Causes of Driver Distraction and Incidents at Signalized Intersections

by

ZANNATUL FERDOUS LABONY

THESIS

Submitted in partial fulfillment of the requirements for the degree of
Master of Science in Civil Engineering at
The University of Texas at Arlington

August, 2021
Arlington, Texas

Supervising Committee:
Stephen P Mattingly, Supervising Professor
Kate Kyung Hyun
Pengfei (Taylor) Li

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to professor Dr. Stephen P Mattingly for providing his support and guiding me in the right direction, both in research and course work. His guidance, mentoring and knowledge helped in the successful accomplishment of this study. Also, I would like to thank Dr. Kate Kyung Hyun and Dr. Pengfei (Taylor) Li for their guidance to improve this study.

I would also like to thank for the support and contribution of the faculty or staff at Virginia Tech Transportation Institute (VTTI) for providing access to the SHRP2 database.

ABSTRACT

Causes of Driver Distraction and Incidents at Signalized Intersections

Zannatul Ferdous Labony, MS
The University of Texas at Arlington, 2021

Supervising Professor: Stephen P Mattingly

Driver distraction causes a major portion of motor vehicle crashes because distractions turn the driver's attention away from driving the driving task, Intersections represent high risk environments because many conflict points within the intersection exist. At intersections with signalized traffic control, drivers may be more likely to become distracted than at other intersections. While distraction during the red indication may not seem to be a significant concern, this study investigates the role distraction plays in crashes at signalized intersections. Using the SHRP2 naturalistic driving study data, this study focuses on investigating the frequency and types of driver distraction, the causes of driver distraction, and factors affecting crashes and conflicts at signalized intersections.

The statistical modeling and decision trees developed in this thesis indicate that many factors significantly influence distraction, but age or years of driving experience appears as a critical factor in the likelihood of distraction. Driver familiarity and mild congestion levels also appear to increase the probability of distraction. The reductions in distraction relate to factors that induce a greater focus on the driving task like weather, age (older adults), or vehicle position. Uncongested conditions appear to decrease distractions and the risk of a crash or near crash event. Distraction from an object inside the vehicle poses a significant crash or near crash risk at signalized intersections. Technology (cell phone) related distractions pose a safety risk even for drivers queued at a signalized intersection, and the high frequency of this distraction among all drivers makes it a significant concern.

TABLE OF CONTENTS

List of Figures and Tables

CHAPTER 1
INTRODUCTION

The cost of motor vehicle crashes exceeds $1 Trillion a year in the U.S. (1). Distracted driving represents a significant concern for road safety because according to the National Highway Traffic Safety Administration (NHTSA) (2015), a total of 34,439 fatal crashes occurred in the United States in 2016 and about 9.2% involved distracted drivers. The fatalities included 562 non-occupants (i.e., pedestrians, bicyclists, and others) killed in distraction-affected crashes. Intersections represent another road safety issue because crashes and conflicts often occur there. The Federal Highway Administration reports 2.5 million intersection crashes nationwide each year and forty percent of all traffic crashes occur at an intersection.

Distractions usually divert drivers' attention, which worsen driving performance and create safety issues. Crash statistics indicate an increasing percentage of fatalities and injuries because of distracted driving (2). Research studies related to driver distraction and probability of crashes, can be categorized into crash data analysis, driving simulator studies, and naturalistic studies.

Crash data analysis represents a straightforward method to study the impact of distracted driving on safety. In 1995, the Crashworthiness Data System (CDS) introduced a new data variable named driver distraction/inattention to driving (DD/ID). Using the 1995 CDS data, Wang et al. (3) studied driver inattention and its involvement in crashes (2). However, crash data alone may fail to identify risky behavior because crashes occur infrequently, and baseline behavior must be collected for identifying risk probability.

Simulator studies represent another common approach to examine distracted driving behavior, given the crash risk of drivers under distraction in real road experimental conditions. A driving simulator study can simulate traffic and roadway conditions and test different types or combinations of distractions; however, these studies have many disadvantages. De Winter et al. (4) summarized that driving simulators have limited physical, perceptual, and behavioral fidelity. Käppler (5) stated that the original risk and consequence of actions never happen in a driving simulator, which may give rise to a false sense of safety, responsibility, or competence. Evans (6) questioned the ability of experiments using "make-believe equipment" to improve because the driving simulator has a reset button that can instantly erase all damage to people and property. The main limitation of using a driving simulator to research the impact of distracted driving on crashes or crash risk is that it cannot simulate real-world crashes, injuries, or fatalities. That is why all simulator studies use surrogate measurements, like speed, lane keeping, and reaction time, to describe crash risk. At the same time, participants know that they will be observed, which may bias their behavior.

Naturalistic observation represents a nonexperimental, primarily qualitative research tool to study subjects in their natural settings. The Strategic Highway Research Program 2 (SHRP2) Naturalistic Driving Study (NDS) study produced data that researchers use to examine driving behaviors including the impacts of driving distractions on safety. As part of the SHRP2 naturalistic driving study, Victor (25) carried out an analysis of people's driving behavior and crash risk using the Roadway Information Database (RID) and SHRP2 data. The RID contains information about roadway characteristics, while the SHRP2 data investigates driver behavior and driving before crash or near crash events. A naturalistic study's observational approach makes it more likely to capture actual distracted driving behaviors, which makes it more appropriate for examining the crash risk of distracted driving compared to crash data analysis and driving simulator studies. With properly designed analysis, naturalistic studies can quantify the crash risk of distraction behaviors and their influences of many factors, such as traffic and roadway conditions and sociodemographic characteristics on distraction behavior.

This study focuses on the frequency and types of driver distractions, causes of driver distraction and factors including driver distraction affecting crashes or near crashes at signalized intersections using the SHRP2 naturalistic driving study database.

CHAPTER 2
LITERATURE REVIEW

While driving, drivers must remain aware and vigilant of not only their status in the roadway environment but also the location and behavior of other dynamic actors like other vehicles and road users. Unfortunately, drivers often engage in secondary tasks that require concentrating on some event or object/person within or outside the vehicle while driving (7). For ordinary driving, drivers need to pay attention to the roads and surrounding areas frequently to maintain awareness of their driving environment (8). Distracted driving can increase the probability of crashes or near crashes, which represent events where vehicles almost collide with a fixed object, pedestrian, or another vehicle. As the use of electronic media becomes more prolific and the infotainment systems in vehicles provide more features, driving distraction appears likely to continue to increase and require further investigation.

Many studies investigated the factors that influence driver distraction. Wu and Xu (9) found driver age, traffic density, alignment, presence of an intersection, and hands on the wheel highly related to driver distraction. They also observed greater distraction in young age drivers (16-19 years old) than other age groups. However, Calvo et al. (10) found a higher crash risk when older and middle-aged drivers engage in distracted driving. Wu & Xu (9) found that distracted driving behavior increased on familiar roads, and Das et al. (11) found that weather conditions may also distract drivers and contribute to crashes and near crashes. According to Kidd et al. (12), holding (5.1%) or talking on (4.2%) a hand-held cellphone, eating or drinking (3.1%), and talking or singing with a passenger (2.7%) represented the most common secondary behaviors. Lee et al. (2018) additionally found that on road radio tuning represented a dangerous distracted driving behavior for drivers, and Sheykhfard & Haghighi (13) observed that digital billboards may also distract drivers.

