Characterizing Technology Impacts on Driving Behaviors, Crash Risks, and Infrastructure Performances

Jobaidul Boni

Characterizing Technology Impacts on Driving Behaviors, Crash Risks, and Infrastructure Performances

By

Jobaidul Alam Boni

Dissertation

Submitted to the Faculty of

The University of Texas at Arlington

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Transportation Engineering

in the Department of Civil and Environmental Engineering

The University of Texas at Arlington

December 2023

Supervising committee:

Dr. Kate Kyung Hyun, Supervising Professor

Dr. Stephen Mattingly

Dr. Nelson Gomez-Torres

Dr. Seri Park

Abstract

In recent years, an increase in driver distraction appears due to the rise in smartphone usage and the introduction of social media. Researchers put significant efforts in examining the impacts of distracted driving, mostly focused on distraction like texting or phone call. However, scant research exists to identify the underlying factors causing the distracted driving particularly caused by social media or showing their safety implications at complex geometries such as intersections and highways. This dissertation offers three independent studies reviewing the impact of technology on drivers' actions using field tests and simulation experiments. The first field tests conducted at three Texas intersections revealed that cell phones are the primary cause of most distractions, resulting in up to 6.6 second of lost time. The second paper uses machine learning classifiers to identify various distractions based on driving behaviors. The resulting behavioral analysis also indicates that social media distractions significantly impact drivers' angular and lateral patterns. Among the different models, the Multi-Layer Perception classifiers demonstrated a strong capacity for detection, achieving over 75% accuracy. The third chapter examined the relationship between distracted driving and risky driving practices, showing that more engaging social media activities like watching videos can lead to more lateral driving errors. In addition, GPS poses a greater risk than less-engaging social media activities like checking feeds. Demographic factors also play an important role in causing particular actions, for example, lower income women tend to keep shorter headways when they engage in social media activities. This dissertation offers valuable insights into behavioral attributes of distracted driving patterns, paving the way for targeted interventions and countermeasures to mitigate the risks associated with distracted driving caused by social media.

Keywords: Distracted Driving, Startup Lost Time, Social Media, Machine Learning, Association Rule Mining, Driving Behavior

Acknowledgements

I humbly begin this dissertation by acknowledging that all my achievements are attributed to the grace of the almighty Allah.

Earning a Ph.D. has been a monumental dream, and I owe a great deal to many people who helped me get to this crucial stage in my life. Firstly, I sincerely thank my mentor and supervisor, Dr. Kate Hyun, whose unwavering guidance has been invaluable to me throughout my Ph.D. journey. Her unwavering support and wise advice helped me overcome difficult times and academic obstacles, guiding me through academic endeavors and personal hardships. Dr. Kate gave me a unique chance to excel in transportation engineering and generously shared her knowledge, skills, and techniques with me along the way. Collaborating with her is an unforgettable and pivotal chapter in my life, marked by receptiveness to novel ideas and abundant resources to actualize them. I'll always carry profound gratitude for everything she's done.

I extend my sincere gratitude to Dr. Stephen Mattingly, an invaluable mentor throughout my academic journey at the University of Texas at Arlington. Over the past six years, collaborating with him on multiple projects has enriched me with invaluable knowledge. I will forever cherish his guidance. Additionally, I wish to recognize the contributions of Dr. Nelson Gomez-Torres and Dr. Seri Park, members of my committee, whose support and insightful suggestions greatly propelled my research. I would also like to thank Dr. Jim Williams for his constant support and guidance during my entire graduate program.

Infinite gratitude to my family – my parents, in-laws, and my little sister for always being my powerhouse and being there with me no matter what the situation is. I believe I didn’t let them down. I would like to express my sincere gratitude to my wife, Farzana Rahman Chowdhury, for her steadfast assistance and inspiration during trying times. Her understanding and patience have been crucial in assisting me in reaching this goal.

In addition, my deepest appreciation to a few dearest lab mates, including but not limited to Ziaur Rahman, Mintu Miah, Jaesik Choi, Mehrdad Arabi, Perez Juana, Farah Naz, Sina Famili and Troyee Saha for their endless encouragement and compassionate ear that made my Ph.D. journey pleasant.

Dedication

In the name of the Most Gracious and Merciful Allah,

I dedicate this thesis to Farzana Rahman Chowdhury, my dear wife, who has inspired me throughout my life. Your unwavering love, support, and encouragement have anchored my journey.

To my cherished parents, Shampa Alam and Mofiqul Alam, whose unwavering faith, selfless love, and sacrifices have inspired me to pursue my goals. My path is still being shaped by your teachings and values.

Above all, I want to express my sincere gratitude to the Almighty Allah, whose favor, direction, and blessings have illuminated every step of this academic journey. This trip would not have been possible without His mercy and direction.

Contents

List of Tables

List of Figures

Chapter 1: Introduction

Driving is one of the most complex human activities as delicate interactions occur between the driver, the vehicle, and the environment (road characteristics and properties, driver intention, incidental effects, etc.). Even though driving demands constant attention to the road and moving traffic, secondary tasks like talking on the phone, reading a billboard, speaking, eating, and adjusting the radio frequently cause drivers to become distracted. Driver behaviors affected by distractions due to all these distinct elements may lead to aggressive driving, lane drifting, or harsh/frequent braking (Ghandour et al., 2021). According to NHTSA, distracted driving was the leading cause in 14% of all traffic crashes recorded by the police in the USA in 2018, and 8% of those crashes resulted in fatalities. The police reported 2,841 fatalities due to distraction in 2018, and their report shows that losing concentration is the cause of 61% of those crashes. In particular, 21% of those crashes involved other passengers, and 17% involved pedestrians and bicyclists (NHSTA, 2020).

Driver distraction represented a primary concern for traffic engineers and planners before the availability of technology in personal vehicles (Stutts et al., 2001). However, the proliferation of cellular telephones, vehicle navigation systems, wireless internet capabilities, wireless messaging, and other in-vehicle technology greatly concerned the NHTSA; they expressed concern for traffic safety and performance (Stutts & Hunter, 2003). Accordingly, distracted driving has been examined in the past several years from various perspectives related to traffic engineering and transportation. Studies examined how distracted driving affected driver performance, traffic flow, and capacity. Rachel et al. (2019) provided a comprehensive review on distracted driving due to mobile phones, inbuilt vehicle infotainment systems, and wearable devices, and found that handheld electronic devices distracted drivers more easily than devices that only require voice interactions. The expansive expansion of the cellphone market is frequently linked to the burgeoning demand for smartphones. Presently, approximately 95% of the U.S. population possesses a cellphone, and the proportion of smartphone users among cellphone owners has surged from 35% in 2011 to 81% in early 2019 (PRC, 2023).

Studies found that using a cell phone while driving appeared riskier than any other distracting activity for drivers (Drews et al., 2009). In 2013, the Centers for Disease Control and Prevention reported that about 69% of US drivers talked on their cell phones, and 31% emailed or used text messages while driving (CDC, 2013). Cell phone use and drivers' perceived reaction times have been the subject of studies which found a significant increase in drivers' perception and reaction time while texting and talking (Alshabibi, 2021). However, relatively few studies investigated the effects of distracted driving caused by cell phones on intersection capacity. A few studies showed that a cell phone may increase start-up loss time and average saturation flow because it is one of the visual, cognitive, and manual distractions that drivers experience when responding to changes in traffic signals. A longer queue delay decreases the signal capacity and raises the level of service (LOS). Another critical issue about distracted driving is, until now, most of the researchers concentrated on conventional distracted activities: talking on the phone and texting. These studies largely overlook new and emerging distracted behaviors such as playing music players, using navigational tools, and engaging social networking (McNabb & Gray, 2016).

The introduction of different social apps increased the use of smartphones during driving maneuvers. Social media has transformed the way people connect, leveraging Internet-based communication services. Various platforms like Twitter, Facebook, LinkedIn, and YouTube attracts millions of users worldwide. For example, Twitter alone records 313 million active users monthly, generating a staggering 500 million tweets globally each day (Li et al., 2020). Although social media apps offer usefulness, efficiency, safety, and entertainment in various scenarios, their use while driving leads to adverse effects. Pokémon Go, an augmented reality game encouraging driving for catching Pokémon, resulted in increased traffic crashes, physical injuries, and fatalities. In a 148-day period in an Indiana county, Pokémon Go usage caused damages estimated at $5 to $25 million, along with 31 injuries and two deaths (Faccio and McConnell, 2020). Expectedly, using or browsing a social media app had the same negative impact as texting on a driver’s performance (Hashash et al., 2019).

Current research faces a significant challenge in fully gauging the risks posed by distracted driving in comparison to focused driving. The Center for Disease Control and Prevention categorizes distracted driving into three main types: visual (eyes off the road), manual (hands off the wheel), and cognitive. Cell phone use, especially texting, encompasses all three types of distraction (CDC, 2011). An additional critical aspect involves understanding if drivers are aware of the extent of their errors while distracted. Horrey et al. (2008) conducted a study where drivers were asked to use cell phones while driving in a simulator. The findings revealed numerous mistakes, including swerving, inconsistent speed, and delayed reactions, all of which significantly elevate the risk of crashes. However, when these drivers were asked to assess their performance, most failed to recognize their errors and couldn't recall the potential consequences of their driving mistakes (Horrey et al., 2008).

A significant portion of risky driving research relies heavily on self-reported data from drivers or information obtained from police crash reports, as well as video-recorded data analysis. These sources, however, have limitations. Self-reported and police-reported crash data may not provide an accurate picture of crash patterns and driving errors. Similarly, while data analysis from recordings provides insight, it frequently falls short of capturing the nuances of various cell phone-related distractions, highlighting a significant gap in current literature. Given that younger drivers are frequent cell phone users, there is an urgent need to investigate the specifics of social networking apps. These apps go beyond text-based communication, necessitating a more thorough examination of their impact on driving behavior.

Moreover, previous research on distracted driving has primarily concentrated on general roadway environments with no complex geometry or roadway condition. For example, Vieira and Larocca's (2017) found that drivers struggled to replicate the same error-free driving behavior when switching from normal driving to dealing with a simple mental workload. This emphasizes the importance of understanding how drivers interact in a variety of roadways and traffic environments.

Another limitation of distracted-driving research is its emphasis on vehicle crashes that result in property damage or injury. There has been insufficient focus on measuring the negative effects of distracted driving on a driver's response to anticipated stimuli and roadway performance. In conclusion, the current literature on risky driving frequently relies on limited sources, such as self-reported data and crash reports, while failing to thoroughly investigate the complexities of cell phone-related distractions, social networking apps, diverse roadway environments, and the broader spectrum of impacts on driver responses beyond crashes and injuries.

To fill these gaps, The study will try to address the following key questions using advanced techniques encompassing field observations, driving simulations, and driver surveys:

- How does new distractions such as electronic devices and in-vehicle entertainment systems characterize drivers' behavior at intersections, and how much is intersection capacity reduced due to drivers?
- How does distracted driving impact driving performance, and are there differences in the impact by the types of distraction sources?
- How risky is distracted driving, and how can we evaluate the risk level of distracted driving behavior?

This thesis is divided into three distinct chapters representing the relevant literature review, research gap, methodologies, and analysis to answer the research questions above.

Chapter 2: Characterizing Technology's Influence on Distractive Behavior at Signalized Intersections

2.1 Introduction

State and local agencies placed traffic control devices at intersections to address roadway operations and safety (FHWA, 2013). Properly installed and operated traffic signals played an essential role in achieving optimal performance at an intersection by assigning vehicular and pedestrian rights of way. In urban transportation, the performance of signalized intersections required attention because they represented the primary source of delay for urban roads. Common measures to evaluate intersection performance included delay, queue length, and cycle failure (Zheng et al.,2013). While demand modeling and simulation might provide reasonable estimates of turning movements and queue length at intersections, factors impacting the saturation flow rate and lost time appeared more uncertain.

Turning movement counts, saturation flow rate, lost time, and queue length represented essential parameters for planning, designing, and controlling a signalized intersection. These factors and other operating parameters, traffic conditions, roadway parameters, and environmental conditions might influence a signalized intersection's optimal timing and performance (Hadiuzzaman et al., 2008). According to the Highway Capacity Manual 2010 (HCM, 2010), capacity represented a planning level estimate that incorporated the effect of lost time and typical saturation flow rates. The factors that affected the startup lost time included vehicle type and road gradient, pedestrians in the intersection, perception-reaction time, which varied from driver to driver, and psychological factors. Under ideal conditions, physiological conditions caused most startup lost time variance; however, any distraction might also significantly impact a driver's perception reaction and response time, which influenced both saturation flow and startup lost time (Çalışkanelli et al., 2017). Many researchers already investigated the relationships between saturation flow and intersection geometry, such as lane type, peak hour, queue length, and green time (Alembo, 2014; Bivina et al., 2016; Shawky & Al-ghafli, 2016). However, the impact of new distractions, such as electronic devices and in-vehicle entertainment systems, on distracted driving and the startup lost time at intersections remained less investigated. In addition, the Highway Capacity Manual (2016) identified an average control delay of 10 to 15 seconds for a Level of Service B (HCM, 2016). The emergence of new technology (i.e., smartphones and in-vehicle entertainment systems) might increase driver distraction and delay at traffic signals.

The saturation flow rate characterized the initial capacity of an approach, even a highly congested approach could not achieve this flow rate throughout a green indication. At the initiation of the green phase, the first driver in the queue observed and reacted to the signal change and accelerated through the intersection from a stand-still, which created a relatively long first headway. This process continued until a certain period when the startup reaction and acceleration no longer influenced the headways. Hence, the startup lost time represented the additional time in seconds that the first few vehicles in a queue at a signalized intersection used beyond the saturation headway (Rouphail et al., 2001).

Saturation flow represented an essential input for optimal signal timing; therefore, a slight variation in saturation flow could significantly change the optimal cycle and phase lengths, which affected the efficiency and operation of an urban street system. As a macro performance measure of intersection operations, saturation flow indicated the potential capacity under ideal operating conditions assuming no heavy vehicles or pedestrians/cyclists and a single movement type (i.e., only straight movement or only turning movement) (HCM, 2016). The HCM (2010) prescribed an ideal saturation flow rate of 1,900 passenger cars per hour per lane, which equated to a saturation headway of about 1.9 seconds per vehicle (HCM, 2010). Startup lost time represented another important parameter in signalized intersection performance. It measured the additional time the first few vehicles consumed in a queue at a signalized intersection above and be- yond the saturation headway (FHWA, 2013). The HCM 2010 indicated that the first four vehicles in a queue generally lose two seconds of green time combined to accelerate to their desired speed. However, vehicle type and gradient, pedestrians in the intersection, perception/reaction time (which varies from driver to driver), and psychological factors also affected startup lost time.

A study in 1965 confirmed that listening to the radio caused driver inattention and distraction, which resulted in prolonged responses during complex maneuvers on the road (Brown, 1965). A study by the National Highway Traffic Safety Administration (NHTSA) in 1997 estimated that 35-50 percent of police-reported crashes at an intersection involved some form of driver's inattention. The inattentions resulted from fatigue, "lost in thoughts', eating or drinking, watching outside objects, talking inside the car with an occupant, and other forms of distraction (Goodman et al., 1999). A study in 2000 suggested that watching an outside object, person, or event (30%), moving any object inside the car or adjusting the radio (15%), and talking to a person in the vehicle (11%) represented the most common driver distractions (Stutts et al., 2001). Choudhury and Velaga (2020) designed a simulated environment to understand the impact of the most common distracted activities - eating and drinking - on drivers' stop and cross decisions at the onset of yellow indication. They used a scenario consisting of six urban signalized intersections and showed up to 12% increase in crossing time and 7% reduction in stopping time compared to non-distracted driving (Choudhary & Velaga, 2020). The proliferation of cellular telephones, vehicle navigation systems, wireless internet capabilities, wireless messaging, and other in-vehicle technology greatly concerned the NHTSA; they expressed concern for traffic safety and performance (Stutts & Hunter, 2003).

Several studies focused on driver performance at inter- sections when they were engaged with their cell phone in a simulated or controlled environment. Several researchers discovered that drivers using a cell phone had a delayed response to stop at traffic signals (Beede & Kass, 2006; Irwin et al., 2015; Papantoniou et al., 2016; Strayer & Johnston, 2001). A study comparing drunk drivers and drivers using their cell phone (Strayer et al., 2006) found that drivers using their cell phone had slower reaction time and were more likely involved in crashes. Consiglio et al. (2003) found that drivers increased their reaction time by 72 milliseconds when using a cell phone. Haque and Washington (2014) also confirmed that reaction time at an intersection rises by 40% when a driver used a cell phone.

Only a few previous research studies investigated the impact of technology-induced distracted driving using a field investigation that could more accurately capture the impacts on technology-induced distraction in real-life environment. Brumfield and Pulugurtha (2011) conducted a field investigation at four intersections in Charlotte, North Carolina, and found 54% higher startup loss times when drivers text. A recent study by Alshababi (2021) investigated the impact of cell phones at 24 signalized intersections in Saudi Arabia and confirmed a significant increase of startup loss time of 0.7 seconds. Moreover, Huth et al. (2015) showed that most distractions were initiated at a red indication, and half of the drivers used their phones even before completely stopping at the intersection. Driver distraction represented a significant concern for most transportation authorities and represented a focus of US state governments since 2007 when Washington banned texting (GHSA, 2020). Technology's role in causing distracted behaviors might become more important with the advancements in technology within vehicles due to their uncertain but substantial impacts.

2.2 Research Gap and Objectives

As seen from the literature study, the prior research had the following limitations:

- Saturation flow and start up loss time is one of the two most important indicators of the performance of signalized intersection. Researchers and Transportation Professionals requires the updated indicators like distracted driving that can impact these two components and incorporate the effect of this indicators into their analysis on understanding the performance of a signalized intersection.
- Despite having plenty of research related to distracted driving there are very few research that is analyzing the impact of distraction on saturation flow and start up loss time.
- Several simulation research has evaluated the changes in reaction time for drivers under distracted condition. But simulated environment does properly give an idea about how this reaction time impacts drivers lost time at a signalized intersection.
- Previous field studies only focus on the reaction time or start up loss time of cell phone induced distracted driving only. But total increase/ change in startup loss time is a combination of all the distraction.

To overcome the research gaps and limitations described above the study will focus on the following objectives-

- A comprehensive literature review to gather state-of-art investigations on driver distraction and startup lost time
- Designing a field test, develop a startup lost time data analysis framework, and performed the field test at three intersections in Arlington and Grand Prairie, Texas
- A descriptive analysis and hypothesis testing to understand distraction behaviors and their impacts on individual and aggregated startup lost time at locations with different land use and drivers.

2.3 Data Collection and Data Processing

This study selected three data collection sites in Arlington and Grand Prairie, Texas as shown in Table 1. The first and second sites represented a mixed land-use area (commercial and residential) with high intensity and high traffic volumes (i.e., long traffic queue and volume/capacity ratio). The third site, South Belt Line Road, represented an industrial area and observed a high truck or heavy vehicle volume of about 11.4%. Heavy vehicles were defined as FHWA (Federal Highway Ad- ministration) Truck Class 8 and above in the US, weighing over 4.5 tons (Koonce et al., 2008). The study collected vehicle data using a Sony Camcorder and tripod from three through lanes on one approach at each site during the afternoon peak period (4 pm-6 pm). Two camcorders recorded vehicle queues and signal indications for data collection, and three observers recorded driver distractions, as shown in Figure 2.1. The camera angle captured the moment that a vehicle bumper crosses the stop bar on the three though lanes. This process ensured that the research team could calculate vehicle headways crossing the stop bar when post- processing the videos.

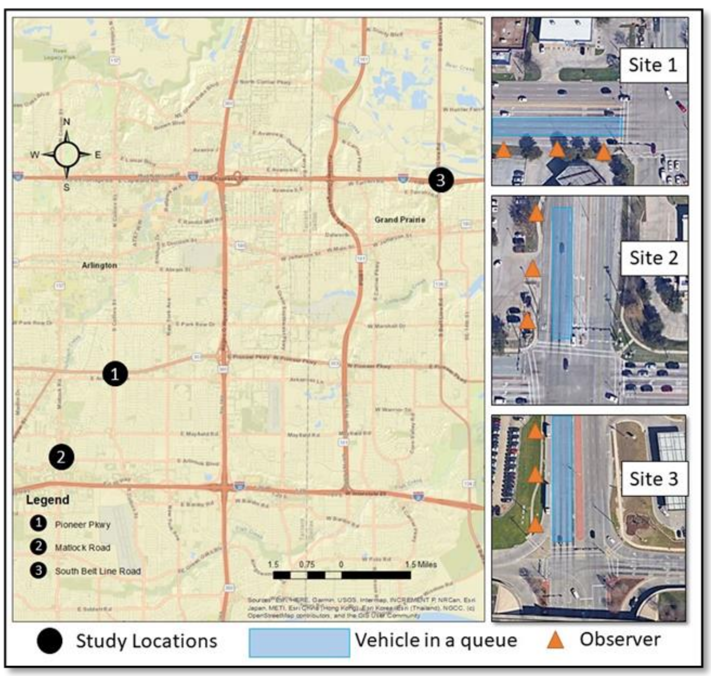


Figure 2.1: Data Collection Location

This study used the 'Visual-Manual on NHTSA Driver Dis- traction Guidelines for In-Vehicle Electronic Devices' published by NHTSA (NHTSA, 2012) to develop a data collection plan and train observers. Before beginning the actual data collection, the researchers trained three observers using the NHTSA guidelines and conducted a trial data collection at a similar intersection. During the trial data collection, the observers identified distracted drivers for the first three vehicles in a queue and the type of distraction. The observers cross-validated their data to evaluate whether all the observers reported the same distracted behaviors. This repeatability test showed 86 percent matching rate (repeatability) in detecting the same distraction type and 97 percent matching rate in detecting a distraction.

During the actual study, the first person observed the first and second vehicles of the queue for the three through (main) lanes, and the second observer covered the third and fourth vehicles in the queue while the third person observed the remaining vehicles in the queue (if any). The observers recorded (i) the types of distraction behavior during a red phase and (ii) whether the distraction persisted into a green phase. To mark the driver's location for later identification, the observer marked the car's location, color, and type. If the distraction recorded during the red interval persisted after the signal turned green, the observers added another mark to the observation sheet to document if this distraction resulted in extra startup delay. The study included six types of distractions – using technology such as cell phones or other electronics, eating or drinking, talking to a passenger, handling objects in the vehicle, grooming, and watching outside person. The cell phone or other electronics included hand-held and hand free devices and other activities such as using stand-alone navigation devices and changing radio stations. This study focused on through lanes only to isolate the impact of distraction from any exogenous factors such as traffic or extra decision–making time by drivers. For example, due to crossing or intersecting traffic movements, vehicles in the right turn lane must determine when to turn and ensure their safe movement, which can cause delays unrelated to distractions.

Previous studies adopted three different data collection/ analysis methods; these included manual observation (Alshabibi, 2021; Huth et al., 2015), video playback (Roy and Saha, 2017) and image conversion or processing using a software (Hurwitz et.al., 2013, Brumfield and Pulugurtha, 2011; Shawky et al., 2016) to calculate time headway between vehicles. Several researchers pointed out that using video footage to identify time headway between vehicles was cumbersome and prone to miscalculation (Brumfield and Pulugurtha, 2011; Shawky et al., 2016) since discerning milliseconds in the footage represented a challenge. Therefore, the study used open-source software, ‘video to jpg’, to generate still images from videos to reduce human error in reading timestamp from videos. The software generated 29 frames per second from the video. The team converted the video into images for every frame, which produced an image every 34.48 milliseconds. The team only kept the images that capture the moment when the front bumper of each vehicle passed the stop line.

Then, the research team used R-programming to assign the timestamp of individual vehicles as a file name (e.g., an im- age taken at 12:00:0034 on April 4th at Matlock recorded as M10412_120000034), and the programming read the file names to calculate the headway between vehicles. These automated processes helped the research team compute the start-up lost time quickly and accurately without repeatedly playing back videos. The analysis considered the images of the first nine vehicles in a queue. The observers also identified the distracted drivers re- corded in the field based on the timestamp of the images. Overall, the researchers processed 17 hours of video that converted and sorted almost 8,000 images capturing the moment vehicles passed the stop line. Field observers recorded 1,350 distraction behaviors at the study sites during 371 signal cycles. This study noted that not distracted drivers could be impacted by a distracted driver located in front of them in the queue. For example, a driver distraction in the 3rd position in the queue included drivers at the 4th and later queue positions to assess the impact of this distracted driver. Figure 2.2 showed the entire data processing and data analysis procedure.

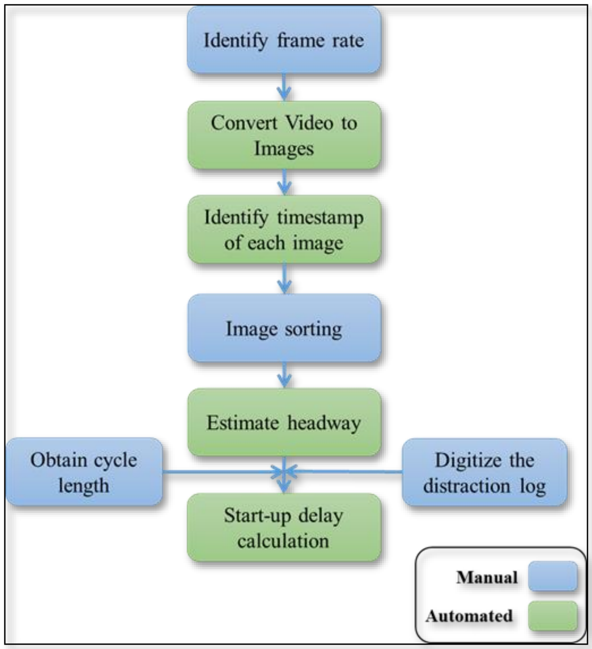


Figure 2.2: Data processing and data analysis procedure

2.4 Descriptive Analysis of Startup Lost Time

At least 13 percent of drivers appeared distracted while waiting during a red indication in this study; however, this distraction rate remained closer to 20 percent for the Pioneer and Matlock locations, as shown in Table 2.1. Among those distracted during the red indication, more than 25% remained distracted during the green indication and contributed to distraction delay at the signalized intersection. The total distraction during the green indication ranged between 5 and 7% at these three sites.

