ADVANCES IN TRANSPORTATION STUDIES
An International Journal

Editor in Chief: Alessandro Calvi

Vol. LXI November 2023

Contents

aracne

Advances in Transportation Studies: an international Journal 61 (2023)

ISBN 979-12-218-0919-0 – ISSN 1824-5463 – DOI 10.53136/97912218091901 – pag. 3-18

Examining in-vehicle distraction sources in relation to crashes using a Bayesian Multinomial Logit model

B. Kutela[1] E. Kidando[2]
A.E. Kitali[3] S. Mwende[4] N. Novat[2]

[1]Roadway Safety Program, Texas A&M Transportation Institute,
1111 RELLIS Pkwy, Bryan, TX 77807
email: b-kutela@tti.tamu.edu
[2]Department of Civil and Environmental Engineering, Cleveland State University,
2121 Euclid Ave. FH 107, Cleveland, OH 44115
email: e.kidando@csuohio.edu
[3]School of Engineering and Technology, University of Washington Tacoma,
1900 Commerce Street Tacoma, WA 98402-3100
email: akitali@uw.edu
[4]Department of Civil and Environmental Engineering, Western Michigan University
4601 Campus Drive, Room G-249, Kalamazoo, MI 49008
email: siamwende95@gmail.com

subm. 24th November 2022 *approv. after rev. 15th March 2023*

Abstract

It is well understood that most crashes are the result of human errors. Among human-related errors, distracted driving, particularly related to cellphones, has received significant attention. Conversely, the underlying factors associated with in-vehicle distractions that are non-cellphone use have not been fully explored. Thus, this paper uses data from driver distraction-related crashes to examine various in-vehicle distraction sources. A Bayesian Multinomial Logit (BMNL) model was developed using 5,078 distracted-driving related crashes from Iowa. Four in-vehicle distraction sources - cellphone use, non-cellphone electronic devices, passengers, and reaching in-vehicle fallen objects - were investigated to determine factors that increase their odds of occurrence. The results suggest that drivers under the influence of alcohol are more likely to be involved in crashes associated with the distraction from cellphones. Furthermore, older drivers are less likely to be involved with distracted driving due to passengers. As expected, the more people in the vehicle, the higher the likelihood a driver can be distracted by passengers. Moreover, the association of driver distraction and speed limit, time of the day, vehicle's age, among others, were evaluated. This study provides useful information for developing and implementing strategies that minimize distractions from all in-vehicle sources.

Keywords - distracted driving, human-related errors, in-vehicle distractions, Bayesian Multinomial Logit (BMNL) model, distracted-driving related crashes

1. Background

Defined as the performance of secondary tasks when driving, distracted driving has been extensively studied. In a ten-year study (1999-2008), researchers [1] found a nation-wide decline in distracted driving-related fatalities from 2003 to 2005 then a rise thereafter. Another study [2] evaluated the changes in driver distraction in Northern Virginia within a span of five years (2014-2018). Their study reported no statistically significant difference in phone use but a significant increase in the proportion of drivers engaging in non-cellphone secondary behaviors.

Studies have shown that distraction sources can originate within or outside a vehicle. Distraction outside the vehicle can include looking at a roadside object, looking at a crash/incident scene, or scanning for emergency/police vehicles, to mention a few. On the other hand, in-vehicle distraction sources that have predominantly been considered are cellphone related (e.g., talking on a phone, dialing a phone, reaching for a phone, or sending text messages). Other in-vehicle distraction sources include adjusting the radio, eating, drinking, reaching for a fallen object, and distraction from passengers [1-9]. Researchers have associated driver distractions with crash frequency and severity [4, 8], demographic factors [5, 6, 10, 11], among other things. Recent studies have focused on the distractions originating from cellphones while driving [12], smartwatch usage [13], and even the impact of texting and web surfing on driving behavior [14]. However, there is limited knowledge of the other in-vehicle distractions and how they contribute to crashes.

Although the investigation of in-vehicle driver distraction sources has been a topic of interest in recent years, researchers have concentrated more on cell phone-related distractions. Therefore, little knowledge is known on the other in-vehicle distractions. The authors did not perform in-depth comparisons among distraction sources for the few studies involving in-vehicle. Besides, most studies have been using data collected through survey questionnaires, which greatly depends on the driver's honesty. Therefore, this study evaluates the factors affecting drivers' distractions for in-vehicle sources of distractions. The current study uses actual crash data, in which drivers were asked whether they were distracted and the distraction sources. Four in-vehicle distraction sources—cellphone use, use of non-cellphone electronic devices (navigation device, DVD player, etc.), distraction from passengers, and distraction from in-vehicle fallen objects—are compared. The study findings will show the similarities and differences of the factors affecting the in-vehicle driver distraction sources. The similarities and variations of these factors towards driver distraction sources will pave the way for a better approach to combating each of the driver distraction sources studied. The paper's remainder is organized as follows: the next part presents the literature review, followed by the methodology, whereby data description and modeling methodology are discussed; the results and discussion follow; and lastly, the conclusion and recommendation are presented.

2. Study objectives

This study aims to achieve the following objectives.

1. To examine the various in-vehicle distraction sources in relation to crashes.
2. To determine the factors that increase the likelihood of different in-vehicle distraction sources.
3. To develop a Bayesian Multinomial Logit model to analyze the data from driver distraction-related crashes.

3. Literature review

Distracted driving has been a topic of interest over the years. One of the early studies was performed in 2003 [15], focusing on distracted and drowsy driving using a nation-wide survey. Since then, the Highway Traffic Safety Administration has been documenting driver distraction statistics derived from the National Occupant Protection Use Survey (NOPUS). The presented statistics have been used in evaluating the trends in driver distractions [1], whereby a ten-year study (1999-2008) showed a decline in distracted driving from 1999 to 2005, then a rise thereafter.

Distracted driving has been linked to traffic safety and traffic flow [1, 3, 4, 8, 16]. Klauer et al [4] evaluated the association of in-vehicle distraction sources to crashes and near-crashes occurrence event using 42 and 109 novice and experienced drivers, respectively. Data collection was through accelerometers, cameras, global positioning systems, and other sensors installed in the vehicles. The study found that the risk of a crash or near-crash for novice drivers increased significantly if they dialed cellphones, sent/received texts, reached for objects, looked at roadside objects, or ate while driving. The risk of crashes and near-crashes increased for experienced drivers when they were dialing a cellphone. Over time, the prevalence of being distracted has increased among novice drivers, but not for experienced drivers. Further, the impacts of distracted driving on traffic flow [3] have also been evaluated. The results show that fluctuation in speed, fewer lane changes, and longer duration to complete a task (change lane, turning) were observed when a driver was texting. Moreover, when drivers were either texting or talking on phones, other simulated vehicles passed them more frequently than when they were undistracted. However, the results were not statistically significant between the age groups. In another study, which used a national crash database [8], the likelihood of severe crashes was reported to increase more for teen drivers if they were distracted by either cellphones or passengers than by other in-vehicle devices.