Many studies use the SHRP2 data to identify the most important factors that increase crashes and crash severity; these studies hope to reduce the number of crashes in the future. Arvin et al. (1) reveal that distracted and aggressive driving correlate with driving volatility (speed instability) and have substantial indirect effects on crash intensity. Wang et al. (8) indicate that speeding, visual distractions, curve design elements, and pavement surface conditions affect the likelihood of a driver’s crash involvement. Bakhit et al. (14) analyze the increased crash or near-crash risk associated with different secondary tasks and demonstrate that reaching for objects, manipulating objects, reading, and cell phone texting represent the highest secondary task crash risk factors. According to Simons-Morton et al. (15), 16–17-year-old drivers experience higher crash rates than older drivers, but no differences between males and females exist within the 16–17-year-old cohort. Huisingh et al. (16) show a 3.79 times higher risk of a major crash event with cell phone use than the risk with no cell phone use for older drivers (drivers aged ≥70 years); they also identify that glance into the interior of the vehicle causes an increase in the risk of near-crash involvement for older drivers. Seacrist et al. (17) find that near crash rates significantly decrease with increasing age. Their study also demonstrates that young drivers exhibit greater rear-end and road departure near crashes and older drivers experience more intersection near crashes. While the crash angle of opposing movements at intersections often results in more dangerous crashes, few studies investigate crashes and near crashes at intersections.

At an intersection, two or more roads cross and many conflict points occur due to left, through, right, and pedestrian movements. Wu & Xu (9) find that vehicle type, traffic signal status, conflicting traffic, conflicting pedestrian, and driver age group represent the top five influencing factors on right-turn driver behavior. When drivers complete a Right-Turn-On-Red (RTOR) maneuver, they often pose a danger to other roadway users because they exhibit high accelerations and low observation frequencies. However, they do not evaluate the relationship between driver behavior and crash risk. Chandran (18) reconstructs crashes in intersections to gain a better understanding of the driver's gaze behavior using SHRP2 data and finds that drivers appear to see and continue to track the threat from the theoretical point of no return until the crash itself and drivers fail to engage in evasive maneuvers until too late. Dinakar & Muttart (19) try to simulate the response of through drivers to left turning vehicles and investigate the time to contact from when the turning driver began to turn. According to their research, time to contact significantly influences driver response time, but age, gender, and secondary task engagement usually do not influence response times. Morgenstern et al. (20) examine secondary task engagement while stopped at a red light using European naturalistic driving data and identify texting as one of the most problematic forms of distraction. In many instances, the drivers continue texting after the green phase begins, which may impact both operations and safety. From the previous studies, more research needs to investigate the factors influencing secondary (distraction) tasks at intersections and their effect on crashes or near crashes.

This study uses the US SHRP2 data and includes an investigation of the influence of demographic characteristics and traffic condition on secondary task engagement. This study also analyzes the factors contributing to crashes and near crashes at signalized intersections. Previous studies have not investigated left turning, right turning and through vehicles all together in terms of factors influencing crashes, near crashes and driver distraction at intersection using the SHRP2 data.

CHAPTER 3
DATA

This research uses a dataset from the Strategic Highway Research Program 2 (SHRP2) NDS database. The SHRP2 database consists of Naturalistic Driving Study (NDS) data and the Road Information Database (RID) data. The entire database includes 3,100 volunteers at six different sites in the United States: Tampa, Florida; Central Indiana; Durham, North Carolina; Erie County, New York; Central Pennsylvania; and Seattle, Washington spread over more than a 3-year period. All drivers have a valid driving license in their states. Almost half of these volunteers are male, and rest are female (9). The NDS data includes time-series records from eight different sensors installed on the vehicles and multi-direction video clips. Figure 1 shows an example of the SHRP2 NDS videos. From those multi-direction video clips, users can extract real-time speed, acceleration, weather conditions, road features and Global Positioning System (GPS) location. The SHRP2 NDS database is operated by The Virginia Tech Transportation Institute (VTTI) (21). The VTTI-developed Next Generation data acquisition system (DAS) gathered the NDS data. The DAS usually included multiple video images and a still image of the cabin, which it intentionally blurred to protect the privacy of passengers who had not given their consent. The DAS included four types of views such as (i) Front view (ii), Rear view, (iii) Face view and (iv) Dash view (24).

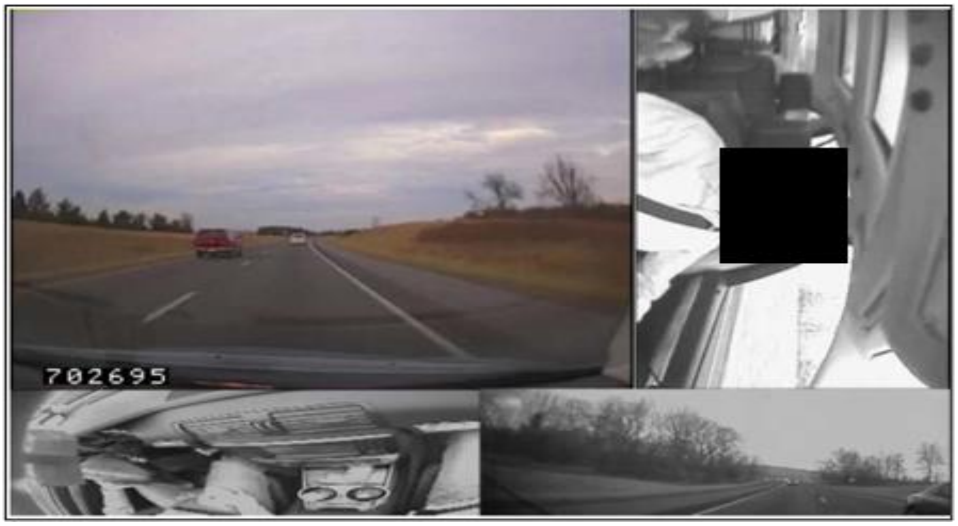

Figure 1: Strategic Highway Research Program 2 (SHRP2) NDS Videos

Data Pre-processing

For analyzing driver distraction and influencing factors for crashes, the study examines a total of 4606 events at signalized intersections from the SHRP2 NDS database. The events include front videos, time-series data, and event details table. This study develops a new variable for vehicles behind a heavy vehicle in the queue to investigate if queuing behind a heavy vehicle contributes to distraction or crash and near crash events. Video annotation is used to identity this binary variable for the 1,257 events where the vehicle speed is 5 mph or less. The study converts those events' videos into images using video to JPG converter software and extract only the images where the speed is 5 mph or less. Using those images, the researchers identify only 24 events of the 1,257 events where the subject vehicle queues behind a heavy vehicle.

To investigate driver distraction at intersection, the study considers various driver demographic data, traffic conditions and secondary task involvement data (Table 1).

Table 1: List of the variables with variable type and their brief description

Independent variable	**Variable Type**	**Brief description**
Secondary Task	Categorical	Observable driver engagement in any secondary tasks
Age Group	Categorical	Age group of the driver
Marital Status	Categorical	Marital Status of the driver
Gender	Binary	The gender of the driver
Traffic Flow	Categorical	Roadway design, including the presence or lack of a median, present at the start of the Precipitating Event. If the event occurs at an intersection, the traffic flow conditions just prior to the intersection are recorded
Contiguous Travel Lanes	Categorical	The total number of contiguous travel lanes at the time of the Precipitating Event start
Through Travel Lanes	Categorical	The number of through lanes present in the subject vehicle's direction of travel at the time of the Precipitating Event
Vehicle Lane occupied	Categorical	A number indicating which lane the subject vehicle is in at the time of the Precipitating Event. Lanes are numbered by starting with the left-most through lane closest to the median or double yellow line (direction of travel only) and starting with "1"
Traffic Density	Categorical	The level of traffic density at the time of the start of the Precipitating Event.
Traffic Control	Categorical	Type of traffic control applicable to the subject vehicle's direction of travel at the time of the start of the Precipitating Event. In this research only traffic signal is applied.
Average annual mileage	Categorical	Average annual mileage of the driver's vehicle
Work Status	Categorical	Work status of the driver
Education	Categorical	Education level of the driver
Years of driving	Continuous	Driving experience of the driver, in years
Alignment	Categorical	Description of the roadway curvature in the subject vehicle's direction of travel that best suits the condition at the time of the start of the Precipitating Event
Grade	Categorical	Description of the roadway profile (e.g., uphill, downhill) in the subject vehicle's direction of travel that best suits the condition at the time of the start of the Precipitating Event
Income	Categorical	Income range of the driver
Surface condition	Categorical	The type of roadway surface condition that would affect the vehicle's coefficient of friction
Locality/State	Categorical	Best description of the surroundings that influence or may influence the flow of traffic at the time of the start of the precipitating event
Weather Condition	Categorical	Weather condition at the time of the start of the Precipitating Event
Lighting	Categorical	Lighting condition at the time of the start of the Precipitating Event
Hands on the wheel	Categorical	A description of how many and/or which hands the driver had on the steering wheel

Datasets

Among 4,606 events in the SHRP2 database, 2,682 events include a secondary task. Thus, 58% of events experience driver distraction while the remaining 42% events represent non-distracted cases as shown in Figure 2.

