Accident Analysis and Prevention 157 (2021) 106162

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Mining patterns of near-crash events with and without secondary tasks

Xiaoqiang Kong [a], Subasish Das [b,*], Yunlong Zhang [a]

[a] Zachry Department of Civil & Environmental Engineering, Texas A&M University, 3136 TAMU, College Station, TX, 77843-3136, United States
[b] Texas A&M Transportation Institute, 3500 NW Loop 410, San Antonio, TX, 78229, United States

ARTICLE INFO

Keywords:
Near-crash
Safety-critical event
Secondary task
Association rules mining

ABSTRACT

The engagement of secondary tasks, like using a phone or talking to passengers while driving, could introduce considerable risks to driving safety. This study utilizes a near-crash dataset extracted from a naturalistic driving study to explore the patterns of near-crash events with or without the involvement of secondary tasks as a surrogate approach to understand the impact of these behaviors on traffic safety. The dataset contains information about driver behaviors, such as secondary tasks, vehicle maneuvers, other conflict vehicles' maneuvers before and during near-crash events, and the driving environment. The patterns for near-crashes with or without the involvement of secondary tasks are mined by adopting the apriori association rule algorithm. Finally, the mined rules for the near-crash events with or without the involvement of the secondary tasks are analyzed and compared. The results demonstrate that near-crashes with the involvement of secondary tasks often occur with drivers in a relatively stable and presumably predictable environment, such as an interstate highway with a constant speed. This type of near-crash is highly associated with the leading vehicle's sudden slowing or stopping since there is no expectation of any interruptions for these drivers performing the secondary tasks. The most common evasive maneuver in this kind of emergency is braking. Near-crashes without the involvement of secondary tasks is often associated with lane-changing behavior and sideswipe incidents. With shorter reaction time and awareness of the driving environment, the drivers in this type of near-crash can often make more complex maneuvers, like braking and steering, to avoid a collision. Understanding the patterns of these two types of near-crash incidents could help safety researchers, traffic engineers, and even vehicle designers/engineers develop countermeasures for minimizing potential collisions caused by secondary tasks or improper lane changing behaviors.

1. Introduction

Traditionally, the approach adopted by researchers to analyze road safety is to observe crash statistics through state-based police-reported crash databases or national databases such as the Fatality Analysis Reporting System (FARS). These crash databases are usually well managed, and information on crashes is well documented. However, the information recorded in the database only describes the location features and characteristics of crashes after the crash occurrence, such as the number of people injured and the number of fatalities. In some relatively new and well-maintained databases, there might be a few pieces of information that can be related to prior-crash characteristics of vehicles, drivers, and the driving environment, including if the driver was under the influence of alcohol or drugs or if cell phone use was involved (TXDOT, 2020). However, it is generally hard to retrieve detailed information about drivers, vehicles, and driving environments before the crash occurred. Moreover, research has shown that drivers tend to omit certain prior-crash behaviors (e.g., cell phone browsing) in their reports, especially when these behaviors led to the incident (National Safety Council, 2013; Regev et al., 2017).

The data to support the analysis of drivers' behaviors is lacking, so not enough is known about the impacts of various behaviors such as performing secondary tasks – browsing a phone, texting, talking to passengers, interacting with in-vehicle devices, and etc. and vehicle maneuvers before and during a safety incident. This study utilized a near-crash data extracted from a naturalistic driving study to bridge this gap. The naturalistic driving dataset was collected through the Virginia Connected Corridor 50 Elite Vehicle Naturalistic Driving Study, and 235 near-crash events were identified (Kim et al., 2020). Twelve categorical variables were used to describe the driver's behavior, vehicle maneuver, and the conflicted vehicle's maneuver before and during near-crash events. Moreover, the variables also contain information about

* Corresponding author.
E-mail addresses: x-kong@tti.tamu.edu (X. Kong), s-das@tti.tamu.edu (S. Das), yzhang@civil.tamu.edu (Y. Zhang).

https://doi.org/10.1016/j.aap.2021.106162
Received 30 January 2021; Received in revised form 2 April 2021; Accepted 26 April 2021

locality, weather, and if the automated driving system (i.e., adaptive cruise control and the lane-keeping assistance supporting system) was active before the incident. Most importantly, the data reduction team from Virginia Tech Transportation Institute (VTTI) manually checked video recordings to report if the driver was conducting secondary tasks before the incident. Near-crash occurrences have been a reliable surrogate source for understanding the nature of crashes. Previous studies have shown that near-crashes share many of the same elements of crashes (Guo et al., 2010b; Wu et al., 2014). Therefore, the availability of these unique categorical data provides an invaluable opportunity for us to understand the patterns of near-crashes with or without the involvement of the secondary tasks.

The findings of this research can help transportation safety researchers understand the impact and patterns of secondary tasks on near-crash events. Patterns that lead to near-crash events could be affected by the involvement of secondary tasks. A combination of various factors may pose a great risk to drivers who perform secondary tasks while driving, while the same combination of factors may not affect a focused driver at all. These in-depth understandings could also help mitigate near-crash or crash events by introducing more advanced driving safety features, such as secondary task detection devices, to vehicles.

2. Literature review

Most of the existing safety studies utilized crash databases that often contain detailed information about crash characteristics that are collected after the crash occurred. However, a conventional crash database provides broader contents of driver and vehicle related information such as driver gender, driver age, and vehicle type. In many cases, these datasets do not provide detailed drivers' behaviors during the pre-crash condition, such as secondary tasks, vehicles' maneuvers, and other vehicles' movement information before and during the incident. The driver's behaviors, such as performance of a secondary task, vehicle's maneuver (e.g., going straight, turning, or decelerating), and other potential conflict vehicle's maneuver before the incident could be a crucial indicator of the incident. Several studies have tried to predict this behavior using either data collected after the crash occurred or via driving simulators. However, the results from the driving simulator data are questionable as they may or may not reflect drivers' actual behaviors prior to a real crash (Cheng et al., 2011; Wynne et al., 2019). The use of naturalistic data bypasses this step as researchers can proactively and accurately collect data with vehicle kinematics variables (e.g., speed, acceleration, yaw rate, etc.) and variables related to the driver, weather, area, etc. Naturalistic data can also provide valuable information when the crash data are inadequate due to rare crash occurrences or the construction of a new facility (Guo et al., 2010a).

In such a case, there is a need for a crash surrogate as the number of crashes with prior-crash information is low, and naturalistic data provides ample options for a surrogate choice. A wide range of surrogate measures has been explored within the current literature. Some studies have used a traffic conflict-based technique to measure safety. However, the validity and the reliability of this surrogate measure have been questionable (Guo et al., 2010a), and research has indicated that it has a weak relationship with the crash rate (Gettman et al., 2008; Gettman and Head, 2003; Zheng et al., 2014a). Nevertheless, it remains a popular method and is still used in some more recent studies (Sayed et al., 2013; So et al., 2015; Uzondu et al., 2018). Some of the other popular crash surrogates used by highway safety researchers are the time to collision (Jiang et al., 2015; Nadimi et al., 2020; Vogel, 2003), post encroachment time (Chandrapp et al., 2016; Paul and Ghosh, 2020; Tang and Kuwahara, 2011), deceleration rate (Strauss et al., 2017), extreme value theory (Songchitruksa and Tarko, 2006; Wang et al., 2018; Zheng et al., 2014b), and counterfactual approach (Bärgman et al., 2015; Davis et al., 2008). In the present study, we used near-crash data as a surrogate measure for naturalistic driving data.