Table 2. 1 Data Observed at Study Sites

	Number of Cycles	**Number of Vehicles observed**	**Number of distractions during red indications**	**Number of distractions during red indications persisted to green**
Site 1: Pioneer Pkwy	98	1,876	419 (22% of Total)	103 (5% of total) (25% of total red)
Site 2: Matlock Road	131	2,960	555 (19% of Total)	200 (7% of total) (35% of total red)
Site 3: South Belt Line Rd	142	2,995	363 (12% of Total)	154 (5% of total) (42% of total red)

This study assessed the headways for the first nine vehicle positions to calculate the saturation flow. As previous literature suggested (HCM, 2010), this study estimated the saturation headway using the vehicles located at the 5th position and after in the queue without any distracted drivers. The aver- age saturation headways for Pioneer, Matlock, and South Belt Line sites showed 1939, 2143, and 2143 milliseconds, respectively. Startup lost time represented the additional headway of the first four vehicles in the queue. As shown in Figure 2.3, the average startup lost times for Pioneer, Matlock, and South Belt Line intersections without distracted drivers in the queue were 3083, 2213, and 2730 milliseconds, respectively. However, a distracted driver in the queue increased the mean lost time by about 600 to 950 milliseconds, and the standard deviation by about 300 to 550 milliseconds. South Belt Line showed a few excessive startup lost time values because the heavy vehicles present at this location caused extra startup loss time due to their slow acceleration.

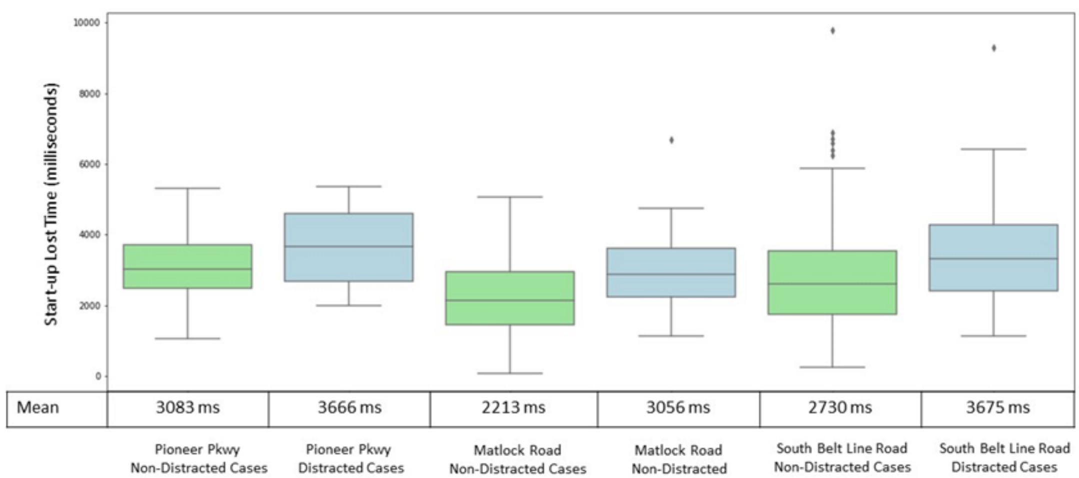


Figure 2.3: Start-up Lost Time

To illustrate the differences in headways along the queue, this study analyzed and compared the individual vehicles' startup lost time behind a distracted driver using one set of queues at three study intersections. Figure 2.4 illustrated a graphical representation of distraction scenarios and the observed startup lost time of the corresponding vehicles in the graphic. For example, a first red vehicle in the queue indicated a distracted driver. This distracted driver may cause additional delay for the drivers of the blue vehicles in a box because they started behind the distracted driver in the queue. Drivers of the blue vehicles without a box represented the undistracted drivers who remained unaffected by the distracted driver since they started in front of the distracted driver in the queue.

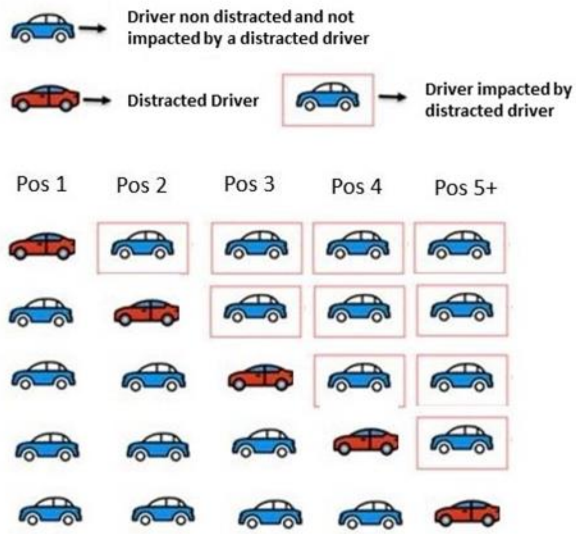


Figure 2.4: Impacts of distraction along the queue

Table 2.2 showed the impact results for trailing drivers due to distraction for all three intersections. A distracted driver in the first queue position created an average startup lost time of 1,417 ms. Undistracted drivers behind these distracted drivers in the first queue position showed an average startup lost time of 1,462 ms at Pioneer. The drivers at positions 3 and 4 in the queue (two and three behind the distracted drivers) showed an average startup loss time of 1,023 ms and 269 ms, higher than the average startup loss time of non-distracted drivers (533 ms and 258 ms for these positions). A driver behind a distracted driver in the second queue position experienced an average start-up lost time of 1,046 ms (513 ms greater than average non-distracted cases) at Pioneer. These results demonstrated that in many cases the driver behind a distracted driver showed an increase in startup lost time.

Table 2. 2 Start up lost time behind a distracted driver along the queue for all three intersections.

	Site 1: Pioneer Pkwy					Site 2: Matlock Road					Site 3: South Belt Line Road				
Position	1	2	3	4	5+	1	2	3	4	5+	1	2	3	4	5+
1	**1417**	1462	1023	269	332	**1584**	742	296	524	131	**2427**	1189	505	366	139
2		**1883**	1046	403	355		**1477**	554	51	63		**1596**	505	285	67
3			**1311**	443	179			**930**	8	194			**1176**	356	100
4				**952**	549				**765**	68				**1196**	229
5					**818**					**570**					**818**
No Distraction	**1066**	**1218**	**533**	**258**	**0**	**807**	**950**	**350**	**153**	**0**	**857**	**1096**	**466**	**184**	**0**

Moreover, the impact appeared to propagate along the queue and affected drivers far from the distracted driver. Similar patterns appeared for the rest of the positions, and a distraction at position 2 in the queue increased the startup lost time up to position five or more. Similarly, at South Belt Line, a distracted driver in the first queue position driver (startup loss time of 2427 ms) affected the drivers in the fourth and fifth queue positions affected since their headways (366ms and 139ms) remained much higher than the drivers at the same positions without any distracted drivers in front of them (Table 2). At Matlock, even though not all vehicle positions showed similar headways to the other two intersections, drivers behind a distracted driver at positions two and four showed higher lost time than non-distracted drivers. These results indicated that the slow response of a distracted driver to a green indication affected the drivers behind them and increased their headways.

Statistical Analysis of Startup Lost Time

The study conducts two statistical analyses using t and F tests. The first analysis shows the increase in startup lost time due to distractive behavior, while the second analysis investigates a distraction's effect on the overall queue for distracted drivers in the first to the fourth position in the queue. This study also focuses on distraction location and evaluates its role on startup lost time. The t and F tests assume the following hypothesis:

t-test

$$H_0 : \mu_1 \le \mu_2 \quad (1)$$
$$H_1 : \mu_1 > \mu_2$$

Where μ_1= average headway for distracted cases; μ_2= average headway for non-distracted cases

F-test

$$H_0 : \sigma_1 \le \sigma_2 \quad (2)$$
$$H_1 : \sigma_1 > \sigma_2$$

Where σ_1 = standard deviation for distracted cases; σ_2 = standard deviation for non-distracted cases

Table 2.3 showed the lost time analysis for the individual queue positions. The results compared the total sample size, descriptive statistics, t-test, and F test between the distraction and no distraction cases. The startup lost time was significantly higher for distracted drivers regardless of site location or vehicle position and had values as high as 1,570 ms. Either the first (Matlock and South Belt Line) or the second (Pioneer) driver in the queue displayed the highest distracted induced lost time. Statistical tests demonstrated that distracted drivers always showed higher lost time at a 90% confidence. The tests indicated that the distracted behaviors, even during a red indication, likely caused significant impacts on drivers' awareness and responses for all queue positions. The F-test confirmed that distracted cases showed a significantly higher standard deviation of the startup lost time except for position one at sites 1 and 2. This result might indicate that the increase in lost time in the first position experienced less variability even though the rise in the mean lost time remained significant.

Table 2.3: Lost Time Analysis for the First Four Positions in the Queue

		Site1: Pioneer Pkwy		**Site2: Matlock Road**		**Site3: South Belt Line Road**	
Position Lost Time							
Vehicle Position	**Stats**	**Distraction**	**No distraction**	**Distraction**	**No distraction**	**Distraction**	**No distraction**
Position 1	Sample Size (N)	10	230	17	305	18	299
	S.D (σ)	σ_1=719	σ_2= 666	σ_1= 750	σ_2= 691	σ_1=2099	σ_2=728
	Mean (μ)	μ_1= 1417	μ_2= 1065	μ_1= 1584	μ_2= 807	μ_1= 2427	μ_2= 857
	T test for different sample means						
	P value	0.08008*		0.00029**		0.00284**	
	F Test for different Sample standard deviations						
	P value	0.31682		0.28307		0.00000**	
Position 2	Sample Size (N)	23	216	22	328	26	332
	S.D (σ)	σ_1= 892	σ_2= 454	σ_1= 869	σ_2= 654	σ_1= 961	σ_2= 668
	Mean (μ)	μ_1= 1883	μ_2= 1218	μ_1= 1477	μ_2= 950	μ_1= 1596	μ_2=1096
	T test for different sample means						
	P value	0.00091**		0.00513**		0.0074**	
	F Test for different Sample standard deviations						
	P value	0.00000**		0.02151**		0.00231**	
Position 3	Sample Size (N)	24	198	21	310	21	292
	S.D (σ)	σ_1= 663	σ_2= 481	σ_1= 984	σ_2= 511	σ_1=1049	σ_2=613
	Mean (μ)	μ_1= 1311	μ_2= 533	μ_1= 930	μ_2= 350	μ_1=1176	μ_2=466
	T test for different sample means						
	P value	0.00000**		0.00704**		0.00296**	
	F Test for different Sample standard deviations						
	P value	0.01002**		0.00000**		0.00004**	
Position 4	Sample Size (N)	15	162	40	282	7	269
	S.D (σ)	σ_1= 624	σ_2= 463	σ_1= 1168	σ_2= 537	σ_1= 297	σ_2= 609
	Mean (μ)	μ_1= 952	μ_2= 258	μ_1= 765	μ_2= 153	μ_1= 1196	μ_2= 184
	T test for different sample means						
	P value	0.00039**		0.00111**		0.00003**	
	F Test for different Sample standard deviations						
	P value	0.03988**		0.00000**		0.23753	
****significant at p<0.05; *significant at p<0.1**							

The second analysis investigated the impacts of a single driver's distracted behavior on the entire queue. Table 2.4 displayed the aggregated startup lost times grouped by the distracted driver's position. For example, a case of vehicles 1 to 4 represented the aggregated startup lost time for the first four vehicles in the queue where the first driver was distracted, and the following three vehicles were affected by distraction. Similarly, a case of vehicles 2 to 4 indicated the aggregated startup lost time for these three vehicles where the second driver in the queue was distracted. In this case, the first vehicle in the queue was not used for the analysis since the distraction of the second driver did not affect the first vehicle. No distraction cases indicated that all drivers in the queue were not distracted. All nine cases showed that the aggregated lost times were significantly higher with a distracted driver at a 90% confidence level. The F test also showed higher standard deviations for the distracted cases at a 90% confidence level for all the locations. This result clearly described that a distracted driver, regardless of its location in the queue, significantly increased the total delay and delay variance.

Table 2.4: Aggregated Start-up Lost Time

		Site1: Pioneer Pkwy		Site2: Matlock Road		Site3: South Belt Line Road	
		Aggregated Lost Time					
Types	**Stats**	**Distraction**	**No distraction**	**Distraction**	**No distraction**	**Distraction**	**No distraction**
Vehicle 1-4	Sample Size (N)	14	118	17	218	16	278
	S.D (σ)	σ1=1129	σ2= 883	σ1=1342	σ2=1057	σ1= 1960	σ2= 1384
	Mean (μ)	μ_1=3666	μ_2=3083	μ_1=3056	μ_2= 2213	μ_1= 3675	μ_2= 2730
	T test for different sample means						
	P value	0.04071**		0.00904**		0.03495**	
	F Test for different Sample standard deviations						
	P value	0.08535*		0.06738*		0.01506**	
Vehicle 2-4	Sample Size (N)	21	117	22	218	21	271
	S.D (σ)	σ1= 1412	σ2= 743	σ1= 1216	σ2= 904	σ1=1442	σ2= 1083
	Mean (μ)	μ1= 3394	μ2=2062	μ1= 2145	μ2= 1313	μ1=2257	μ2=1825
	T test for different sample means						
	P value	0.00018**		0.00239**		0.09617*	
	F Test for different Sample standard deviations						
	P Value	0.00000**		0.01916**		0.02342**	
Vehicle 3-4	Sample Size (N)	24	117	21	218	19	271
	S.D (σ)	σ1= 767	σ2= 614	σ1= 1244	σ2= 721	σ1=1197	σ2=831
	Mean (μ)	μ_1= 1752	μ_2= 845	μ_1= 1185	μ_2= 486	μ_1= 1408	μ_2=695
	T test for different sample means						
	P Value	0.00000**		0.00968**		0.00978**	
	F Test for different Sample standard deviations						
	P Value	0.06493*		0.00004**		0.00716**	
****significant at p<0.05; *significant at p<0.1**							

Distraction Types and Impacts

Figure 2.5 presented the proportions of the six different dis- traction types (1. Using technology such as a smartphone;2. eating or drinking; 3. talking to a passenger; 4. handling objects inside the car; 5. grooming and 6. watching outside person) observed during a red indication at intersections by lane where the first lane referred to the rightmost lane. Technology-based distractions, such as cell phones and navigation, represented the most distractive behaviors in all three locations. Talking to other people in the car represented the second-most frequently observed (9% -15%) distraction while waiting in the queue. Eating or drinking occurred the third most frequently and ranged from 3% to 17%. These categorized patterns also showed that the lane position (inner or outer lane) had little to do with the distraction pattern; however, lane one experienced greater distractions from outside persons or objects than other lanes, mostly because of its proximity to sidewalks and pedestrians.

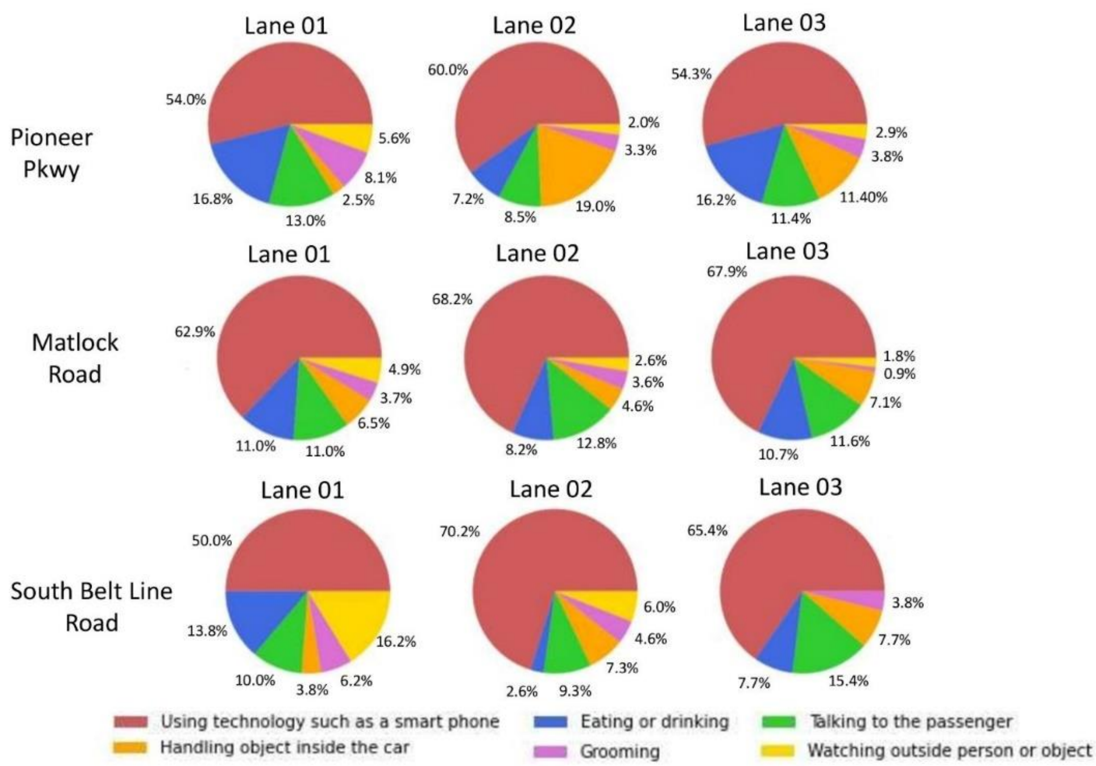


Figure 2.5: Distraction Proportion in Different Lane

Figure 2.6 compares the headway distributions of distracted drivers observed from the study sites. This study used the most commonly observed distraction types, including using technology, eating or drinking, talking to a passenger, and non-distracted cases for comparisons. The headway distributions of handling objects in a car, grooming, and watching outside person are not included in the analysis as they have very few observations compared to the other three types of distraction. In general, drivers engaging with technology show higher average headways while the drivers who talk to other passengers show a higher variance of headway based on a wider interquartile distribution of the boxplot. Interestingly, technology-related distraction shows many outliers outside the interquartile range between 5000 to 9000 milliseconds. It indicates that technology distraction may affect startup delay and possibly cause unusually higher delays. Some outliers occur for the no distraction cases; these outliers result from the heavy vehicles at South Belt Line Road, which may create larger headways due to slow accelerations.

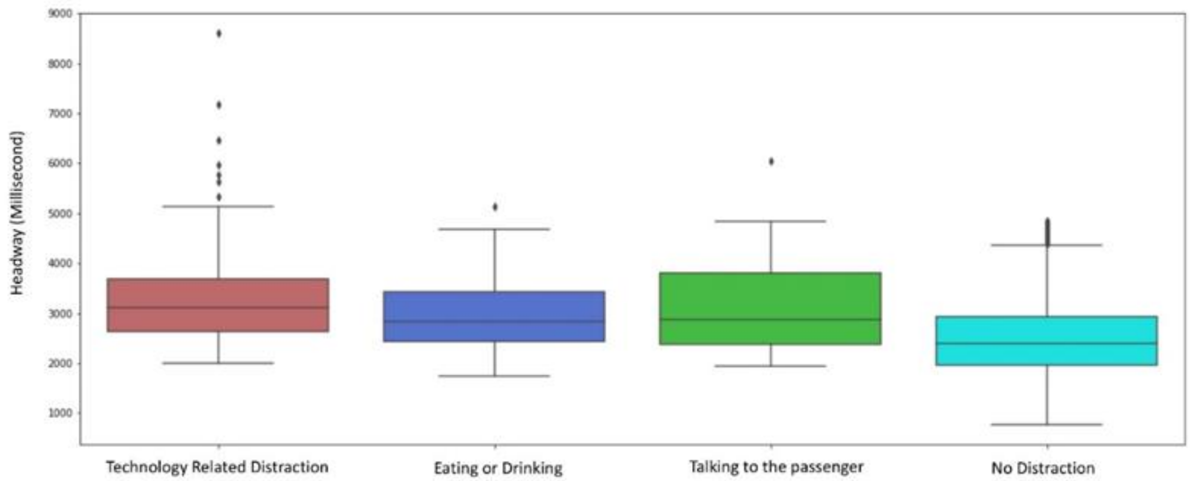


Figure 2.6: Headway Distributions of the Most Common Distractive Behaviors

This study used the South Belt Line location to investigate the effect on distraction behaviors behind a heavy vehicle in a queue. A driver in a passenger car right behind a heavy vehicle showed more distracted behavior. On average, 14% of drivers showed distracted behaviors behind a passenger car; however, the distraction increased to 20% for a vehicle behind a heavy vehicle. This study found only two types of distractions at this location, including using technology and talking to passengers for the driver's behind a heavy vehicle. Over 85% of these drivers appeared to use their cell phones or electronic devices (Figure 2.7). However, this result might not reflect the whole population of distraction types because the limited data collection within the study only noted two distraction types at one intersection (South Belt Line Road).

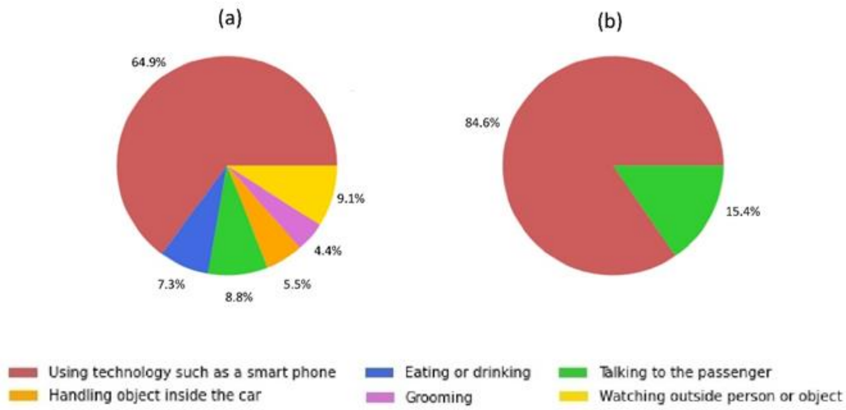


Figure 2.7: Distraction Proportion at South Belt Line Road (a) with no heavy vehicle in front of the driver, (b) driver is waiting behind a heavy vehicle in the queue

2.5 Conclusions

This study conducted data collection at three intersections in Texas and performed statistical analysis to understand the types of distractive behaviors and their effects on saturation flow and startup lost time at an intersection. While two seconds represented the typically assumed startup lost time under ideal condition, this study observed three seconds of startup lost time on average and as high as 6.6 seconds. This study showed that distraction clearly affected the intersection's lost time and saturation flow. The results indicated that, on average, 15% of drivers were distracted during a red indication, and more than 60% of these drivers were distracted due to a cell phone. Statistical analysis showed that distraction caused significantly higher headway and total lost time than non-distraction conditions, which clearly affected saturation flow rates and intersection capacity. The study also found that drivers behind a heavy vehicle in the queue experienced distraction more often than drivers behind another passenger vehicle. The result showed that people behind a truck tend to have more cell phone- related distractions (85%) than a driver who sits behind a car (50-65%). This did not conclusively show that drivers used cell phones more when they waited behind a truck because of limited data, but these drivers might use their phones more often when their view was blocked by heavy vehicles. This study showed that the distraction type varied by vehicle location in the queue. The level and the depth of impacts also varied by the distraction types. Therefore, understanding the frequency and effect on lost time and intersection capacity by distraction type appeared critical to properly calibrate the phase timing and maximize operational efficiency. The technology-induced delay showed a higher standard deviation, which implied technology-induced delay varied from case to case more than non-technology-induced delay. Distracted drivers might continue to increase in the future as technology such as cell phones and in-vehicle entertainment continue to evolve.

Chapter 3. Analyzing Behavioral Patterns of Using Social Media While Driving

3.1 Introduction

Interpersonal communication methods have seen a significant transformation in recent decades due to the development of information technology, particularly with the rapid spread of Internet-based social media. Easier and more widespread access to the internet and social media platforms has caused unreasonable and excessive use of social media, resulting in the potential for social media addiction (Hou et al., 2019, Starcevic, 2013). As a result, drivers frequently use their cell phones and tablets to read and post on social media. Several activities, such as watching feeds and recording short videos, have become everyday driving activities. Many young generations fear missing out if they do not check their social media several times a day or even an hour. It creates another form of distracted driving different from the conventional ones engaging with eating, drinking, and grooming. The National Highway Traffic Safety Administration (NHTSA, 2020) defined three types of distracted driving that caused visual, manual, and cognitive impairments (NHTSA, 2020). Engaging in social media while driving integrates all three types of distractions because a driver needs to look at their phone rather than the road with holding the phone and scrolling (instead of holding the steering wheel) and thinking about the contents while making decisions for driving.

Social media platform usage has increased exponentially and quickly in the modern era. A recent survey found that 69 percent of Americans regularly use Facebook, while 40 percent actively use other apps like Instagram and 25 percent use Snapchat daily (PRC, 2021). The increased use of social media by drivers has led to increased incidents in which they use these platforms while operating a vehicle. The broad acceptance of social media across various demographics has profoundly affected people's daily lives, including driving. A survey by telecommunication company AT&T 2015 showed that 7 out of 10 people engage in smartphone activities while driving, and almost 4 out of 10 people engage in social media activities. In particular, the survey indicated 28% surfing the internet, 27% using Facebook, 17% taking a selfie, 28% using Twitter or Instagram, 12% trying to shoot a video, and 10% conducting video chat while driving (AT&T , 2015).

This prevalent impact of social media has presented planners and engineers with an ever-more complex challenge as it has brought about a new wave of distractions. These professionals must reconsider and reinterpret traditional road safety and infrastructure planning methods in light of the complex difficulty presented by this changing environment. Nevertheless, existing literature reflects limited research examining the impact of drivers' distractions attributed explicitly to social media. While studies exploring distracted driving exist, the distinct influence of social media remains largely unexplored territory. Notably, investigations by Basacik et al. (2012) and Hasash et al. (2019) have predominantly centered on the consequences of passive social media browsing during driving, highlighting the possibility of degradation in driving performance under this particular form of social media distraction. However, it is essential to consider the more engaging and interactive features of different social networking sites, which leaves a significant knowledge gap about the broader range of distractions these distractions can cause when driving.