Moreover, distracted driving was assessed per several personal attributes, including age and gender [5, 6, 10]. One closed-road course study with a sample of 11 younger and nine older drivers was performed [5]. The driving performance of these two groups of drivers while text messaging via handheld mobile phones and an in-vehicle texting system was evaluated. The study found that texting from an in-vehicle system resulted in slightly better driver performance than a handheld mobile phone in terms of degraded steering measures. Steering control was also the assessment measure for elderly and middle-aged drivers when evaluated in an instrumented vehicle, with a controlled auditory-verbal processing load [6]. With a sample of 86 elderly and 51 middle-aged drivers, the study [6] revealed that, compared to no-task, both groups reduced steering control in the presence of an instructed task. A decrease and high variability in speed were observed more with the elderly than middle-aged drivers. Another study focused on cell phone use while driving among adolescents and emerging adults [10]. Utilizing logistic regression on survey questionnaire data, the study concluded that participants whose peers frequently texted while driving were more likely to text while driving the following year. One of the most recent studies [7], found no significant influence in cellphone distraction by gender, but driver age was associated with cellphone distraction. These findings are contrary to those reported in a ten-year study of trends in distracted driving [1]. Their study revealed that males driving alone were more often distracted driving crash victims than females. In a study conducted by Kim et al. (2019), it was found that using an in-vehicle infotainment system significantly increased the risk of distracted driving, especially when the task involved entering a destination into the navigation system. This highlights the need for car manufacturers to design infotainment systems that are less distracting and safer to use while driving. Another study conducted by [17] analyzed the effects of different types of distractions, including cognitive, manual, and visual distractions, on driving performance. The results showed that cognitive distractions, such as listening to the radio, had the least impact on driving performance, while manual distractions, such as reaching for a phone, had the greatest impact. This highlights the importance of reducing manual distractions while driving to improve road safety. In a study by [18], the effects of different types of smartphone usage, such as texting, talking, and web browsing, on driving performance were evaluated. The results showed that texting had the greatest negative impact on driving performance, followed by talking and web browsing. These findings suggest that texting and other forms of smartphone use should be discouraged while driving to minimize the risk of crashes. The impact of wearable technology, such as smartwatches, on driving performance has also been studied. In a study by [13], it was found that using a smartwatch while driving had a similar impact on driving performance as using a smartphone. This suggests that wearable technology can also be a source of distraction while driving and should be used with caution. Finally, a study by [14] evaluated the impact of texting and web surfing on driving behavior and safety in rural roads. The results showed that texting and web surfing while driving had a negative impact on driving behavior and increased the risk of crashes. This highlights the need for drivers to be aware of the dangers of distractions while driving, regardless of the type of road they are on.

In conclusion, these studies demonstrate the need for continued research on in-vehicle distractions to understand the impact they have on driving and crashes. This information can be used to design safer vehicles and promote safer driving habits, reducing the number of crashes and improving road safety.

To summarize, distracted driving has been extensively investigated, whereby its association to traffic safety and flow has been established. Moreover, in-vehicle and outside vehicle distraction sources have been identified and evaluated per personal attributes. Most of the previous studies' focus was cell phone-related distractions; meanwhile, other distraction sources have received less attention. Furthermore, most previous studies used data collected through either survey questionnaires, simulation, or instrumented vehicles. Additionally, the similarities and differences of the factors associated with in-vehicle driver distraction sources have not been explored.

4. Study data description

This study used three-year data of distracted driving-related crashes collected and stored by the Iowa Department of Transportation (IOWA DOT). IOWA DOT has an open data policy where the collected data are released to the public for research [19].

The authors downloaded the crash data. Within crash data, the major causes of the crashes were identified, including driver distraction. Different sources of driver distractions were listed in the dataset, which includes in-vehicle and out of vehicle distraction sources. The authors were interested in determining the similarities and differences of the factors affecting the distractions that originate from inside the vehicle. These include distractions from the cellphones (handheld and hands-free), passengers, radio and non-cellphone electronic devices, and fallen objects. Other distraction sources listed in the dataset included inattentive and unrestrained animals, among others. The dataset had the cellphone distraction-related crash data for 2015, 2016, and 2018. Therefore, the authors selected three years for analysis. Moreover, radio and non-cellphone electronics were grouped. After data cleaning for the few missing observations, a total of 5078 observations were available for further analysis.

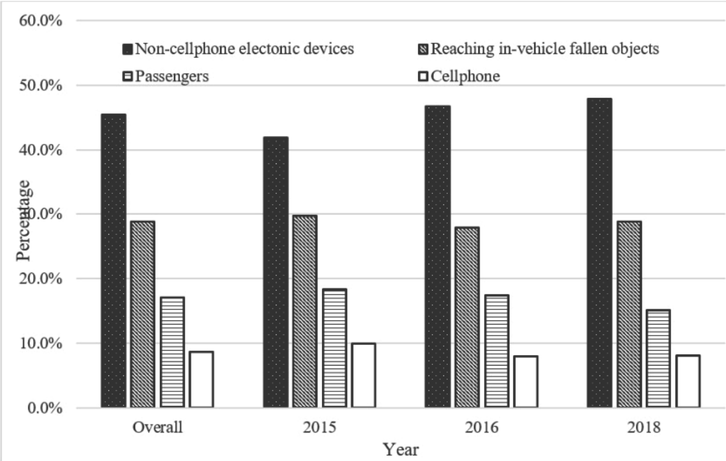

Fig. 1 - Distribution of the number of crashes per year by distraction source

Figure 1 shows the distribution of the number of crashes according to the distraction sources. It can be observed that the non-cellphone electronic devices have the largest proportion of all crashes. On average, 45% of crashes in this study are due to distraction from non-cellphone electronic devices inside the vehicle. The second source of distraction that contributed to a large number of crashes is reaching for the fallen object. The distraction from the passengers accounts for between 15% and 18% of all crashes, which is much higher that the percentage reported in the a previous meta-analysis [20]. On the other hand, cellphone use accounts for the smallest portion of the crashes.

Table 1 shows the variables used in this study. These variables were selected based on the knowledge gained through the literature review and engineering judgment. The authors grouped the variables into driver characteristics, vehicle characteristics, roadway, and temporal characteristics. The number of observations and the percentages across the categories is provided. According to Table 1, male drivers account for a large percentage of distracted driving for all sources of distractions except for distraction from passengers. The distribution of the driver's age per distraction sources shows that drivers aged 18-24 years have the highest percentage of crashes related to distraction from non-cellphone electronic devices and reaching for the fallen object. The group of drivers aged 25-34 years leads in terms of cellphone use and passengers distraction sources. Most of the distracted related crashes occurred when a driver was going straight, while a relatively significant proportion is parking-related crashes. The distribution per crash type shows that passenger car/SUVs have the largest percentage across all four distraction sources. The vehicle year distribution shows that the percentage of distracted related crashes is relatively small for newer (2015-2018) vehicles. However, this might be due to a low overall population of vehicles manufactured within that time, which were operating. The passengers revealed a different insight. As the number of occupants increases, the proportion of crashes due to distraction from cellphones, fallen objects, and non-cellphone electronic devices decreases. Conversely, the vehicles with two peoples have the largest proportion of crashes resulting from passengers' distraction. The speed limit distribution shows a higher proportion of that driver distracted-related crashes at low speeds. However, for cell phone use, a significant proportion of crashes (22.7%) is observed for 50 & 55 mph.