Near-crashes can provide insights into driver behavior during critical crash scenarios and help differentiate key elements of successful crash avoidance maneuvers versus unsuccessful ones (Seacrist et al., 2020). In previous studies (Guo et al., 2010a; Wu et al., 2014), it was found that near-crashes share many of the same elements as crashes, and there is a strong positive relationship between the variable frequencies of crashes and near-crashes. As highlighted in Table 1, some studies have used near-crashes as a surrogate measure for naturalistic data and found many variables that significantly affect crash risk. Wu et al. (2014) found that drivers under the age of 25 are more likely to be engaged in near-crashes and crashes. A similar result was obtained by Seacrist et al., who found near-crash rates to decrease with increasing age considerably (Seacrist et al., 2020). Guo et al. (2013) found the driver's experience to be a more sensitive variable and suggested that newly licensed drivers should be monitored and limited to specific driving conditions. Some researchers found roadway characteristics such as the presence of a median (Wu and Jovanis, 2012), the presence of a vertical curve (Hamzeie et al., 2017), or the road type (Naji et al., 2018) to be influential, while others found traffic characteristics such as traffic density (Tian et al., 2013) or road congestion (Su et al., 2017) to be more significant. Similarly, as indicated in Table 1, many factors were found to influence crash-risk using near-crash as a surrogate measure.

One of the most prominent factors that affect near-crashes and crashes, as revealed by many of the published studies, is distracted

Table 1
Studies investigating crash risk and factors affecting it using near-crash as a surrogate.

Factor affecting crash risk	Database	Papers
Vehicle kinematics	100-Car NDS	(Taccari et al., 2018; Wu and Jovanis, 2012)
	SHRP2	(Hamzeie et al., 2017; Osman et al., 2019; Perez et al., 2017)
	Other	(Naji et al., 2018; Perez et al., 2017; Su et al., 2017; Wang et al., 2015)
Roadway characteristics	100-Car NDS	(Wu and Jovanis, 2012)
	SHRP2	(Hamzeie et al., 2017)
	Other	(Naji et al., 2018; de Rome et al., 2018)
Traffic characteristics	100-Car NDS	(Jovanis et al., 2011; Tian et al., 2013)
	Other	(Su et al., 2017)
Environmental/ lighting conditions	100-Car NDS	(Jovanis et al., 2011; Klauer et al., 2006; Wu and Jovanis, 2012)
	Other	(Su et al., 2017)
Alcohol/Drug Use	SHRP2	(Arvin and Khattak, 2020)
	Other	(Beck et al., 2019; Ogeil et al., 2018)
Driver Age / Driver experience	100-Car NDS	(Jovanis et al., 2011; Wu et al., 2014)
	SHRP2	(Hamzeie et al., 2017; Seacrist et al., 2020, 2018)
	Other	(Guo et al., 2013; Lee et al., 2011; Naji et al., 2018)
Driver behavior / characteristics	100-Car NDS	(Guo and Fang, 2013)
	SHRP2	(Huisingh et al., 2017; Markkula et al., 2016)
	Other	(Ashouri et al., 2018; Cheng et al., 2011)
Distracted driver / Secondary task	100-Cas NDS	(Klauer et al., 2014, 2006; Liang et al., 2014, p., 2012; Tian et al., 2013)
	SHRP2	(Arvin and Khattak, 2020; Bakhit et al., 2018; Bálint et al., 2020; Dingus et al., 2016; Huisingh et al., 2019; Ye et al., 2017)
	Other	(Bakiri et al., 2013; Cunningham and Regan, 2018; Ersal et al., 2010; Esfahani et al., 2019; Farmer et al., 2015; Klauer et al., 2015; Miller et al., 2015; Oviedo-Trespalacios et al., 2017; Simons-Morton et al., 2014; Tivesten and Dozza, 2014; Wandtner et al., 2018)

Note: NDS stands for Naturalistic Driving Study.

driving. Dingus et al. (2006) found distracted driving to be a factor in approximately 80 % and 65 % of all crashes and near-crashes, respectively. The National Highway Traffic Safety Administration (NHTSA) also estimated that nearly 23,000 deaths in the U.S. from 2012 to 2018 occurred because of distracted driving (NHTSA, 2019). A major source of distraction for drivers is engagement in secondary tasks (Regan et al., 2008; Stutts et al., 2001). Ranney estimated that an average driver is involved in performing secondary tasks for almost 30 % of their driving time (Ranney, 2008). Involvement in secondary tasks like cell phone holding/dialing (Bálint et al., 2020; Klauer et al., 2014), texting (Arvin and Khattak, 2020; Bakhit et al., 2018; Bálint et al., 2020), reaching for an object (Arvin and Khattak, 2020; Bakhit et al., 2018; Bálint et al., 2020), or manipulating objects (Bakhit et al., 2018) are shown to reduce driver's attentiveness to the road and subsequently increase crash-risk. The present study utilizes the database from Virginia Connected Corridor 50 Elite Vehicle Naturalistic Driving Study or VCC50 Elite NDS (Kim et al., 2020), which covers all of the above factors.

Since naturalistic driving data can consist of a large plethora of information with many variables that influence crash risk, researchers are increasingly adopting various Machine Learning (ML) techniques because of their ability to handle complex multi-dimensional data, high training and testing accuracy, and low prediction time. Ersal et al. (2010) and Ye et al. (2017) deployed Artificial Neural Networks (ANNs) for analyzing driver performance while performing secondary tasks. Wang et al. (2015) used classification and regression tree (CART) to establish a relationship between crash-risk and road, vehicle, and driver characteristics. Using Logistic regression on survey data, Oviedo-Trespalacios et al. (2017) inspected the influence of phone usage on road safety. Similarly, Mousa et al. (2019) and Osman et al. (2019) analyzed naturalistic data using Extreme Gradient Boosting (XGBoost) and AdaBoost, respectively, to identify various factors contributing to near-crashes.

Even with the increasingly accessible naturalistic driving data, datasets that provide detailed information about drivers, vehicles, and other potential conflict vehicles before and during the near-crash incidents are still rare. This study utilized a valuable near-crash dataset extracted from a naturalistic driving study and recorded videos provided by VTTI. This dataset contains rich information about the driver, vehicle maneuver, and other vehicles' maneuver before and during the near-crash incidents. To mine associations among these categorical features regarding the involvement of secondary tasks, the Aprorio algorithm has been adopted for this research.

This paper is presented in the following manner. First, the paper provides the introduction to identify the research gap and elaborate on the uniqueness of the research. Then, the literature review section describes previous studies that explain the necessity of studying near-crash events and secondary tasks, the utilization and popularity of naturalistic driving data, the findings of previous near-crash studies, and the uniqueness of the used dataset. The following methodology section discusses the dataset and association rules. Then, the results of mined association rules are presented and explained. After this section, the summary of findings and comparison between two sets of rules are given to provide more insights. Finally, the conclusion and limitations are discussed.