A notable research void also exists concerning the underutilization of advanced modeling techniques, such as machine learning models, to comprehend the complexities of social media-induced distracted driving. While such models have proven valuable in various transportation engineering disciplines, their potential in analyzing the behavior of distracted drivers, particularly those involving social media, still needs to be explored. In earlier studies related to distracted driving, most machine learning algorithms compared the characteristics of distracted and non-distracted driving, and scant studies have compared the features of various forms of distracted driving. Furthermore, understanding which specific forms of distraction lead to erratic and hazardous driving patterns remains an unexplored realm within this field. As the landscape of social media usage continues to expand rapidly, a significant knowledge gap persists in differentiating social media distractions from conventional distractions that impact driving. This disparity emphasizes the necessity of thorough studies using cutting-edge modeling methods to identify, classify, and understand the unique effects of distractions—especially those resulting from social media use—on driving behavior.

Some research investigated distracted behaviors at horizontal and vertical alignments (Tractinsky et al., 2013) with varying traffic volumes (Arnold & Van Houten, 2011). Their analysis shows that roadway environment and traffic conditions have a complex relationship and influence driving and mobile phone tasks. It is essential to acknowledge a research need to understand how various road traffic circumstances affect vehicle control while distracted by social media, which is an entirely new challenge for traffic safety. But most existing research has primarily explored distracted driving in specific roadway environments, traffic conditions, or roadway geometries. This approach provides only a limited understanding of the intricate relationships between these factors and the types of distractions that drivers experience.

To address these research gaps and contribute to the field, the study aims to achieve two key research objectives. Firstly, it will explore the interrelations between social media distractions and driving behaviors by understanding important driving behavior attributes affected by the distraction. The study will evaluate the most common social media activities at different roadway and traffic conditions and their corresponding impacts on drivers performances. Secondly, the research will focus on developing machine classifiers that are capable of accurately identifying various distraction types, considering the diverse range of roadway environments and traffic conditions that drivers may encounter. This step is vital in creating effective interventions and countermeasures to mitigate the risks of distracted driving.

3.2 Literature Review

According to National Highway Traffic Safety Administration (NHTSA), more than 42,000 people died in motor vehicle-related crashes in the USA in 2021 and 8.2% of them are responsible for distracted driving (Stewart, 2023). Distracted driving may influence many primary tasks, such as steering, braking, accelerating, and changing gears and slowing down a direct vehicle response. In order to understand the impact of distracted driving on drivers' responses and driving performances, previous studies have conducted different types of experimental designs using instrumented vehicles, naturalistic driving data, driving simulators, and self-reported surveys. Even though naturalistic driving data provide higher accuracy and validity among all different experimental setups, a smaller sample size is known as a significant limitation to capture diverse driving behaviors and safety issues during the experiment (McCartt et al, 2006). An easy and straightforward approach to gathering information regarding driver behavior is through surveys; however, it usually carries some inherent bias because of underreporting and information loss (Oviedo-Trespalacios et al., 2016). Driving simulators have recently gained popularity as a means of experimental design because they give researchers the ability to control driving conditions and ensure participant safety. (Oviedo-Trespalacios et al., 2016).

Previous driving simulator experiments have examined various functional parameters of driving behavior using speed, acceleration, lane position, steering angle, and headway distance under distracted driving conditions. Speed and speed variability are major functionality parameters evaluated in most literature. Literature showed that different mobile phone tasks such as conversing on the phone (Tractinsky et al., 2013), holding a mobile phone, reaching a mobile phone, using a navigation system (Christoph et al, 2013), reading a text (Rudin-Brown et al, 2013) and answering or dialing a call (Tractinsky et al., 2013) influence speed control. Most studies have found that speed tends to decrease when drivers use a phone for conversation or texting, but few studies have found an increase in speed (Garrison & Williams, 2013, Liu & Ou, 2011). According to some other studies drivers distracted by their phones showed high-speed variability, faster throttle acceleration or deceleration, and sudden non-directional accelerations. In addition, Rudin-Brown et al. (Rudin-Brown et al, 2013) found a significant difference in lane deviation among distracted drivers concerning non-distracted drivers. They designed a simulated experiment where drivers had to drive a tunnel and a freeway scenario under distracted conditions with texting and non-distracted condition. Their analysis showed an increase in lane deviation for drivers with texting which is more prominent in tunnel sections than freeway. This study summarized that under complex work conditions like texting, drivers’ were less able to adequately monitor the position of the vehicle. Stavrinos et al. (2013) found increased lane variation while texting compared to talking or without distraction cases. Headway is another critical parameter because distracted drivers tend to keep a longer following distance and exhibit more significant gap variability (Oviedo-Trespalacios et al., 2016, Bergen et al., 2013).

Very few studies assess drivers' behavioral performance under social media distraction. Basacik et al. (2012) investigated the effect of Facebook browsing on driving performance in 2012. In this experiment, they asked the participants to update their Facebook statuses and send an instant message through the Facebook messaging system. Their analysis for both distracted conditions showed a significant increase in reaction time, an increased number of lane departures showing an inability to stay in a lane, and higher variability in headway. Another simulator study by McNabb and Gray (2016) showed that drivers using social media messaging for texting had a similar negative effect on drivers' performance as texting using a cell phone. Hashash et al. (2019) designed a simulated experiment to understand the performance effects of social media browsing (Facebook browsing) to texting or no distraction. Their evaluation showed that browsing while driving harms drivers' performance in mean speed and lane variation while texting has a more detrimental effect on drivers' performance than social media browsing. Social networking apps contain a range of multiple methods of interaction beyond reading texts and sending text messages, including exchanging images and videos (such as on Snapchat, Instagram, and Vine) and recording and watching videos (such as on short reels or Boomerangs), and all of these activities contribute different memory and attentional demands. For these reasons, the studies on social media browsing and texting discussed above may not be directly predictive of the impacts of social networking on driving. Consequently, a dearth of knowledge exists concerning the impact of engaging in social media interactions on driver behavior. Consequently, there exists a substantial lack of knowledge regarding how engaging in social media interactions affects driver behavior.

One of the major challenges of the distracted driving behavior analysis is to capture the unique behavioral characteristics of each driver who has a unique driving style under various environmental and traffic conditions (Lattanzi & Freschi, 2021). Statistical regression models or descriptive statistical analyses are the most popular and established techniques to examine the relationships between distracted driving and behavioral factors. The significant correlations between the many components contributing to driving behavior can be achieved using these classic methods, however, they require a sound model assumption and well-established predefined relationships, which are difficult to obtain because of the non-linearity and interaction effects among the behavioral factors (Niu et al, 2021, Darzi et al., 2018). Also, the consequences of distracted driving extend to a phenomenon known as "inattentive blindness," wherein drivers gradually become less aware of their surroundings, leading them to overlook critical information on the road. This development unfolds gradually over a specific period during which a driver's cognitive load and attentional resources become compromised. To comprehensively grasp and tackle the issues of distracted driving and inattentive blindness, a deeper examination of driving behavior patterns over time is crucial (Ye et al., 2017).

Machine learning models (ML) have recently contributed to detecting predictive characteristics in various behavior and engineering disciplines. ML techniques are flexible and adaptable to identify linear and non-linear correlations among dependent and independent variables with a processing capacity of big data (Tang et al., 2019). To characterize distracted driving behaviors, researchers used many different algorithms including Support vector machines (SVM), Logistic Regression, Decision trees (DT), Gradient Boosting, K-Nearest Neighbors (KNN), Multilayer Perception (MLP) classifiers and Neural Networks (NN). These techniques are intended to completely utilize data value and find prospective laws that have been difficult to collect. (Abou Elassad et al., 2020). For example, the decision tree algorithm offers exact performance results, optimum splitting parameters, and effective tree pruning methods. Simplicity and the easy-to-establish conceptual rule are major advantages of the DT algorithm. (Charbuty & Abdulazeez, 2021). Gradient Boosting is another popular tree-based algorithm that computes the direction of gradient or the direction of trees where the model can be approved. To lower the bias of the model rather than the variance, each subsequent tree seeks to get the model closer to the target. One of the advantages of gradient boosting is it can handle large data sets in short time (Biau et al., 2019). Another popular supervised machine learning technique is Logistic Regression. It is a useful linear regression method for binary classification problems. It does not necessarily require a linear relationship between variables and applies nonlinear log transformation for any data (Subasi, 2020). Support Vector Machines (SVM) on the other hand identifies patterns by inferring non-linear correlations between variables. SVM has good generalization capabilities which helps to stop any overfitting issue and small change of data does not affect the hyperplane which is very important for data classification (Cristianini & Shawe-Taylor, 2000). The k-nearest neighbor (KNN) algorithm is easily interpretable because it does not make any assumptions about the data but groups the data based on the points around it (Triguero et al., 2019).

Artificial Neural Networks (ANN) or Neural Networks (NN) are developed based on the human nervous system, which can process a large number of interconnected elements (Tango et al., 2010). ANN can handle large amounts of data sets, identify non-linear complex relationships between dependent and independent variables, and evaluate the connection between predictor variables (Gharehbagh, 2016). Multilayer Perception (MLP) classifier is a form of feedforward artificial neural network which is a popular machine learning technique for binary classification. This particular neural network architecture does not have any feedback loops and only allows information to travel in one direction, from the input layer to the output layer. This particular neural network architecture does not have any feedback loops and only allows information to travel in one direction, from the input layer to the output layer. The basic structure of a Multilayer Perception (MLP) neural network consists of an input layer, an output layer and multiple hidden layers between them. Multiple neurons, also called as units or nodes, make up each hidden layer. These neurons use the input data to learn complicated patterns and representations (Bishop, 1995, Goodfellow et al., 2016, Al-Naymat et al., 2016).

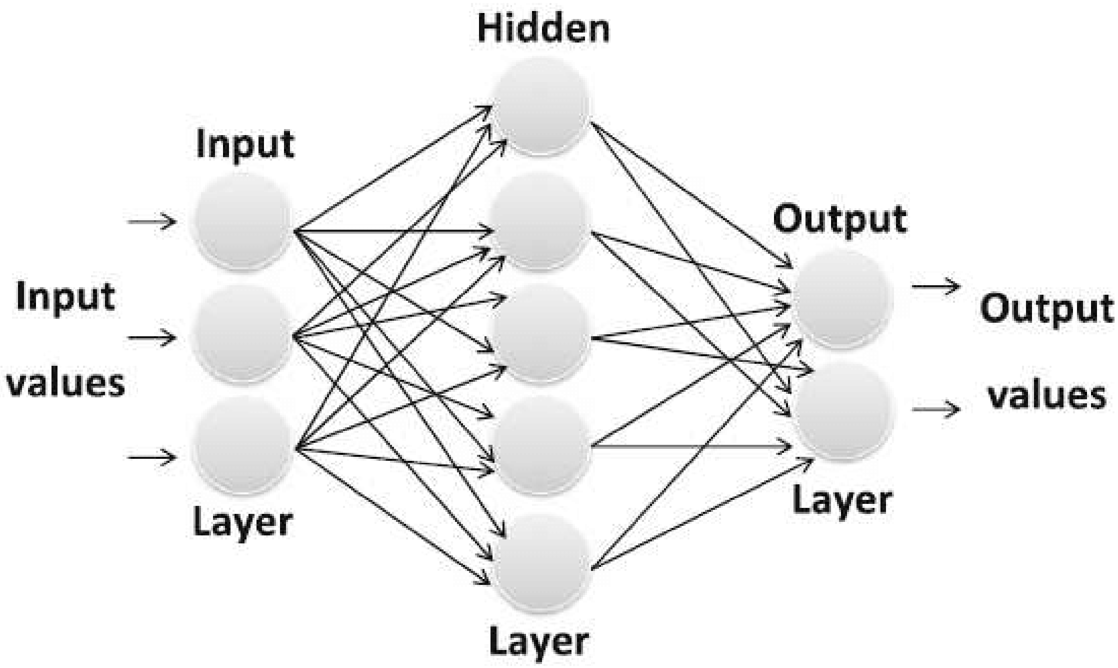


Figure 3.1: Multilayer Perception (MLP) structure (Al-Naymat et al., 2016).

Investigating all these various machine learning models based on their characteristics will provide their unique strengths and weaknesses, which can eventually help choose the suitable machine learning model for this study that requires an accurate understanding of behavior patterns.

3.3 Data Collection

Participant Information

The study designed a driving simulator experiment and invited the University of Texas at Arlington students to participate. A pre-screening survey identified 49 candidates who use social media platforms frequently. Reading and speaking English and having a driver's license are other eligibility criteria used by the study. Table 3.1 shows participants' demographic information a.

Table 3.1: Demographics of the participants

Demographics		**Percent (%)**	**Count(N)**
Gender			
	Male	65%	32
	Female	35%	17
AGE			
18	Min	Mean	26.8
26	Median	SD	5.971
45	Max		
Hispanic, Spanish, Latino origin or decent			
	Yes	18%	9
	No	80%	39
	N/A	2%	1
Race and/or Ethnicity			
	White or Caucasian	24%	12
	Black or African American	8%	4
	American Indian or Alaska Native	4%	2
	Asian	59%	30
	Other	6%	3
Annual household income before taxes			
	Less than $10,000	6%	3
	$10,000 to $19,999	16%	8
	$20,000 to $29,999	20%	10
	$30,000 to $39,999	6%	3
	$40,000 to $49,999	6%	3
	$50,000 to $74,999	14%	7
	$75,000 to $99,999	4%	2
	$100,000 to $149,999	10%	5
	$150,000 or more	4%	2
	Prefer not to answer	12%	6
Have a vehicle			
	Yes	71%	35
	No	29%	14
Type of vehicle driving most often			
	Car (Sedan, Coupe, Sports car, Hatch bag, etc.)	86%	42
	Van or Minivan	2%	1

	Motorcycle	0%	0
	Pickup Truck	6%	3
	Sport Utility Vehicle	6%	3
Years of driving experience			
	Less than 1 year	8%	4
	1 to 3 years	27%	13
	4 to 10 years	53%	26
	More than 10 years	12%	6

Experimental Setup

As mentioned above, one of the primary objectives of the study is to understand the effects of social media distracted driving on different traffic and roadway conditions. This study examines the impact of various road conditions and traffic scenarios within highway settings, illustrated in Figure 3.2 below.

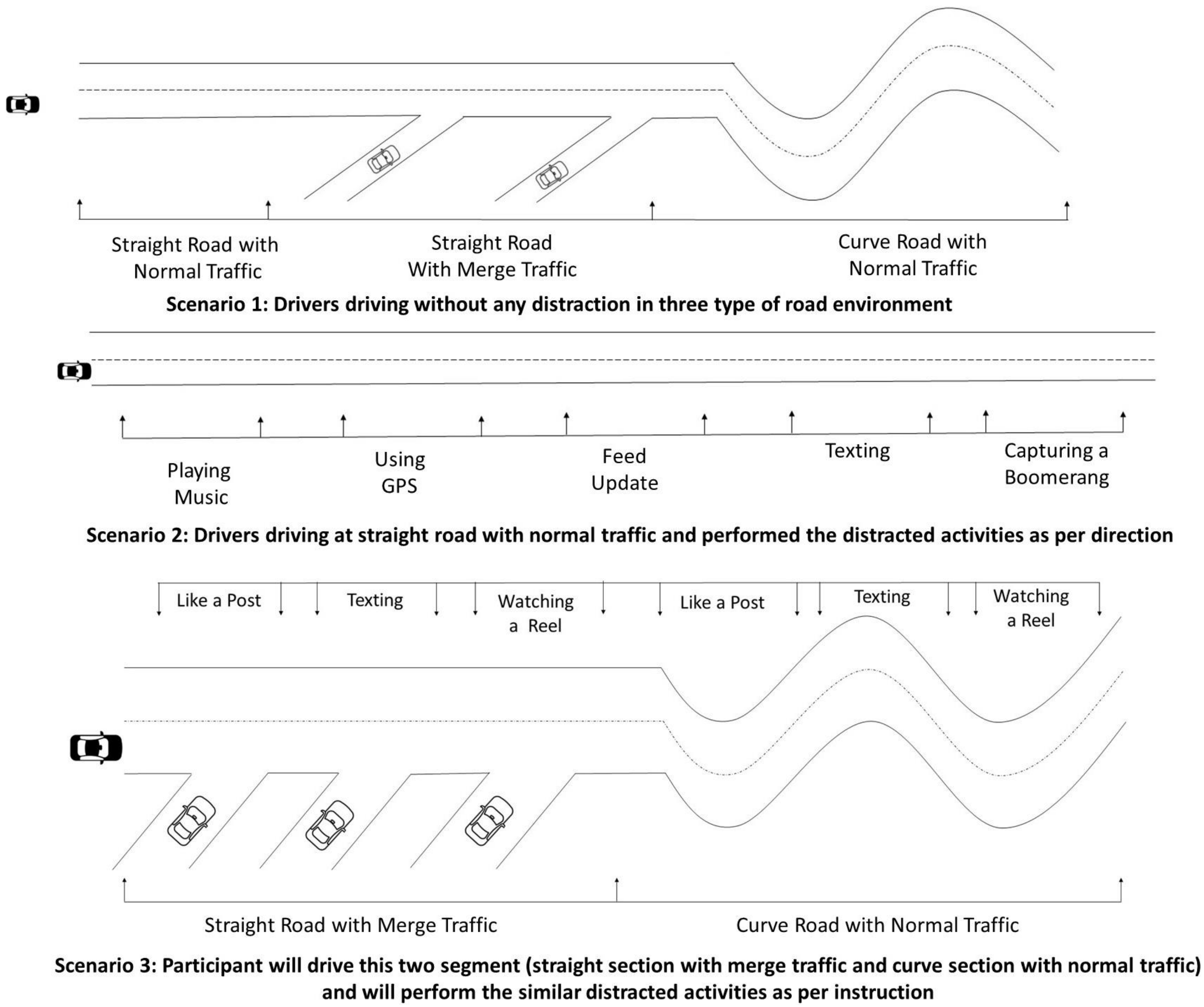


Figure 3.2: Graphical explanation of three scenarios and the experimental setup at a driving simulator

This study chose the following distraction activities as the most common social media activities to understand their impacts on driving behaviors:

- Playing a song in any music app (e.g., Youtube, I-tunes, etc.)
- Using GPS to find a location
- Texting using a social media app
- Browsing using any social media app

- Liking a post on any social media app
- Creating a Boomerang (short video) using Instagram or another social media app
- Watching a reel or video in a social media app.

Two researchers conducted the experimental simulation process. Before commencing the experiment, participants underwent a 5-minute familiarization phase using the driving simulator to ensure comfort and prevent any potential discomfort, such as nausea or dizziness. Once participants felt at ease, the experiment proceeded. The experimenters clearly instructed the participants regarding the starting, stopping, and executing distraction tasks while navigating the general highway scenario, adhering to a speed limit of 65 mph. In scenario 1, participants were given to participants to refrain from using cell phones to establish their natural or baseline driving conditions.

The first experimenter recorded the simulator times for the distraction scenarios and guided the participants accordingly. Simultaneously, the second experimenter utilized a stopwatch to record the start and end times of each tasks notifying the first experimenter upon the completion of each task by a participant. For scenarios 2 and 3, participants used their own smartphones and received instructions regarding their secondary tasks from the experimenter. In scenario 2, participants engaged in five secondary tasks, comprising three conventional distractions (Music, GPS, and Text) and two social media distractions (Feed and Video). For instance, tasks included using the music app to select a song, browsing a location on the GPS app, and responding to a question from the experimenter via text, ensuring a complete sentence without abbreviations or acronyms. During scenario 3 involving weaving and curves, participants completed one conventional distraction (texting) and two social media distractions: liking a post on an Instagram account set up for the experiment by the experimenter and watching a reel (a video of up to 30 seconds) chosen by the experimenter. The sequence of secondary tasks was randomly assigned to each participant, with the initial task assigned when the driver reached a saturated flow rate at the onset of both distracted scenarios. Before the experiment, the experimenters ensured that participants were familiarized with various secondary tasks and confirmed their prior experience with them.

Feature extraction and feature pre-processing

The driving simulator provides more than 20 feature attributes of driving parameters every 1.67 s as shown in Figure 3.3. The study used scenario 1 as the base scenario to understand each participant's no-distraction driving behavior. We removed the first 10 seconds and last 10 seconds of time series data to ensure their driving reached a saturation flow state (for the start of their driving) to remove intentional stopping processes at the end of the experiment.

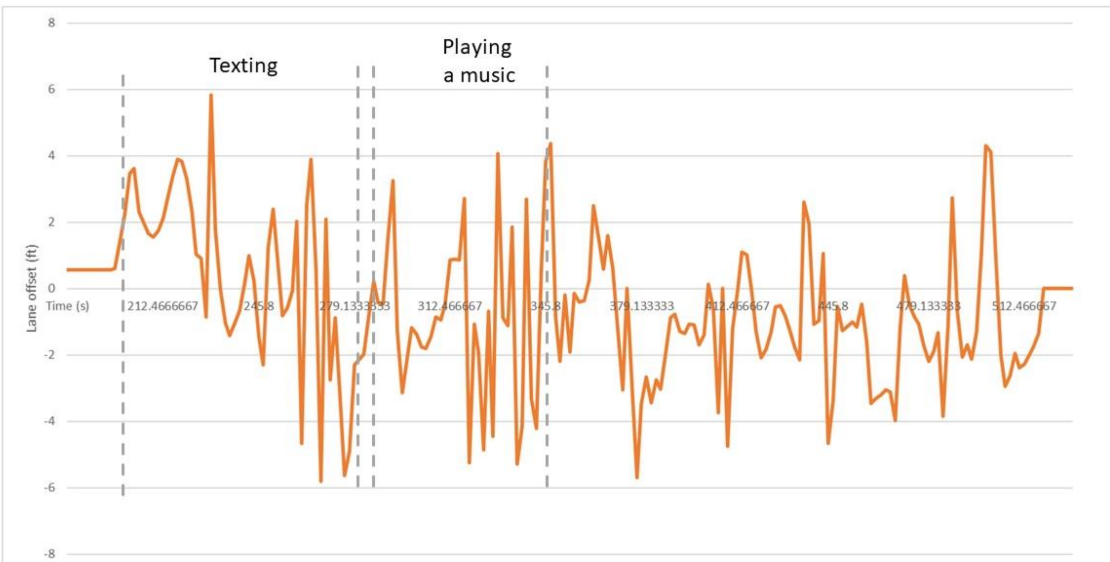


Figure 3.3: Example of a time series data showing lane offset.

To properly handle inattentive blindness, the study considered the data as a time series and analyzed each performance attribute over a 20s-time window of 12 data points, as shown in Figure 3.4 below. Then the window moves one step closer to calculate the second set of performance attributes. This is because independent observations may not explain how distractions impact driving performance comprehensively.

Time(s)	142.47	144.13	145.8	147.47	149.13	150.8	152.47	154.13	155.8	157.47	159.13	160.8	162.47	164.13	165.8	167.47
Speed (mph)	67	67	68	68	68	68	68	68	69	69	70	71	72	73	74	75

Figure 3.4: Example of moving window for speed as an attribute

Using the literature review (Rudin-Brown et al., 2013, Garrison & Williams, 2013, Liu & Ou, 2011), this study adopted the following performance attributes to describe drivers' behavior while distracted.

1. Acceleration: Acceleration of the vehicle
2. Throttle positioning angle: This is the angle of the pedal position.
3. Headway: This is the headway distance between the CG of the vehicle to the CG of the vehicle ahead.
4. Heading Error: Heading is a value in degrees (deg) representing the heading of the vehicle. Heading error is the angle (deg) between road path and current heading.
5. Brake Pedal Force: Brake force. Generally a value between 0 and 170 representing N.
6. Lane Offset: A value in meters or feet of the position of the vehicle from the center of the lane
7. Steering wheel position: A value in radians (rad) describing the steering wheel position.
8. Speed: Speed of the vehicle.
9. Lateral Speed: Speed of the car at lateral direction
10. Lane Changing Indicator: Indicator signal confirmation.

For the aforementioned metrics, this study used mean, standard deviation, maximum, minimum, and maximum absolute differences in driving parameters between successive points in a time series for all the parameters except Brake Pedal Force, Lane changing indicator, and Lateral Speed. For Brake Pedal Force, this study calculated how many times a participant pressed the brake during a task or a time window. Brakes are a vital tool for preventing collisions, and how frequently they are used can tell us how often distracted drivers are put in dangerous circumstances. Additional brakes may indicate that distractions create more urgent situations that call for quick responses to prevent crashes.

Distracted driving can often lead to reduced situational awareness and cognitive overload, hampering a driver's decision-making abilities. One of the common errors in drivers' decision-making abilities is not using turn signals properly. Turn signals can be a predictive analyzing feature in identifying distracted behavior. Therefore, in addition to the aforementioned performance measures, this study estimated the number of lane changes with (Lchn) and without the lane change indicator (UnLchn) to understand situational awareness while changing lanes. A detailed description of the calculation of each variable is available in table 3.2.

Table 3.2: Overview of all features:

Feature name	Mean (mean)	Std dev (std)	Max (max)	Min (min)	Max Rolling Diff (rdiff)	Frequency (fr)
Acceleration (Acc)	X	X	X	X	X	
Throttle positioning angle (Throttle)	X	X	X	X	X	
Heading Error (Herror)	X	X	X	X	X	
Brake Pedal Force (Brake)						X
Headway distance from front vehicle (Headway)	X	X	X	X	X	
Lane offset (Loffset)	X	X	X	X	X	
Steering Wheel Position (Stpos)	X	X	X	X	X	
Speed (Speed)	X	X	X	X	X	
Lateral Speed (Lspd)			X			
Lane changing indicator signalized (Lchn) or unsignalized (UnLchn)						X
*The symbol in parenthesis will be used to represent attributes in figures and tables						

Data Normalization and Standardization

This study pre-processed all of the driving behavior features using data normalization and standardization. These techniques improve the performance and convergence of machine learning algorithms, especially when dealing with features that have different scales or units. Normalization ensures that all the feature values lie within a common range, making the data more manageable and allowing the machine learning algorithms to converge faster. However, it does not handle outliers well, as it squeezes the entire data range into a smaller interval. Standardization is robust to outliers and ensures that the features have zero means, which is helpful for certain algorithms like gradient-based optimization techniques. However, it does not impose a specific range for the data, and the transformed feature values can have positive or negative values. (Bishop et al., 2006, Raschka & Mirjalili, 2019)

This study used data normalization to rescale the features of independent datasets to a range between 0 and 1. The normalization formula for a feature x is given by:

$$\text{Normalized feature x} = \frac{x - \text{Min}\ (x)}{Max(x) - Min\ (x)} \tag{1}$$

where min(x) is the minimum value of the feature, and max(x) is the maximum value of the feature in the dataset.

