp2nds.us/documents/shrp2_background.pdf, January 7, 2018.

24. Dingus, T. A., F. Guo, S. Lee, J. F. Antin, M. Perez, M. Buchanan-King, and J. Hankey. “Driver Crash Risk Factors and Prevalence Evaluation Using Naturalistic Driving Data.” Proceedings of the National Academy of Sciences, 2016.

25. NHTSA, “Driver Electronic Device Use in 2016.” National Highway Traffic Safety Administration, Washington DC, 2017.

26. A.K. Huemer, M. Schumacher, M. Mennecke, M. Vollrath, Systematic review of observational studies on secondary task engagement while driving, Accid. Anal. Prev. 119 (2018) 225–236. doi:10.1016/j.aap.2018.07.017.

27. Huth, V., Sanchez, Y. and Brusque, C. “Drivers’ phone use at red traffic lights: A roadside observation study.” Accident Analysis and Prevention, Vol. 74, No. 5, 2015, pp. 42–48.

28. Sullman, M. J. M., Prat, F., and Tasci, D. K. “A Roadside Study of Observable Driver Distractions.” Traffic Injury Prevention, vol. 16, No. 6, pp. 552-557.

29. Huisingh, C., Griffin, R. and Mcgwin, G.”The Prevalence of Distraction Among Passenger Vehicle Drivers: A Roadside Observational Approach.” Traffic Injury Prevention, vol. 16, No. 2, pp. 140-146, 2014.

30. Autolabs, G. Which States Ban Cellphone Use While Driving? January 2018. https://www.germanautolabs.com/blog/which-states-ban-cellphone-use-while-driving, Accessed January 5, 2018.

31. Trempel R. E., Kyrychenko, S. Y., and Moore, M. J. “Does Banning Hand-held Cell Phone Use while Driving Reduce Collisions?” Chance, Vol. 34, pp 6–11, 2004.

32. Lim S. H., and Chi J. “Cellphone Bans and Fatal Motor Vehicle Crash Rates in the United States.” Journal of Public Health Policy, Vol. 34, pp197–212, 2013.

33. Abouk R, and Adams S. “Texting Bans and Fatal Accidents on Roadways: Do They Work? Or Do Drivers Just React to Announcements of Bans?” Am Econ J Appl Econ., vol. 5, pp179–99, 2013.

34. Lumley, T. and Miller, A., n.d. Package ‘leaps’: Regression Subset Selection.

35. Fisher, R. A. “On the Interpretation of χ 2 from Contingency Tables, and the Calculation of P.”, J. R. Stat. Soc. Vol. 85 No. 87, 1922. doi:10.2307/2340521.

36. Kim, H.Y. “Statistical notes for clinical researchers: Chi-squared test and Fisher’s exact test.” Restor. Dent. Endod. Vol. 42, No.152, 2017. doi:10.5395/rde.2017.42.2.152.

37. Mehta, C. R., Patel, N. R. “A Network Algorithm for Performing Fisher Exact Test in r X c Contingency Tables.” J. Am. Stat. Assoc., Vol 78, pp 427–434, 1983. doi:10.2307/2288652.

38. Agresti A. “A Survey of Exact Inference for Contingency Tables.” Stat. Sci. Vol. 7 , pp 131–153, 1992. doi:10.1214/ss/1177011454.

39. Cabral, D. R. de L., and de Barros, R. S. M., “Concept drift detection based on Fisher’s Exact test.” Inf. Sci. (Ny), pp 442–443, 2018. doi:10.1016/j.ins.2018.02.054

40. Mangiafico, S. “rcompanion: Functions to Support Extension Education Program Evaluation”, 2018. http://rcompanion.org.

41. R Core Team, R: A Language and Environment for Statistical Computing, (2018). http://www.r-project.org/.

42. Noldus. FaceReader. N.d. https://www.noldus.com/human-behavior-research/products/facereader?gclid=CjwKCAjwt7PcBRBbEiwAfwfVGLRcrAN30BTuTznIe8oZirmYLus3rLFoOhLRyqCGEhzjt8rdGfSZuxoC2okQAvD_BwE, June 5, 2018.

43. NHTSA. Model Minimum Uniform Crash Criteria Guideline (4th edition). National Highway Traffic Safety Administration, Washington DC, 2012.

44. NHTSA. Model Minimum Uniform Crash Criteria Guideline (5th edition). National Highway Traffic Safety Administration, Washington DC, 2017.

45. NHTSA. Model Minimum Uniform Crash Criteria Guideline. National Highway Traffic Safety Administration, Washington DC, 2012.

46. NHTSA. 2016 FARS / CRSS Coding and Validation Manual. National Highway Traffic Safety Administration, Washington DC, 2017.

47. Starkey, N. J., and Isler, R. B. "The role of executive function, personality and attitudes to risks in explaining self-reported driving behaviour in adolescent and adult male drivers." Transportation Research Part F: Traffic Psychology and Behaviour, vol. 38, pp. 127-136, 2016.

48. EladioJiménez-Mejías, Prieto, C. A., Martínez-Ruiz, V., delCastillo, J. d. D. L., Lardelli-Claret, P., and Jiménez-Moleón, J. J. "Gender-related differences in distances travelled, driving behaviour and traffic accidents among university students." Transportation Research Part F: Traffic Psychology and Behaviour, vol. 27, pp. 81-89, 2014.

49. Sagberg, F., Selpi, Piccinini, G. F. B., and Engström, J. "A Review of Research on Driving Styles and Road Safety." Journal of Human Factors, Vol. 57, No. 7, pp. 1248-1275, 2015.

50. Ericsson, E. "Variability in urban driving patterns." Transportation Research Part D: Transport and Environment, Vol. 5, No. 5, pp. 337-354, 2000.

51. Li, W., Gkritza K., and Albrecht, C. "The culture of distracted driving: Evidence from a public opinion survey in Iowa." Transportation Research Part F: Traffic Psychology and Behavior, Vol. 26, pp. 337-347, 2014.

52. NHTSA.Traffic Safety Facts Research Note: Young Drivers Report the Highest Level of Phone Involvement in Crash or Near-Crash Incidences. National Highway Traffic Safety Administration, Washington DC, 2012

53. NHTSA. Teen Distracted Driver Data 2016. https://crashstats.nhtsa.dot.gov/Api/Public/Publication/812504, Accessed April 12, 2018.

54. Engelberg, J. K., Hill, L. L., Rybar J., and Styer, T. "Distracted driving behaviors related to cell phone use among middle-aged adults." Journal of Transport & Health, vol. 2, no. 3, pp. 434-440, 2015.