3. Methodology

The purpose of this research is to mine patterns among near-crash incidents. To detect the associations between near-crash incidents with or without the involvement of the secondary tasks and available categorical variables, association rule mining is an appropriate candidate for data mining. As the study is designed for an unsupervised learning framework, apriori association rule mining method serves the purpose well. The purpose of this study is to mine the patterns among the near-crash incidents with or without secondary tasks involvement by exploring a categorical dataset. Apriori algorithm is a great candidate for finding association among categorical variables. The other reasons are the applicability and interpretability. This study initially examined other rules mining algorithms such as 'Eclat' algorithm. As apriori provides different performance measures such as support, confidence, and lift, this study considered apriori as the suitable approach. Rules mining method has some advantages compared to statistical or machine learning model. This method can provide evidence of higher probability scores of some hidden patterns with a combination of variable categories, which is difficult to attain in a statistical model. The rules mining method is not a black box method like many machine learning algorithms. Thus, association rule mining is an appropriate candidate for data mining to detect the associations between near-crash incidents with or without the involvement of the secondary tasks and available categorical variables. Agrawal and Srikant proposed the first association rule mining algorithm in 1994, known as the *Apriori algorithm*. In recent years, this method has been used in many research fields, including transportation research. For example, Xu and Luo (2020) utilized this method for risk prediction and early warning of air traffic controllers' unsafe acts in 2020. Kong et al. (2020) also used this algorithm to mine the associations between speeding behavior and roadways' geometric features (Kong et al., 2020), and Das et al. (2019) explored association rules among the hit and run crashes.

It is important to understand the term "transaction" in order to understand the *Apriori algorithm.* A transaction is considered an event, including a set of items. For example, for a shopping event, all the shopping items on the shopping receipt are considered a set of items consisting of a transaction. For crash data analysis, all variables associated with a crash can be considered as a set of items, and one crash can be considered as one transaction. This algorithm can detect patterns if the categorical variables frequently occur together under certain pre-set thresholds.

The categories of each near-crash event are considered as a set of items: $I=\{i_1, i_2, \dots i_m\}$. The transaction dataset is $T = \{t_1, t_2, \dots, t_m\}$, where each t_i consists of a subset of categories from I. The other two terms to understand the association rule are Antecedent and Consequent, or X and Y. Mined association rule can be written as $X \rightarrow Y$, Where X and Y are disjointed categories from item sets. Many combinations of rules can be mined through the categorical dataset. Fortunately, not every one of them is meaningful or remarkable. Three critical measurements are used for evaluating the performance and filtering out meaningful association rules. The measurement metrics are *Support*, *Confident*, and *Lift*. The *Support* in this study represents the percentage of transactions containing the combination of X and Y in all transactions. In other words, *support* says how often these X and Y appear together. N in the following equation equals the total number of transactions.

$$S(X \rightarrow Y) = \frac{count(X \cup Y)}{N}$$

Confidence means the percentage of transactions containing $X \rightarrow Y$ in all transactions containing X.

$$C(X \rightarrow Y) = \frac{S(X \cup Y)}{S(X)}$$

The critical performance measure is Lift. Lift describes the independence between X and Y. A lift value higher than 1 indicates the strong dependence among the antecedent and consequent. Otherwise, a small life value demonstrates the weak association between them. While during the rules mining process, the definition of a 2-item rule is that an association rule contains a 1-item antecedent and 1-item consequent. Similarly, a 3-item rule is a combination of a 2-item antecedent and a 1-item consequent. In this study, the consequent is always the {secondary task = yes} for near-crash events with the involvement of secondary tasks dataset and {secondary task = no} for the near-crash event without the involvement of secondary tasks dataset.

3.1. Dataset overview

The naturalistic driving dataset is collected through the Virginia Connected Corridor 50 Elite Vehicle Naturalistic Driving Study (Kim et al., 2020). The original VCC50 Elite dataset was collected from 50 connected vehicles with automated driving features, such as adaptive cruise control and lane-keeping assistance. The recruited drivers came from and commuted in the Washington, DC metro area for 12 months. The age of these drivers is from 24 to 76. In 12 months of the data collection period, 684, 931 miles were driven by all participants. The NDS equipment was able to collect data from driver, vehicle, and environmental factor perspectives. The kinematic thresholds of defining the near-crash event adopted by the data reduction team from VTTI, that conducted multiple NDSs studies (Dingus et al., 2016; Scofield, 2015; Simons-Morton et al., 2014). Broadly, a near-crash event is defined as a non-crash incident with an evasive braking maneuver reaching -0.3 g if the object being avoided is extremely close. For light 4-wheeled vehicles, which is the vehicle type of the recruited drivers, the criteria are a less-than-2-second “Time to Collision” measurement (Scofield, 2015). Elevated g-force events were also assessed, including longitudinal deceleration/hard braking (<-0.45 g), hard left (<-0.50 g) and hard right turn (>0.50 g), and yaw rate higher than 6 degrees in 3 s (Simons-Morton et al., 2014). To accurately identify these near-crash events, using the criteria mentioned above and kinematic data to identify the potential near-crash events is the first step. After identifying these potential near-crash events, trained data reductionists inspected all video recordings (30-second epoch) to visually verify the occurrence of the events and also collected related information, such as secondary tasks, evasive behaviors, reaction time, etc. (Kim et al., 2020a, 2020b). These behavioral-related variables, such as maneuvers, secondary tasks, reaction times, and driving environment-related variables, such as an object on the road, other vehicle-turning, are documented by the trained data reductionists based on the most recent published coding protocols and data dictionaries (Russell et al., 2018; Scofield, 2015).There are 235 safety-critical events – near-crash incidents extracted from the whole NDS dataset and manually coded by the VTTI data reduction team (Kim et al., 2020). VTTI researchers code a series of categorical variables about driver, vehicles, and the driving environment to describe the characteristics of the near-crash events.

3.2. Used variables

There are 12 categorical variables in the dataset to describe all near-crash events (see Table 2). The raw data published by VTTI contains several additional variables that were removed during analysis due to heavily skewed distribution or with a large percentage of missing values. For example, the data contains a variable to describe the road surface conditions (dry or wet). 216 out of 235 events were on the roadways with dry surface conditions. There are also two variables describing the multiple secondary tasks performed by the drivers at the same time. These variables are described as secondary task 1, secondary task 2. There are only about 10 out of 235 drivers who performed secondary task 2. To avoid generating misleading results, the secondary task 2 variable was removed. These variables can be categorized into vehicle-related driver-related, driving-environment related variables. Vehicles-related variables are related to the maneuver of the vehicle before and during the near-crash events and the automated driving system’s status, like if adaptive cruise control or lane-keeping assistance was active when the near-crash event happened, the maneuver of the vehicle before and during the event. Driver-related variables are driver’s behavior before the near-crash events, such as distracted, reaction time to the emergency, or if the driver was performing a secondary task if the driver had both hands on the wheel, and the driver’s reaction time. Secondary task is the key variable for this study. This secondary task variable describes the drivers’ undertaking tasks, except driving, before the near-crash occurred. These tasks include cell phone browsing, calling, texting, hand-hell talking, talk to passengers, interact with pet, eat, drink or smoke, operate in-vehicle devices, etc. The driving environment-related variables are other vehicles’ maneuvers before the near-crash, locality, and weather. Several variables are describing the nature of the near-crash event, such as incident type.