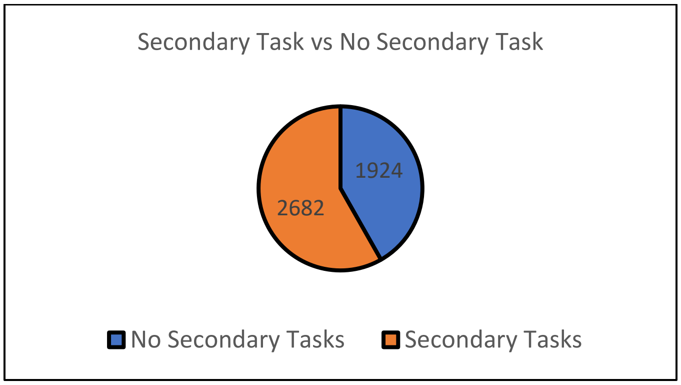

Figure 2: Frequency of Secondary tasks and No Secondary tasks

The SHRP2 NDS Data Access Website provided 63 types of distracted driving behaviors (21). Among the 4,606 signalized intersection events, 51 types of secondary tasks occurred. The authors simplified those 51 types into 10 groups and created a pie chart for better understanding of those secondary task involvements as shown in Figure 3. The grouping details of secondary tasks are shown in the appendix.

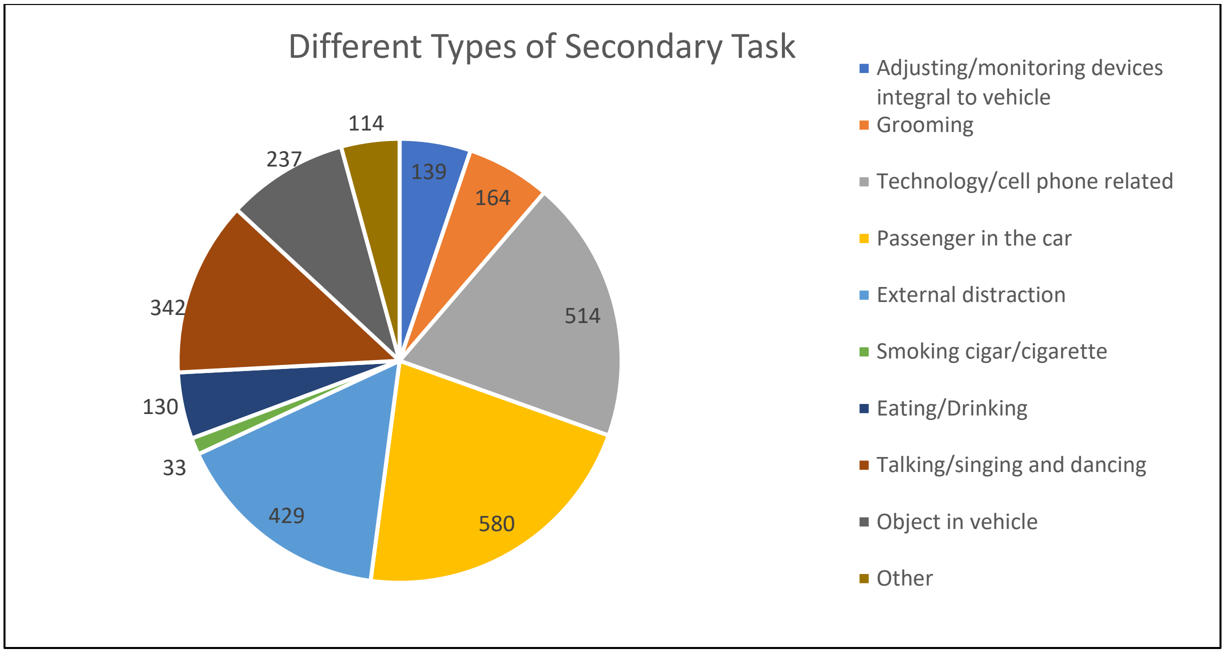

Figure 3: Different types of secondary tasks at signalized intersections

Among the 2,682 distracted events, about 22% of the total distracted events relate to a distraction caused by a passenger in the car. The second highest distraction events (19.2%) relate to technology/cell phone. External distractions (16%) and talking/singing and dancing (12.8%) also represent frequent distractions. These four distractions account for about 80% of the total distractions.

CHAPTER 4

MODELING AND ANALYSIS

The study develops models to (i) investigate the factors that contribute to distraction at signalized intersections and to (ii) identify the factors contributing to crashes and near-crashes at signalized intersections. The Python 3.8.5 programming language is used to estimate the models. From the 4,606 intersection events, 921 events are randomly selected as the testing dataset, and the remaining 3,685 events fall into the training dataset.

Logistic Regression

This paper investigates two dependent variables using Logistic Regression: (a) distracted (labelled as 1) vs. not distracted (labelled as 0) and (b) crash/near crash (labelled as 1) vs. no crash (labelled as 0). For each of these responses, the study uses a significance level of 0.10 to retain a variable in the model. In a Logistic Regression model, the probability that an observation is true is defined by the following equation:

$$P = \frac{1}{1+e^{-y}} \text{ where, } y_i = \beta x_i + \beta_0 \text{ for } i = 1 \text{ to } n \qquad (1)$$

Here, n is the number of events, and y_i is a utility function of the binary dependent variable, which takes the independent variables x_i. Here, β are the regression coefficients of the utility function, and P is the probability that the response is true (11, 23).

Decision Tree

In addition to the logistic regression models, this study develops classification decision trees to investigate the dependent variables. The classification decision tree model uses a tree structure to split the sample data into subsets. A parent node splits into exactly two child nodes and each child node may work as a parent node and split again. The topmost decision node in a tree corresponds to the best predictor called the root node. Gini measures the impurity of the node. A node is pure when all of its records belong to the same class (11). The study investigates the same dependent variables as the logistic regression model. To avoid overfitting, the researchers use pruning by setting the alpha value as 0.001 and the minimum leaf size at 5.

The study develops two decision tree models (dummy and numeric). The dummy model creates individual dummy variables for the n-1 categories in each categorical variable. In the numeric model, the researchers order the variables in each category according to either the percent distraction or their existing ordinal structure (e.g., age, years of driving experience, level of service).

CHAPTER 5
RESULTS

Distraction model

Logistic Regression

The logistic regression uses the categories with the largest number of occurrences as the reference cases and all independent variables are significant for $\alpha = 0.10$. Table 2 shows the coefficients and odd ratios of the significant predictors. The odd ratios and estimates indicate the influence the predictor exerts on driver distraction at a signalized intersection.