Tab. 1 - Descriptive analysis of the variables

	Non-cellphone electronic devices		Cellphone use		Passengers		Reaching fallen objects	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
Driver characteristics								
Driver's gender								
Female	1026	44.5%	208	47.2%	463	53.5%	637	43.5%
Male	1281	55.5%	233	52.8%	402	46.5%	828	56.5%
Driver's age								
Less than 18 years	247	10.7%	16	3.6%	80	9.2%	104	7.1%
18-24 years	644	27.9%	101	22.9%	151	17.5%	340	23.2%
25-34 years	505	21.9%	103	23.4%	227	26.2%	254	17.3%
35-44 years	333	14.4%	68	15.4%	131	15.1%	220	15.0%
45-54 years	252	10.9%	68	15.4%	108	12.5%	225	15.4%
55-64 years	205	8.9%	52	11.8%	101	11.7%	203	13.9%
65 and over	121	5.2%	33	7.5%	67	7.7%	119	8.1%
Driver's condition								
No DUI	2186	94.8%	407	92.3%	837	96.8%	1440	98.3%
DUI	121	5.2%	34	7.7%	28	3.2%	25	1.7%
Driver action								
Going straight	1366	59.2%	260	59.0%	491	56.8%	855	58.4%
Turning left, right U-turn	188	8.1%	55	12.5%	98	11.3%	91	6.2%
Decelerating	185	8.0%	20	4.5%	65	7.5%	93	6.3%
Braking	10	0.4%	15	3.4%	19	2.2%	4	0.3%
Lane change	28	1.2%	8	1.8%	13	1.5%	14	1.0%
Parking related	530	23.0%	83	18.8%	179	20.7%	408	27.8%
Vehicle characteristics								
Vehicle type								
Other vehicles	85	3.7%	30	6.8%	13	1.5%	89	6.1%
Light truck/pick-up	335	14.5%	70	15.9%	136	15.7%	249	17.0%
Passenger car SUV	1762	76.4%	318	72.1%	642	74.2%	1027	70.1%
Bus/van	125	5.4%	23	5.2%	74	8.6%	100	6.8%
Vehicle year								
Before 2000	290	12.6%	54	12.2%	108	12.5%	192	13.1%
2000-2004	566	24.5%	99	22.4%	209	24.2%	394	26.9%
2005-2009	638	27.7%	130	29.5%	260	30.1%	416	28.4%
2010-2014	551	23.9%	116	26.3%	202	23.4%	318	21.7%
2015-2018	262	11.4%	42	9.5%	86	9.9%	145	9.9%
Number of occupants								
One person	1814	78.6%	368	83.4%	233	26.9%	1197	81.7%
Two people	369	16.0%	59	13.4%	413	47.7%	182	12.4%
Three people	95	4.1%	10	2.3%	163	18.8%	53	3.6%
Four or more	29	1.3%	4	0.9%	56	6.5%	33	2.3%
Road and temporal characteristics								
Speed limit								
Less than 30 mph	588	25.5%	133	30.2%	252	29.1%	507	34.6%
30 & 35 mph	721	31.3%	113	25.6%	306	35.4%	427	29.1%
40 & 45 mph	265	11.5%	38	8.6%	103	11.9%	161	11.0%
50 & 55 mph	491	21.3%	100	22.7%	161	18.6%	223	15.2%
60 & 65 mph	168	7.3%	28	6.3%	28	3.2%	91	6.2%
Over 65 mph	74	3.2%	29	6.6%	15	1.7%	56	3.8%
Temporal factors								
Non-peak hour	466	20.2%	111	25.2%	239	27.6%	402	27.4%
Night time	684	29.6%	131	29.7%	189	21.8%	295	20.1%
Peak hours	1157	50.2%	199	45.1%	437	50.5%	768	52.4%
Day of the week								
Weekday	1703	73.8%	362	82.1%	622	71.9%	1137	77.6%
Weekend	604	26.2%	79	17.9%	243	28.1%	328	22.4%

5. Modeling methodology

To study the complex issue of in-vehicle distraction, a robust and comprehensive statistical model was needed. In this research, a multinomial logit model was used to investigate the different types of distraction, such as cellphones, passengers, and non-cellphone electronic devices, and their associated risk factors. This model was chosen because the response variable was unordered in nature. Bayesian inference was used to estimate the posterior distributions of the model parameters, offering several advantages over traditional frequentist methods.

The use of Bayesian inference allowed for the incorporation of prior knowledge in the analysis, which improved the accuracy of the results. The interpretation of the results was also easier and more intuitive with Bayesian inference [21-23].

In addition, the coefficients estimated using Bayesian inference were in the form of distributions, allowing for updates to be made in the future with new data. This made the model more flexible and adaptable over time [22, 23]. By considering various risk factors and incorporating prior knowledge, this approach offers valuable insights into the issue of distracted driving and has the potential to inform future policy and intervention strategies aimed at reducing the number of related crashes. This research demonstrates the power and utility of Bayesian inference in modeling complex.

In modeling the multinomial logit model, one response category is usually selected for use as the reference or baseline category in the parameter estimations [26, 27]. In this case, the non-cellphone electronic devices category, which includes television screens, GPS devices, and tablets, was used as a base category. Assume that a distraction variable, Y has K total number of observed categories; the expression in Equation 1 can estimate each distraction category's probability.

$$Prob(Y = i) = \frac{exp(\lambda_i)}{\sum_{h=1}^{K} exp(\lambda_{hi})} \qquad (1)$$

where:

$$\lambda_i = \beta_i + \sum_{h=1}^{H} \beta_{ih} \boldsymbol{X}_{ih}$$

where β_i and β_{ih} are vector of estimable coefficients and $\boldsymbol{X}_{ih}$ is the vector of explanatory variables.

Note that the estimated posterior distributions of coefficients result from the likelihood of the observed data and the prior distributions, as presented in Equation 2. The function in Equation 1 defines the likelihood of observed data in this study. The prior distribution (in Equation 2) is the probability of each coefficient before data has been used in the model [28, 29]. This likelihood can be either a strong prior distribution that is usually obtained from the previous findings or weakly informative priors in the absence of strong priors [30]. Since there is no study with a similar set of analyses as the current study, the authors used weakly informative priors to estimate parameters. It is important to understand that weakly informative priors in Bayesian inference have a negligible influence on the estimates, and data characteristics usually dominate in the estimates [30].

$$posteriordistribution, P(\theta|d) \propto prior,\ P(\theta) \times\ likelihood,\ L(d|\theta|) \qquad (2)$$

where,
$P(\theta)$ is the prior distribution of the parameter θ, $L(d|\theta)$ is the likelihood function, and $P(d|\theta)$ is the posterior distributions.

The prior distributions for regression coefficients were assigned to follow a Gaussian distribution, a weakly informative prior that is commonly used in Bayesian inference [29, 30]. The estimation of the parameters was implemented using the "brms" package [31] in R version 4.0.0 environment [32]. The No-U-Turn Sampler (NUTS) sampling step was applied, in which the initial burn-in phases were set to 1,000 iterations, and subsequently, 2,000 iterations were used for inference. As with the Bayesian inference, the model convergence was evaluated using the Gelman-Rubin Diagnostic statistic. For a model to achieve convergence, the difference between chain variances, which is the Gelman-Rubin Diagnostic statistic, has to be equal to 1 [33]. Moreover, the number of effective samples in each parameter was evaluated. Visual plots of chains, such as the autocorrelation plot, density, and trace plot of each parameter, were used in the assessment.

The discussion of the significant factors influencing distracted driving in the next section is based on the relative risk ratio (RRR). The RRRs are risk factors that are calculated by exponentiation of the estimated coefficients. The outcome implicitly indicates the ratio between a particular category's predicted probability to the base category. In this case, the relative probability of the cellphone distraction or passenger ($Y=i$) to the non-cellphone electronic devices ($Y = 1$) can be estimated as presented in Equation 3.

$$RRR = e^{coef} = \frac{P(Y=i)}{P(Y=1)} \tag{3}$$

where:
$coef$ Represents the estimated coefficient in the BMNL model.