Table 2
Variable description.

Variable	Attribute	Description
Premaneuv	changelane	changing lanes
	curve	negotiating a curve
	start_stop	starting/stopped in traffic lane
	straight_acc	going straight, accelerating
	straight_cons	going straight, constant speed
	straight_dec	going straight, decelerating
	turning	turning right/left
Preevent	objinroad	object in roadway
	otherveh_langchange	other vehicle lane change
	otherveh_slow/stop	other vehicle slowed and stopped
	otherveh_turning	other vehicle turning
	subject_lanchange	subject lane change
	subject_turning	subject turning
Eventnature	conflict_adjacent	conflict with a vehicle in an adjacent lane
	conflict_following	conflict with the following vehicle
	conflict_lead	conflict with a lead vehicle
	conflict_obj	conflict with an object
	conflict_turning	conflict with a turning vehicle
	conflict_unknown	unknow conflict
Incidenttype	others	others
	rear-end	rear-end, struck
	sideswipe	sideswipe
Evasivemaneuv	brakedandsteered	braked and steered
	brakedonly	braked only
	noreaction	no reaction
	steered	steered
Driverbehav	distracted	distracted
	improperdriving	cutting in, improper lane change, aggressive, right of way error
	none	none
Handsonwheel	both	both hands on the wheel
	none	no hands on the wheel
	onehand	one hand on the wheel
Weather	clear/partly cloudy	clear, partly cloudy weather
	overcast/raining	overcast, raining weather
Locality	business/industrial	business, or industrial areas
	interstate/divided hwy	interstate or divided highway
	residential	residential areas
ACC_LKA	no	adaptive cruise control and lane-keeping assistance not active
	yes	adaptive cruise control or lane-keeping assistance, or both active
Reacttime	shorter than 0.5 s	0−0.5 seconds from conflict started to start to react
	0.5−1 sec	0.5−1 second from conflict started to start to react
	longer than 0.5 s	longer than 1 s from conflict started to start to react
Secondary task	no	no secondary task was observed before the near-crash event occurred
	yes	Some kind of secondary task was observed before the near-crash occurred. Secondary tasks include cell phone browsing, calling, texting, hand-held talking, talk to passengers, interact with pet, eat, drinking or smoking, etc.

3.3. Distribution of key attributes

Note that rules mining is very effective in identifying trends from a large dataset. However, rules mining has been widely used based on the complexity of the data regardless the sample size. For example, our data contains 12 variables with several categories in each variable. The proportion distribution of all of these categories for two separate datasets makes the data worthy for investigation by using association rules mining.

Table 3 describes the data in two categories: *secondarytask* = *yes* and *secondarytask* = *no*. There are 235 near-crash events in this dataset. One hundred and thirty near-crash events involve performing secondary tasks. One hundred and five near-crash events are without the involvement of any secondary tasks. For each selected variable, the count across different levels of each category are not evenly distributed. For example, the variable "premanuev" describes the vehicle maneuver behavior before a near-crash event. For the near-crash event with the involvement of secondary tasks, the majority of vehicles in this category were going straight with constant speed. In another category, the near-crash without the involvement of secondary tasks, vehicles may go straight with constant speed or perform lane changes.

Table 3
Distribution of Key Attributes.

Variable	Atrribute	Secondarytask = yes		Secondarytask = no	
		count	%	count	%
Premaneuv	changelane	12	11	35	27
	curve	11	10	12	9
	start_stop	11	10	7	5
	straight_acc	9	9	12	9
	straight_cons	37	35	39	30
	straight_dec	17	16	13	10
	turning	8	8	12	9
preevent	objinroad	7	7	8	6
	otherveh_langchange	31	30	65	50
	otherveh_slow/stop	40	38	27	21
	otherveh_turning	5	5	8	6
	subject_lanchange	15	14	18	14
	subject_turning	7	7	4	3
Eventnature	conflict_adjacent	39	37	78	60
	conflict_following	6	6	4	3
	conflict_lead	40	38	36	28
	conflict_obj	5	5	8	6
	conflict_turning	10	10	12	9
	conflict_unknown	5	5	2	2
Incidenttype	others	8	8	10	8
	rear-end	79	75	82	63
	sideswipe	18	17	38	29
Evasivemaneuv	brakedandsteered	31	30	49	38
	brakedonly	61	58	63	48
	noreaction	7	7	4	3
	streered	6	6	14	11
Driverbehav	distracted	36	34	11	8
	improperdriving	40	38	64	49
	none	29	28	55	42
Handsonwheel	both	36	34	50	38
	none	9	9	1	1
	onehand	60	57	79	61
Weather	clear/partly cloudy	72	69	32	25
	overcast/raining	33	31	98	75
Locality	business/industrial	33	31	34	26
	interstate/divided hwy	59	56	83	64
	residential	13	12	13	10
ACC_LKA	no	85	81	95	73
	yes	20	19	35	27
Reacttime	shorter than 0.5sec	38	36	73	56
	0.5-1sec	40	38	19	15
	longer than 0.5sec	27	26	38	29

4. Results

To explore the patterns for near-crash events with and without performing secondary tasks, the Apriori association rules algorithm has been applied to the near-crash dataset with supervised right-hand side item – "secondary task = yes" or "secondary task = no." Moreover, this study mines the patterns with a different number of items, which are 2-item, 3-item, and 4–5-item. In these mined rules of near-crash events with secondary tasks, there is one fixed item – "secondary task = yes." For the mined rules of near-crash events without secondary tasks, the fixed item is" secondary task = no." Investigating the patterns through multiple iterations based on the total number of items in the mined rules could provide more insights than mining rules with only one iteration, which mines all rules with various items together.

4.1. Summary of parameter setting and parameters in outcomes

Table 4 introduces the chosen parameter values for the Apriori algorithm utilized in this study. The parameter values for mining 2-item rules and 3-item rules are the same. The minimum support threshold is set as 0.05, and the minimum confidence value is set as 0.1. For 4–5-item rules, the support value threshold is 0.1, and the value of confidence keeps the same. Many rounds of testing and evaluation were done before defining final thresholds. The criteria for choosing proper thresholds for the parameters is to mine reasonable rules and ensure the stability of the model performance. This process may suffer a certain level of subjectivity. Thus, a thorough understanding of the purpose of these parameters is necessary for finding reasonable thresholds. First, the minimum support value should an optimized value to ensure the pattern (an itemset) presented in a reasonable number of observations. For example, for 2-item rules, 0.05 minimum support indicates, in this analysis, an itemset only be considered as a candidate itemset if it shows in at least 0.05*235 ≅ 12 out of 235 observations. 12 out of 235 may appear to be a small amount. However, an association rule containing this itemset would be worthy of reporting if the lift value of this rule is high. A high lift value suggests the strong dependency between this itemset and Consequent (secondary = yes or no, in this case). This is also one of the advantages of using this method to mine the meaningful patterns through a categorical dataset with some less frequent but important items. For example, an itemset *{Eventnature = conflict_lead + Reacttime = longer than* 1 s*}* may not be a frequent itemset and has a minimum support value 0.01. However, this rule is worth reporting if its lift value is high for {secondary take = yes}., which indicates that this itemset is highly associated with secondary tasks. In plain words, it may less often to observe the conflict with a leading vehicle and take a relatively long reaction time. However, this could often occur if the driver was performing secondary tasks and being distracted. With the minority presentation in the dataset, traditional statistical methods, such as logistic regression, may not be capable of identifying this pattern.