The predictors with a positive value increase and negative values decrease the likelihood of driver distraction. The odds ratio indicates that failure to use a seat belt indicates a 2.89 times higher probability of distraction. Two geometric conditions, which may make drivers more comfortable, increase the likelihood of distraction; drivers on divided roads appear 1.35 times more likely to experience distraction, and drivers on roads with 2-way left turn lanes become distracted 1.59 times more often. Only level of service C increases distraction (1.58 times higher); this shows that greater levels of congestion likely cause drivers to focus more on the driving task and lower congestion levels may not experience enough signal delay to support secondary tasks. Locations in less dense areas like churches and moderate residential locations increase the likelihood of distraction while urban areas with more external distractions also appear more distracting than business/industrial areas. A single income category appeared significant; the probability of driver distraction increases for income from \$50,000 - \$69,000. However, part-time status also increases the probability of distraction, and it may serve as a proxy for income. Construction signs/warnings represent a possible exception to other factors influencing distraction because drivers seem to become more distracted when faced with unfamiliar construction warnings; however, they may also make passengers more concerned and distracting and external objects may also become more distracting in a construction zone. Driver familiarity and mild congestion levels appear to increase the probability of distraction.

Some factors decrease the likelihood of distraction. Drivers appear to focus more on the driving task during rainy conditions because drivers appear about 1.3 times less likely to become distracted. The age results appear to align with the previous literature. All age groups over sixty years old experience a decrease in probability of distraction, which also appears to indicate they focus more on the driving task. The 25 to 29 year old age group appears to be more likely than the 20-24 age group to engage in secondary tasks. The lane the vehicle currently occupies appears to decrease distractions possibly by limiting the external distractions; those vehicles in a dedicated left turn land appear almost 1.19 times less likely to be distracted and those in lane 3 appear almost 1.38 times less likely to be distracted. The reductions in distraction appear to relate to greater focus on the driving task caused by weather, age, or vehicle position.

Table 2: Significant variables and associated estimates for Logistic Regression model (Distraction model)

Significant variables	**Reference Categories**	**Categories**	**Coefficient (P value)**	**Odd Ratios**
Driver seat belt use	Lap/shoulder belt properly worn	None used	0.7356 (0.000)	2.886734
Weather	No Adverse Conditions	Raining	-0.2571(0.088)	0.773270
Traffic Flow	Not divided – simple 2-way trafficway	Divided (median strip or barrier)	0.2998(0.000)	1.349687
		Not divided-center 2-way left turn lane	0.4107(0.002)	1.587865
Vehicle lane occupied	1	3	-0.3190(0.098)	0.726874
		Dedicated left turn lane	-0.1712(0.055)	0.842692
Traffic Density	Level-of-service B: Flow with some restrictions	Level of service C: Stable flow, maneuverability and speed are more restricted	0.4087(0.000)	1.584883
Traffic control	Traffic signal	Construction sign/warnings	0.5184(0.026)	1.679381
Locality	Business/Industrial	Church	0.6513(0.001)	1.917955
		Moderate Residential	0.3093(0.012)	1.362439
		Urban	0.2508(0.049)	1.285865
Work Status	Full-time	Part-time	0.3934(0.000)	1.481944
Income	Under $29,000	$50,000 to $69,999	0.2266(0.013)	1.254378
Age Group	20-24	25-29	0.2579(0.031)	1.294260
		50-54	-0.4673(0.006)	0.626681
		60-64	-0.4199(0.016)	0.657133
		65-69	-0.4270(0.009)	0.652473
		70-74	-0.8423(0.000)	0.430713
		75-79	-0.5818(0.000)	0.558879
		80-84	-0.6187(0.000)	0.538649
		85+	-0.4297(0.072)	0.652669

Decision Tree: Dummy model

The dummy DT tree has 12 leaves and selects years of driving as the root node. Drivers with less than 31.5 years of driving experience appear more likely (63.9%) to be distracted than those with more than 31.5 years of driving experience (45.6%). This closely aligns with the age variable in the logistic regression model. The experienced drivers appear more likely to be distracted when the level of service is C (58.7%) rather than other traffic conditions (43.8According the tree, less experienced drivers suffered more distraction (81.1% vs. 63.2%) when not using seat belt, which matches the results from the logistic regression model. The less experienced drivers using seatbelts are less distracted during rainy conditions (50.4%) than other conditions (63.9%). The less experienced drivers appear less likely to be distracted when in a dedicated left lane. Less experienced drivers without college degrees appear to be more distracted. Income, lane occupied, and years of driving experience also appear in the tree multiple times. Many of the variables in the tree appear similar with the variables identified in the logistic regression model, but locality and construction and traffic control do not appear.

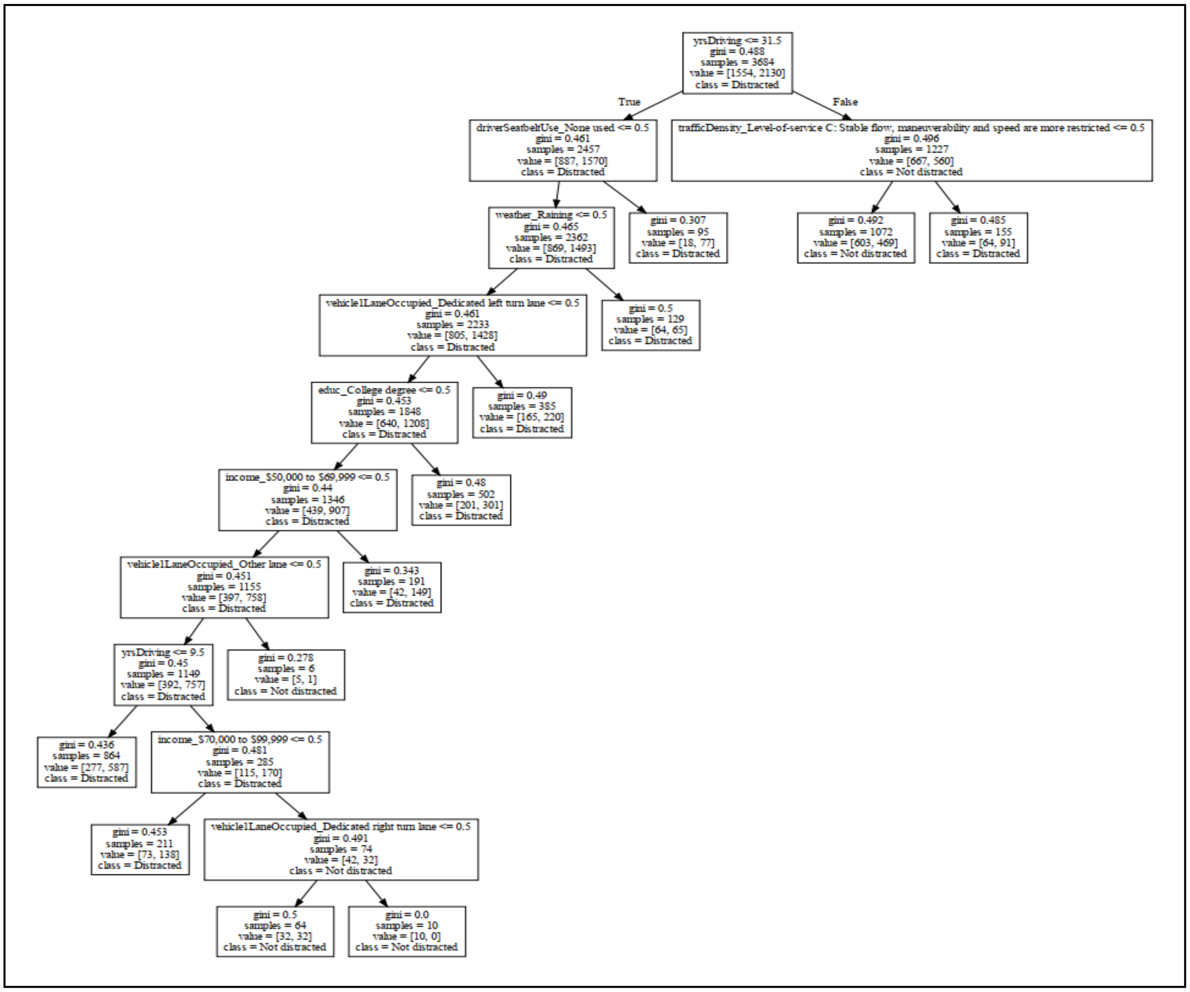

Figure 4: Decision Tree dummy model for distraction

Decision Tree: Numeric model

The numeric DT model also selects years of driving expereince as the root node; the less experienced drivers experienced more distraction (64.3% vs. 46.8%). Less experienced drivers appear more influenced by roadway geometry than more experienced drivers. The less experienced drivers also seem to be impacted by weather conditions and education level, which matches the dummy DT. In this tree, more experienced drivers appear less likely to be distracted during levels of service A (39.5%) than other levels of service (51.8%). The more experienced drivers in higher levels of congestion appear less likely to suffer distraction if their annual mileage is under 15,000 (48.4%). Those experienced drivers with more than 15,000 annual miles may also be less susceptible to distraction at night. The numeric model adds annual miles driven and lighting as independent variables to the variables selected by the dummy tree and drops income. The numeric model allows similar categories to group more easily than the dummy model.