This study used data standardization to rescale the features of a dataset to have a mean of 0 and a standard deviation of 1. The standardization formula for a feature x is given by:

$$\text{Standardized x} = \frac{x - \text{Mean}\ (x)}{Std\ (x)} \tag{2}$$

where mean(x) is the mean of the feature, and std(x) is the standard deviation of the feature in the dataset.

Selecting Features for Modelling

The study uses a distraction indicator (e.g., distracted vs. not distracted) as a dependent variable in a classification model. A correlation analysis was employed to understand the linear relationships between the dependent variables (distraction or no distraction) and the independent variables (driving attributes). Driving attributes with a strong correlation to distractions were identified as the initial attributes. Additionally, this study applied feature importance tools using a random forest algorithm to identify critical attributes and ultimately used essential features with a feature importance greater than 0.1 for modeling. To address multicollinearity among the driving attributes, the study utilized the Variable Inflation Factor (VIF) and attributes with a VIF value exceeding 5 were removed from the dataset. Furthermore, a correlation check was conducted among the remaining attributes to ensure the validity of the analysis.

Modelling

The study used six machine learning algorithm models including K- nearest neighbor (KNN), Logistic regression, Support Vector Machine (SVM), Gradient Boosting Classifier, Deep Neural Network (NN) and Multilayer Perception (MLP) classifier to detect distraction using driving parameters.

To avoid any overfitting issue, data sets were randomly split into training data consisting of 80% of the participant and test datasets of 20% of the participants. As shown in Figure 3.5, the study employed a four-fold cross-validation method. The process involved randomly selecting 30 participants' data to train the model and the remaining ten participants' data as validation sets. Four models were trained and assessed for each hyperparameter and calculated their mean performances. Subsequently, the study identified the hyperparameter that yielded the best mean performance across all training and validation data sets. Once the best hyperparameter was determined, the study used all the available training data from 40 participants to develop the final model.

All Data (50 Participants)	
Training Data (40 Participants)	Test Data (10 Participants)

	Fold1 (10 Participant)	Fold2 (10 Participant)	Fold3 (10 Participant)	Fold4 (10 Participant)	Train/Validation Performance
Split1	Train	Train	Train	Validation	Performance 1
Split2	Train	Train	Validation	Train	Performance 2
Split3	Train	Validation	Train	Train	Performance 3
Split4	Validation	Train	Train	Train	Performance 4
	Mean Train/Validation Performance				(1/4)* Σ Performance

Figure 3.5: Four-fold cross-validation procedure

The final accuracy and precision scores were calculated as below.

$$\text{Accuracy}=\frac{TP+TN}{TP+FP+FN+TN} \tag{3}$$

$$\text{Precision}=\frac{TP}{TP+FP} \tag{4}$$

TP = True Positive

FP = False Positive

TN = True Negative

FN = False Negative

3.4 Result Discussion

Distracted Driving Detection for a Simple Roadway Geometry and Traffic Environment

The study compared distractions in Scenario 2 and the straight segments from the base scenario (Scenario 1). The correlation analysis in Figure 3.6 revealed that drivers faced challenges in maintaining their vehicle's lateral position regardless of the distractions they were involved. Several attributes related to lateral control, such as Heading error, Lane offset, and Steering wheel position, exhibited higher correlations with distracted driving. Notably, capturing a Boomerang (video) while driving consistently showed the highest maximum correlation value with distraction among these lateral attributes. Additionally, an interesting finding emerged where Boomerang also demonstrated the highest correlation with unsignalized lane change frequency. This suggests that drivers experienced higher cognitive inattention during video recording and posting and increased their Lateral Speed (Lspd_max) as an overcompensation response.

The impact of drivers' longitudinal properties, such as speed and acceleration are less associated with distraction. Furthermore, there is a negative correlation with acceleration attributes to any secondary tasks, suggesting that drivers tend to control or reduce their speed while distracted. On the other hand, mean and maximum headway (Headway_mean and Headway_max) are strongly connected with distracted driving. These characteristics show a consistent inverse relationship to distracted driving, indicating as drivers become more distracted, their headway, which represents the distance maintained between their vehicle and the vehicle in front, tends to decrease.

	Music	Texting	Browsing	GPS	Boomerang
Herror_std	0.357	0.275	0.361	0.34	0.415
Loffset_std	0.408	0.417	0.348	0.418	0.635
Stpos_std	0.566	0.494	0.404	0.547	0.541
Acc_mean	-0.306	-0.356	0	-0.284	-0.408
Throttle_mean	-0.318	-0.352	0	-0.216	0
Herror_mean	-0.297	-0.2	-0.3	-0.252	-0.324
Speed_mean	0	0	0	0	0.237
Headway_mean	-0.402	-0.417	-0.334	-0.246	-0.359
Acc_max	0	-0.312	-0.237	-0.272	-0.422
Throttle_max	0	-0.268	0	-0.234	-0.218
Herror_max	0.649	0.618	0.443	0.645	0.606
Loffset_max	0.261	0.271	0.235	0.29	0.449
Stpos_max	0.473	0.416	0.348	0.422	0.534
Speed_max	0	0	0	0	0.211
Headway_max	-0.451	-0.468	-0.428	-0.315	-0.435
Acc_min	-0.2	0	0	0	0
Throttle_min	-0.27	-0.24	0	0	0
Herror_min	-0.347	-0.261	-0.347	-0.315	-0.408
Loffset_min	-0.294	-0.332	-0.291	-0.303	-0.521
Stpos_min	-0.513	-0.499	-0.375	-0.493	-0.478
Speed_min	0	0	0	0	0.268
Headway_min	-0.335	-0.309	-0.255	0	-0.283
Brake_fr	0	0.237	0.229	0	0
Herror_rdiff	0.371	0.294	0.37	0.352	0.428
Loffset_rdiff	0.314	0.34	0.298	0.303	0.538
Stpos_rdiff	0.488	0.454	0.373	0.458	0.508
UnLchn_fr	0.407	0.357	0.371	0.338	0.546
Lspd_max	0.393	0.351	0.341	0.261	0.407

Figure 3.6: Correlation between driving behavior attributes and distracted activities for straight road

Utilizing the most relevant driving attributes (those with an absolute correlation greater than 0.2), the study conducted a feature importance analysis to further identify crucial factors in detecting distracted driving (Figure 3.7). Among all distracted conditions, Maximum Heading Error (Herror_max) emerged as the most influential variable, followed by Maximum Acceleration (Acc_max). Interestingly, in the case of the Boomerang (video recording and posting) distraction, Maximum Lateral Speed (Lspd_max) was identified as a significant factor. Moreover, Maximum Headway (Headway_max) is a crucial attribute for most distracted scenarios, except for the Boomerang distraction, where Minimum Headway stood out as one of the most critical attributes. It suggests Boomerang is a more engaging distracted activity, which reduces the distance between the front vehicle and increases the risk factor.

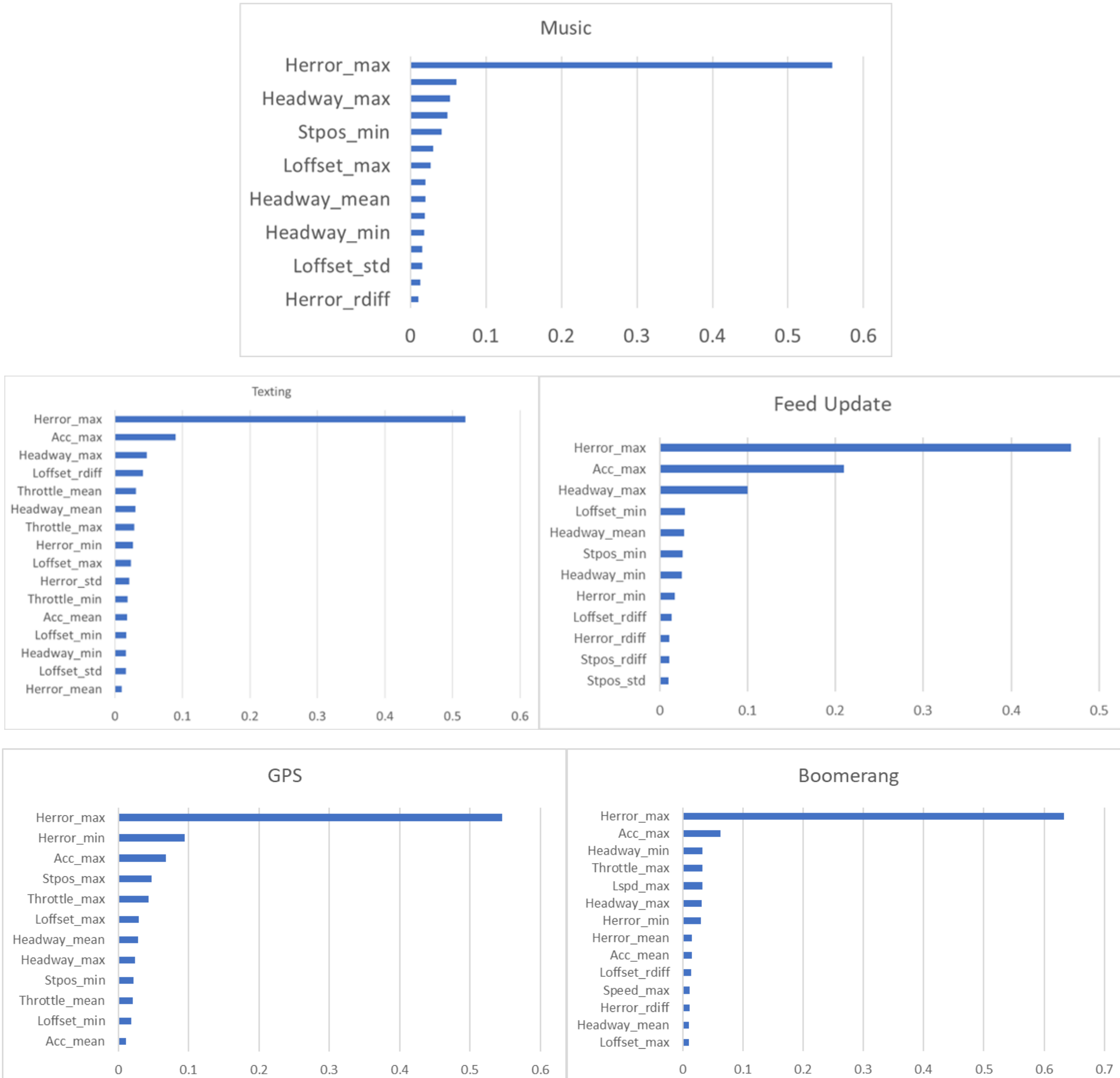
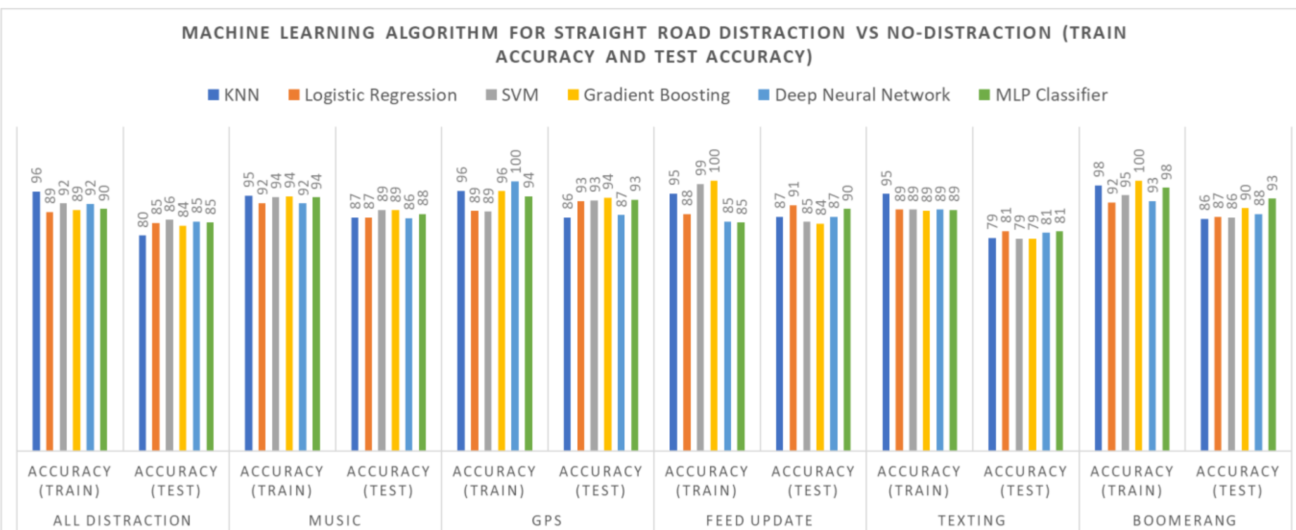


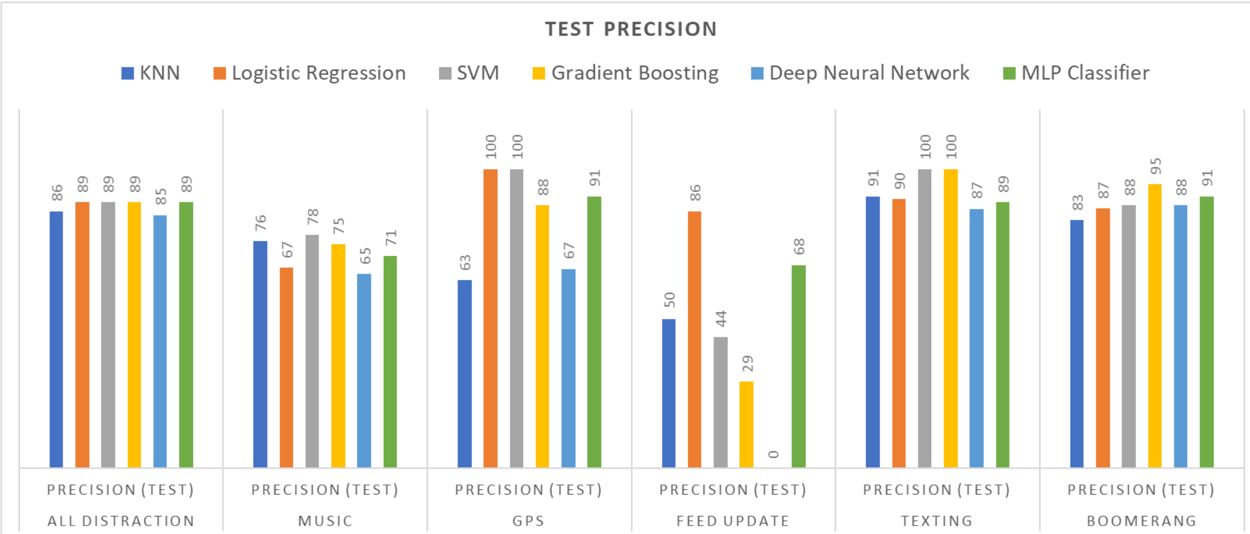
Figure 3.7: Feature importance using Random Forrest Classifier for distraction at straight road

Figure 3.8(a) and 3.8(b) presents the model performance for identifying different distraction types compared to baseline (no distraction) cases. The study included training accuracies to better understand potential overfitting issues. Note that, the precision score shows the accuracy of correctly identifying distracted driving. A high precision score indicates a low rate of false positives, ensuring that the model's predictions of driver distraction are likely to be accurate. By comparing accuracy and precision together, the study ensures the predictions are reliable and trustworthy.

For the all-distraction model where all of the secondary tasks (individual distractions) are combined as distracted driving, both SVM and MLP Classifier demonstrated higher precision and accuracy. For both non-social media distraction types (i.e., Playing Music and Using GPS), SVM and MLP Classifier again demonstrated superior precision and accuracy. However, this study found that drivers tended to engage in "social media browsing’ provided a small number of data points for a very short time interval. This small dataset may result in the model's predictions becoming towards ‘no distraction' which is the dominant sample, causing a higher accuracy score but a lower precision score. For more engaging distraction types requiring longer duration to complete like texting and Boomerang, SVM, Gradient Boosting, Deep Neural Network and MLP classifiers all perform very well with a high precision score.



(a)



(b)

Figure 3.8: (a) Training and Test Accuracy of machine learning models for Straight Road (b) Test Precision score developed for the same machine learning model.

Distracted Driving Detection for Complex Road Geometries

The study aimed to detect distracted driving activities for more complex roadway geometries including (i) merge traffic in a straight segment of highway (Scenario 2) and curved road (Scenario 3). Both segments in Scenarios 2 and 3 compare with 'no distraction' cases from the baseline scenario (Scenario 1). The correlation heatmap analysis as shown in Figure 3.9 reveals that for both curved roads and straight roads with merge traffic, acceleration, and speed do not significantly influence the detection of distraction behaviors. Instead, the driver's lateral positional behavior correlates more strongly with distraction behaviors across both roadway geometries. Regardless of the roadway geometry and distraction types, lateral behavioral attributes, such as standard deviation, maximum and maximum sudden changes in lane position, and steering wheel angle, show the highest correlations among attributes. The social media distractions "Like Post" and "Watching Reel" showed a considerably stronger correlation with those attributes. Furthermore, drivers tend to make unintentional lane-changing behavior possibly due to their cognitive and physical impairments while being distracted. As curved roads require more precise hand-eye coordination, and any secondary tasks lead to higher correlations with maximum Lateral Speed (Lspd_max), indicating drivers' attempts to compensate for lane positioning errors, which increases the chance of lateral crashes or near-miss scenarios.

	Distracted Activities at Straight Segment with Merge Traffic			Distracted Activities at Curve Road			
	Like Post	Texting	Watching Reel	Like Post	Texting	Watching Reel	
Throttle_std	0.000	0.000	0.000	0.000	0.000	-0.274	Throttle_std
Herror_std	0.342	0.000	0.000	0.435	0.000	0.274	Herror_std
Loffset_std	0.291	0.237	0.282	0.448	0.333	0.333	Loffset_std
Stpos_std	0.388	0.262	0.341	0.456	0.311	0.357	Stpos_std
Headway_std	0.228	0.000	0.000	0.000	0.000	0.000	Headway_std
Throttle_mean	0.000	-0.225	-0.208	0.000	0.000	0.000	Throttle_mean
Herror_mean	-0.392	0.000	0.000	-0.387	0.000	-0.220	Herror_mean
Headway_mean	0.000	0.000	-0.236	0.000	0.000	0.000	Headway_mean
Throttle_max	0.000	0.000	0.000	0.000	0.000	-0.294	Throttle_max
Herror_max	0.000	0.000	0.000	0.413	0.207	0.292	Herror_max
Loffset_max	0.263	0.241	0.320	0.269	0.257	0.288	Loffset_max
Stpos_max	0.341	0.213	0.236	0.342	0.000	0.350	Stpos_max
Headway_max	0.000	0.000	-0.217	0.000	0.000	0.000	Headway_max
Herror_min	-0.361	0.000	0.000	-0.417	0.000	-0.253	Herror_min
Loffset_min	0.000	0.000	0.000	-0.284	0.000	0.000	Loffset_min
Stpos_min	-0.395	-0.248	-0.343	-0.397	0.000	-0.233	Stpos_min
Headway_min	0.000	0.000	-0.213	0.000	0.000	0.000	Headway_min
Throttle_rdiff	0.000	0.000	0.000	0.000	0.000	-0.211	Throttle_rdiff
Herror_rdiff	0.285	0.000	0.000	0.425	0.000	0.267	Herror_rdiff
Loffset_rdiff	0.229	0.000	0.213	0.357	0.318	0.272	Loffset_rdiff
Stpos_rdiff	0.367	0.218	0.281	0.483	0.300	0.333	Stpos_rdiff
Headway_rdiff	0.306	0.000	0.000	0.000	0.000	0.000	Headway_rdiff
UnLchn_fr	0.380	0.000	0.000	0.529	0.408	0.382	UnLchn_fr
Lspd_max	0.399	0.000	0.000	0.441	0.249	0.257	Lspd_max

Figure 3.9: Correlations between driving behavior attributes and distracted activities on a straight road with merging traffic and curve road

(a) Straight Section with Merge Traffic (b) Curve Road

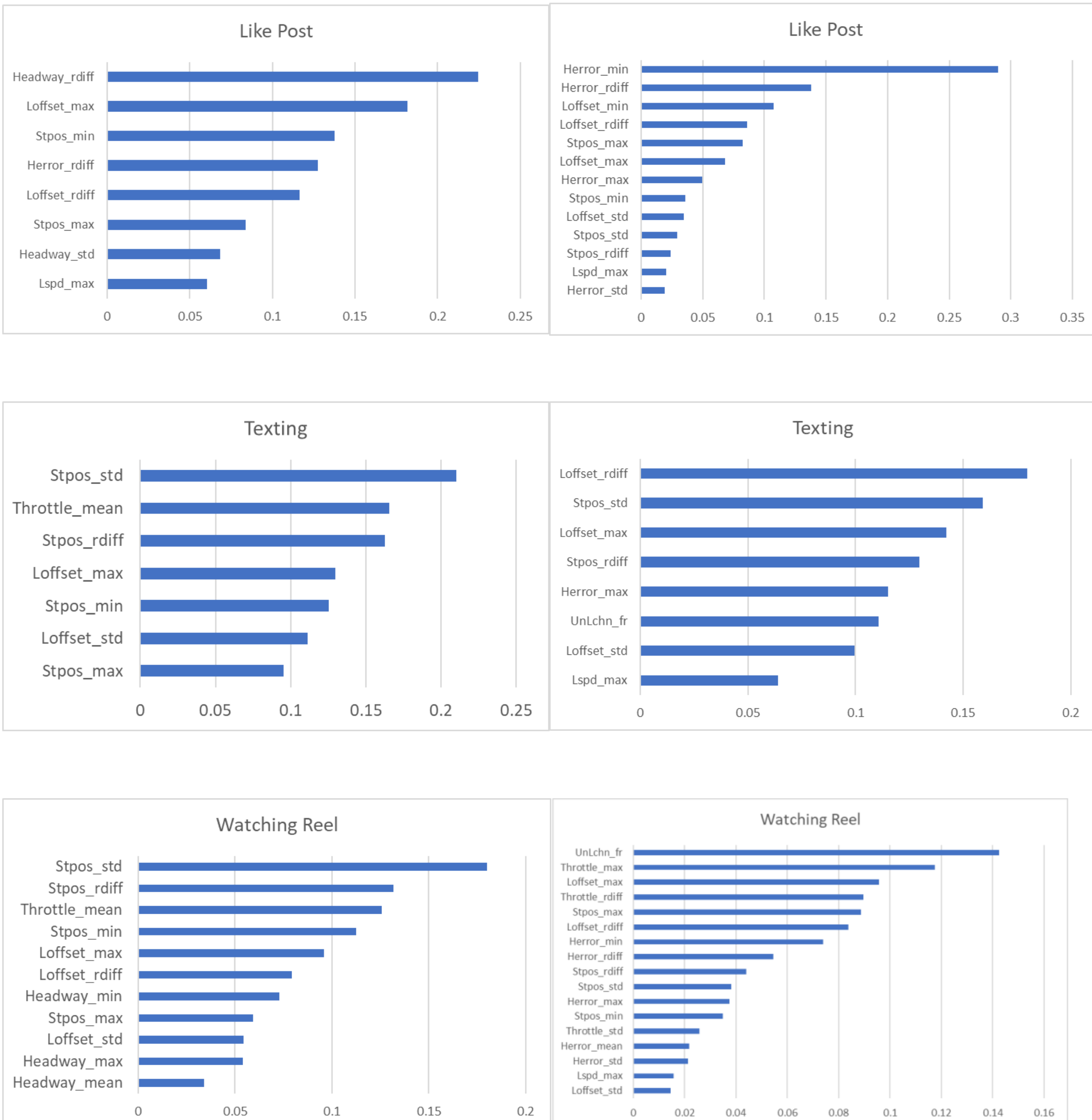
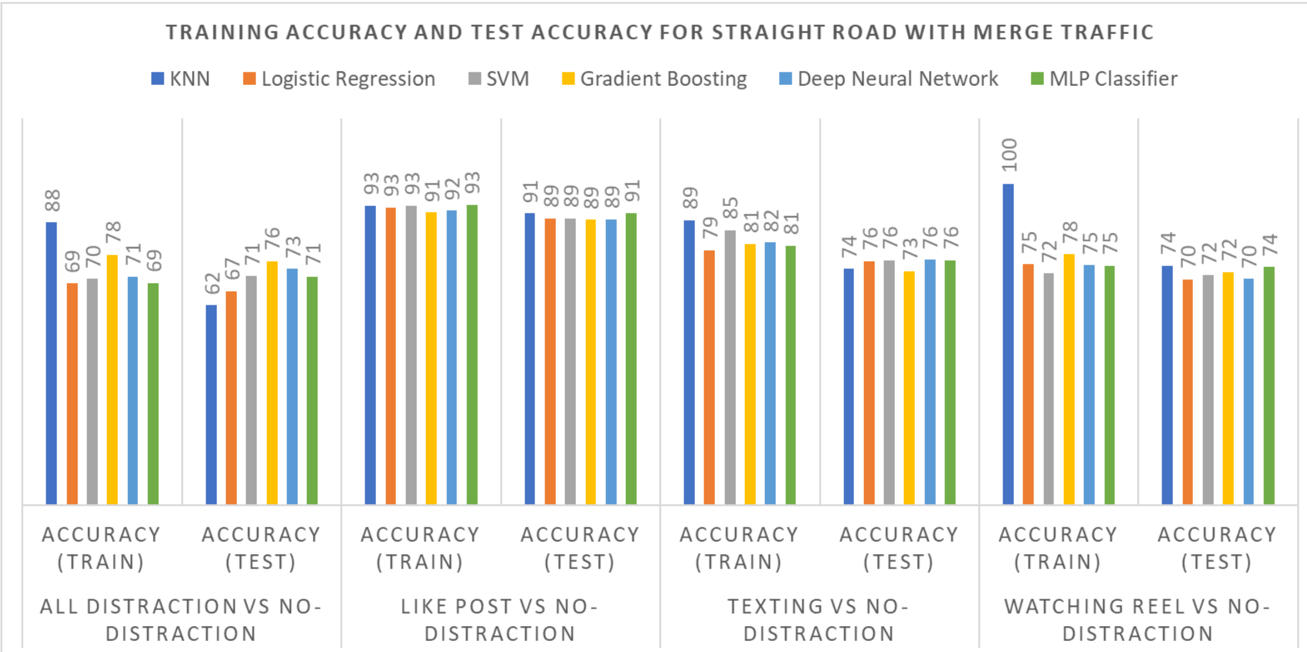