Table 5 shows a summary of the values of the evaluation parameters of all mined association rules. The number of rules and mean, min., and max. The support, confidence, and lift values are reported. There are 31 and 27 *2-item rules* mined for near-crash events without and with secondary tasks, separately. There are 189 and 164 *3-item rules* mined for near-crash events without and with secondary tasks. For 4–5-item rules from the near-crash events without and with secondary tasks, the number are 184 and 71. The mean support value of all rules is above 0.1. The mean confidence value of all rules is higher than 0.45, and the mean of the lift value is above 1.02. The maximum value of support, confidence, and lift among all rules are 0.42, 0.92, and 2.07.

Table 4
Parameters values selected for rules mining.

Rule Types	Min. Support	Min. Confidence	Minlen	Maxlen
2-item rules	0.05	0.1	2	2
3-item rules	0.05	0.1	3	3
4–5-item rules	0.1	0.1	4	5

4.2. Performance evaluation

Table 5 lists the number of rules for each of the datasets. It is nearly impossible to explain each and every rule. To get an overall understanding of the number of rules and associated performance matrices, two-key plots are excellent data visualization tools (see Fig. 1 and Fig. 2). The x-axis is the value of support of mined rules, and the y-axis is the value of confidence of mined rules. Fig. 1 indicates the distribution of these parameter values of mined rules for near-crash events with secondary tasks. Fig. 1 shows that the two-key plot of 2-item rules shows the range of support value of all 2-item rules is from 0.05 to 0.4, and the majority of these rules have more than 40 percent confidence. For the two-key plot of the 3-item rules, the range of support value is from 0.05 to 0.3, and the majority of these rules have more than 40 percent confidence. A similar trend is seen for 4–5-item rules. The 4-item rules are in purple color, and the 5-item rules are in red color.

Fig. 2 shows the two-key plots for rules of near-crash events without secondary tasks. The ranges of support value across 2-item, 3-item, and 4-item groups are 0.05 – 0.4, 0.05 – 0.3, and 0.05–0.25. The ranges of the support value are similar to the above Fig. 1. However, the confidence value for the majority of rules of near-crash without secondary tasks is higher than 50 percent, which is generally higher than rules of near-crash events with secondary tasks.

4.3. Interpretation of rules with high lift measures

As shown in Table 5, there are 58 2-item association rules, 353 3-item association rules, and 255 4–5-item rules mined from the near-crash dataset. It is impossible and unnecessary to interpret all of these mined rules. All of these rules satisfied the support and confidence thresholds. The ones with the highest lift value are the ones most interesting. Tables 6 and 7 contain 26 association rules for near-crash events with secondary tasks and near-crash events without secondary tasks. In these 26 rules, 6 rules from 2-item rules, 10 rules from 3-item rules, and 10 rules from 4–5-item rules are selected. These rules are selected based on their high lift values and usefulness for identifying meaningful patterns. Although the Apriori is an unsupervised rule mining algorithm, the function still provides an alternative to fix the right-hand side (RHS) item. Thus, the rule mining process becomes less random and more supervised because the fixed RHS item ensures that all mined rules are associated with or contained in this item. The rest of the rules (without this designated RHS item, which is mentioned in the subsection heading and rule descriptions) are described below.

4.3.1. Patterns for Near-Crash events with secondary tasks

4.3.1.1. 2-item rules. The first six rules in Table 6 are 2-item rules. One item is the fixed RHS item – *{secondary task = yes}*. The first rule with the highest lift value is {driverhehav = distracted, secondary task = yes}. It is intuitive to understand this rule. Most secondary tasks would distract drivers' attention rather than focus on driving. The corresponding evaluation indices are support = 15 %, confidence = 77 %, and lift = 1.71. To understand these indices, a detailed explanation can be found in the methodology section. Here is a brief interpretation. The value of support indicates 15 % near-crash events has these two items *{driverhehav = distracted}* and *{secondary task = yes}*. The value of confidence means that out of all near-crash events has {driverhehav = distracted}, 77 % of them were performing secondary tasks. The value of lift is 1.71, which says two items in this rule are highly dependent. In other words, the distracted behavior of a driver is highly associated with performing secondary tasks. As mentioned earlier, when the lift value higher than 1,

Table 5
Summary of the parameter values of mined rules.

Consequent (Secondary Task)	# items	# rules	Support			Confidence			Lift		
			Mean	Min.	Max.	Mean	Min.	Max.	Mean	Min.	Max.
No	2	31	0.19	0.05	0.42	0.56	0.39	0.74	1.02	0.71	1.35
Yes	2	27	0.16	0.05	0.36	0.45	0.26	0.77	1.02	0.57	1.71
No	3	189	0.12	0.05	0.30	0.60	0.32	0.86	1.09	0.58	1.55
Yes	3	164	0.10	0.05	0.27	0.50	0.22	0.92	1.11	0.48	2.07
No	4–5	184	0.13	0.10	0.22	0.67	0.38	0.86	1.21	0.69	1.55
Yes	4–5	71	0.12	0.10	0.20	0.61	0.32	0.92	1.38	0.72	2.07

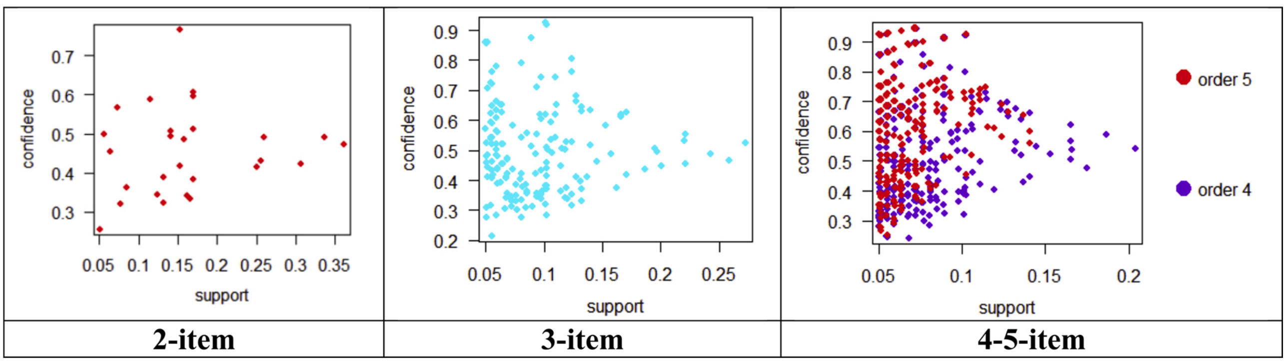

Fig. 1. Two-key plots for 'secondary task = yes'.