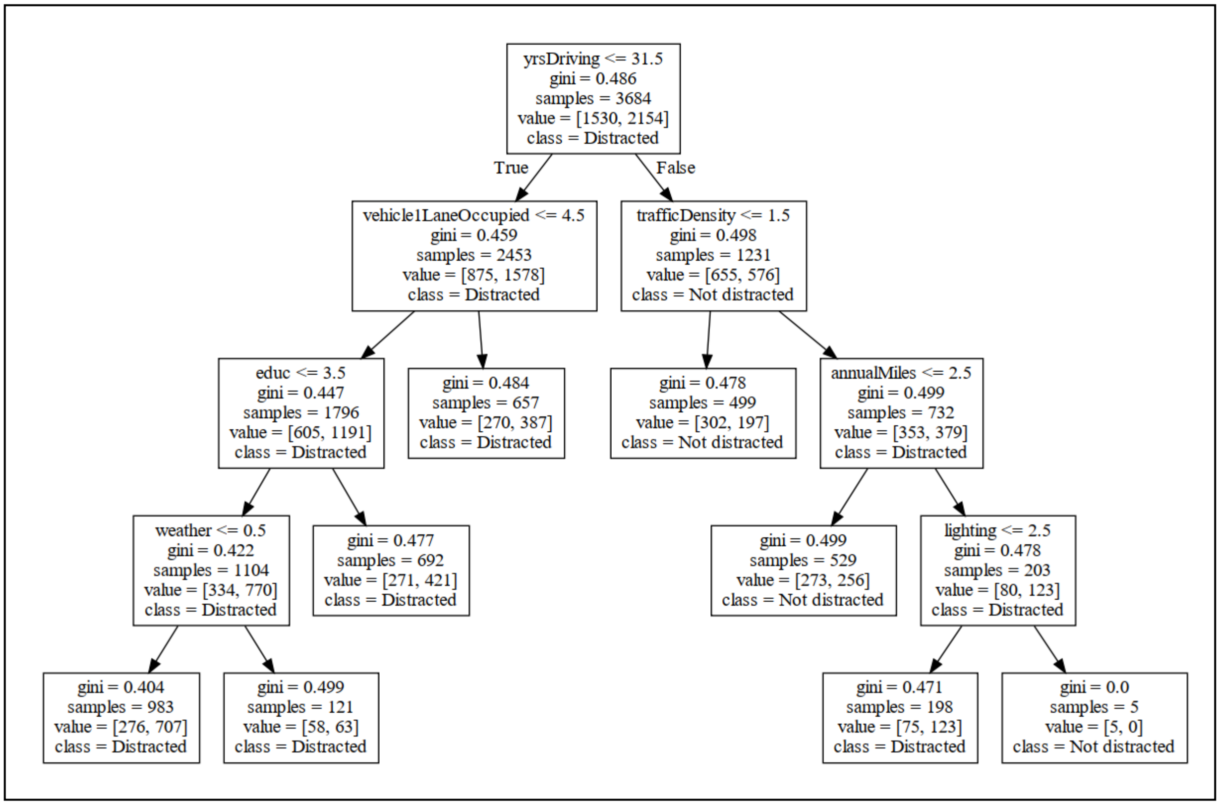

Figure 5: Decision Tree numeric model for distraction

Performance Measurement

The study uses confusion matrix metrics to evaluate model performance; the metrics include true positive (TP), true negative (TN), false positive (FP) and false negative (FN).

Confusion Matrix

Table 3 shows the confusion matrix for training and testing data performance for the distraction model. The logistic regression model correctly predicts 504 non-distracted events of 1,545 (accuracy 32.6%). On the other hand, the model correctly predicts 1743 distracted events of 2,139 (accuracy 81.49%). Thus, this model works well for predicting distracted events, although the model does not appear as strong at predicting non-distracted events. For the testing dataset, it generates similar accuracies for non-distracted events (accuracy 33.51%), and distracted events (81.4% accuate). Since the testing result is similar to the training result, this model appears properly fit.

Precision and Recall

Precision is computed as following:

$$\frac{\text{True positive}}{\text{Number of predicted positive}} = \frac{\text{True positive}}{\text{True positive+False positive}} \tag{2}$$

The regression model precision is 0.63 for training and 0.64 for testing data.

Recall was computed as below:

$$\frac{\text{True positive}}{\text{Number of actual positive}} = \frac{\text{True positive}}{\text{True positive+False negative}} \tag{3}$$

The logistic regression model recall is 0.81 for both training and testing data. According to the recall value, this model works really well; however, improvement in the precision value would be desirable.

For both decision trees, the precision values remain almost the same as the logistic regression. The numeric model has a slighly higher recall value than the dummy model, but they both fall below the recall value for the logistic regression model. The testing values for the tree models appears close to the training values but smaller for recall.

Table 3: Experimental results of driver distraction model

	Logistic Regression		**Decision Tree (Dummy)**		**Decision Tree (Numeric)**	
	Training	**Testing**	**Training**	**Testing**	**Training**	**Testing**
Accuracy	0.61	0.62	0.62	0.60	0.62	0.61
Confusion Matrix	504 1041 396 1743	127 252 101 442	650 904 502 1628	155 215 155 397	580 950 453 1701	151 243 117 411
Precision	0.63	0.64	0.64	0.65	0.64	0.63
Recall	0.81	0.81	0.76	0.72	0.79	0.78

Crash model

Logistic Regression

For the crash model, this logistic regression again uses the categories with the largest number of occurrences as the reference cases and all independent variables are significant for $\alpha = 0.10$. Table 4 shows the coefficients and odd ratios of the significant predictors. The odd ratios and estimates indicate the influence of different factors on crashes or near crashes at a signalized intersection.

After controlling for the confounding effects related to age groups, traffic density, grade, seat belt use, weather, surface conditions, alignment, grade, contiguous travel lanes, location, marital status, and income, some secondary tasks represent a significant factor in crashes and near crashes at signalized intersection. The odds ratio indicates a 4.10 times higher probability of crashes or near crashes for drivers distracted by an object inside the vehicle, which represents the largest causal factor in the model. The model identifies three other secondary tasks as causal factors for crashes and near crashes at signalized intersections. Technology/cell phone related distraction causes the increase the probability (1.74 times higher) of a crash or near crash to increase 1.74 times higher than the no distraction case. Adjusting/monitoring devices integral to vehicle and grooming also increase the probability (1.44 times higher for both) of crash and near crashes at signalized intersection. The distraction cases that require more individual attention inside the vehicle to complete appear to contribute significantly to crashes and near crashes at intersections while distractions from passengers and outside the vehicle appear insignificant.