When the estimated RRR of a variable is greater than 1 (RRR > 1), the risk ratio increases, while when it is less than 1 (RRR < 1), the risk ratio decreases relative to a base category [34]. More specifically, a unit change in the explanatory factor leads to the RRR of the predicted category to change (increase or decrease) relative to the reference group, given the other variables in the model are held constant.

6. Model results and discussion

The model results are presented in Table 2. The results interpretation is based on the 95% Bayesian credible intervals (BCI). The 95% BCI shows the credibility of the obtained RRR. The entire 95% BCI should be on either side of a unit RRR for the estimated RRR to be 95% credible interval. The results discussion section is divided into three segments; driver characteristics, vehicle characteristics, and roadway characteristics and temporal factors.

6.1. Driver characteristics

The driver's age, gender, DUI, and action/maneuver before crash occurrence are the characteristics of interest in this part. The gender of the drivers was not a statistically significant contributing factor for distracted driving. Compared to the distractions from non-cellphone electronic devices such as GPS devices and tablets, male drivers are less likely to be involved in crashes resulting from distraction by either cellphone use (RRR=0.77) or passengers (RRR=0.73). The results are consistent with one of the previous studies, which found that male drivers were less likely to be distracted by the passengers [35], but more likely to be involved in severe crashes that involve cellphone distraction [36]. A comparison between the distraction from non-cellphone electronic devices and the fallen objects on the vehicle shows that gender is not a statistically significant factor at a 95% BCI. The findings are consistent with a study that found that female drivers are less likely to be distracted by cellphones and are more likely to talk or sing with passengers [2].

Considering the driver's age, the results in Table 2 show that the likelihood of involvement in distracted driving crashes increases as the driver's age increases, irrespective of the distraction source. However, the magnitudes of the RRR are larger for cellphone distraction than other sources. More specifically, crashes for drivers aged 65 years and above are about 4.1 times more likely to be due to cellphones distraction than non-cellphone electronic devices in the vehicles. On the other hand, the crashes for the same age group are about 62% and 70% more likely to be associated with the distraction from passengers and reaching for the fallen objects compared to distraction by non-cellphone electronic devices. The drivers aged 18-44 years didn't show a statistically significant difference at a 95% BCI for cell phone use only as well as drivers of age 25-34 years for fallen objects only. These results are not consistent with previous studies' findings [37, 38]. According to their findings [37-39], young drivers are more likely to be distracted by cellphones than older drivers. However, their studies did not investigate the influence of age on distraction from passengers. Contrarily, [40] found no significant differences between young and middle-aged adults on distracted driving behaviors. However, the study did not distinguish different sources of distractions that are shown in the current study.

The relationship between driving under the influence (DUI) which is the act of operating a vehicle while under the influence of alcohol or drugs and driver distraction revealed a positive association with cellphone use, and a negative association with the other two distraction sources, compared to non-cellphone electronic devices in the vehicle. DUI drivers are about 45% more likely to be involved in crashes that are due to distractions from cellphone use (RRR=1.45) than non-cellphone electronic devices. The observation suggests that dialing and conversing while driving impedes focus on the road leading to crashes. Similar findings were found in Choudhary & Velaga [41]. On the other hand, these drivers are about 17% (RRR=0.83) and 59% (RRR= 0.41) less likely to be distracted by passengers and fallen objects in the vehicle, respectively.

The association between driver distraction and drivers' actions before crash occurrence was also assessed. Turning movements, braking, and deceleration were found to be statistically significant at 95% BCI. The study found that compared with the use of non-cellphone electronic devices, drivers that were braking were more than seven times likely to be involved in crashes whose distraction source is the cellphone use (RRR=7.17). Studies Choudhary & Velaga, [42] and Nasr Esfahani et al., [43] also found that phone use during driving causes a decrease in situation awareness and delays response to the events happening in the driving environment, which may lead to accidents. Furthermore, these drivers were more than two times likely to be associated with crashes due to passengers' distraction (RRR = 2.18). On the other hand, the braking was about 65% less likely to be associated with fallen objects distraction. Turning movements and lane change were more likely to be associated with cellphone use and passengers distraction but less likely to be associated with the distraction from dropped objects. There is not enough evidence to conclude the association between distraction from passengers and turning movements since the variable category has a unit RRR. When either decelerating or parking, drivers are less likely to be involved in crashes that are due to either of the distraction sources (RRRs are less than 1).

6.2. Vehicle characteristics

The vehicle type was also a variable of interest in this study. Four types of vehicles—light truck/pickups, passenger car/SUV, Bus/vans, and other vehicles (truck, tractor, trailer) were compared. Table 2 shows a negative correlation between vehicle type and distraction from cellphones and fallen objects. However, strong evidence for such an association is between passenger cars and fallen objects. In this case, the drivers in passenger cars/SUVs are about 42% less likely to be distracted by the fallen objects (RRR = 0.58). For other distraction sources as well as other vehicle types, no enough supporting evidence is available to conclude.

The vehicle's age has revealed a strong and positive association with the cell phone distraction compared to non-cellphone electronic devices. The results in Table 2 show that the drivers of the most recent versions of the vehicles are less likely to be involved in crashes whose distraction sources are cellphones than non-cellphone electronic device use. However, the association is not statistically significant at 95% BCI. The drivers of the vehicles produced in 2015-2019 vehicles are 30% less likely to be distracted by cellphones as compared to drivers of vehicles manufactured before 2000 (RRR = 0.70). On the other hand, drivers of the vehicles manufactured between 2000 and 2004 are only 6% less likely to be distracted by cellphones as compared to drivers of vehicles manufactured before 2000. Further, results show that there is a statistically significant negative association between the most recent version of vehicles (2015-2019) and crashes due to drivers reaching for the fallen objects (RRR=0.74).

As it was expected, the number of passengers is associated with the crashes that are due to distraction from passengers. The more the number of passengers, the higher is the magnitude of the RRRs.

Tab. 2 - Multinomial Logit results (*continue*)