Figure 3.10: Feature importance using Random Forrest Classifier for distraction at (a) straight road with merge traffic and (b) curve road

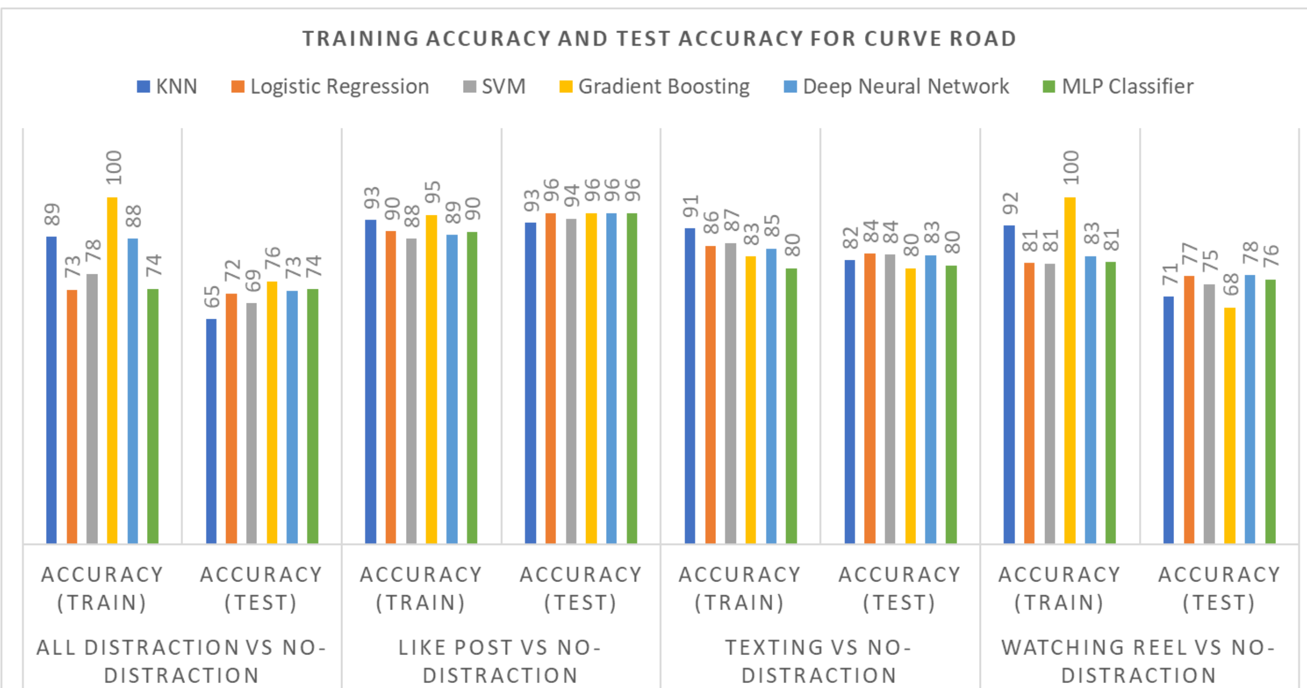
The feature importance tool revealed varying behavioral importance factors for different distraction types in Figure 3.10 (a) for straight road with merge traffic and Figure 10 (b) for curve road. ‘Like Post’ made drivers fail to notice the front vehicle location properly, leading to a sudden maximum change in headway (Headway_rdiff) the most significant variable. Conversely, the same distraction on curved roads showed Heading Error (Herro_min and Herror_rdiff) as the first two most critical behavioral attributes. Moreover, when comparing important attributes for distracted activities on straight roads with merge traffic, headway attributes proved essential for detecting distractions, except for texting. However, headway had no impact on curved roads on identifying any distractions. The challenges of maintaining the vehicle’s position in the center of a lane while distracted led to unwanted lane changes or excessive swerving, making the frequency of unsignalized lane changes an essential factor for both ‘texting’ and ‘watching reel’ distraction types.

Figures 3.11 (a) and (b) illustrate the training and test data set accuracies for the straight road with merge traffic and curve road, respectively. Figure 12 displays the test precision scores for the corresponding models. Combining all distractions for both straight roads with merge traffic and curve roads led to lower accuracy and precision rates than straight roads, indicating more unpredictable behavioral patterns on complex roadway geometries. However, SVM and MLP Classifiers demonstrated the best performance across both environments, achieving over 70 percent accuracy for straight roads with merge traffic and 74 percent for curve roads. Gradient Boosting and Deep Neural Networks exhibited better performance for straight roads with merge traffic, achieving more than 70 percent accuracy for all distraction types.

Machine learning models showed promising results in detecting 'texting' with higher accuracy and precision. However, the models performed less effectively detecting drivers engaged in 'watching reel' on straight roads with merge traffic, showing lower precision. Gradient Boosting and MLP Classifier algorithms slightly outperformed other models in this aspect. Conversely, for the same distraction on curve roads, SVM, Deep Neural Network, and Logistic Regression performed better, achieving an accuracy of up to 78 percent and an accuracy of 86 percent.

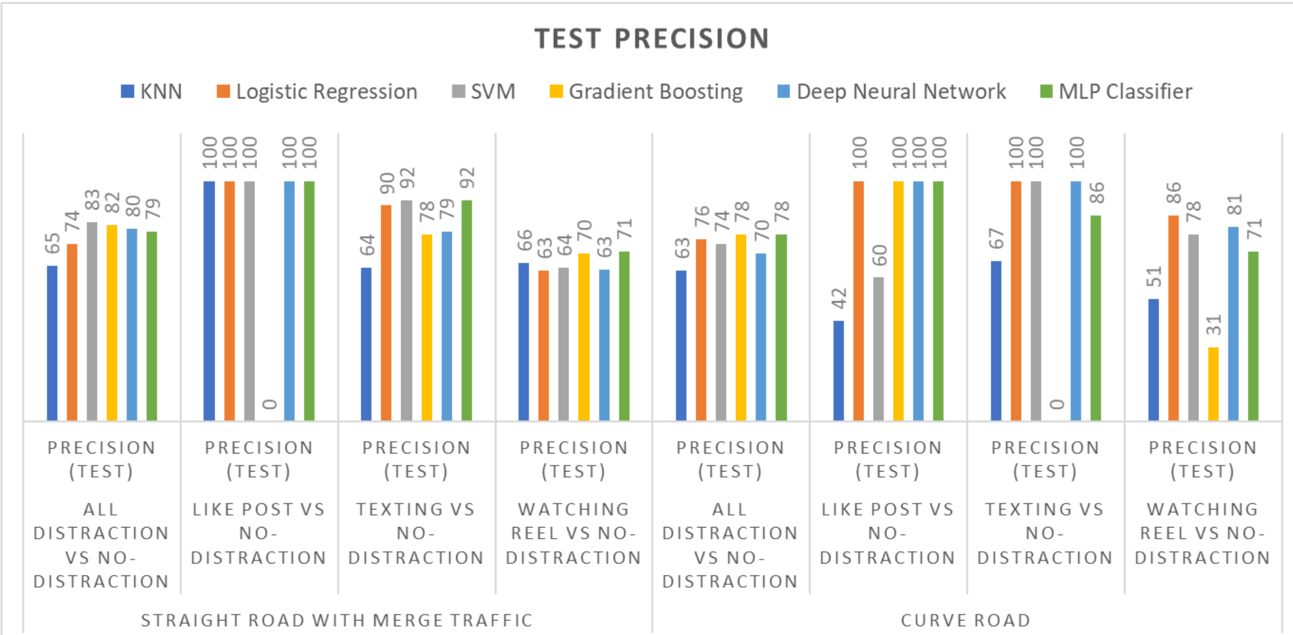


(a)



(b)

Figure 3.11: Training and Test Accuracy for Machine Learning Algorithms for (a) straight road with merge traffic and (b) curve road.



(a) (b)

Figure 3.12: (a) Test Precision score developed for the same machine learning models for straight road with merge traffic as shown in figure 11(a); (b) Test Precision score developed for the same machine learning models for curve road in figure 11(b)

3.5 Conclusion

This study focuses on making a notable contribution to the existing research on social media-based distractions by going beyond the conventional focus on common simple distractions like browsing using social media apps. Instead, it introduces other prevalent social media-related distractions, such as taking a Boomerang, liking a post, and watching a video or reel. Unlike previous studies that primarily relied on statistical analysis and had limited performance attributes, this research took a more comprehensive approach. The study investigates various behavioral characteristics to obtain a better understanding and offers reasons for distracting events due to social media. The approach made it possible to explore the effects of distracted driving on drivers' behavior in more detail. The findings from correlation and feature importance analyses are particularly noteworthy. They revealed that distracted driving significantly affects drivers' angular and lateral properties, rather than longitudinal driving behavior attributes. Among the longitudinal attributes, headway attributes emerged as the critical indicators uniquely affected by distracted driving.

Additionally, this study revealed that different roadway conditions exert varying impacts on driving behavior. On straight roads, drivers' angular properties, such as Headway Error, were predominantly affected by distractions, followed by a negative relationship with acceleration, suggesting a tendency to slow down while distracted. However, when distracted drivers encountered merging traffic, they focused more on lateral positional changes, leading to a more significant influence on steering wheel position and lane position. Consequently, their car-following behavior was affected, making headway an influential factor. Similarly, lateral driving attributes played a crucial role on curved roads, and the additional stress of negotiating curves led to unsafe lane-changing behavior, including sudden lane changes without signals.

The study also developed several distracted driving detection models using six machine learning algorithms. Proper hyperparameter tuning and crossed validation supported identifying the best machine learning algorithms. The multi-Layer Perception (MLP) classifier exhibited the most desirable performance, with close training/testing accuracy, and high precision for most secondary tasks. On the other hand, the K-nearest neighbor (KNN) demonstrated inferior performance due to a high overfitting issue among the models. A significant finding of this study is that despite the driving environment's complexity, the models could detect distracted driving with more than 75 percent accuracy and 85 percent precision for most cases. Even when combining all distractions into one category, the models achieved at least 75 percent accuracy and 78 percent precision.

As a potential application, the developed models could be used for crash investigation to determine the likely engagement of drivers in secondary tasks and activities prior to crashes. One proper application of employing algorithms could serve a dual purpose: alerting drivers or even potentially turning off their phones instead of solely investigating post-crash incidents. However, directing these efforts toward preventing crashes before they occur would be a great application of this technology. However, one of the major limitations of this study is the availability of data for certain distractions due to shorter durations required to finish those activities. Despite this limitation, the study offers valuable insights into the use of behavioral attributes for identifying distracted driving patterns and highlights the potential of machine learning models in enhancing road safety.

Chapter 4: A risky and distracted driving behavior recognition based on the driving simulator trajectory data.

4.1 Introduction

Operating a vehicle demands constant visual and cognitive attentions. Secondary tasks range from reading a newspaper at an intersection, grooming for work, and engaging in a conversation with passengers add an extra burden to drivers. and significantly diminish driving capabilities and heighten the risk of crashes or near-misses. Furthermore, modern vehicle technology and the integration of advanced wireless communication devices have introduced an additional layer of distraction (Bakhit et al., 2018). Over the past few decades, cell phones have become significant sources of distraction with various forms of impairments, such as physical impairments with holding a phone or cognitive/visual impairments with hands-free calls, texting, or using social media (Rahman et al., 2021). According to the World Health Organization (WHO), texting and conversing increases the risk of a crash by 2% to 9% (WHO, 2011).

The frequency of distracted driving by cell phone varies among age groups. Younger drivers particularly have drawn attention due to their higher incidence of texting while driving. An online safety study by Tucker et al. (2015) among young drivers stated that male cellphone users are more prone to speeding while texting. Additionally, the research noted a decrease in texting among teen drivers as they gained more experience behind the wheel. Hossain et al. (2022) identified young male drivers who drive light trucks as more vulnerable to crashes when texting and browsing rather than talking or listening. A national study in 2003 involving 449,049 teenage driver crashes used a model to predict the likelihood of three common crash types based on four distraction categories: cognitive, cell phone-related, in-vehicle, and passenger-related distractions. The findings demonstrated how particular distractions raised the likelihood of particular crash types. For instance, rear-end crashes are more likely to occur at intersections due to cognitive and passenger-related distractions (Neyens et al., 2007). Simon et al. (2019) conducted a simulator experiment involving young drivers. They discovered that texting increased the duration drivers spent not looking at the road by 400 percent and raised the likelihood of lane variability by up to 50 percent which leads to high lateral crash possibility. The majority of the research on younger adults and distracted driving has focused on the distractions caused by texting on mobile phones. To better understand the broader implications and associated risks of these social media distractions on the younger generation, it is crucial to investigate their impact on driving behaviors. There is a critical need to identify and assess how distractions from social media specifically influence drivers among this young generation, shedding light on the extent of their riskiness and potential effects on road safety.

Several studies have attempted to dissect drivers' psychophysiological states, cognitive abilities, and adherence to traffic regulations using questionnaire surveys and demographic information to quantify their inclination toward risky driving behaviors. Vanlaar et al. (2008) developed a framework, for example, that classifies driving behavior into aggressive and assertive patterns and defines unsafe driving behaviors based in psychology. The driving habits and demographics of drivers can significantly influence their inclination towards high-risk driving behaviors. This is caused by a number of factors, including age. Younger drivers, particularly those in their teens and early twenties, frequently exhibit riskier behaviors because of a combination of inexperience and overconfidence (Williams, 2003). Gender differences are also evident, as men are more likely than women to drive aggressively and fast. Experienced drivers tend to exhibit more cautious habits, while novice drivers are more likely to engage in risky behaviors (Laapotti et al., 2003). Developing targeted interventions and policies aimed at improving road safety and reducing risky driving practices requires an understanding of these interconnected factors. Essentially, research on the effects of daily routines and demographics on driving risks is insightful, but it may not provide the depth and accuracy needed to fully address the risks associated with distracted driving brought on by electronic devices or other emerging distractions in today's technologically advanced society.

Distractions make it difficult for a driver to concentrate on the task at hand, which is driving carefully and safely. This decreased ability to focus, along with increased stress and emotions, may create an atmosphere that encourages drivers to engage in aggressive driving behaviors on the road, which can result in a variety of hazardous scenarios (Hancock et al., 2003). Researchers have analyzed and identified different contributing factors of aggressive or risky driving. According to NHTSA – speeding, sudden or erratic lane speed changes, unpredictable and uncertain lane changes, illegal driving on the median, sidewalk, or shoulder, improper car following, failure to obey a traffic signal, signs, and instructions, and prohibited passing are one of the significant signs of aggressive driving. Speeding is the number one factor and has been responsible for one-third of the primary motor vehicle fatalities for the last two decades (NHTSA, 2020). Field studies conducted back in the 1930s and 1940s tried to understand the safe speeds and acceleration for drivers until they felt discomfort. A study revealed that people tend to feel most comfortable traveling at an average speed of 35 mph on straight roads, 25 mph on gentle curves, and 15 mph on sharp curves. Additionally, they observed that the average comfortable acceleration stood at 4 ft/s^2 on straight roads, 3 ft/s^2 on gentle curves, and 2 ft/s^2 on sharp curves. (DeSilva, 1942). Chuan et al. (2015) noted that speed-changing behavior is one of the high-risk behaviors because of their volatile nature. Some studies focus on drivers' lane change behavior to define aggressive driving. In Beijing, Tian (2016) gathered data on lane-changing behaviors from 38 car drivers across various highways and arterials. Employing hierarchical cluster analysis and K-means clustering techniques, he examined drivers' lane-changing patterns, focusing on shorter lane change times. The research revealed that smaller lane change values were associated with a more aggressive driving approach. The study categorized drivers into five types based on their behaviors, ranging from cautious to highly aggressive drivers (Tian, 2016). While these studies provide insightful information about specific risky driving behaviors within predefined events or historical contexts, there is still a critical knowledge gap regarding the complex relationships between these behaviors and distracted driving. It is imperative that more research be done to determine the precise relationships between distracted driving and different aggressive or risky driving traits. By closing this gap, researchers can better understand the complex interactions between particular behavioral patterns and distracted driving, which will improve literature of the underlying causes of risky driving behaviors.

This research includes two main goals – the primary goal is to measure the frequency of unsafe driving practices or driving errors from different types of distractions among the younger drivers. Second, by identifying the most significant correlations between various driving behavioral errors, the study aims to understand drivers' behavioral patterns and demographics in connection to these errors. By exploring these goals, the study seeks to clarify how distractions affect driving errors and reveal more profound relationships between driving behaviors and drivers' traits and daily routines.

4.2 Literature review

Risky Secondary Tasks

Previous studies have analyzed driving performances to develop a risk measurement of distracted driving in several longitudinal and lateral dimensions (such as speed control, reaction time, lane variation, etc.). Papantoniou et al. (2019) conducted a driving simulator study and compared the performance of two types of distraction- conversation on a cell phone and without a cell phone. They developed a structural equation model (SEM) to understand driving performance based on the driving errors (sudden brakes, speed limit violations, outside lane violation, hit of sidebars, etc.) of drivers under the distracted condition with variable traffic and environmental conditions. Li et al. (2019) examined the effect of hands-free and hand-held mobile phone usage and no phone conditions in a simulated experiment. Based on drivers' collision events and possibilities, a two-step cluster analysis classifies the drivers into two groups-low risk and high risk. Then they developed a logistic regression model to identify the relationship between drivers' characteristics, mobile phone use, collision avoidance performances, and their involvement in the collision risk. According to the findings, drivers using a hand-held phone had a slower brake reaction time and a higher chance of getting in a crash than those using a hands-free or not using any phone. By regulating car-following speed and distance (i.e., Time-to-collision (TTC)) in a distracted state, drivers take compensation behavior for lessening the risk of safety-critical incidents. Another study examined the effect of hands-free and hand-held mobile phone usage concerning no-phone conditions focusing on the rare end collision (Li et al., 2016). They developed a simulator experiment where 42 drivers drove a car following a situation with a possibility of rear-ending collisions by the abrupt deceleration of the following vehicle. The study used a minimum distance between the participants and the leading car to measure the possible collision risk measurement model. The analysis shows hand-held cell phone usage provides a higher collision risk than the hands-free situation. Pawar and Patil (2018) assess the driver's risk profile due to distracted behaviors by cell phones based on response time before possible conflict (RTPC), average speed, and deceleration rate. They have designed a 10.4 mile simulator experiment comprised of semi-urban areas with many possible conflicting events at unsignalized intersect to understand drivers' instant reactions time under distracted vs. no distraction case. The analysis shows that the top 90th percentile data for RTPC was 9.63s for non-distracted cases and 12.57s for distracting circumstances. Choudhury and Velaga (2019) used Generalized Estimating Equations (GEE) to statistically model crash probability under distracted behavior caused by eating, drinking, and texting. Eighty-nine participants approached the simulated unsignalized intersection through the secondary road for this purpose and came into contact with several vehicles on the major road. The analysis shows that a 1m/s increased approach speed can lead to a 26% chance of crash risk. Their result shows an interesting finding that distracted driver accepts higher gap acceptance for oncoming vehicle resulting in a 76% decrease in crash risk for texting to eating and drinking task.

Types of Unsafe Driving

Several studies have developed different methodologies to identify drivers' risky driving behaviors. The threshold method is one common and recognized method for defining risky driving behavior. Previous studies have defined different threshold values of drivers' driving characteristics and developed a methodology to label a driving behavior as safe or dangerous. Dingus et al. (2006) derived threshold values for lateral and longitudinal accelerations (0.6g or greater), alongside yaw rates (4 degree change in heading within 3 second and headway (2 s), based on data from 100 drivers. Exceeding these thresholds in a vehicle's kinematic characteristics flags the possibility of a collision event. Paefgen et al. (2012) assessed risky driving behavior via a mobile app, defining *risky acceleration* as 0.1g or $1 m/s^2$ and deceleration as -0.1g or $-1 m/s^2$. Baldwin et al. (2004) designed a driver monitoring system around sudden accelerations and decelerations of $1.5 m/s^2$. Osafune et al. (2016) analyzed accelerometer data to distinguish safe from risky drivers, identifying thresholds for acceleration over $2.4 m/s^2$, deceleration above $1.4 m/s^2$, and left acceleration surpassing $1.1 m/s^2$. Their classifier achieved 70% classification accuracy and detected risky drivers with 83% accuracy. Considering road type, Fitch et al. (2013) set thresholds like longitudinal acceleration exceeding 0.3g for a vehicle at 64 km/h on a highway. Bao et al. (2021) conducted K-means clustering to understand different risky driving categories based on naturalistic driving data. It identified a sudden deceleration of close to 1.6 m/s2 as risky and can lead to crashes or near-crashes.

Speeding poses a significant danger on the roads by increasing the chances of crashes and worsening their outcomes. When vehicles surpass the designated speed limits, drivers have less time to react, making it harder to avoid collisions. Moreover, higher speeds amplify the impact force during crashes, leading to more severe injuries or even fatalities. Research by Elvik (2015) exemplifies this, highlighting that a 50 percent increase in impact speed from 40 to 60 mph results in a staggering 125 percent increase in the energy that needs managing during a crash. Managing this additional energy becomes difficult, which impacts the vehicle's structure and increases the risk of serious injuries. Craig and Jack's study in Australia (1998) underscored that the risk of being involved in a casualty crash rises significantly when drivers exceed posted speed limits. For instance, going from 60 km/h to 65 km/h doubles the risk, and going from 60 km/h to 70 km/h quadruples it. Bao et al. (2021) identified risky driving behaviors through speed analysis, categorizing drivers going over 8 km/h (5 mph) and 17 km/h (10 mph) as high-risk clusters. Elvik (2009) conducted a meta-analysis examining the effects of speed variations on rural and freeway road safety. It demonstrated a strong correlation: a 1% increase in mean speed causes an increase in collisions with injuries of 2%, severe crashes of 3%, and fatal crashes of 4%. The Insurance Institute for Highway Safety (IIHS) discovered a sharp rise in motor vehicle fatalities during its 1996 investigation into the effects of raising interstate highway speed limits. According to their research, increasing speed dramatically increases the likelihood of fatal crashes; an increase of just ten mph doubles the likelihood of dying in a crash (IIHS, 2023).

Tailgating and keeping an excessively close distance are examples of inappropriate headway and are regarded as unsafe driving practices. Too little headway slows down reaction times and makes it more difficult to react to abrupt alterations in traffic. Due to the limited space available for safe braking and maneuvering, this behavior raises the possibility of rear-end collisions. In order to determine drivers' comfort thresholds in automated driving scenarios, Li et al. (2019) conducted a driving simulator study. Their research revealed that when headway deviated by 5% to 9% from the desired spacing, drivers felt uneasy and turned off automated driving modes. The National Highway Traffic Safety Administration (NHTSA) recommends a minimum headway of 2 seconds for passenger cars under favorable weather and road conditions. This guideline suggests that drivers should count at least two seconds from when the vehicle ahead passes a fixed point until their vehicle reaches the same point (NHTSA, 2020). Research by Dingus et al. (2011) revealed that safer headway averages over 1.4 seconds in naturalistic driving studies. According to the study, the risk of a crash or near-crash increased exponentially when headway fell below one second.

Unsignalized lane changes and lane deviations are dangerous driving practices that involve changing lanes without checking blind spots or flashing warning lights. This action increases the chance of a crash, which could result in serious injuries or fatalities. It entails changing lanes abruptly, increasing the risk of colliding with oncoming traffic, pedestrians, or bicyclists in nearby lanes (Li et al., 2022). Li et al. (2017) present a logistic regression-based crash risk prediction model for lane-changing behavior on urban expressways. Using data from a driving simulator, they examine variables that affect the likelihood of a collision during lane changes, such as the distance from the trailing vehicle, the speed difference between lanes, and the frequency of lane changes. Their results show that a maximum lane deviation of more than one meter considerably increases the crash risk. Unsignalized lane changes are a significant contributor to auto crashes, especially at intersections, according to Sen et al. (2003). Crashes may result from drivers failing to yield the right-of-way or misjudging the movements of other road users. Due to speed differences between vehicles, this behavior also increases the risk of rear-end collisions, frequently resulting in abrupt braking or swerving.

Methodologies used to define unsafe driving

Risky and unusual driving occurrences have been recognized through different studies by research on risk assessment methodologies and safety intervention. Chen et al. (2021) proposed a distribution-based and box-plot-based method for defining the risky driving behavior of cars and heavy traffic. In addition to traditional driving characteristics, they focused on three types of risky driving behavior- speed-unstable driving, serpentine driving, and risky car-following driving. The results show considerable disparities between threshold values for light vehicles and large trucks when the risky behavior consists of speed, unstable driving and dangerous car following. Toledo et al. (2008) developed a driver's risk index based on drivers' risky behavioral data, such as speeding, emergency braking, and sharp lane changes. They provide weights to each of the behaviors based on the duration of the video. Simsek et al. (2013) evaluated speed violations, vehicle idle duration, and fuel cost collected from a logistic firm's employees and developed a conceptual framework for drivers' performance evaluation. They used a Global Positioning System (GPS) terminal to gather drivers' behavior data and used statistical process control (SPC) to analyze the data. Based on the result, they provide a guideline for the company to discover the performance appraisal of the drivers. To recognize unsafe driving behaviors like speeding, rapid acceleration, emergency braking, and sharp turns, Castignani et al. (2015) presented a fuzzy approach. By accumulating the frequency of risky driving incidents over a predetermined distance, researchers gave scores ranging from 0 (worst) to 100 (best) to drivers. However, they only take into account a subset of driving events, such as acceleration, braking, turning, and lane changing. Other types of driving events, such as speeding, tailgating, or weaving, may be relevant for driver profiling. A study in Sydney, Australia, by Ellison et al. (2015) evaluated the driving data of 133 drivers collected from GPS. It proposed a standard assessment of driving behavior based on the risk of a fatal crash provided by driver behavior profiles. The study used fixed values from the literature to quantify the dangerous behavior brought on by speed, acceleration, and deceleration. Based on the contribution of each activity to crash risk, the study weighted the score for each risky behavior.