Table 4: Significant variables and associated estimates for Logistic Regression model (Crash model)

Significant variables	Reference Categories	Categories	Coefficient (P value)	Odd Ratios
Driver Seat belt Use	Lap/shoulder belt properly worn	None Used	0.3612(0.037)	1.435069
Weather	No Adverse Conditions	Mist/Light Rain	0.3041(0.083)	1.355356
Surface Condition	Dry	Snowy	0.7721(0.046)	2.164261
Contiguous Travel Lanes	3	4	-0.2497(0.003)	0.779054
		6	-0.3251(0.023)	0.722426
Traffic Density	Level-of-service B: Flow with some restrictions	Level-of-service A1: Free flow, no lead traffic	-1.3645(0.000)	0.255508
		Level-of-service A2: Free flow, leading traffic present	-0.6303(0.000)	0.532409
		Level-of-service C: Stable flow, maneuverability and speed are more restricted	0.6103(0.000)	1.840978
		Level of service D: Unstable flow -temporary restrictions substantially slow driver	0.8829(0.000)	2.417913
Traffic Control	Traffic signal	No traffic control	-0.5002(0.000)	0.606381
Alignment	Straight	Curve left	0.3937(0.039)	1.482494
Grade	Level	Grade Down	0.7349(0.000)	2.021997
		Grade Up	0.7041(0.000)	2.085374
Locality	Business/Industrial	Urban	0.2334(0.088)	1.262938
Average annual mileage	10,000 – 15,000 miles	5,000 – 10,000 miles	-0.3880(0.000)	0.678384
Marital Status	single	divorced	-0.2897(0.072)	0.748519
		married	-0.2780(0.000)	0.757316
income	Under $29,000	$30,000 to $39,999	-0.2941(0.011)	0.745174
		$40,000 to $49,999	-0.2798(0.036)	0.755953
Age Group	20-24	30-34	0.3237(0.054)	1.382242
		35-39	0.3305(0.092)	1.391644
		75-79	-0.4160(0.014)	0.659687
Secondary Task	No Secondary Tasks	Adjusting/monitoring devices integral to vehicle	0.3660(0.068)	1.441917
		Grooming	0.3660(0.060)	1.441988
		Object in vehicle	1.4123(0.000)	4.105327
		Technology/cell phone related	0.5580(0.000)	1.747257

Decision Tree: Dummy model

Decision tree models are run without marital status and income as their importance appears limited according to the logistic regression and their role seems unlikely to be a causal factor. The Dummy DT tree selected the traffic density as the root node; the probability of crashes or near crashes at signalized intersections increase (21.76%) when the level of service is A1. Two secondary tasks play significant roles in characterizing the data samples. A distraction from an object in the vehicle separates 62.7% of the samples, and a technology/cell phone related distraction separates 46.9% of the samples in two different branches. In non-free-flow conditions, an object in the vehicle distraction results in a crash or near crash 75.2% of the time while the no distraction case results in a crash or near crash 44.8% of the time. Less congested conditions (A2) decrease the impact of a technology or cell phone related secondary task. It only results in a crash or near crash 48.7% of the time while LOS B and worse results in a crash or near crash 62.6% of the time. The adjusting/monitoring devices integral to vehicle and grooming secondary tasks do not appear in the dummy decision tree, but passenger in car and talking/singing or dancing split 7.7% and 4.6% of the events.

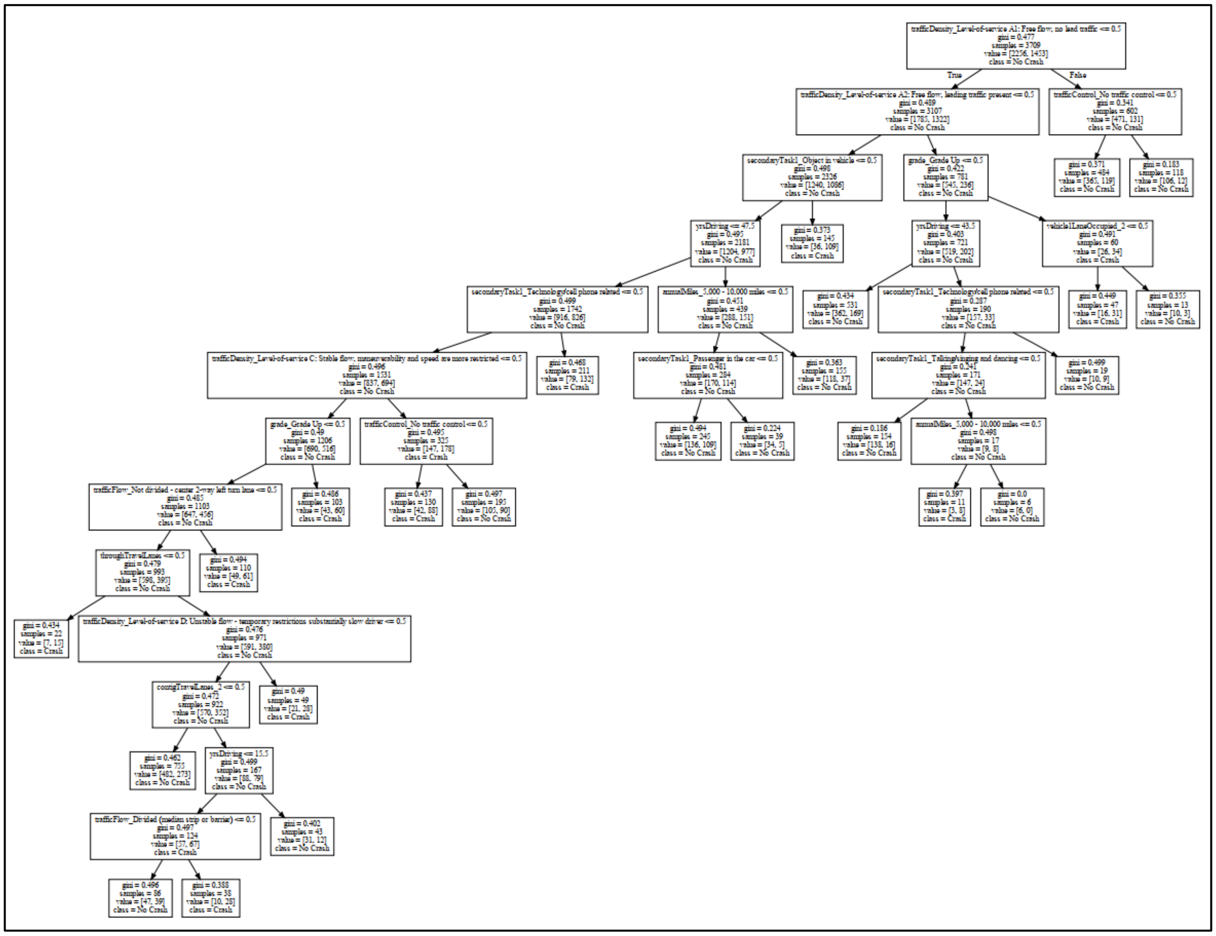

Figure 6: Decision Tree dummy model for crash

To provide a better description of the importance of distraction in the complex decision tree, Table 5 provides the variable importance of the predictors to illustrates the variables' impacts on crashes/near crashes. After free flow conditions, the object in vehicle secondary task represents the most important variable. The technology/cell phone related distraction is the fifth most important variable. The passenger in the car distraction appears as the tenth most important variable and talking/singing and dancing remains more important than some traffic densities and all travel lane variables.