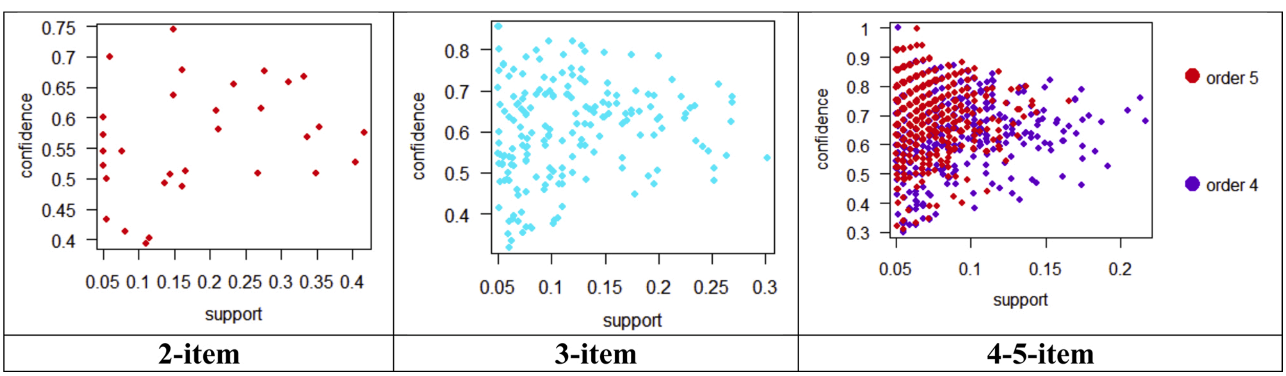

Fig. 2. Two-key plots for 'secondary task = no'.

it indicates a strong or high association among the items. Otherwise, it indicates a weak association between the items in the rule.

Other rules in this 2-item rule set also demonstrate interesting associations. The second rule

{Eventnature = conflict_lead, secondary task = yes} indicates that near-crash event with secondary tasks is highly associated with conflicting with the leading vehicle. Third rule *{Preevent = otherveh_slow/stop, secondary task = yes}* says this conflict often occurs with the other vehicle suddenly slowed or stopped. For drivers who perform secondary tasks while driving, it is very likely that they conduct secondary tasks when the driving environment is predictable, from their perspectives. Therefore, these unexpected sudden slow down or stop behaviors of the leading vehicle would trigger them to hard break and being involved in these crash or near-crash events. The fourth rule *{Reacttime = longer than 1 s, secondary task = yes}* and the sixth rule *{Reacttime = 0.5−1 sec, secondary task = yes}* also associate the near-crash with secondary tasks with a relatively longer reaction time to the emergency(considering 3 reaction categories in the dataset: shorter than 0.5 s, 0.5−1.0 s and long than 1 s). The fifth rule *{Premaneuv = straight_dec, secondary task = yes}* associates the near-crash event with a secondary task with the maneuver (going straight and deceleration) of the vehicle before the near-crash event. The high association of this combination of a vehicle going straight and decelerating, a driver performing a secondary task, and a near-crash event indicates that a common near-crash scenario is insufficient breaking behavior caused by performing the secondary task and poor judgment on the driving environment.

4.3.1.2. 3-item rules. Rules 7–16 are ten 3-item rules. A 3-item rule contains two items in the antecedent column and a fixed RHS item – *{secondary task = yes}*. The rule with the highest lift value is rule 7 –

{Preevent = otherveh_slow/stop, Driverbehav = distracted, secondary task = yes}. The corresponding indices are support = 10 %, confidence = 90 %, and lift = 2.07. Out of the near-crash events in the whole dataset contain the antecedent item *{Preevent = otherveh_slow/stop, Driverbehav = distracted}*, 92 percent of them performed the secondary tasks. The association of these three items is strong. The near-crash often occurs when the leading vehicle suddenly slowed or stopped while the driver is distracted by the secondary tasks. For near-crashes with secondary tasks, the rules frequently show the drivers' evasive maneuver was braking only. This pattern presents in rules 8, 12, 13, and 14. The dominant evasive response of these drivers performing the secondary tasks is to brake only to avoid a collision. Performing secondary tasks often

Table 6
Selected rules for near-crash events with secondary tasks.

	Antecedent	S	C	L
No.	**2-item rules ("supervised" item: secondary task = yes)**			
1	Driverbehav = distracted	0.15	0.77	1.71
2	Eventnature = conflict_lead	0.17	0.61	1.36
3	Preevent = otherveh_slow/stop	0.17	0.60	1.34
4	Reacttime = longer than 1 s	0.11	0.59	1.31
5	Premaneuv = straight_dec	0.07	0.57	1.27
6	Reacttime = 0.5−1 sec	0.17	0.51	1.15
	3-item rules ("supervised" item: secondary task = yes)			
7	Preevent = otherveh_slow/stop + Driverbehav = distracted	0.10	0.92	2.07
8	Evasivemaneuv = brakedonly + Driverbehav = distracted	0.09	0.88	1.96
9	Eventnature = conflict_lead + Reacttime = longer than 1 s	0.05	0.86	1.92
10	Incidenttype = rear-end + Reacttime = longer than 1 s	0.10	0.74	1.66
11	Hansonwheel = both + Reacttime = longer than 1 s	0.06	0.72	1.62
12	Eventnature = conflict_lead + Evasivemaneuv = brakedonly	0.13	0.68	1.53
13	Preevent = otherveh_slow/stop + Evasivemaneuv = brakedonly	0.13	0.67	1.49
14	Evasivemaneuv = brakedonly + Reacttime = longer than 1 s	0.06	0.65	1.46
15	Premaneuv = straight_cons + Preevent = otherveh_slow/stop	0.06	0.63	1.40
16	Locality = interstate/divided hwy + Reacttime = longer than 1 s	0.07	0.59	1.31
	4−5-item rules ("supervised" item: secondary task = yes)			
17	Preevent = otherveh_slow/stop + Eventnature = conflict_lead + Driverbehav = distracted	0.10	0.92	2.07
18	Incidenttype = rear-end + Driverbehav = distracted + ACC_LKA = no	0.10	0.80	1.79
19	Preevent = otherveh_slow/stop + Eventnature = conflict_lead + Evasivemaneuv = brakedonly + ACC_LKA = no	0.11	0.75	1.68
20	Preevent = otherveh_slow/stop + Incidenttype = rear-end + Evasivemaneuv = brakedonly + Weather = clear/partly cloudy	0.10	0.73	1.63
21	Preevent = otherveh_slow/stop + Eventnature = conflict_lead + Incidenttype = rear-end + Hansonwheel = onehand	0.11	0.66	1.47
22	Incidenttype = rear-end + Locality = business/industrial + ACC_LKA = no	0.11	0.63	1.40
23	Preevent = otherveh_slow/stop + Eventnature = conflict_lead + Incidenttype = rear-end	0.17	0.62	1.39
24	Incidenttype = rear-end + Evasivemaneuv = brakedonly + Hansonwheel = onehand + ACC_LKA = no	0.12	0.61	1.36
25	Incidenttype = rear-end + Evasivemaneuv = brakedonly + Locality = interstate/divided hwy + ACC_LKA = no	0.11	0.52	1.17
26	Evasivemaneuv = brakedonly + Locality = interstate/divided hwy + ACC_LKA = no	0.11	0.50	1.12

* S-support, C-confidence, L-lift.