The literature mentioned above however overlooks a major crash factor: distraction types, if any. It is imperative to establish a link between crash probabilities and risky behaviors, with a focus on the interactions between these variables and driver distractions to comprehend this relationship and develop safety interventions. Understanding the relationship between different types of distractions and specific risky driving behaviors is important because not all distracted driving situations lead to the same dangerous behaviors. Developing these links can provide a more in-depth understanding of crash-avoidance tactics adapted to various distractions and unsafe driving practices. Research on risky or aggressive driving often neglects a prevalent distraction pattern: social media-induced distraction. This type of distraction has become increasingly frequent in today's driving scenarios, yet there is a significant gap in the literature addressing risky driving behavior associated with social media distractions. Therefore, there is a critical need to spotlight social media-related distractions as a focal point in driving behavior research.

Historical literature on Application of Association Rule Mining on Distracted Driving Research

Association rule mining can be a useful tool to understand the complex relationships between various factors in the distracted driving. since it reveals latent patterns and associations among various distractions and the ensuing driving behaviors. This technique can assist in identifying complex relationships by highlighting the distractions that frequently coincide with unsafe driving behaviors. Rahman et al. (2019) conducted an observational study among 3,727 drivers and found that male drivers were notably involved in cellphone manipulation during peak hours in urban areas and engaged in cellphone conversations during rural peak hours. Conversely, female drivers are more associated with using cell phones at urban intersections across peak and off-peak hours. During peak hours at rural intersections, they are more likely to use devices while conversing and driving along continuous segments. Hossain et al. (2022) identified that cellphone distraction and no seatbelt usage are frequently visible in confirmed injury crash scenarios. Cellphone crashes of novice teenagers at intersections has strong association with talking/listening rather than texting/browsing/dialing and reaching for/ answering/locating. Jazayeri et al. (2021) adopted specific data collection methods to yield distinct sets of variables grouped into three categories. The first comprises driver-related factors (age, experience, cognitive abilities) and measures of distraction (physiological changes, eye movements, brain activity). The second involves vehicle or simulator dynamics (speed, acceleration, steering behavior). Lastly, the third category encompasses environmental variables, detailing internal or external distractions and environmental characteristics like traffic signs and road layout. The result showed that the most common co-occurring behaviors are distracted driving and improper turns. Xing et al. (2023) designed a natural driving experiment and applied association rule mining between distracted driving and multi-source features (including emotion, valence, arousal, driving tasks, cumulative driving hours, velocity, longitudinal acceleration, lateral acceleration, heading rate, presence of intersections, traffic status, and driving time of day). The results show that the emotional state of drivers has intriguing links to distracted driving, with cognitive and operational distractions associated with positive driving emotions. In contrast, auditory distractions caused by passenger interference can be linked to negative driving emotions.

Nevertheless, a critical component of these studies is absent: a connection between particular distracted driving behaviors and the associated risky driving actions. These studies primarily concentrate on general distraction and its relationship to behavior. Comprehending this relationship is essential to understanding driving behaviors in general. Further research into how demographic data and everyday driving behaviors impact these driving patterns is also essential to understanding the implications for road safety.

4.3 Data Collection and Methodology

Participant Information

The research conducted a driving simulator experiment, with students from the University of Texas at Arlington. Initially, a pre-screening survey selected 50 candidates using eligibility criteria such as proficiency in reading and speaking English and possessing a valid driver's license. The study gathered various demographic details from participants, such as age, gender, income level, vehicle ownership, and daily travel patterns. Additionally, we collected information on participants' distracted activities within the seven days following the experiment. Table 4.1 presents the questionnaire-based information gathered from the participants for a comprehensive overview.

Table 4.1: Drivers demographic information and weekly driving activities

Demographic Information	Count	Percentage of Total Participant
Male	33	66.0%
Female	17	34.0%
Age		
25 years and less	16	32.0%
26-30 years old years old	26	52.0%
More than 30 years old	8	16.0%
Low-Income Household	23	46.0%
Does not own a Vehicle	15	30.0%
High Frequency trips	6	12.0%
Daily Longer trips (More than 45 mins)	14	28.0%

Self-reported distracted Activities within seven days range prior to Experiment		
Making or Receiving Phone Call	38	76.0%
Texting	34	68.0%
Using Social Media App	28	56.0%
Using Vehicle Navigation System i.e. GPS	45	90.0%
Eating or Drinking	35	70.0%
Get Distracted with the Distracted by Passengers	34	68.0%
Grooming	16	32.0%
Cognitive distraction (thinking, day- planning, etc.)	41	82.0%

Experimental Setup

As mentioned above, one of the primary objectives of the study is to understand the effects of social media distracted driving on different traffic and roadway conditions. This study examines the impact of various road conditions and traffic scenarios within highway settings, illustrated in Figure 4.1 below.

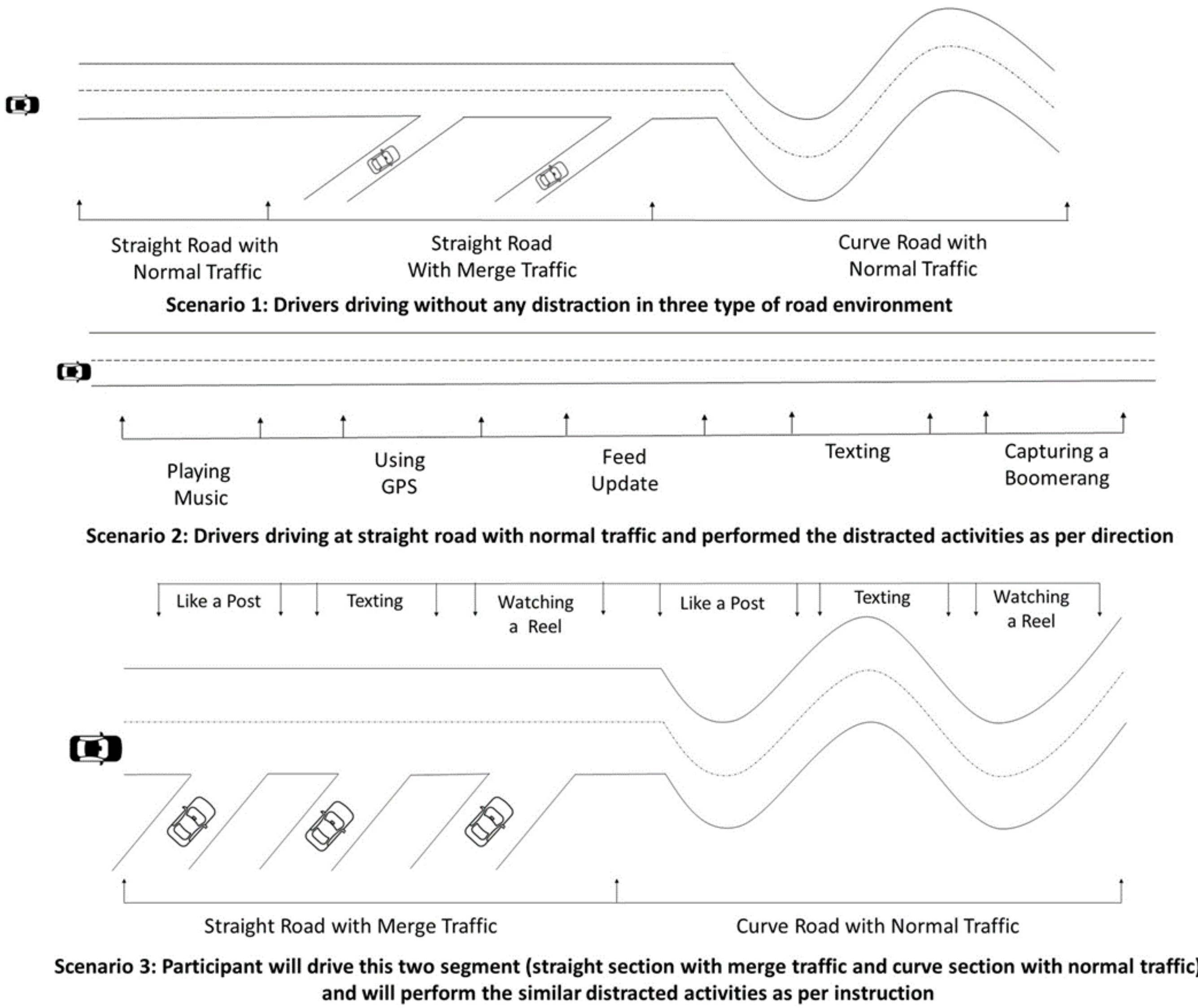


Figure 4.1: Graphical explanation of three scenarios and the experimental setup at a driving simulator

The study chose the following distraction activities as the most common social media and Nonsocial media activities to understand their impacts on driving behaviors:

- Music: Playing a song in any music app (e.g., Youtube, I-tunes, etc.)
- GPS: Using GPS to find a location
- Texting: Texting using a social media app
- Feed Update: Browsing using any social media app
- Like: Liking a post on any social media app
- Boomerang: Creating a Boomerang (short video) using Instagram or another social media app
- Reel: Watching a reel or video in a social media app.

The study initially classified distractions into two distinct categories: non-social media distractions (involving activities like listening to music and using GPS) and social media distractions. Within the realm of social media distractions, a further classification was made based on the level of active participation and the attention-demanding nature of the activities. These social media distractions were then subcategorized into two groups: less engaging social media distractions and more engaging social media distractions. Less engaging activities included actions like updating a feed or liking a post, which typically requires minimal user interaction and attention. In contrast, more engaging social media distractions encompassed activities such as capturing a boomerang, watching a reel, and texting, demanding higher levels of interaction and attention from users due to their more immersive and engaging nature (NBFC, 2018).

Two researchers conducted the experimental simulation process. Before commencing the experiment, participants underwent a 5-minute familiarization phase using the driving simulator to ensure comfort and prevent any potential discomfort, such as nausea or dizziness. Once participants felt at ease, the experiment proceeded. The experimenters clearly instructed the participants regarding the starting, stopping, and executing distraction tasks while navigating the general highway scenario, adhering to a speed limit of 65 mph. In scenario 1, participants were given to participants to refrain from using cell phones to establish their natural or baseline driving conditions.

The first experimenter recorded the simulator times for the distraction scenarios and guided the participants accordingly. Simultaneously, the second experimenter utilized a stopwatch to record the start and end times of each tasks notifying the first experimenter upon the completion of each task by a participant. For scenarios 2 and 3, participants used their own smartphones and received instructions regarding their secondary tasks from the experimenter. In scenario 2, participants engaged in five secondary tasks, comprising three conventional distractions (Music, GPS, and Text) and two social media distractions (Feed and Video). For instance, tasks included using the music app to select a song, browsing a location on the GPS app, and responding to a question from the experimenter via text, ensuring a complete sentence without abbreviations or acronyms. During scenario 3 involving weaving and curves, participants completed one conventional distraction (texting) and two social media distractions: liking a post on an Instagram account set up for the experiment by the experimenter and watching a reel (a video of up to 30 seconds) chosen by the experimenter. The sequence of secondary tasks was randomly assigned to each participant, with the initial task assigned when the driver reached a saturated flow rate at the onset of both distracted scenarios. Before the experiment, the experimenters ensured that participants were familiarized with various secondary tasks and confirmed their prior experience with them.

Feature extraction and feature pre-processing

The driving simulator provides more than 20 feature attributes of driving parameters, as introduced earlier, every 1.67 s as shown in Figure 4.2. Like chapter 2, the study used scenario 1 as the base scenario to understand each participant's no-distraction driving behavior.

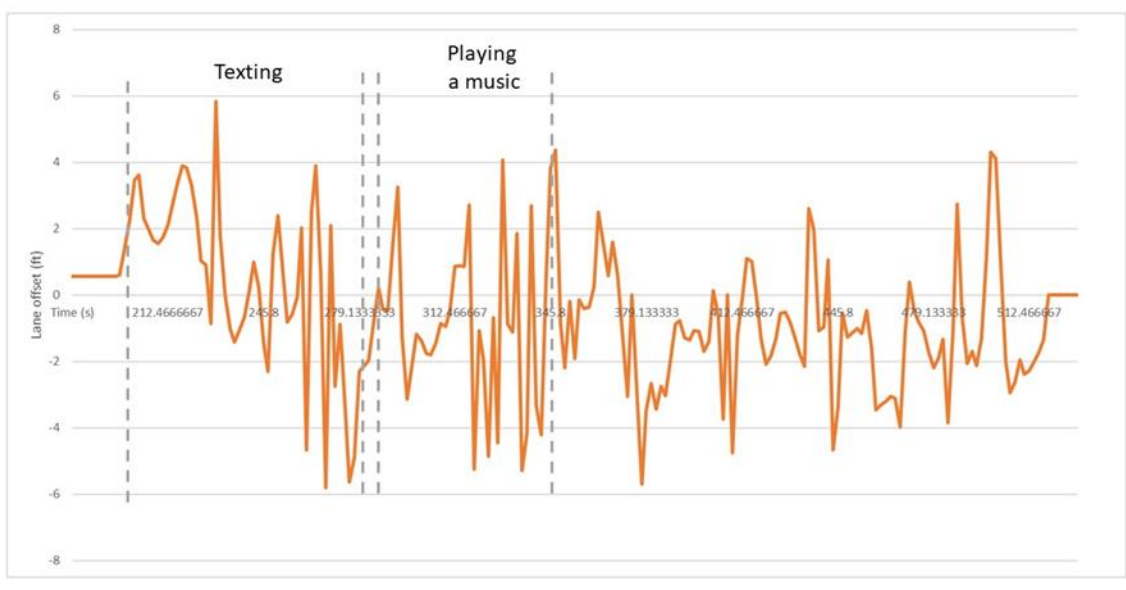


Figure 4.2: Example of a time series data showing lane offset.

Performance Metrices and Threshold Selection

Based on the above literature review, this study implemented specific performance attributes to characterize drivers' behavior under various distractions for each participant:

1. Sudden Deceleration: The simulator records acceleration and deceleration values. This study focuses on instances of maximum deceleration. Sudden deceleration of 1.4m/s^2 or greater is categorized as a sudden deceleration error (Osafune et al., 2016).
2. Risky Headway: The simulator provides headway distance between the vehicles. Risky headway is identified when the minimum headway between each participant's vehicle and the vehicle ahead is less than 75 meters (Dingus et al., 2011).
3. Lane Excursion Error: Using Lane offset data from the simulator (in meters), risky lane excursions are determined. Any lane offset or excursion exceeding 1 meter from the lane center in either direction is considered a risky or error lane offset (Li et al., 2017).
4. High-Speeding Error: The simulator records vehicle speed in mph. As the experimental highway's speed limit is 65 mph, any speed exceeding 75 mph is classified as a speeding error (Bao et al., 2021).
5. Changing Lane without Signal Error: The simulator indicates when a participant activates a left or right turning indicator. This study examines the frequency of lane changes with

and without the indicator. Any lane change made without the signal indicator is marked as a Changing Lane without Signal Error.

These performance attributes provide a structured framework for this study to assess and categorize driving errors or risky behaviors associated with different distractions, aiding in the systematic analysis of drivers' behaviors under varying distraction scenarios.

Association Rule Mining

Association rule mining is an advanced and descriptive data mining tool that operates within a rule-based framework. Its main purpose is to locate and understand the important relationships between different variables. Unlike traditional methods, such as hypothesis-driven analyses, the association rule approach does not rely on predetermined hypotheses for investigation. The strength lies in its exceptional robustness in extracting meaningful insights and an ability to handle missing or incomplete data seamlessly. Because of its unmatched flexibility in revealing hidden relationships between various factors influencing distracted driving, this method stands out as an effective way to conduct multifaceted risky driving analysis (Kong et al., 2021).

The study employed the Apriori algorithm to explore the correlation between distracted driving, driving errors, and participant demographics. This algorithm utilizes a bottom-up hash tree structure and level-wise search techniques to efficiently extract frequent item-sets from the data. The association rules extracted take the form of S→T, where S (the antecedent or left-hand-side, LHS) and T (the consequent or right-hand-side, RHS) are distinct item-sets. The Apriori algorithm uses predefined threshold values based on various criteria to identify these frequent item-sets. These rules are evaluated primarily based on three parameters: support, confidence, and lift. (Gu et al., 2022)

Support is an indication of the frequency of a combination of items in the dataset. If the antecedent S is a combination of items of variables (different risky driving behavior and drivers' demographic) and the consequent T is the targeted item (different distracted type), in the association rule S → T – the support of S i.e. Supp(S) is the proportion of observations of the complete dataset that contains the item S. Similarly, the support of the rule S → T, i.e. Support (S → T) is the proportion of observations of the complete dataset that contains the combination of items S and T in the antecedent and consequent, respectively. Therefore, an equation for the Supp(S) will be

$$\text{Support } (S \rightarrow T) = P(S \cap T) = \frac{\#(S \cup T)}{N} \qquad (1)$$

Where N is the total number of distracted drivers and #(S ∪ T) is the co-occurrence of item sets S and T.

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

$$\text{Confidence } (S \rightarrow T) = \frac{Supp(S \cup T)}{Supp\ (S)} \qquad (2)$$

The lift of a rule S → T is the confidence of the rule divided by the expected confidence, assuming S and T are independent of each other (Kong et al., 2021).

$$\text{Lift } (S \rightarrow T) = \frac{Supp(S \cup T)}{Lift\ (S)} \qquad (3)$$

Prior research has set benchmarks for both Support and Confidence, with ranges of 1% to 20% and 30% to 80%, respectively. Furthermore, the lift's lower limit was utilized from 1 to 1.5. Based on this knowledge, certain thresholds were used in this study: a 1% support, a 60% confidence level, and a 1.4 lift lower limit. As demonstrated by earlier research, these thresholds for Confidence and lift were purposefully set marginally higher to reduce the likelihood of an excessive production of redundant association sets (Kong et al., 2021; Gu et al., 2022; Weng and Li, 2019).

The study employed an association rule approach in two distinct stages. Initially, we focused on identifying associations between various distractions and risky or faulty driving behaviors. Subsequently, the research delved deeper, scrutinizing the most correlated rules pinpointed between distractions and risky driving behaviors concerning drivers' demographics and their daily distracted behaviors. By employing these thresholds and staged analyses, the study aimed to strike a balance between capturing significant associations and minimizing redundant or inconsequential patterns. This approach allowed for a comprehensive exploration of relationships between distractions, driving errors, and the broader context of drivers' characteristics and behaviors.

4.4 Results and Discussion

The Most Common Driving Errors

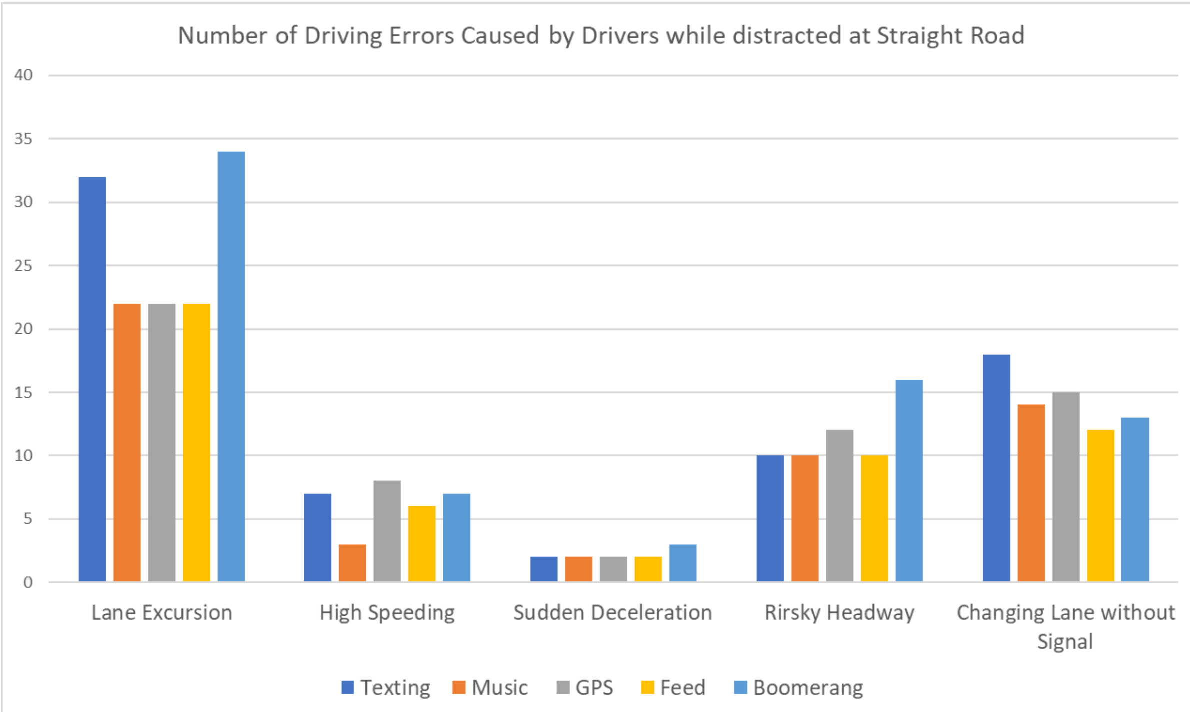


Figure 4.3: Number of driving errors while distracted on straight road section

The study carefully examined drivers' unsafe driving practices while engaging in various distracted activities. First, the analysis concentrated on errors on straight roads with no merging traffic in Figure 4.3. The most common error was lane excursion, which was especially common when drivers were preoccupied with capturing boomerang footage. During this distracted activity, 34 participants (68 %) had lane excursion errors. The second common error was to switch lanes without using the signal indicator. Notably, the distraction linked to the most significant number of participants making this mistake turned out to be texting, followed by using GPS (15 participants, 30 %) Interestingly, most participants (N=16) distracted by boomerang drove extremely closely behind the leading vehicle. As lane excursion and risky headway are the most prominent errors for distracted drivers by Boomerang, it has the possibility of high lateral and longitudinal crash probability. Surprisingly, participants distracted by GPS showed the greatest number of high-speed driving errors (8 participants), more than those distracted by boomerangs (7 participants) or texting. The analysis revealed fewer sudden deceleration errors, primarily linked to boomerang distractions, observed in 2 participants. In summary, more engaging social media distractions, boomerang and texting, are associated with the highest participant errors. Surprisingly, using a GPS caused more errors than just listening to music or scrolling through feeds, which may be due to the high demand for cognitive attention as it involves the processing, understanding, and making a directional decision.

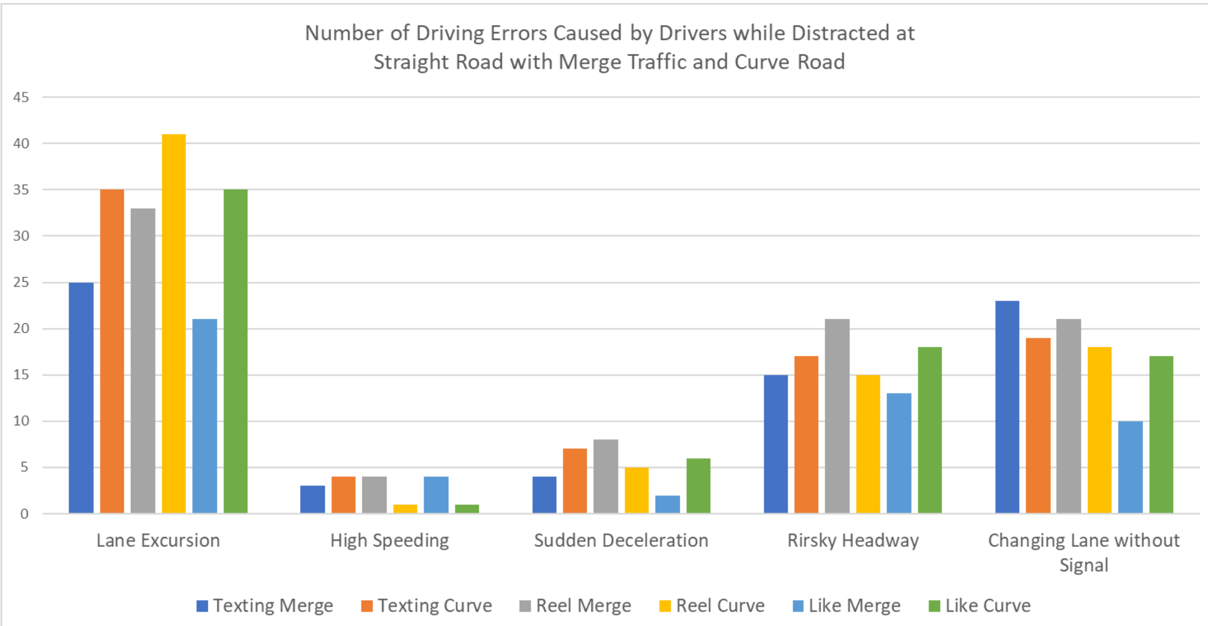


Figure 4.4: Number of driving errors while distracted at straight road with merge traffic and curved road sections.

Figure 4.4 provides a side-by-side comparison of participant errors associated with the same distractions observed on straight roads with merging sections and curved roads. Remarkably, lane excursion remains the predominant error participants commit on both types of roads. However, an interesting finding appears; drivers exhibited a higher tendency to experience lane excursion while navigating curved roads compared to straight roads with merge sections across various distraction types. Particularly, watching a reel or short video while driving on curved roads led to the highest number of lane excursion errors. Over 80% of participants struggled to maintain the center of their lane in this scenario. Similarly, while texting or liking a post on curved roads, more than 70% of participants encountered lane excursion errors, whereas these proportions dropped to 50% for texting and 66% for watching a reel on straight roads with merge sections.

Changing lanes without a signal emerged as the second most prevalent error, similar to straight roads. Observation of distinct patterns appears when comparing distractions like texting and watching a reel on merge roads, causing participants to change lanes without signaling. Interestingly, liking a post on a curved road resulted in more errors than liking one on a straight road with merge traffic. Moreover, when watching a reel on merge roads, drivers exhibited a higher frequency of errors related to risky headway, sudden deceleration, and high speeding compared to the same distraction on curved roads. This suggests that watching a reel while navigating straight roads with possible merge sections exerts a more pronounced negative influence on drivers' errors and risky driving behaviors.