Table 5: Importance of Variables for Dummy Decision Tree Model

Significant variables	**Categories**	**Importance**
Traffic Density	Level-of-service A1: Free flow, no lead traffic	20.38
	Level-of-service A2: Free flow, leading traffic present	14.84
	Level-of-service C: Stable flow, maneuverability and speed are more restricted	3.44
	Level-of-service D: Unstable flow – temporary restrictions substantially slow driver	1.57
Secondary Task	Object in vehicle	11.73
	Technology/cell phone related	6.93
	Passenger in the car	3.16
	Talking/singing and dancing	1.93
Years of Driving	Continuous Variable	10.33
Grade	Grade Up	6.77
Traffic Control	No traffic control	5.23
Annual Miles	5,000 – 10,000 miles	4.40
Traffic Flow	Not divided – center 2-way left turn lane	2.28
	Divided (median strip or barrier)	1.98
Vehicle1LaneOccupied	Lane 2	1.75
Contiguous Travel Lanes	Lane 2	1.59

Decision Tree: Numeric model

The numeric DT model also selects traffic density as a root node. Drivers appear more influenced by secondary tasks in congested conditions. For non-free flow conditions, the probability of a crash or near crash is 46.8% and the probability decreases to 26.8% for free-flow conditions. For non-free flow conditions distraction from an object in vehicle and technology or cell phone creates a crash/near crash probability of crash or near crash by 62% versus 43.6% for all other distraction and no distraction cases. The object in the vehicle again poses the greater risk (73.33% versus 56.3%. For free flow conditions, distraction from adjusting/monitoring devices integral to vehicle, talking/singing and dancing, and eating/drinking as also appear as risk factors for crashes or near crashes 32.7% versus 20.4%.

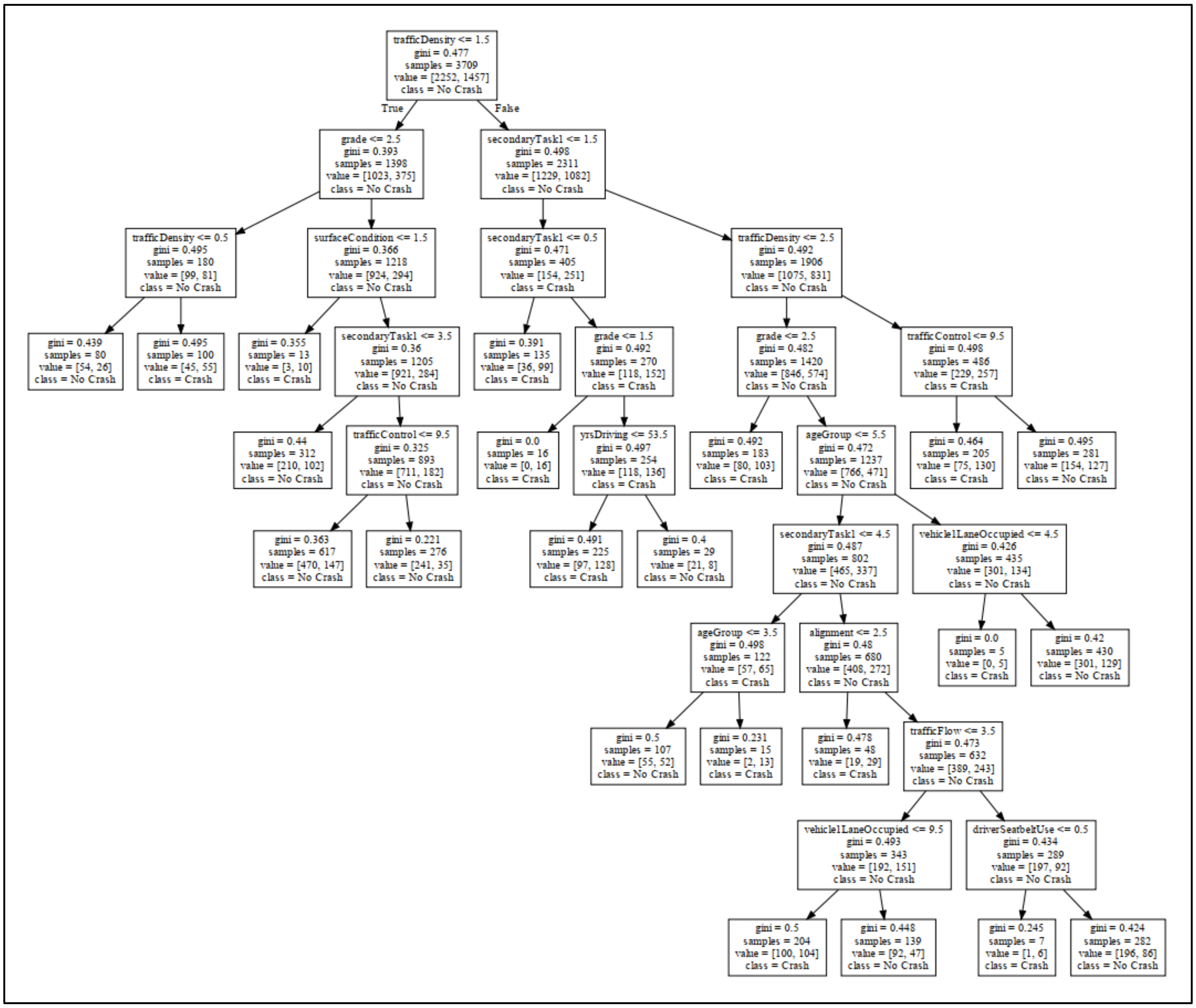

Figure 7: Decision Tree Numeric model for crash

Variable importance of predictors in Table 6 explain the general influence of different variables on crashes or near crashes. Traffic density represents the most importance independent variable with 40.16 importance while distraction from secondary tasks represents the second most important variable with 18.08 importance.

Table 6: Importance of Variables for Numeric Decision tree Model

Significant Variables	Importance
Traffic Density	40.16
Secondary Task	18.08
Grade	14.45
Traffic Control	5.93
Age Group	5.13
Vehicle1 Lane Occupied	4.57
Surface Condition	3.44
Traffic Flow	2.19
Years of Driving	2.07
Alignment	2.02
Driver Seat belt Use	1.96

Performance Measurement

Table 7 shows the confusion matrix for the training and testing data performance for the crash model. The logistic regression medel correctly predicts 1,857 no-crash events of 2,248 (accuracy 82.6%); however, the model only correctly predicts 632 crash events of 1,461 (accuracy 43.3%). For the testing dataset, it generates similar accuracies for the no-crash events (accuracy 78.9%), and crash events events (40.67% accuate). Since the testing result is similar to the training result, this model also appears properly fit. The precision value is 0.62 for training and 0.55 for testing. The recall value is 0.43 for training and 0.41 for testing.

Table 7: Experimental results of Logistic Regression model for Crash

	Logistic Regression		**Decision Tree (Dummy)**		**Decision Tree (Numeric)**	
	Training	**Testing**	**Training**	**Testing**	**Training**	**Testing**
Accuracy	0. 67	0.64	0.68	0.65	0.67	0.68
Confusion Matrix	1857 391 829 632	449 120 213 146	1950 306 893 560	472 89 237 130	1794 458 759 698	467 98 195 168
Precision	0. 62	0.55	0.65	0.59	0.60	0.63
Recall	0. 43	0.41	0.39	0.35	0.48	0.46

The numeric model has a considerably higher recall value than other models, and the numeric DT model appears to outperform the other models especially for the testing dataset.

CHAPTER 6

CONCLUSIONS

This study focuses on the causes of diver distraction and influencing factors for crashes and near crashes at signalized intersection using naturalistic driving data from SHRP2 database. This study identifies that traffic density, age, driving experience, seat belt use, weather and education have an extensive impact on driver distraction. The level of service on the approach legs and distraction from secondary tasks, especially the presence of an object in vehicle and technology or cell phone related distractions, represent the most influential factors for determining crash/near crash risk at signalized intersection. From previous research, it was found that reaching for objects and manipulating objects may increase the probability of crashes or near creases. In this study, it is found that this statement is also true for signalized intersection. Adverse weather condition specifically raining decreases driver distraction, but it may increase the probability of crashes or near crashes at signalized intersection. However, this study can be improved with more data to show vehicle interactions with different vehicle types such as heavy vehicles and clearly separating the crash models for those vehicles queued at the signalized intersection and those vehicles that do not queue at the intersection because the crash related factors may vary significantly between these two cases. Future research will also investigate the influence of vehicle position in the queue at the signalized intersections on secondary tasks and vehicle conflicts.