	Cellphone use					Passengers					Reaching fallen objects				
	Mean	SD	RRR	[95% CI]		Mean	SD	RRR	[95% CI]		Mean	SD	RRR	[95% CI]	
Driver characteristics															
Driver's gender															
Female*															
Male	**-0.26**	**0.12**	**0.77**	**0.62**	**0.97**	**-0.32**	**0.10**	**0.73**	**0.60**	**0.88**	-0.04	0.08	0.96	0.83	1.12
Driver's age															
Less than 18 years*															
18-24 years	**0.84**	**0.28**	**2.32**	**1.35**	**4.18**	0.04	0.18	1.04	0.74	1.49	**0.36**	**0.14**	**1.43**	**1.09**	**1.90**
25-34 years	**1.19**	**0.28**	**3.29**	**1.93**	**5.93**	**0.57**	**0.17**	**1.77**	**1.27**	**2.51**	0.28	0.15	1.32	0.99	1.75
35-44 years	**1.18**	**0.30**	**3.25**	**1.90**	**6.05**	**0.37**	**0.19**	**1.45**	**1.00**	**2.10**	**0.53**	**0.16**	**1.70**	**1.28**	**2.32**
45-54 years	**1.41**	**0.30**	**4.10**	**2.32**	**7.61**	**0.70**	**0.20**	**2.01**	**1.38**	**2.97**	**0.83**	**0.16**	**2.29**	**1.70**	**3.06**
55-64 years	**1.33**	**0.31**	**3.78**	**2.10**	**7.10**	**0.92**	**0.21**	**2.51**	**1.70**	**3.78**	**0.90**	**0.16**	**2.46**	**1.77**	**3.32**
65 and over	**1.42**	**0.33**	**4.14**	**2.20**	**8.00**	**0.96**	**0.24**	**2.61**	**1.67**	**4.18**	**0.86**	**0.18**	**2.36**	**1.67**	**3.35**
Driver's condition															
No DUI*															
DUI	**0.37**	**0.22**	**1.45**	**0.93**	**2.23**	-0.19	0.25	0.83	0.50	1.36	**-0.88**	**0.23**	**0.41**	**0.26**	**0.66**
Driver action															
Going straight*															
Turning left, right U-turn	**0.40**	**0.17**	**1.49**	**1.06**	**2.08**	0.24	0.16	1.27	0.92	1.73	**-0.31**	**0.14**	**0.73**	**0.55**	**0.96**
Decelerating	**-0.60**	**0.26**	**0.55**	**0.32**	**0.89**	-0.20	0.18	0.82	0.58	1.15	**-0.36**	**0.14**	**0.70**	**0.53**	**0.90**
Braking	**1.97**	**0.45**	**7.17**	**2.94**	**17.64**	0.78	0.45	2.18	0.94	5.42	-1.05	0.62	0.35	0.10	1.13
Lane change	0.31	0.43	1.36	0.58	3.13	0.09	0.40	1.09	0.50	2.44	-0.33	0.35	0.72	0.36	1.39
Parking related	-0.23	0.15	0.79	0.59	1.06	-0.52	0.12	0.59	0.47	0.76	-0.05	0.09	0.95	0.80	1.14
Vehicle characteristics															
Vehicle type															
Other vehicles*															
Light truck/pick-up	-0.36	0.27	0.70	0.41	1.20	**0.67**	**0.35**	**1.95**	**1.00**	**4.10**	-0.30	0.19	0.74	0.51	1.06
Passenger car SUV	-0.44	0.25	0.64	0.39	1.05	0.25	0.34	1.28	0.68	2.53	**-0.54**	**0.18**	**0.58**	**0.40**	**0.82**
Bus/van	-0.57	0.34	0.57	0.28	1.09	0.18	0.37	1.20	0.60	2.56	-0.35	0.22	0.70	0.45	1.09
Vehicle year															
Before 2000*															
2000-2004	-0.04	0.19	0.96	0.66	1.40	-0.03	0.16	0.97	0.72	1.32	0.10	0.12	1.11	0.88	1.39
2005-2009	0.07	0.19	1.07	0.74	1.55	-0.05	0.16	0.95	0.70	1.31	0.01	0.12	1.01	0.79	1.27
2010-2014	0.03	0.19	1.03	0.71	1.48	-0.10	0.16	0.90	0.66	1.23	-0.21	0.13	0.81	0.63	1.03
2015-2019	-0.36	0.24	0.70	0.44	1.08	-0.15	0.19	0.86	0.59	1.25	**-0.30**	**0.15**	**0.74**	**0.55**	**1.00**
Number of occupants															
One person*															
Two people	-0.19	0.16	0.83	0.61	1.13	**2.24**	**0.11**	**9.39**	**7.61**	**11.70**	-0.22	0.10	0.80	0.66	0.98
Three people	-0.56	0.35	0.57	0.28	1.08	**2.70**	**0.16**	**14.88**	**10.91**	**20.09**	-0.06	0.18	0.94	0.65	1.34
Four or more	-0.32	0.57	0.73	0.21	2.01	**2.91**	**0.26**	**18.36**	**11.25**	**30.27**	0.60	0.27	1.82	1.04	3.13

Road and temporal characteristics															
Speed limit															
Less than 30 mph*															
30 & 35 mph	**-0.35**	**0.15**	**0.70**	**0.53**	**0.95**	-0.18	0.12	0.84	0.66	1.06	**-0.47**	**0.09**	**0.63**	**0.52**	**0.76**
40 & 45 mph	**-0.43**	**0.21**	**0.65**	**0.43**	**0.97**	-0.27	0.16	0.76	0.55	1.05	**-0.48**	**0.12**	**0.62**	**0.49**	**0.79**
50 & 55 mph	-0.12	0.16	0.89	0.65	1.23	**-0.44**	**0.14**	**0.64**	**0.49**	**0.84**	**-0.71**	**0.11**	**0.49**	**0.40**	**0.60**
60 & 65 mph	-0.38	0.24	0.68	0.42	1.08	**-1.13**	**0.24**	**0.32**	**0.20**	**0.51**	**-0.62**	**0.15**	**0.54**	**0.39**	**0.73**
Over 65 mph	0.38	0.27	1.46	0.86	2.46	**-0.82**	**0.33**	**0.44**	**0.22**	**0.84**	**-0.43**	**0.20**	**0.65**	**0.44**	**0.97**
Temporal factors															
Non-peak hour*															
Night time	-0.22	0.16	0.80	0.58	1.09	**-0.54**	**0.14**	**0.58**	**0.44**	**0.76**	**-0.62**	**0.10**	**0.54**	**0.44**	**0.66**
Peak hours	**-0.37**	**0.14**	**0.69**	**0.53**	**0.91**	**-0.27**	**0.11**	**0.76**	**0.61**	**0.94**	**-0.24**	**0.09**	**0.79**	**0.67**	**0.93**
Day of the week															
Weekday*															
Weekend	**-0.54**	**0.14**	**0.58**	**0.44**	**0.76**	**-0.13**	**0.10**	**0.88**	**0.72**	**1.07**	**-0.11**	**0.08**	**0.90**	**0.76**	**1.05**
Intercept	-1.62	0.43	0.20	0.09	0.43	-1.95	0.42	0.14	0.06	0.31	0.39	0.25	1.48	0.90	2.46

Note: *Denotes a base category; SD =Standard Deviation; RRR= Relative Risk Ratio; Bolded numbers means significant at 95% BCI; the BCI intervals (2.5%-97.5% are for the RRR's).

The association of the number of passengers and crashes due to distraction from either cellphone or fallen objects showed no statistical significance difference at 95% BCI. For the fallen objects, the possible reason might be that passengers can reach for the fallen object; thus, let the driver continue focusing on the road. Studies by [44], [45] also concluded that more passengers are associated with a higher likelihood of a driver being distracted due to a higher degree of occupants interaction.

6.3. Roadway characteristics and temporal factors

Speed limit, day of the week, and peak hours are the variables that are presented in this section.

The association between the speed limit and different sources of distractions shows mixed results. However, in general, as the posted speed limit increases, drivers are less likely to be distracted by either passengers or fallen objects. The RRR for these two distraction sources is less than 1 for all posted speed limits intervals. Similarly, at lower posted speed limits, drivers are less likely to be involved in crashes whose distraction source is the cellphone. In contrast, the results in Table 2 show that at higher speed, drivers are more likely to be involved in crashes whose distraction source is the cellphone. In fact, at 65mph or higher speed limit, drivers are about 46 % (RRR = 1.46) more likely to be involved in distraction-related crashes whose source is cellphone use than non-cellphone electronic devices. However, that variable category is not statistically significant at 95% BCI. The observations suggest that, at higher speeds, the distracted driving-related crashes are likely to be due to the use of non-cellphone electronic devices. Several studies [46], [47] also report the high crash frequency with the increase in the speed limit, although they do not explicitly describe the nature of distraction leading to these crashes.