An intriguing observation is that liking a post, considered a less engaging task than texting or watching a reel, surprisingly led to significantly more errors regarding risky headway (18 instances) and sudden deceleration (6 instances) on curved roads. This can be because curved roads require more complex driving maneuvers than straight roads. On curved roads, the cognitive and physical demands of steering and adjusting speed are greater. Drivers can be less cautious with a simple distraction like liking a post, potentially leading to more errors. Overall, distractions on curved roads tend to induce more errors. Figure 4.5 demonstrates that while watching a reel triggered the highest number of mistakes, texting (82 errors), watching a reel (80 errors), and liking a post (77 errors) on curved roads follow as the subsequent sources of errors. Engaging distractions demand simultaneous use of hands, eyes, and cognitive functions, which causes a big challenge for drivers on curved roads, which already require heightened hand-eye coordination. It leads to the highest error rates and elevated crash probabilities at the curve while distracted compared to various roadway environments.

Consequently, distractions like Boomerangs, which demand substantial hand-eye-cognitive coordination, rank fifth in total errors. Conversely, less engaging tasks such as liking a post (except on curved roads), playing music, browsing feeds, and using GPS resulted in fewer errors. Nevertheless, the notable finding emerges that using GPS led to more errors (59 errors) than expected, demonstrating that this widely used yet non-social media distraction can elevate errors and crash probabilities beyond typical assumptions.

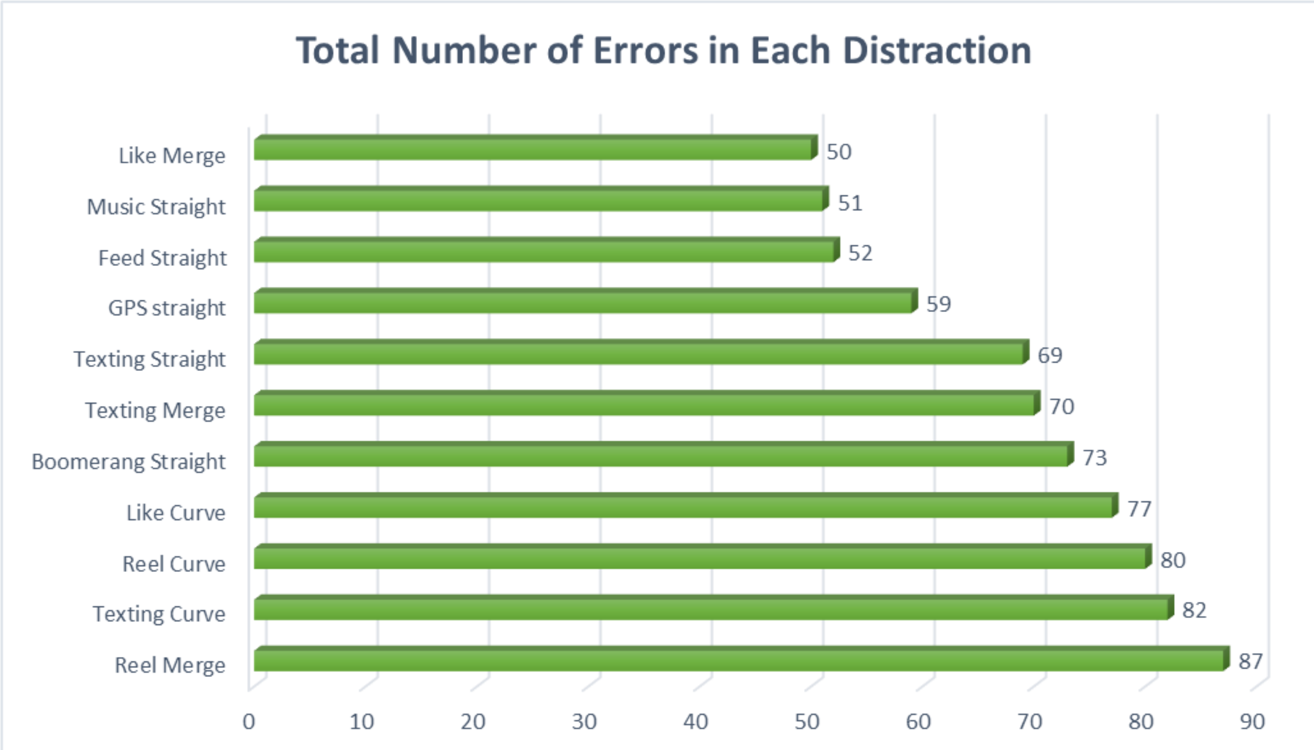


Figure 4.5: Total number of errors at each distraction at different roadway type.

Rules for Distracted Driving

This study focused on understanding the associations between different forms of distraction, risky behavioral errors, and demographics. It conducted an association rule method at three groups of distracted driving behaviors:

1. Social Media and Nonsocial media distractions
2. Engaging Social Media, Less Engaging Social media and Nonsocial media distractions
3. Individual Seconary Tasks

1. Social media and nonsocial media distraction

Figure 6 illustrates the primary two association rules correlating social media distraction and nonsocial media distraction with distinct driving errors. This graphic shows the stronger association between social media distraction and several driving behavioral errors, such as unsignalized lane change (less than five times and five or more times), risky headway, sudden deceleration, and lane excursion. Compared to nonsocial media distractions, social media related tasks can lead to a wider range of driving errors although lane deviation is the most frequent driving errors from both distractions. Additionally, drivers distracted by any secondary tasks other than social media tend to speed more often than those distracted by social media.

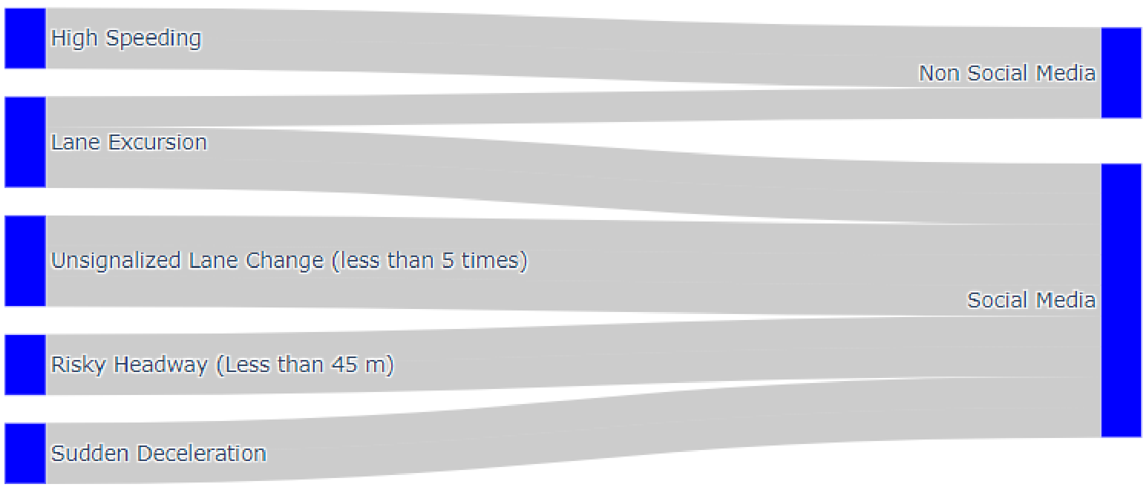


Figure 4.6: Association rule among distracted drivers by social media and non-social media with their common driving errors

The study conducted a secondary set of association rule analyses, delving into drivers' distracted behaviors and their corresponding risky actions alongside diverse demographic information. These analyses unveiled intriguing associations among various demographic attributes and daily activities, shedding light on distinctive patterns. One of the primary rules stated that when male drivers who are 25 years or younger and generally engage non-social media activities while driving, they are more likely to make high-speeding or lane excursion errors. On the other hand, the second association rule revealed that low-income female drivers between the ages of 26 and 30 are more likely to engage in dangerous headway behaviors when distracted by social media. Moreover, regardless of age or gender, individuals who undertake frequent weekly travels are to associated with causing lane excursion errors or unsignalized lane change errors. This signifies that social media distractions tend to induce more lateral errors than distractions unrelated to social media.

The association rules also highlighted typical distracted driving behaviors and their impacts on one's driving ability while distracted. Making phone calls while driving a vehicle is more strongly linked to several errors, especially when the distraction comes from nonsocial media activities. Drivers who use social media while driving, talk to passengers, or eating and drink are more likely to make mistakes when distracted by social media-related activities.

Table 4.2: Top three association rules among distracted driving by social media and nonsocial media distractions

Association Rule for Nonsocial media distraction as consequent			
	Rule Association 1	***Rule Association 2***	***Rule Association 3***
Driving behavior	*High Speeding*	*High Speeding*	*Lane Excursion, High Speeding*
Sociodemographic features	*Male and 25 or less years old*	*Male and 25 or less years old*	*Male*
Driving experiences			
Weekly distracted behavior while driving	*Making a phone call, Cognitive Distraction*	*Making a phone call, GPS*	*Cognitive Distraction*
Association Rule for social media distraction as consequent			
Driving behavior	*Risky Headway (less than 45m)*	*Risky Headway (less than 45m)*	*Lane Excursion and Unsignalized Lane Change (less than five times)*
Sociodemographic features		*Female, 26-30 years old and low income*	
Driving experiences	*High Frequency trip*		*High Frequency trip*

Weekly distracted behavior while driving	*By passenger, social media and cognitive distraction*	*Eating/drinking, making a phone call, texting and GPS*	*Social media and Passengers*

2. Engaging social media, Less Engaging Social media and Nonsocial media distraction

The subsequent study aimed to examine the association rules between various levels of engagement in social media activities, encompassing active interactions like texting, creating boomerangs, and watching reels, as well as less engaging activities such as scrolling through feeds or liking posts. These were compared with nonsocial media distractions concerning their correlation with specific driving errors, as depicted in Figure 4.7. This investigation highlighted distinct driving errors associated with different types of distractions. Similar to previous findings, nonsocial media distraction strongly correlated with high-speeding errors. Conversely, actively engaging with social media activities showed a higher association with sudden deceleration or abrupt braking, lane excursion, and unsignalized lane changes (occurring less than five times). Interestingly, unsignalized lane changes (less than 5 times) emerged as a common driving error across various social media distractions. However, risky headway behaviors were notably prevalent when individuals were distracted by less engaging social media activities. The data illustrates that highly engaging social media distractions tend to result in a higher frequency of lateral errors, while less engaging distractions have more car-following errors.

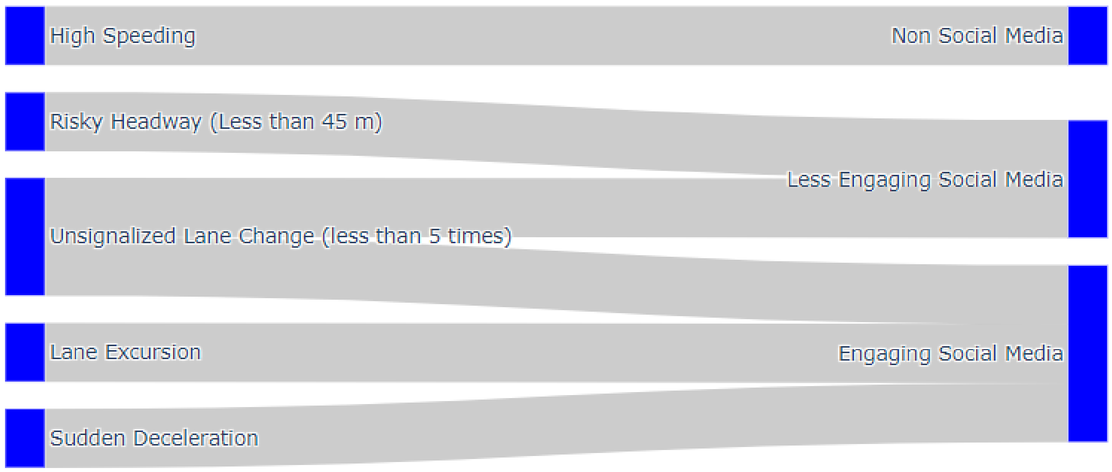


Figure 4.7: Association rule among distracted drivers by social media and non-social media with their common driving errors

In the second set of association rules analysis, the study highlighted the most prominent associations for each type of distraction (refer to Table 4.3). Nonsocial media distractions exhibited similar characteristics to those identified in the previous section. Increased occurrence of errors stemming from highly engaging social media distractions is independent of any specific sociodemographic feature or driving experience. This finding suggests secondary tasks are more important contributors of risky actions such as sudden deceleration, lane excursion, or unsignalized lane change while distracted, rather than drivers' demographics or driving experiences. .

On the other hand, individuals who do not own a car, indicating those who drive fewer times than their counterparts, show a stronger correlation with unsafe headway behaviors (less than 45 meters). Regarding self-reported distracted behaviors they engaged in the prior week, social media distractions appear to occure more with people who tend to eat/drink while driving as well as those who used GPS while driving. This highlights the similarities in the relationships between specific behaviors and incidents involving distracted driving, highlighting the overlap in some distracted behaviors across various types of distractions.

Table 4.3: Top association rule among distracted driving by social media and nonsocial media distraction

	Rule Association for Non-Social Media Distraction	***Rule Association for More Engaging Social Media Distraction***	***Rule Association for less Engaging Social Media Distraction***
Driving behavior	High Speeding	Sudden Deceleration, Lane Excursion, Unsignalized Lane Change	Risky Headway (Less than 45 m)
Sociodemographic features	Male and 25 or less		Does not own a vVehicle
Driving experiences			
Weekly distracted behavior while driving	Making a phone call, Cognitive Distraction	Eating/drinking or GPS	Eating/drinking and GPS

3. Association rule among each distraction type

Similar to the previous two sections, this study analyzed two sets of association rules, initially focusing on driving errors and subsequently examining the most correlated rules with other behavioral factors. As shown Table 4.4, females aged between 26 and 30 exhibit a stronger association with lane excursion errors related to Boomerang-related activities. In contrast, males show a higher association with lane excursion errors when watching reels regardless of age. Moreover, individuals between the ages of 26 and 30 who take longer daily drives exhibit a stronger correlation with text-related lane excursion errors.

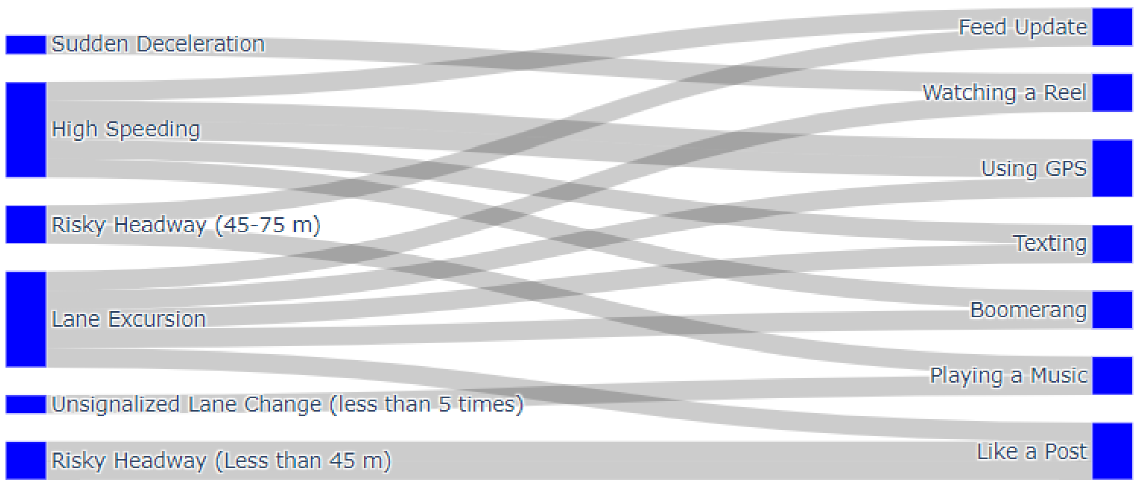


Figure 4.8: Association rule among distracted drivers by different distraction media with their common driving error

As shown in Table 4.5, male participants exhibit a stronger association with risky headway errors when distracted by scrolling through feeds. Conversely, individuals without a vehicle display a higher association with risky headway errors while liking posts. Participants who use phone calls are more associated with risky headway errors for feed updates, and participants who are distracted by eating/drinking while driving are more associated with risky head errors for a post. Interestingly, nonsocial media distractions such as playing music show no discernible associations with specific sociodemographic features or driving experiences concerning their most associated driving error, unsignalized lane change. However, males demonstrate a higher likelihood of high-speeding errors when using GPS, as highlighted in Table 4.6.

Table 4.4: More engaging social media distraction- Boomerang, Texting and Watching a Reel

The most association rule by each type of distraction			
	Boomerang	***Texting***	***Watching a Reel***
Driving behavior	Lane Excursion Error	Lane Excursion Error	Lane Excursion Error
Sociodemographic features	Female, 26-30 years old	26-30 years old	Male
Driving experiences		Daily Longer trips (More than 45 mins)	
Weekly distracted behavior while driving	Making a Phone Call	Distracted by Passenger	Distracted by GPS

Table 4.5: Less Engaging Social Media Distraction: Feed Update and Liking a post

The most association rule by each type of distraction		
	Feed Update	***Like a Post***
Driving behavior	Risky Headway Error (45-75 m)	Risky Headway Error (45-75 m)
Sociodemographic features	Male	No Vehicle
Driving experiences		
Weekly distracted behavior while driving	Distracted by Passenger, Phone call	Distracted by Eating/ Drinking

Table 4.6: Non social media distraction: playing a music and GPS

The most association rule by each type of distraction		
	Playing Music	***Using GPS***
Driving behavior	Unsignalized Lane Change (less than 5 times)	High Speeding Error
Sociodemographic features		Male
Driving experiences		
Additional distractive behaviors	Distracted by a Phone call , eating/drinking, grooming	Distracted by Eating/Drinking

4.5 Conclusion

The study thoroughly analyzed different distracted driving behaviors and the corresponding causes of related errors. The study found distractions have a significant impact on a variety of driving errors, highlighting their significance in compromising road safety. This study's findings can help create practical solutions and beneficial guidance to lower the risk of driving errors. One of the study's essential findings is to address drivers' more likely lane excursion errors while engaging in social media activities such as watching reels or boomerangs. The study also highlighted some important associations of demographic variables with distracted driving risky behavior. For example, specific age groups, such as young male drivers prone to high-speeding errors when distracted by nonsocial media or low-income females engaged in risky headway behaviors due to social media distractions. These results could be crucial in developing effective interventions like education or public campaigns. Moreover, recognizing the significant impact of GPS-related distraction, a widely used yet often underestimated nonsocial media distraction, calls for revisiting assumptions about its influence on driving errors. It emphasizes the need to educate drivers about the risks associated with GPS use while driving and implement strategies to minimize distractions without compromising navigational assistance.

The findings from this study can build a pathway to a safer driving environment by implementing focused driver education programs, using technological tools to minimize distractions, and creating rules or regulations to address particular distraction-related mistakes. Ultimately, the study's findings provide legislators, driving instructors, and tech developers with practical advice on developing interventions that reduce the risks associated with distracted driving. By customizing interventions to target particular distractions and related mistakes, it is possible to work together to promote safer driving practices and lower the startlingly high numbers of incidents involving distracted drivers.

Chapter 5: Conclusion

Transportation engineers have been paying close attention to the issue of distracted driving over the past few decades because of how important it is to safety and mobility. While many studies have examined the relationship between smartphone use and driving, few have investigated the more recent secondary tasks involving social media. These distractions substantially impact traffic flow, road safety, and driver performance. They encompass many activities beyond making phone calls or sending texts while driving. Because of this, planners and traffic engineers find it extremely difficult to understand how these social media activities affect drivers and the surrounding environment of the roadway. The three independent studies in this dissertation not only significantly address these issues but also provide important new information that could significantly improve road performance and safety.

The first paper of this dissertation used a field observation study conducted at three Texas intersections to analyze how different distractions affected intersection capacities. The result shows that distractions caused longer startup lost times than typically assumed. Distractions delayed startup times by an average of 3 seconds, up to 6.6 seconds. About 15% of drivers are distracted at red lights, primarily by phone use. This diversion significantly increased delays, affecting intersection capacity and traffic flow rates. Following large vehicles resulted in higher distraction rates, especially related to cell phone use (85%) as opposed to regular cars (50-65%). Moreover, technology-induced delays exhibited more variability compared to non-technology-induced distractions. Even though this offers a valid framework to estimate intersection capacities and startup lost time at a single intersection, a future study could broaden the scope and expand to assess the impacts of distracted behaviors along a corridor.

The second chapter examines social media distractions like liking posts or recording/posting Boomerangs. Unlike earlier studies focusing on conventional secondary tasks with more straightforward statistical modeling, this research thoroughly explores how distractions affect driver behaviors. This study shows that distracted driving substantially impacts angular and lateral driving properties rather than longitudinal behaviors. The result also shows how various road conditions influence drivers' levels of distraction. The study's machine learning models can identify distractions with over 75% accuracy and 85% precision. These models can undoubtedly aid in crash investigations and reduce crashes by warning or restricting users by alerting them about using their phones. Traffic engineers, transportation planners, and auto manufacturers can use the information and analysis in this chapter to create strategies for safety enforcement and intervention.

The third chapter looked at distracted driving behaviors and emphasized the serious driving mistakes brought on by distracted activities. The results of this study provide groundwork for focused interventions aimed at lowering driving errors. The analysis results indicated high lane deviation errors caused by boomerangs or social media reel-watching, highlighting the need for drivers to be more aware of the risks associated with these activities. This study also found correlations between demographic characteristics and distracted driving. For example, young men are prone to speeding while distracted. The study also highlights the overlooked effects of GPS-related distractions. Policymakers, educators, and tech developers can use the study's practical findings to develop interventions to reduce distracted driving risks. Interventions designed to target specific distractions and associated errors collaborate to promote safer driving practices and reduce the alarmingly high number of crashes caused by distracted driving.

References

- Abou Elassad, Z. E., Mousannif, H., Al Moatassime, H., & Karkouch, A. (2020). The application of machine learning techniques for driving behavior analysis: A conceptual framework and a systematic literature review. Engineering Applications of Artificial Intelligence, 87, 103312.
- Alembo, Z. A. (2014). *User delay costs and uncertainty in the traffic forecast for road projects* [KTH Royal Institute of Technology]. https://www.diva-portal.org/smash/get/diva2:739736/FULLTEXT01.pdf
- Al-Naymat, G., Al-Kasassbeh, M., Abu-Samhadanh, N., & Sakr, S. (2016). CLASSIFICATION OF VOIP AND NON-VOIP TRAFFIC USING MACHINE LEARNING APPROACHES. *Journal of Theoretical & Applied Information Technology*.
- Alshabibi, N. (2021). Impact Assessment of Driver Distraction by Cellphone on Start-up Lost-time and Average Saturation Headway at Signalized Intersections Based on Vehicle Position in the Queue. *Periodica Polytechnica Transportation Engineering*, *49*(4), 359–368. https://doi.org/10.3311/PPTR.16785
- Arnold, M. L., & Van Houten, R. (2011). Increasing following headway with prompts, goal setting, and feedback in a driving simulator. Journal of applied behavior analysis, 44(2), 245-254.
- At&T(2015). Smartphone Use While Driving Grows Beyond Texting to Social Media, Web Surfing, Selfies, Video Chatting. Retrieved from https://about.att.com/story/smartphone_use_while_driving_grows_beyond_texting.html
- 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.
- Baldwin, K. C., Duncan, D. D., & West, S. K. (2004). The driver monitor system: A means of assessing driver performance. *Johns Hopkins APL technical digest*, *25*(3), 269-277.
- Bao, Q., Tang, H., & Shen, Y. (2021). Driving behavior based relative risk evaluation using a nonparametric optimization method. *International journal of environmental research and public health*, *18*(23), 12452.
- Basacik, D., Reed, N., & Robbins, R. (2012). *Smartphone use while driving: a simulator study* (No. PPR592).

- Beede, K. E., & Kass, S. J. (2006). Engrossed in conversation: The impact of cell phones on simulated driving performance. *Accident Analysis and Prevention*, *38*(2), 415–421. https://doi.org/10.1016/j.aap.2005.10.015
- Bergen, B., Medeiros-Ward, N., Wheeler, K., Drews, F., & Strayer, D. (2013). The crosstalk hypothesis: why language interferes with driving. *Journal of experimental psychology: general*, *142*(1), 119.
- Biau, G., Cadre, B., & Rouvìère, L. (2019). Accelerated gradient boosting. *Machine learning*, *108*(6), 971-992.
- Bishop, C. M. (1995). Neural networks for pattern recognition. Oxford university press.
- Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4, No. 4, p. 738). New York: springer.
- Bivina, G. R., Landge, V., & Kumar, V. S. S. (2016). Socio Economic Valuation of Traffic Delays. *Transportation Research Procedia*, *17*(December 2014), 513–520. https://doi.org/10.1016/j.trpro.2016.11.104
- Blincoe, L., Miller, T. R., Zaloshnja, E., & Lawrence, B. A. (2014). The economic and societal impact of motor vehicle crashes, 2010. *Motor Vehicle Crashes: Economic and Societal Impact*, *In Depth*, *2015*(May), 1–285.
- Brown, I. D. (1965). Effect of a car radio on driving in traffic. *Ergonomics*, *8*(4), 475–479. https://doi.org/10.1080/00140136508930828
- Brumfield, R. M., & Pulugurtha, S. S. (2011). Effect of driver cell phone use on queue discharge patterns at signalized intersections. *Transportation Research Record*, *2257*, 71–79. https://doi.org/10.3141/2257-08
- Çalışkanelli, S. P., Coşkun Atasever, F., & Tanyel, S. (2017). Start-up Lost Time and its Effect on Signalized Intersections in Turkey. *PROMET - Traffic&Transportation*, *29*(3), 321–329. https://doi.org/10.7307/ptt.v29i3.2214
- Carney, C., Harland, K. K., & McGehee, D. V. (2016). Using event-triggered naturalistic data to examine the prevalence of teen driver distractions in rear-end crashes. *Journal of safety research*, *57*, 47-52.
- Castignani, G., Derrmann, T., Frank, R., & Engel, T. (2015). Driver behavior profiling using smartphones: A low-cost platform for driver monitoring. *IEEE Intelligent transportation systems magazine*, *7*(1), 91-102.