Table 7
Selected rules for near-crash events without secondary tasks.

	Antecedent	S	C	L
	2-item rules ("supervised" item: secondary task = yes)			
1	Premaneuv = changelane	0.15	0.74	1.35
2	Evasivemaneuv = steered	0.06	0.70	1.27
3	Incidenttype = sideswipe	0.16	0.68	1.23
4	Preevent = otherveh_lanechange	0.28	0.68	1.22
5	Eventnature = conflict_adjacent	0.33	0.67	1.21
6	Reacttime = shorter than 0.5 s	0.31	0.66	1.19
	3-item rules ("supervised" item: secondary task = yes)			
7	Eventnature = conflict_adjacent + Evasivemaneuv = steered	0.05	0.86	1.55
8	Premaneuv = changelane + Preevent = otherveh_lanechange	0.12	0.82	1.49
9	Premaneuv = changelane + Driverbehav = improperdriving	0.10	0.82	1.48
10	Eventnature = conflict_adjacent + Reacttime = shorter than 0.5 s	0.20	0.78	1.42
11	Hansonwheel = both + Reacttime = shorter than 0.5 s	0.13	0.75	1.36
12	Evasivemaneuv = brakedandsteered + Reacttime = shorter than 0.5 s	0.11	0.69	1.25
13	ACC_LKA = yes + Reacttime = shorter than 0.5 s	0.09	0.69	1.25
14	ACC_LKA = no + Reacttime = shorter than 0.5 s	0.23	0.65	1.17
15	Weather = overcast/raining + Reacttime = shorter than 0.5 s	0.08	0.64	1.16
16	Driverbehav = improperdriving + Reacttime = shorter than 0.5 s	0.15	0.64	1.16
	4−5-item rules ("supervised" item: secondary task = yes)			
17	Eventnature = conflict_adjacent + Locality = interstate/divided hwy + ACC_LKA = no + Reacttime = shorter than 0.5 s	0.10	0.86	1.55
18	Eventnature = conflict_adjacent + Weather = clear/partly cloudy + Locality = interstate/divided hwy + Reacttime = shorter than 0.5 s	0.12	0.85	1.53
19	Premaneuv = changelane + Eventnature = conflict_adjacent + Locality = interstate/divided hwy	0.11	0.84	1.53
20	Hansonwheel = both + Weather = clear/partly cloudy + Reacttime = shorter than 0.5 s	0.10	0.83	1.50
21	Eventnature = conflict_adjacent + Driverbehav = improperdriving + Hansonwheel = onehand	0.10	0.83	1.50
22	Preevent = otherveh_lanechange + Weather = clear/partly cloudy + Locality = interstate/divided hwy + Reacttime = shorter than 0.5 s	0.10	0.83	1.50
23	Premaneuv = changelane + Preevent = otherveh_lanechange + Eventnature = conflict_adjacent	0.11	0.82	1.48
24	Driverbehav = none + Weather = clear/partly cloudy + Reacttime = shorter than 0.5 s	0.11	0.82	1.48
25	Driverbehav = none + ACC_LKA = no + Reacttime = shorter than 0.5 s	0.11	0.81	1.47
26	Preevent = otherveh_lanechange + Eventnature = conflict_adjacent + Locality = interstate/divided hwy + Reacttime = shorter than 0.5 s	0.14	0.80	1.45

* S-support, C-confidence, L-lift.

distracts drivers, and distracted drivers have relatively longer reaction times (in rules 9, 10, 11, 14, and 16). Longer reaction time leads to less time for the driver to properly respond, like brake and steer maneuver. This could explain why the *{Evasivemaneuv = brakedonly}* dominantly presented in many rules. Rule 16 shows the *{Locality = interstate / divided hwy, Reacttime = longer than* 1 s*}* is strongly associated with the secondary tasks. This indicates that the near-crash events with secondary tasks occurring on the interstate/divided highway often take longer than 1 s reaction time. This combination suggests that drivers on the highway who perform the secondary tasks may not be cautious enough of the potential risks. The traffic status on the highway is often stable and predictable in a short time period, and a longer reaction time is required when the near-crash events occurred.

4.3.1.3. 4–5-item rules. Rules 17–26 are rules with 4 or 5 items. These ten rules show the patterns discussed above more clearly. The combination of different items shows strong association with near-crash events with secondary tasks. For example, rule 17 *{Preevent = otherveh_slow / stop, Eventnature = conflict_lead, Driverbehav = distracted, secondary task = yes}* shows that the near-crash event with a secondary task occurs when the other leading vehicle suddenly slowed or stopped and drivers were distracted, and the event nature is often classified as conflict with leading vehicle (rules 17, 19, 21, 23). Other rules also demonstrate that near-crash with secondary tasks are highly associated with rear-end incident type (rules 18, 20–25). Most near-crashes with leading vehicles are classified as rear-end incidents. Another interesting rule is the rule 21 *{Preevent = otherveh_slow / stop + Eventnature = conflict_lead + Incidenttype = rear-end + Handsonwheel = onehand, secondary task =*

yes}. This rule indicates that the hands-on-wheel status is one hand for these near-crash events with secondary tasks. A majority of secondary tasks require the involvement with at least one hand. This could introduce the extra risk of less control of the vehicles, especially when the reaction time becomes longer for these drivers distracted by performing the secondary tasks. The complex combination of items shows these safety-critical events like near-crashes are often a combination of many factors, such as driving on a highway with a presumably stable and predictable traffic condition, performing secondary tasks, being distracted, one hand on the wheel, leading vehicle suddenly slowed or stopped.

4.3.2. Patterns for Near-Crash events without secondary tasks

4.3.2.1. 2-item rules. The first six rules in Table 7 are 2-item rules with high lift values, which state the strong association between the antecedent and fixed RHS item *{secondary task = no}*. The first rule is *{Premaneuv = changelane, secondary task = no}*. Its evaluation indices are support = 15 %, confidence = 74 %, and lift = 1.35. These indices mean that 15 percent of near-crash events in the data contains these two items. Out of these events containing lane change behavior item, 74 percent of them are near-crash events without any secondary task involvement. The lane change behavior is highly associated with the near-crash events without secondary tasks. Rule 4 also points out that other vehicle lane change behavior is highly associated with this type of near-crash event. Since this type of near-crash dominantly associated with lane change behaviors the participant vehicles or other vehicles, it is expected to find out this type of near-crash also highly associated with the sideswipe incident type (rule 3 *{Incidenttype = sideswipe}*) and event nature as conflict with adjacent vehicle (rule 5 *{Eventnature = conflict_adjacent}*). Another interesting rule in this 2-item rule set is the sixth rule *{Reacttime = shorter than* 0.5 s*, secondary task = no}*. The shortest reaction time category, less than 0.5 s, is highly associated with the near-crash events without secondary tasks. Without the distraction from the secondary tasks, the reaction to the emergency events tends to be quicker.