The nighttime, peak hours, and weekends show less likelihood association with either cellphone use, passengers, or fallen objects in the vehicles. The RRRs for all these variables are less than one. However, no statistically significant difference at a 95% BCI between nighttime and non-peak hour with regard to cellphone use while driving. Similarly, a study by Zhang et al. [48] suggests that crashes showed little difference between peak hours, light and dark time periods, implying that factors other than the discussed distraction variables play the dominant role in these crashes.

7. Conclusion and future studies

It is well understood that more than 90 percent of all crashes are associated with human errors. Among many other human-related errors, distracted driving's contribution to crash occurrence, particularly related to cellphone use, has been extensively studied in the literature. However, the underlying factors associated with in-vehicle distractions other than cellphone use have not been fully explored. Thus, this paper contributes to the body of knowledge by evaluating distracted driving's contributory factors. The evaluated factors included driver and vehicle characteristics, roadway characteristics, and temporal factors. These contributory factors were assessed using the Bayesian Multinomial logit (BMNL) model. This study used 5,078 observations, whose crashes were linked with distracted driving in Iowa. Four in-vehicle distraction sources—cellphone use, non-cellphone electronic devices, and passengers, and fallen objects—were investigated to determine the contributory factors that increase their odds of occurrence.

The descriptive analysis shows that about 45% of crashes were due to distractions from non-cellphone electronic devices, 29% from reaching the fallen objects within a vehicle, 17% from non-driver passengers, and 8.7% due to cellphone use. The BMNL model was performed to associate the sources of distracted driving-related crashes and other predictor variables. All the distraction sources were compared to the distraction from non-cellphone electronic device use. The BMNL model results suggest that older drivers are more likely to be involved in crashes that are associated with the distraction from cellphones. Conversely, the older the driver, the less likely it is to be distracted by the passengers and reach for the fallen object. Drivers under the influence of alcohol were more likely to be involved in crashes that are due to cellphone use. Moreover, compared to going straight, either braking or turning drivers were more likely to be involved in cellphone-related crashes. Drivers with newer vehicles were less likely to be involved in crashes involved fallen objects as distraction sources. As was expected, the more people in the vehicle, the higher the likelihood of driving distracted-related crashes due to the passenger distraction. Higher speed limits were associated with crashes due to either cellphone, passengers, or fallen objects. Similarly, day peak-hour periods, nighttime, and weekends were associated with the decrease of the crashes related to either cellphone, passengers, or fallen objects.

These findings suggest that cellphone use accounts for a relatively small percentage of distracted driving-related crashes compared to other sources such as passengers, non-cellphone electronic device use, and fallen objects. Moreover, the model results suggest the distraction sources share a large percent of the predictors. These findings imply that the success of combatting one distraction source would significantly reduce other distraction sources. However, other predictors decrease the likelihood of the crash due to one distraction source but increase the chance of crash occurrence due to other distraction sources. Therefore, in combating distracted driving across all drivers, engineers and planners should be careful in selecting strategies that minimize distractions from all in-vehicle sources.

One of the limitations of this study is the use of self-reported data collected from drivers after the crash had occurred. The accuracy of this information is dependent on the honesty of the drivers. As such, it is important for the reader to keep in mind that the results may be biased if the drivers were not truthful in their reporting. This is a common limitation of self-reported data, and it is a limitation that should be considered when interpreting the results of this study. This can also lead to underreporting of certain distractions, such as cell phone usage, as drivers may be unwilling to admit to engaging in such activities while driving. Therefore, it is important to consider this potential source of bias and its impact on the results when interpreting the findings of this research. Additionally, the study was conducted in a specific location (Iowa) and the results may not be generalizable to other regions with different driving cultures and road conditions. Furthermore, the study only considered four in-vehicle distraction sources and did not include other factors that may have contributed to the crashes. Therefore, it is important to consider these limitations when interpreting the results and applying them to other contexts.

While the findings and conclusions of this study are valuable, there is potential for further research to expand upon these findings. Other factors not considered in this study, such as other roadway geometric conditions, dynamic signs on the road, work zones, and other road incidents, could also play a role in in-vehicle distraction and should be explored in future research. Furthermore, investigating the impact of different interventions aimed at reducing in-vehicle distraction would be valuable in developing effective policies to reduce distracted driving-related crashes. By considering a wider range of contributing factors and interventions, future research can further enhance our understanding of in-vehicle distraction and contribute to making our roads safer for all.

References

1. F. A. Wilson and J. P. Stimpson, "Trends in Fatalities From Distracted Driving in the United States," *Am J Public Health*, vol. 100, no. 11, 2010, doi: 10.2105/AJPH.2009.187179.
2. D. G. Kidd and N. K. Chaudhary, "Changes in the sources of distracted driving among Northern Virginia drivers in 2014 and 2018: A comparison of results from two roadside observation surveys," *J Safety Res*, vol. 68, pp. 131–138, Feb. 2019, doi: 10.1016/j.jsr.2018.12.004.
3. D. Stavrinos *et al.*, "Impact of distracted driving on safety and traffic flow," *Accid Anal Prev*, vol. 61, pp. 63–70, 2013, doi: 10.1016/j.aap.2013.02.003.
4. S. G. Klauer *et al.*, "Distracted Driving and Risk of Road Crashes among Novice and Experienced Drivers From the Virginia Tech Transportation Institute (A bs tr ac t," *n engl j med*, vol. 370, pp. 54–63, 2014, doi: 10.1056/NEJMsa1204142.
5. J. M. Owens, S. B. Mclaughlin, and J. Sudweeks, "Driver performance while text messaging using handheld and in-vehicle systems," *Accid Anal Prev*, vol. 43, pp. 939–947, 2010, doi: 10.1016/j.aap.2010.11.019.
6. K. R. Thompson, A. M. Johnson, J. L. Emerson, J. D. Dawson, E. R. Boer, and M. Rizzo, "Distracted driving in elderly and middle-aged drivers," *Accid Anal Prev*, vol. 45, pp. 711–717, 2011, doi: 10.1016/j.aap.2011.09.040.
7. X. Sun, "Investigating Problem of Distracted Drivers on Louisiana Roadways," Dec. 2018.
8. D. M. Neyens and L. N. Boyle, "The influence of driver distraction on the severity of injuries sustained by teenage drivers and their passengers," *Accid Anal Prev*, vol. 40, no. 1, pp. 254–259, Jan. 2008, doi: 10.1016/J.AAP.2007.06.005.
9. J. K. Caird, S. M. Simmons, K. Wiley, K. A. Johnston, and W. J. Horrey, "Does Talking on a Cell Phone, With a Passenger, or Dialing Affect Driving Performance? An Updated Systematic Review and Meta-Analysis of Experimental Studies," *Hum Factors*, 2018, doi: 10.1177/0018720817748145.
10. N. Trivedi, D. Haynie, J. Bible, D. Liu, and B. Simons-Morton, "Cell Phone Use While Driving: Prospective Association with Emerging Adult Use," 2017, doi: 10.1016/j.aap.2017.04.013.
11. O. S. Fagbemi and K. Pfeffer, "The relationship between chronic sleep deficits and distractions in young adult drivers," *Advances in Transportation Studies*, pp. 57–64, 2007.