REFERENCES

1. Arvin, R., Kamrani, M., & Khattak, A. J. (2019). The role of pre-crash driving instability in contributing to crash intensity using naturalistic driving data. *Accident Analysis & Prevention*, *132*, 105226.

2. Qi, Y., Vennu, R., & Pokhrel, R. (2020). Distracted Driving: A Literature Review. *FHWA-ICT-20-004*.

3. Wang, J. S., Knipling, R. R., & Goodman, M. J. (1996, October). The role of driver inattention in crashes: New statistics from the 1995 Crashworthiness Data System. In *40th annual proceedings of the Association for the Advancement of Automotive Medicine* (Vol. 377, p. 392).

4. De Winter, J., van Leeuwen, P. M., & Happee, R. (2012, August). Advantages and disadvantages of driving simulators: A discussion. In *Proceedings of measuring behavior* (Vol. 2012, p. 8th).

5. Käppler, W. D. (1993, June). Views on the role of simulation in driver training. In *Proceedings of the 12th 31uropean annual conference on Human decision making and manual control* (pp. 5-12).

6. Evans, L. (2004). Traffic Safety; Science Serving Society. *Bloomfield Hills, MI*, *179*.

7. Stutts, J. C., Reinfurt, D. W., Staplin, L., & Rodgman, E. (2001). The role of driver distraction in traffic crashes.

8. Wang, B., Hallmark, S., Savolainen, P., & Dong, J. (2017). Crashes and near-crashes on horizontal curves along rural two-lane highways: Analysis of naturalistic driving data. *Journal of safety research*, *63*, 163-169.

9. Wu, J., & Xu, H. (2018). The influence of road familiarity on distracted driving activities and driving operation using naturalistic driving study data. *Transportation research part F: traffic psychology and behaviour*, *52*, 75-85.

10. Calvo, J. A., Baldwin, C., & Philips, B. (2020). Effect of age and secondary task engagement on motor vehicle crashes in a naturalistic setting. *Journal of safety research*, *73*, 297-302.

11. Das, A., Ghasemzadeh, A., & Ahmed, M. M. (2018). *A comprehensive analysis of driver lane-keeping performance in fog weather conditions using the SHRP2 naturalistic driving study data* (No. 18-06242).

12. Kidd, D. G., Tison, J., Chaudhary, N. K., McCartt, A. T., & Casanova-Powell, T. D. (2016). The influence of roadway situation, other contextual factors, and driver characteristics on the prevalence of driver secondary behaviors. *Transportation research part F: traffic psychology and behaviour, 41*, 1-9.

13. Sheykhfard, A., & Haghighi, F. (2020). Driver distraction by digital billboards? Structural equation modeling based on naturalistic driving study data: a case study of Iran. *Journal of safety research*, *72*, 1-8.

14. Bakhit, P. R., Guo, B., & Ishak, S. (2018). Crash and near-crash risk assessment of distracted driving and engagement in secondary tasks: a naturalistic driving study. *Transportation research record*, *2672*(38), 245-254.

15. Simons-Morton, B. G., Gershon, P., O'Brien, F., Gensler, G., Klauer, S. G., Ehsani, J. P., & Dingus, T. A. (2020). Crash rates over time among younger and older drivers in the SHRP 2 naturalistic driving study. *Journal of safety research*, *73*, 245-251.

16. Huisingh, C., Owsley, C., Levitan, E. B., Irvin, M. R., MacLennan, P., & McGwin, G. (2019). Distracted driving and risk of crash or near-crash involvement among older drivers using naturalistic driving data with a case-crossover study design. *The Journals of Gerontology: Series A*, *74*(4), 550-555.

17. Seacrist, T., Douglas, E. C., Hannan, C., Rogers, R., Belwadi, A., & Loeb, H. (2020). Near crash characteristics among risky drivers using the SHRP2 naturalistic driving study. *Journal of safety research*, *73*, 263-269.

18. Chandran, T. (2018). *Driver Glance Behaviour in Intersection Crashes: A SHRP2 Naturalistic Data Analysis* (Master's thesis).

19. Dinakar, S., & Muttart, J. (2019). *Driver behavior in left turn across path from opposite direction crash and near crash events from SHRP2 naturalistic driving* (No. 2019-01-0414). SAE Technical Paper.

20. Morgenstern, T., Petzoldt, T., Krems, J. F., Naujoks, F., & Keinath, A. (2020). Using European naturalistic driving data to assess secondary task engagement when stopped at a red light. *Journal of safety research*, *73*, 235-243.

21. Lv, B., Yue, R., & Zhang, Y. (2019). The Influence of Different Factors on Right-Turn Distracted Driving Behavior at Intersections Using Naturalistic Driving Study Data. *Ieee Access*, *7*, 137241-137250

22. Rahman, Z., Martinez, D., Martinez, N., Zhang, Z., Memarian, A., Pulipati, S., … & Rosenberger, J. M. (2018). Evaluation of cell phone induced driver behavior at a type II dilemma zone. *Cogent Engineering*, *5*(1), 1436927.

23. Rahman, Z., Memarian, A., Madanu, S., Iqbal, G., Anahideh, H., Mattingly, S. P., & Rosenberger, J. M. (2018). Assessment of the impact of lane width on arterial crashes. *Journal of Transportation Safety & Security*, *10*(3), 229-250.

24. Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences*, *113*(10), 2636-2641.

25. Victor, T., Dozza, M., Bärgman, J., Boda, C. N., Engström, J., Flannagan, C., … & Markkula, G. (2015). *Analysis of naturalistic driving study data: Safer glances, driver inattention, and crash risk* (No. SHRP 2 Report S2-S08A-RW-1).

APPENDIX A

Table A.1: Grouping of different types of secondary tasks

Secondary Tasks	Group
Adjusting/monitoring climate control	Adjusting/monitoring devices integral to vehicle
Adjusting/monitoring other devices integral to vehicle	
Adjusting/monitoring radio	
Inserting/retrieving CD (or similar)	
Applying make-up	Grooming
Biting nails/cuticles	
Brushing/flossing teeth	
Combing/brushing/fixing hair	
Other personal hygiene	
Reaching for personal body-related item	
Cell phone, Browsing	Technology/cell phone related
Cell phone, Dialing hand-held	
Cell phone, Dialing hand-held using quick keys	
Cell phone, Holding	
Cell phone, Locating/reaching/answering	
Cell phone, other	
Cell phone, Talking/listening, hand-held	
Cell phone, Texting	
Tablet device, Viewing	
Child in adjacent seat - interaction	Passenger in the car
Child in rear seat - interaction	
Passenger in adjacent seat - interaction	
Passenger in rear seat - interaction	
Distracted by construction	External distraction
Looking at an object external to the vehicle	
Looking at pedestrian	
Looking at previous crash or incident	
Other external distraction	
Smoking cigar/cigarette	Smoking cigar/cigarette
Reaching for cigar/cigarette	
Drinking from open container	Eating/Drinking
Drinking with lid and straw	
Drinking with lid, no straw	
Drinking with straw, no lid	
Eating with utensils	
Eating without utensils	
Reaching for food-related or drink-related item	

Secondary Tasks	**Group**
Moving object in vehicle	Object in vehicle
Object in vehicle, other	
Removing/adjusting clothing	
Removing/adjusting jewelry	
Removing/inserting/ adjusting contact lenses or glasses	
Reaching for object, other	
Pet in vehicle	
Other known secondary task	Other
Other non-specific internal eye glance	
Reading	
Unknown	
Writing	
Talking/singing, audience unknown	Talking/singing and dancing
Dancing	