- CDC. (2013). *Leading causes of death reports, national and regional, 1999–2013*. National Center for Injury Prevention and Control, Atlanta, GA.
- CDC (2011). *Distracted Driving*. National Center for Injury Prevention and Control, Atlanta, GA. Retrived from https://www.cdc.gov/transportationsafety/distracted_driving/index.html
- Chen, S., Xue, Q., Zhao, X., Xing, Y., & Lu, J. J. (2021). Risky driving behavior recognition based on vehicle trajectory. *International journal of environmental research and public health*, *18*(23), 12373.
- Chen, C., Zhao, X., Zhang, Y., Rong, J., & Liu, X. (2019). A graphical modeling method for individual driving behavior and its application in driving safety analysis using GPS data. *Transportation research part F: traffic psychology and behaviour*, *63*, 118-134.
- Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. Journal of Applied Science and Technology Trends, 2(01), 20-28.
- Choudhary, P., & Velaga, N. R. (2019). A comparative analysis of risk associated with eating, drinking and texting during driving at unsignalised intersections. Transportation research part F: traffic psychology and behaviour, 63, 295-308.
- Choudhary, P., & Velaga, N. R. (2020). Impact of distraction on decision making at the onset of yellow signal. *Transportation Research Part C: Emerging Technologies*, *118*(August), 102741. https://doi.org/10.1016/j.trc.2020.102741
- Christoph, M., van Nes, N., & Knapper, A. (2013). Naturalistic driving observations of manual and visual–manual interactions with navigation systems and mobile phones while driving. *Transportation research record*, *2365*(1), 31-38.
- Chuan, S. U. N., Chao-zhong, W. U., Duan-feng, C. H. U., Zhi-gang, D. U., & Fei, T. I. A. N. (2015). Driving speed behavior clustering for commercial vehicle based on connected vehicle data mining. *Journal of Transportation Systems Engineering and Information Technology*, *15*(6), 82.
- Consiglio, W., Driscoll, P., Witte, M., & Berg, W. P. (2003). Effect of cellular telephone conversations and other potential interference on reaction time in a braking response. *Accident Analysis and Prevention*, *35*(4), 495–500. https://doi.org/10.1016/S0001-4575(02)00027-1

- Cristianini, N., & Shawe-Taylor, J. (2000). An introduction to support vector machines and other kernel-based learning methods. Cambridge university press.
- DeSilva, H.R. (1942). Why We Have Automobile Accidents. New York, NY: John Wiley and Sons.
- Darzi, A., Gaweesh, S. M., Ahmed, M. M., & Novak, D. (2018). Identifying the causes of drivers' hazardous states using driver characteristics, vehicle kinematics, and physiological measurements. Frontiers in neuroscience, 12, 568.
- Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., ... & Knipling, R. R. (2006). *The 100-car naturalistic driving study, Phase II-results of the 100-car field experiment* (No. DOT-HS-810-593). United States. Department of Transportation. National Highway Traffic Safety Administration.
- Dingus, T. A., Hanowski, R. J., & Klauer, S. G. (2011). Estimating crash risk. *Ergonomics in Design*, *19*(4), 8-12.
- Drews, F. A., Yazdani, H., Godfrey, C. N., Cooper, J. M., & Strayer, D. L. (2009). Text messaging during simulated driving. *Human Factors*, *51*(5), 762–770. https://doi.org/10.1177/0018720809353319
- Dula, C. S., Martin, B. A., Fox, R. T., & Leonard, R. L. (2011). Differing types of cellular phone conversations and dangerous driving. Accident Analysis & Prevention, 43(1), 187-193.
- Ellison, A. B., Greaves, S., & Daniels, R. (2013). Capturing speeding behaviour in school zones using GPS technology. *Road & Transport Research: A Journal of Australian and New Zealand Research and Practice*, *22*(4), 30-42.
- Elvik, R. (2005). Speed and road safety: synthesis of evidence from evaluation studies. *Transportation Research Record*, *1908*(1), 59-69.
- Engelberg, Jessa K., Linda L. Hill, Jill Rybar, and Tara Styer. 2015. "Distracted Driving Behaviors Related to Cell Phone Use among Middle-Aged Adults." Journal of Transport & Health 2 (3): 434–40. https://doi.org/10.1016/j.jth.2015.05.002.
- Faccio, M., & McConnell, J. J. (2020). Death by Pokémon GO: The economic and human cost of using apps while driving. *Journal of Risk and Insurance*, *87*(3), 815-849.

- Fitch, G. M., Soccolich, S. A., Guo, F., McClafferty, J., Fang, Y., Olson, R. L., ... & Dingus, T. A. (2013). *The impact of hand-held and hands-free cell phone use on driving performance and safety-critical event risk* (No. DOT HS 811 757).
- FHWA. (2013). *Signalized Intersections: Informational Guide, Publication No. FHWA-SA-13-027*. Federal Highway Administration (FHWA). http://safety.fhwa.dot.gov/intersection/conventional/signalized/fhwasa13027/fhwasa13027.pdf
- Garrison, T. M., & Williams, C. C. (2013). Impact of relevance and distraction on driving performance and visual attention in a simulated driving environment. *Applied cognitive psychology*, *27*(3), 396-405.
- Ghandour, R., Potams, A. J., Boulkaibet, I., Neji, B., & Al Barakeh, Z. (2021). Driver Behavior Classification System Analysis Using Machine Learning Methods. *Applied Sciences*, *11*(22), 10562.
- Gharehbaghi, K. (2016). Artificial neural network for transportation infrastructure systems. In *MATEC web of conferences* (Vol. 81, p. 05001). EDP Sciences.
- GHSA. (2020). *Distracted Driving Laws by State- Governors Highway Safety Association (GHSA)*. https://www.ghsa.org/state-laws/issues/distracted driving
- Gliklich, Emily, Rong Guo, and Regan W. Bergmark. 2016. "Texting While Driving: A Study of 1211 U.S. Adults with the Distracted Driving Survey." Preventive Medicine Reports 4: 486–89. https://doi.org/10.1016/j.pmedr.2016.09.003.
- Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.
- Goodman, M. J., Tijerina, L., Bents, F. D., & Wierwille, W. W. (1999). Using Cellular Telephones in Vehicles: Safe or Unsafe? *Transportation Human Factors*, *1*(1), 3–42. https://doi.org/10.1207/sthf0101_2
- Ghasemzadeh, A., Hammit, B. E., Ahmed, M. M., & Eldeeb, H. (2019). Complementary methodologies to identify weather conditions in naturalistic driving study trips: Lessons learned from the SHRP2 naturalistic driving study & roadway information database. *Safety Science*, *119*, 21-28.

- Gliklich, Emily, Rong Guo, and Regan W. Bergmark. 2016. "Texting While Driving: A Study of 1211 U.S. Adults with the Distracted Driving Survey." Preventive Medicine Reports 4: 486–89. https://doi.org/10.1016/j.pmedr.2016.09.003.

- Gu, C., Xu, J., Gao, C., Mu, M., E, G., & Ma, Y. (2022). Multivariate analysis of roadway multi-fatality crashes using association rules mining and rules graph structures: A case study in China. *Plos one, 17*(10), e0276817.
- Hadiuzzaman, M., Rahman, M., & Karim, M. A. (2008). Saturation Flow Model at Signalized Intersection for Non-lane Based Traffic. *Canadian Journal of Transportation*, *2*(1), 77–90.
- Hancock, P. A., Lesch, M., & Simmons, L. (2003). The distraction effects of phone use during a crucial driving maneuver. *Accident Analysis & Prevention*, *35*(4), 501-514.
- Hashash, M., Abou Zeid, M., & Moacdieh, N. M. (2019). Social media browsing while driving: Effects on driver performance and attention allocation. *Transportation Research Part F: Traffic Psychology and Behaviour*, *63*, 67–82. https://doi.org/10.1016/j.trf.2019.03.021
- Haque, M. M., & Washington, S. (2014). A parametric duration model of the reaction times of drivers distracted by mobile phone conversations. *Accident Analysis and Prevention*, *62*, 42–53. https://doi.org/10.1016/j.aap.2013.09.010
- HCM. (2010). *Highway Capacity Manual*. Transportation Research Board, National Research Council, Washington, DC. https://www.trb.org/Main/Blurbs/164718.aspx
- HCM. (2016). *Highway Capacity Manual, Sixth Edition: A Guide for Multimodal Mobility Analysis*. Transportation Research Board,Washington, DC. https://www.trb.org/Main/Blurbs/175169.aspx
- Horrey, W. J., Lesch, M. F., & Garabet, A. (2008). Assessing the awareness of performance decrements in distracted drivers. Accident Analysis & Preven - tion, 40, 675-682.
- Hossain, M. M., Zhou, H., Rahman, M. A., Das, S., & Sun, X. (2022). Cellphone-distracted crashes of novice teen drivers: Understanding associations of contributing factors for crash severity levels and cellphone usage types. *Traffic injury prevention*, *23*(7), 390-397.
- Hosking, S. G., Young, K. L., & Regan, M. A. (2009). The effects of text messaging on young drivers. *Human factors*, *51*(4), 582-592.
- Hou, Y., Xiong, D., Jiang, T., Song, L., & Wang, Q. (2019). Social media addiction: Its impact, mediation, and intervention. Cyberpsychology: Journal of psychosocial research on cyberspace, 13(1).

- Hurwitz, D. S., Heaslip, K. P., Schrock, S. D., Swake, J., Marnell, P., Tuss, H., & Fitzsimmons, E. (2013). Implications of distracted driving on start-up lost time for dual left-turn lanes. *Journal of transportation engineering, 139*(9), 923-930.
- Huth, V., Sanchez, Y., & Brusque, C. (2015). Drivers' phone use at red traffic lights: A roadside observation study comparing calls and visual-manual interactions. *Accident Analysis and Prevention, 74,* 42–48. https://doi.org/10.1016/j.aap.2014.10.008
- Irwin, C., Monement, S., & Desbrow, B. (2015). The Influence of Drinking, Texting, and Eating on Simulated Driving Performance. *Traffic Injury Prevention*, *16*(2), 116–123. https://doi.org/10.1080/15389588.2014.920953
- Insurance Institute for Highway Safety (2023). Vehicles with higher, more vertical front ends pose greater risk to pedestrians. Retrived from https://www.iihs.org/news/detail/vehicles-with-higher-more-vertical-front-ends-pose-greater-risk-to-pedestrians
- Jazayeri, A., Martinez, J. R. B., Loeb, H. S., & Yang, C. C. (2021). The Impact of driver distraction and secondary tasks with and without other co-occurring driving behaviors on the level of road traffic crashes. *Accident Analysis & Prevention*, *153*, 106010
- Klaus, C., Narvaez, J., Loza, J., & Thimmavajjhala, V. (2018). *2018 Transportation Conformity*. North Central Texas Council of Governments. https://www.nctcog.org/trans/quality/air/federal-air-quality/conformity
- Koonce, P., Rodegerdts, L., Lee, K., Quayle, S., Beaird, S., Braud, C., Bonneson, J., Tarnoff, P., & Urbanik, T. (2008). *Traffic Signal Timing Manual*. Federal Highway Administration (FHWA).
- Laapotti, S., Keskinen, E., & Rajalin, S. (2003). Comparison of young male and female drivers' attitude and self-reported traffic behaviour in Finland in 1978 and 2001. *Journal of safety research*, *34*(5), 579-587.
- Lattanzi, E., & Freschi, V. (2021). Machine learning techniques to identify unsafe driving behavior by means of in-vehicle sensor data. Expert Systems with Applications, 176, 114818.
- Li, Li, Ruth A. Shults, Rebecca R. Andridge, Merissa A. Yellman, Henry Xiang, and Motao Zhu. 2015. "Texting/Emailing While Driving among High School Students in 35 States,

United States, 2015.” Journal of Adolescent Health 63: 701–708. https://doi.org/10.1016/j.jadohealth.2018.06.010

- Liu, Y. C., & Ou, Y. K. (2011). Effects of age and the use of hands-free cellular phones on driving behavior and task performance. *Traffic injury prevention*, *12*(6), 550-558.
- Li, X., Yan, X., Wu, J., Radwan, E., & Zhang, Y. (2016). A rear-end collision risk assessment model based on drivers' collision avoidance process under influences of cell phone use and gender—A driving simulator based study. *Accident Analysis & Prevention*, *97*, 1-18.
- Li, Y., Lu, J., & Xu, K. (2017). Crash risk prediction model of lane-change behavior on approaching intersections. *Discrete Dynamics in Nature and Society*, 2017.
- Li, X., Oviedo-Trespalacios, O., Rakotonirainy, A., & Yan, X. (2019). Collision risk management of cognitively distracted drivers in a car-following situation. *Transportation research part F: traffic psychology and behaviour*, *60*, 288-298.
- Li, Z. N., Huang, X. H., Mu, T., & Wang, J. (2022). Attention-based lane change and crash risk prediction model in highways. *IEEE transactions on intelligent transportation systems*, *23*(12), 22909-22922.
- Malta, L., Ljungaust, M., Faber, F., Metz, B., SaintPierre, G., Benmimoun, M., & Schäfer, R. (2012). Eurofot SP6 D6. 4–Final results: Impacts on traffic safety.
- Martinussen, L. M., Møller, M., & Prato, C. G. (2014). Assessing the relationship between the Driver Behavior Questionnaire and the Driver Skill Inventory: Revealing sub-groups of drivers. *Transportation research part F: traffic psychology and behaviour*, *26*, 82-91.
- McCartt, A. T., Hellinga, L. A., & Bratiman, K. A. (2006). Cell phones and driving: review of research. *Traffic injury prevention*, *7*(2), 89-106.
- McLean, J. R., & Hoffmann, E. R. (1971). Analysis of drivers' control movements. *Human Factors*, *13*(5), 407-418.
- McNabb, J., & Gray, R. (2016). Staying connected on the road: a comparison of different types of smart phone use in a driving simulator. *PLoS one*, *11*(2), e0148555.
- Mobile fact sheet. (2021, April 7). Pew Research Center.
- NeuroGrow Brain Fitness Center. (2018) *What Social Media Does to Your Brain*. Available at: https://neurogrow.com/what-social-media-does-to-your-brain/ (Accessed: 4 December 2023).

- Neyens, D. M., & Boyle, L. N. (2007). The effect of distractions on the crash types of teenage drivers. *Accident Analysis & Prevention*, *39*(1), 206-212.
- NHTSA. (2012). Visual-Manual NHTSA Driver Distraction Guidelines Devices, for In-Vehicle Electronic. In *National Highway Traffic Safety Administration, Docket No. NHTSA-2010-0053*. file:///C:/Users/jxb5105/Downloads/Distraction_NPFG-02162012.pdf
- NHTSA. (2019). Driver electronic device use in 2017 (Research Note. Report No. DOT HS 812 665). In *National Highway Trafic Safety Administration*. https://doi.org/10.1016/j.annemergmed.2012.03.003
- NHTSA (2020). *Traffic Safety Facts Research Note: Distracted Driving 2018;* NHTSA: Washington,DC, USA, 2020.
- NHTSA. (2020). NHTSA 2020 report. U.S. Department of Transportation. Retrieved from https://one.nhtsa.gov/nhtsa/whatis/planning/2020Report/2020report.html
- NHTSA 2020. Facts + Statistics: Aggressive driving. *Driving Behaviors Reported For Drivers And Motorcycle Operators Involved In Fatal Crashes, 2020*. Retrieved from https://www.iii.org/fact-statistic/facts-statistics-aggressive-driving
- NHTSA. (2021). Distracted driving 2019 (Research Note. Report No. DOT HS 813 111). In *National Highway Trafic Safety Administration*. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812381
- Niu, Y., Li, Z., & Fan, Y. (2021). Analysis of truck drivers' unsafe driving behaviors using four machine learning methods. International Journal of Industrial Ergonomics, 86, 103192.
- O'Brien, S. R. (2011). *The psychosocial factors influencing aggressive driving behaviour* (Doctoral dissertation, Queensland University of Technology).
- Oh, C., Jung, E., Rim, H., Kang, K., & Kang, Y. (2012). Intervehicle safety warning information system for unsafe driving events: Methodology and prototypical implementation. *Transportation research record*, *2324*(1), 1-10.
- Osafune, T., Takahashi, T., Kiyama, N., Sobue, T., Yamaguchi, H., & Higashino, T. (2017). Analysis of accident risks from driving behaviors. *International journal of intelligent transportation systems research*, *15*, 192-202.

- Östlund, J., Peters, B., Thorslund, B., Engström, J., Markkula, G., Keinath, A., ... & Foehl, U. (2005). Driving performance assessment-methods and metrics. EU Deliverable, Adaptive Integrated Driver-Vehicle Interface Project (AIDE).
- Oviedo-Trespalacios, O., Haque, M. M., King, M., & Washington, S. (2016). Understanding the impacts of mobile phone distraction on driving performance: A systematic review. *Transportation research part C: emerging technologies*, *72*, 360-380.
- Paefgen, J., Kehr, F., Zhai, Y., & Michahelles, F. (2012, December). Driving behavior analysis with smartphones: insights from a controlled field study. In *Proceedings of the 11th International Conference on mobile and ubiquitous multimedia* (pp. 1-8).
- Papantoniou, P., Yannis, G., Antoniou, C., Papadimitriou, E., Pavlou, D., & Golias, J. (2016). Investigating the Effect of Area Type and Traffic Conditions on Distracted Driving Performance. *Transportation Research Procedia*, *14*, 3839–3848. https://doi.org/10.1016/j.trpro.2016.05.469
- Papantoniou, P., Yannis, G., & Christofa, E. (2019). Which factors lead to driving errors? A structural equation model analysis through a driving simulator experiment. *IATSS research*, *43*(1), 44-50.
- Pawar, D. S., & Patil, G. R. (2018). Response of major road drivers to aggressive maneuvering of the minor road drivers at unsignalized intersections: A driving simulator study. *Transportation research part F: traffic psychology and behaviour*, *52*, 164-175.
- Pew Research Center (2021). Mobile Fact Sheet. Available at https://www.pewresearch.org/internet/fact-sheet/mobile/
- Qian, C., Li, Y., Zuo, W., & Wang, Y. (2020). Analysis of Driving Safety and Cellphone Use Based on Social Media. In *Advances in Human Factors of Transportation: Proceedings of the AHFE 2019 International Conference on Human Factors in Transportation, July 24-28, 2019, Washington DC, USA 10* (pp. 521-530). Springer International Publishing.
- Rachel, G., Mitchell, C., & Anna, C. (2019). Driver Distraction: A Review of Scientific Literature (Issue February). National Transport Commission.
- Rahman, Z., Martinez, D., Martinez, N., Zhang, Z., Memarian, A., Pulipati, S., Mattingly, S. P., & Rosenberger, J. M. (2018). Evaluation of cell phone induced driver behavior at a

type II dilemma zone. *Cogent Engineering*, *5*(1). https://doi.org/10.1080/23311916.2018.1436927

- Rahman, M. A., Sun, X., Sun, M., & Shan, D. (2021). Investigating characteristics of cellphone distraction with significance tests and association rule mining. *IATSS research*, *45*(2), 198-209.
- Raschka, S., & Mirjalili, V. (2019). *Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2*. Packt Publishing Ltd.
- Rouphail, N., Tarko, A., & Jing, L. (2001). Traffic Flow at Signalized Intersections. *Traffic-Flow Theory*, 9-1-9–28. http://web.tongji.edu.cn/~yangdy/its/tft/chap9.pdf
- Roy, R., & Saha, P. (2018). Headway distribution models of two-lane roads under mixed traffic conditions: a case study from India. *European transport research review*, *10*(1), 1-12.
- Rudin-Brown, C. M., Young, K. L., Patten, C., Lenné, M. G., & Ceci, R. (2013). Driver distraction in an unusual environment: Effects of text-messaging in tunnels. *Accident Analysis & Prevention*, *50*, 122-129.
- Sen, B., Smith, J. D., & Najm, W. G. (2003). *Analysis of lane change crashes* (No. HS-809 571). United States. National Highway Traffic Safety Administration.
- Shawky, M., & Al-ghafli, A. (2016). Start-up delay Estimation at Signalized Intersections : Impact of Left-Turn Phasing Sequences. *International Journal of Engineering and Applied Sciences*, *3*(11), 7–12. https://doi.org/10.1007/978-3-319-41682-3
- Simons-Morton, B. G., Guo, F., Klauer, S. G., Ehsani, J. P., & Pradhan, A. K. (2014). Keep your eyes on the road: Young driver crash risk increases according to duration of distraction. *Journal of Adolescent Health*, *54*(5), S61-S67.
- Şimşek, B., Pakdil, F., Dengiz, B., & Testik, M. C. (2013). Driver performance appraisal using GPS terminal measurements: A conceptual framework. *Transportation Research Part C: Emerging Technologies*, *26*, 49-60.
- Starcevic, V. (2013). Is Internet addiction a useful concept?. Australian & New Zealand Journal of Psychiatry, 47(1), 16-19.
- Starcevic, V. (2013). Is Internet addiction a useful concept?. Australian & New Zealand Journal of Psychiatry, 47(1), 16-19.

- Stavrinos, D., Jones, J. L., Garner, A. A., Griffin, R., Franklin, C. A., Ball, D., ... & Fine, P. R. (2013). Impact of distracted driving on safety and traffic flow. *Accident Analysis & Prevention*, *61*, 63-70.
- Stewart, T. (2023). *Overview of Motor Vehicle Traffic Crashes in 2021* (No. DOT HS 813 435).
- Strayer, D. L., Drews, F. A., & Crouch, D. J. (2006). A comparison of the cell phone driver and the drunk driver. *Human Factors*, *48*(2), 381–391. https://doi.org/10.1518/001872006777724471
- Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-Task Studies of Simulated Driving and Conversing on a Cellular Telephone. *Psychological Science*, *12*(6), 462–466. https://doi.org/10.1111/1467-9280.00386
- Stutts, J. C., & Hunter, W. W. (2003). Driver inattention, driver distraction and traffic crashes. *ITE Journal (Institute of Transportation Engineers)*, *73*(7), 34–45.
- Stutts, J. C., Reinfurt, D. W., & Rodgman, E. A. (2001). The role of driver distraction in crashes: an analysis of 1995-1999 Crashworthiness Data System Data. *Annual Proceedings / Association for the Advancement of Automotive Medicine. Association for* ***the*** *Advancement of Automotive Medicine*, *45*(May), 287–301.
- Subasi, A. (2020). *Practical Machine Learning for Data Analysis Using Python*. Academic Press.
- Tang, J., Liang, J., Han, C., Li, Z., Huang, H., 2019. Crash injury severity analysis using a two-layer Stacking framework [Journal Article]. Accid. Anal. Prev. 122, 226–238. https://doi.org/10.1016/j.aap.2018.10.016.
- Tango, F., Botta, M., Minin, L., & Montanari, R. (2010). Non-intrusive detection of driver distraction using machine learning algorithms. In ECAI 2010 (pp. 157-162). IOS Press.
- Tefft, B. C. (2018). Reducing crash risk and improving traffic safety: research on driver behavior and performance. *Institute of Transportation Engineers. ITE Journal*, *88*(8), 30-34.
- Teply, S., & Jones, A. M. (1991). *SATURATION FLOW: DO WE SPEAK THE SAME LANGUAGE?* Transportation Research Record. https://trid.trb.org/view/365597
- Tian, E. (2016). Drivers Classification Based on the Driver Lane Change Behavior Parameters. *Int. J. Mater. Mech. Eng*, *5*, 53-57.

- Toledo, T., Musicant, O., & Lotan, T. (2008). In-vehicle data recorders for monitoring and feedback on drivers' behavior. *Transportation Research Part C: Emerging Technologies*, *16*(3), 320-331.
- Tractinsky, N., Ram, E. S., & Shinar, D. (2013). To call or not to call—That is the question (while driving). *Accident Analysis & Prevention*, *56*, 59-70.
- Triguero, I., García-Gil, D., Maillo, J., Luengo, J., García, S., & Herrera, F. (2019). Transforming big data into smart data: An insight on the use of the k-nearest neighbors algorithm to obtain quality data. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(2), e1289.
- Tucker S, Pek S, Morrish J, Ruf M. 2015. Prevalence of texting while driving and other risky driving behaviors among young people in Ontario, Canada: evidence from 2012 and 2014. Accid Anal Prev. 84:144–152. doi:10.1016/j.aap.2015.07.011
- Vanlaar, W., Simpson, H., Mayhew, D., & Robertson, R. (2008). Aggressive driving: A survey of attitudes, opinions and behaviors. *Journal of safety research*, *39*(4), 375-381.
- Vieira, F. S., & Larocca, A. P. C. (2017). Drivers' speed profile at curves under distraction task. *Transportation research part F: traffic psychology and behaviour*, *44*, 12-19.
- Weng, J., & Li, G. (2019). Exploring shipping accident contributory factors using association rules. *Journal of Transportation Safety & Security*, *11*(1), 36-57.
- Williams, A. F. (2003). Teenage drivers: patterns of risk. *Journal of safety research*, *34*(1), 5-15.
- World Health Organization. (2011). Mobile phone use: a growing problem of driver distraction.
- World Heath Organization. (2016). Road Safety. Retrieved December 5, 2023, from WHO
- Ye, M., Osman, O. A., Ishak, S., & Hashemi, B. (2017). Detection of driver engagement in secondary tasks from observed naturalistic driving behavior. *Accident Analysis & Prevention*, *106*, 385-391.
- Yannis, G., Laiou, A., Papantoniou, P., & Christoforou, C. (2014). Impact of texting on young drivers' behavior and safety on urban and rural roads through a simulation experiment. *Journal of safety research*, *49*, 25-e1.
- Zhang, L., Schonfeld, P., Kastrouni, E., Shayanfar, E., & He, X. (2016). *Estimating the Economic Impacts of Multimodal Transportation Improvements*. State Highway

Adminstration, Maryland Department of Transportation. https://www.roads.maryland.gov/OPR_Research/MD-16-SHA-UM-3-17_Economic-Impacts_Report.pdf

- Zhao, Y., Zhang, W., Lu, J., Zhang, W., & Ma, Y. (2015). Estimation of saturation flow rate and start-up lost time for signal timing based on headway distribution. *Discrete Dynamics in Nature and Society*, *2015*. https://doi.org/10.1155/2015/304823
- Zheng, J., Ma, X., Wu, Y. J., & Wang, Y. (2013). Measuring signalized intersection performance in real-time with traffic sensors. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, *17*(4), 304–316. https://doi.org/10.1080/15472450.2013.771105
- Zhou, T., Zhang, J., 2019. Analysis of commercial truck drivers' potentially dangerous driving behaviors based on 11-month digital tachograph data and multilevel modeling approach. Accid. Anal. Prev. 132, 105256 https://doi.org/10.1016/j.aap.2019.105256.