12. C. Ortiz, S. Ortiz-Peregrina, J. J. Castro, M. Casares-López, and C. Salas, “Driver distraction by smartphone use (WhatsApp) in different age groups,” *Accid Anal Prev*, vol. 117, pp. 239–249, Aug. 2018, doi: 10.1016/J.AAP.2018.04.018.
13. D. Perlman, A. Samost, A. G. Domel, B. Mehler, J. Dobres, and B. Reimer, “The relative impact of smartwatch and smartphone use while driving on workload, attention, and driving performance,” *Appl Ergon*, vol. 75, pp. 8–16, Feb. 2019, doi: 10.1016/J.APERGO.2018.09.001.
14. M. Sekadakis, C. Katrakazas, F. Orfanou, D. Pavlou, M. Oikonomou, and G. Yannis, “Impact of texting and web surfing on driving behavior and safety in rural roads,” *International Journal of Transportation Science and Technology*, Jun. 2022, doi: 10.1016/J.IJTST.2022.06.001.
15. D. Royal, “National survey of distracted and drowsy driving attitudes and behavior,” United States. National Highway Traffic Safety Administration, Mar. 2003.
16. A. R. Mahpour, A. M. Amir, and E. S. Ebrahimi, “Do drivers have a good understanding of distraction by wrap advertisements? Investigating the impact of wrap advertisement on distraction-related driver’s accidents,” *Advances in Transportation Studies*, pp. 19–30, 2019.
17. T. Chen, O. Oviedo-Trespalacios, N. N. Sze, and S. Chen, “Distractions by work-related activities: The impact of ride-hailing app and radio system on male taxi drivers,” *Accid Anal Prev*, vol. 178, p. 106849, Dec. 2022, doi: 10.1016/J.AAP.2022.106849.
18. P. A. Busch and S. McCarthy, “Antecedents and consequences of problematic smartphone use: A systematic literature review of an emerging research area,” *Comput Human Behav*, vol. 114, Jan. 2021, doi: 10.1016/j.chb.2020.106414.
19. IOWADOT, “Iowa Department of Transportation - Open Data,” 2020. https://data.iowadot.gov/ (accessed Apr. 25, 2020).
20. A. Theofilatos, A. Ziakopoulos, E. Papadimitriou, and G. Yannis, “How many crashes are caused by driver interaction with passengers? A meta-analysis approach,” *J Safety Res*, 2018, doi: 10.1016/j.jsr.2018.02.001.
21. N. Novat, E. Kidando, B. Kutela, and A. E. Kitali, “A comparative study of collision types between automated and conventional vehicles using Bayesian probabilistic inferences,” *J Safety Res*, Nov. 2022, doi: 10.1016/J.JSR.2022.11.001.
22. A. Gelman, J. B. B. Carlin, H. S. S. Stern, and D. B. B. Rubin, “Bayesian Data Analysis, Third Edition (Texts in Statistical Science),” *Book*, p. 675, 2014, doi: 10.1007/s13398-014-0173-7.2.
23. D. Makowski, M. S. Ben-Shachar, S. H. A. Chen, and D. Lüdecke, “Indices of Effect Existence and Significance in the Bayesian Framework,” *Front Psychol*, vol. 10, Dec. 2019, doi: 10.3389/FPSYG.2019.02767.
24. D. Makowski, M. S. Ben-Shachar, S. H. A. Chen, and D. Lüdecke, “Indices of Effect Existence and Significance in the Bayesian Framework,” *Front Psychol*, vol. 10, p. 2767, Dec. 2019, doi: 10.3389/fpsyg.2019.02767.
25. M. Ahmed, M. Abdel-Aty, and R. Yu, “Bayesian updating approach for real-time safety evaluation with automatic vehicle identification data,” *Transp Res Rec*, no. 2280, pp. 60–67, 2012, doi: 10.3141/2280-07.
26. D. Chimba, D. Emaasit, and B. Kutela, “Likelihood Parameterization of Bicycle Crash Injury Severities,” *J Transp Technol*, vol. 02, no. 03, pp. 213–219, Jul. 2012, doi: 10.4236/jtts.2012.23023.
27. B. Kutela, E. Kidando, A. E. Kitali, S. Mwende, N. Langa, and N. Novat, “Exploring pre-crash gate violations behaviors of drivers at highway-rail grade crossings using a mixed multinomial logit model,” *https://doi.org/10.1080/17457300.2021.1990348*, 2022, doi: 10.1080/17457300.2021.1990348.

28. J. K. Kruschke, "Bayesian estimation supersedes the t test.," *J Exp Psychol Gen*, 2013, doi: 10.1037/a0029146.
29. E. Kidando *et al.*, "Novel Approach for Calibrating Freeway Highway Multi-Regimes Fundamental Diagram," vol. 2674, no. 9, 2020, doi: 10.1177/0361198120930221.
30. J. K. Kruschke, "Bayesian estimation supersedes the t test.," *J Exp Psychol Gen*, 2013, doi: 10.1037/a0029146.
31. P. C. Bürkner, "Advanced Bayesian multilevel modeling with the R package brms," *R Journal*, vol. 10, no. 1, pp. 395–411, 2018, doi: 10.32614/rj-2018-017.
32. R Core Team, "R: A language and environment for statistical computing," 2018.
33. A. Gelman, J. B. B. Carlin, H. S. S. Stern, D. B. Dunson, A. Vehtari, and D. B. B. Rubin, *Bayesian Data Analysis, Third Edition (Texts in Statistical Science)*. 2014. doi: 10.1007/s13398-014-0173-7.2.
34. A. K. Çelik and E. Oktay, "A multinomial logit analysis of risk factors influencing road traffic injury severities in the Erzurum and Kars Provinces of Turkey," *Accid Anal Prev*, vol. 72, pp. 66–77, 2014, doi: 10.1016/j.aap.2014.06.010.
35. F. Zhang, S. Mehrotra, and S. C. Roberts, "Driving distracted with friends: Effect of passengers and driver distraction on young drivers' behavior," *Accid Anal Prev*, 2019, doi: 10.1016/j.aap.2019.07.022.
36. P. Wu, L. Song, and X. Meng, "Temporal analysis of cellphone-use-involved crash injury severities: Calling for preventing cellphone-use-involved distracted driving," *Accid Anal Prev*, vol. 169, p. 106625, May 2022, doi: 10.1016/J.AAP.2022.106625.
37. Highway Traffic Safety Administration, "Driver Electronic Device Use in 2016," 2016.
38. K. A. Braitman and A. T. Mccartt, "National Reported Patterns of Driver Cell Phone Use in the United States," *Traffic Inj Prev*, vol. 11, no. 6, pp. 543–548, 2010, doi: 10.1080/15389588.2010.504247.
39. M. M. Hossain, H. Zhou, S. Das, X. Sun, and A. Hossain, "Young drivers and cellphone distraction: Pattern recognition from fatal crashes," *Journal of Transportation Safety & Security* , 2022, doi: 10.1080/19439962.2022.2048763.
40. C. Northcutt, T. R. Bell, and D. Stavrinos, "Mechanisms behind distracted driving behavior: The role of age and executive function in the engagement of distracted driving," *Accid Anal Prev*, vol. 98, pp. 123–129, 2017, doi: 10.1016/j.aap.2016.09.030.
41. J. K. Caird, S. M. Simmons, K. Wiley, K. A. Johnston, and W. J. Horrey, "Does Talking on a Cell Phone, With a Passenger, or Dialing Affect Driving Performance? An Updated Systematic Review and Meta-Analysis of Experimental Studies," *Hum Factors*, 2018, doi: 10.1177/0018720817748145.
42. P. Choudhary and N. R. Velaga, "Modelling driver distraction effects due to mobile phone use on reaction time," *Transp Res Part C Emerg Technol*, 2017, doi: 10.1016/j.trc.2017.02.007.
43. H. Nasr Esfahani, R. Arvin, Z. Song, and N. N. Sze, "Prevalence of cell phone use while driving and its impact on driving performance, focusing on near-crash risk: A survey study in Tehran," *Journal of Transportation Safety and Security*, 2019, doi: 10.1080/19439962.2019.1701166.
44. A. Theofilatos, A. Ziakopoulos, E. Papadimitriou, and G. Yannis, "How many crashes are caused by driver interaction with passengers? A meta-analysis approach," *J Safety Res*, 2018, doi: 10.1016/j.jsr.2018.02.001.
45. F. Zhang, S. Mehrotra, and S. C. Roberts, "Driving distracted with friends: Effect of passengers and driver distraction on young drivers' behavior," *Accid Anal Prev*, 2019, doi: 10.1016/j.aap.2019.07.022.

46. J. I. Castillo-Manzano, M. Castro-Nuño, L. López-Valpuesta, and F. V. Vassallo, "The complex relationship between increases to speed limits and traffic fatalities: Evidence from a meta-analysis," *Safety Science*. 2019. doi: 10.1016/j.ssci.2018.08.030.
47. R. A. Boateng, M. D. Fontaine, and Z. H. Khattak, "Driver Response to Variable Speed Limits on I-66 in Northern Virginia," *J Transp Eng A Syst*, 2019, doi: 10.1061/jtepbs.0000236.
48. D. Robb and T. Barnes, "Accident rates and the impact of daylight saving time transitions," *Accid Anal Prev*, 2018, doi: 10.1016/j.aap.2017.11.029.

Advances in Transportation Studies: an international Journal 61 (2023)

ISBN 979-12-218-0919-0 – ISSN 1824-5463 – DOI 10.53136/97912218091902 – pag. 19-36

Prediction of road traffic accident severity based on XGBoost-BP neural network

R.Y. Qian X. Wang

School of Applied Science - Beijing Information Science and Technology University
Xiaoying East Road No.12,100192, Beijing, China
email: wangxin@bistu.edu.cn

subm. 21st November 2022 approv. after rev. 18th March 2023

Abstract

Traffic safety has been of great concern in recent years. The prediction of the severity of traffic accidents is an important part of it. The occurrence of traffic accidents shows the characteristics of uncertainty and non-linearity because of the influence of random factors. However, most of the existing models are single machine learning (ML) models, which have limitations in accuracy and generalization. This study proposes a traffic accident severity prediction model based on a combination of XGBoost (eXtreme Gradient Boosting) and Backpropagation Neural Network (BPNN). Firstly, feature selection is performed using the XGBOOST model. Secondly, the selected feature is used as the input layer of BPNN. In addition, traffic accidents have class imbalance, so the total cost is minimized by using cost-sensitive algorithm. Finally, the precision, recall and area under the curve (AUC) are used to evaluate the prediction results of the model. The 2005-2014 UK traffic accident dataset is used for prediction and compared with other machine learning models. Experiments show that (1) the XGBoost-BPNN model outperformed the single XGBoost, logistic regression (LR), and Support vector machine (SVM) models in terms of AUC, recall, and precision. (2) The number of neurons, the number of hidden layers and the learning rate of a neural network model have a large impact on the prediction accuracy. Increasing the number of neurons appropriately can improve the convergence speed and prediction effect of the model. This study can provide a reference for traffic accident prevention and early warning.

Keywords – neural network, traffic accident risk predicting, imbalanced dataset

1. Introduction

Road traffic injuries are the eighth leading cause of death worldwide. According to the World Health Organization (WHO), road traffic injuries will be the fifth leading cause of death by 2030 [1]. In the UK, 160597 people suffered varying levels of injury in road traffic accidents reported to the police in 2018. Economic losses from traffic accidents account for approximately 3% of annual GDP [2]. Due to the frequency of traffic accidents worldwide, it is possible to analyze the causes based on the predicted results of the severity of the accidents. We can prevent huge losses by summarising the occurrence, development, characteristics and distribution of serious road traffic accidents [3].

Usually, most domestic and international scholars apply statistical models and machine learning models to predict traffic accidents [4-5]. Researchers usually use statistical models to predict the number of collisions with different accident severity. These primarily include ordered probit (OP) models [6], ordered logit models [7], polynomial logit models [8], and logistic regression models [9]. However, statistical models often require some assumptions about the underlying probability distribution of the data and the predetermined relationships between the dependent and independent variables. Violation of these assumptions may produce incorrect estimates and inferences. Recent research has found that machine learning (ML) methods can provide accurate predictive models because they can handle more complex functions. A variety of ML methods have been applied in the study of accident severity, including classification and regression trees (CART) models [10], support vector machine (SVM) models [11-13], and ensemble learning [14-15].

Among them, artificial neural network (ANN) modeling has better results for such problems. ANN does not require any predetermined underlying relationship between the dependent and independent variables, which provides better flexibility and accuracy in dealing with prediction problems [16]. Shohel Amin applied back propagation-artificial neural network (BP-ANN) to model the factors affecting traffic accidents of both older female and male drivers. It was found that the BP-ANN approach was effective in reducing prediction errors [17]. However, due to the structural complexity of the neural network model, the training may fall into local minima. To overcome the redundancy of the model, feature filtering is particularly important [18]. Dong studied the Pearson correlation coefficients of accident influencing factors and developed a BP neural network prediction model for road traffic accidents. Twelve features were selected for the final model, which gave very satisfactory prediction results with an average error of less than 10% [19]. Yang used a grey correlation model to select the eight factors most associated with accidents at intersections. BP neural network was used to predict the number of traffic accidents at urban road intersections based on the eight factors [20]. The above methods work well for understanding the data, but are not necessarily effective for feature optimisation. Tree-based ensemble learning algorithms can quickly filter features based on the Gini coefficient. One of these is the extreme gradient boosting (XGBoost) model. It can identify key features by making a link between behavioural features and the corresponding risk level. This can effectively simplify the complexity of BP neural network models and improve their generalisation ability [21].

However, none of the above scholars have considered the unbalanced nature of the traffic accident data set. In practical applications, this unbalanced nature means that dichotomous or multiclassification training is likely to cause serious bias. Delen analysed US traffic accident data from 1995 to 2000. The study was carried out by transforming a five classification problem into a two classification problem. This approach improved the unbalanced nature of the data [22]. Chen et al. analysed and predicted highway crash severity data for Taiwan, China, from 2015 to 2019. The article regroups the accident data for classification. However, such a classification is too subjective [23]. Xie used oversampling, undersampling and Synthetic Minority Oversampling Technique (SMOTE) methods in his study to improve the imbalance of the data. The results show that random oversampling can greatly improve the accuracy of unbalanced data. However, the source of the data and the characteristics considered are not stated in the article [24]. Therefore, this paper deals with the highly unbalanced classification problem using cost-sensitive learning, which assigns different penalties for different classification errors. No data were duplicated or lost in the process of adjusting the weights [25].

This paper proposes a method combining feature selection and BP neural network to predict accident severity. The XGBoost model is used to analyze the feature importance and replace more raw data with fewer features. Then the selected features are used as the input layer parameters of the BP neural network structure, which improves the iterative speed and computing efficiency of the model. The combined models can complement each other for accurate prediction purposes. The