4.3.2.2. 3-item rules. Rules 7–16 are ten 3-item rules with the highest lift values of the near-crash event without secondary tasks. Rule 7 is *{Eventnature = conflict_adjacent, Evasivemaneuv = steered, secondary task = no}*. The evaluation indices are support = 5%, confidence = 86 %, and lift = 1.55. Out of 235 near-crash events, there are 5 percent of them contain these three items. This amount is not particularly surprising. However, in all events containing the antecedent items *{Eventnature = conflict_adjacent, Evasivemaneuv = steered}*, 86 percent did not perform any secondary tasks. A higher than 1 lift value indicates the high associations between the antecedent and the near-crash event without a secondary task. To interpret it in plain words, the drivers near-crashes that conflict with adjacent vehicles, and the driver were able to steer the wheel to avoid collision are mostly likely without performing any secondary tasks. Rule 8 explains another common near-crash scenario - both participant vehicles and adjacent vehicles change lanes at the same time period. This scenario is highly associated with no secondary tasks. The lane change behavior generally requires more attention from drivers. Thus, it is rare to find that lane-changing drivers still perform secondary tasks.

Rule 11 *{Hansonwheel = both, Reacttime = shorter than* 0.5 s*, secondary task = no}* indicates for drivers without performing secondary tasks often have both hands on the wheel, and the reaction time to the emergency often fast, less than 0.5 s. Rules 13 and 14 states that with or without an automated driving system (ACC_LKA) active, the drivers who did not perform the secondary task were able to react to the incident in a short time, less than 0.5 s. Rule 15 depicts the high association between inclement weather, short reaction time, and near-crash events without secondary tasks. The possible explanation could be that driving in inclement weather, drivers are less likely to perform the secondary tasks, and their reaction time is short in a driving environment filled with uncertainty.

4.3.2.3. 4−5-item rules. Rules 17–26 are the rules with 4 or 5 items for near-crash events without secondary tasks. Rules with more items show more complex patterns. Rule 17 is *{Eventnature = conflict_adjacent + Locality = interstate/divided hwy + ACC_LKA = no + Reacttime = shorter than* 0.5 s*, secondary task = no}*. This rule presents the combination of features that are highly associated with near-crashes without secondary tasks. The items associated with lane changes, *{Eventnature = conflict_adjacent}* and *{Premaneuv = changelane}*, have a dominant presence in 7 rules out of 10 rules (rules 7–19, 21–23, and 26). The dominance of this item in this rule set further highlights the high associations between the lane change behavior and near-crashes without secondary tasks. The locality item {*Locality = interstate/divided hwy}* shows in 4 out of 10 rules (rules17, 18, 19, 26). For example, rule 19 *{Premaneuv = changelane + Eventnature = conflict_adjacent + Locality = interstate/divided hwy, secondary task = no}* demonstrates the strong correlation between the near-crashes without secondary tasks and the combination of lane changing behaviors on the interstate or divided highways. To clarify the possible confusion, this does not suggest that highway is a hot spot of near-crash events or lane changing behavior is the dominant cause of the near-crash event. This rule can tell that many near-crashes without involvements in performing any secondary tasks occur on highways while the lane changing movements occur.

5. Summary and comparison

5.1. Near-crash events with the involvement of secondary tasks

- The drivers are often distracted by performing secondary tasks. Most secondary tasks would distract drivers' attention from focusing on the road.
- The most common cause of this near-crash type is the leading vehicle suddenly slowed or stopped, which leads to the vehicle conflict with the leading vehicle, and the dominant incident type is rear-end near-crash. The drivers who perform any kind of secondary task are normally confident with their driving environment. Their perception of the driving environment is stable and predictable. When the leading vehicle suddenly slowed or stopped, the crash or near-crash events take place. This also explains the prevalence of locality items – interstate or highway in the rules. The traffic conditions often more stable on the interstate highway and divided highways because of fewer interruptions from traffic control devices or unexpected incidents.
- The rules also find that near-crash often has a relatively long reaction time, longer than 1 s. Without full attention on driving, it is reasonable to take a longer time to react to the emergency.
- The rules indicate that the most common evasive maneuver of avoiding the possible crash is braked only. Without enough reaction time to evaluate the driving environment and surroundings, the most direct response is the hard press on the brake.

5.2. Near-crash event without the involvement of secondary tasks

- The most dominant item of this dataset is the lane changing behavior of the participant vehicle or vehicles from an adjacent lane. The rules also found that when both lane changing behaviors occur simultaneously, the likelihood of near-crash is very high. The dominant item associated with this type of crash also leads to another frequent item – sideswipe incident type.
- The rules state the evasive maneuver pattern for this type of crash is either steered or braked and steered.
- A relatively short reaction time, less than 0.5 s, is highly associated with this type of near-crash. While drivers are paying more attention

to the road, instead of conducting secondary tasks, the reaction to the emergency is quick.

- The results also found that with or without the automated driving system active, the reaction time is short if the drivers' attention is on the driving.
- The analysis indicates that the near-crash that occurs during inclement weather often has a short reaction time. With the uncertainties in the driving environment, drivers often are more focused on driving.

5.3. Comparison and countermeasures

- Drivers who do not perform secondary tasks are less likely to be distracted. As more statistics proving the association between distracted drivers and crashes, limiting the frequency of the secondary task should be a more serious and important topic in traffic safety education.
- The majority of near-crashes with the involvement of secondary tasks are because of drivers' overconfidence of the presumably predictable driving environment, such as driving on an interstate highway. The most common event is the leading vehicles suddenly slowed or stopped. However, for near-crashes without the involvement of secondary tasks, the most common trigger is lane-changing behaviors. With the full attention on the roadway, the sudden slow down or stop of the leading vehicles often do not lead to near-crash events for drivers without performing any secondary tasks. The countermeasures for mitigating these two types of near-crashes could be utilizing the advancement of the automated driving system. For example, future vehicles could use wheels or sensors to detect the secondary task occurrence. If the secondary task is detected, since the most common near-crash scenario for the distracted driver is the leading vehicle suddenly slowed or stopped, the vehicle's front distance with the leading vehicle detection sensitivity can be automatically adjusted to the highest level. Thus, the auto-braking system would be activated if an emergency occurs in the first place. For near-crash without secondary tasks involved, the possible solution would be to enlarge the radar detection range. Therefore, the radar sensor does not only detect the leading vehicles but also detects the vehicles in adjacent lanes. The sensor's sensitivity should increase to the highest level while performing lane-changing, especially on interstate or divided highways.
- Drivers not performing secondary tasks often have both hands on the wheel and react faster to emergency incidents, such as collisions. This finding can also be incorporated in traffic safety education to encourage drivers to drive with both hands on the steering wheel.
- Both near-crash events suggest that the interstate highway or divided highway are highly associated with near-crash events. As mentioned above, the traffic conditions on these roadways are often stable and presumably predictable. Drivers may become overconfident about driving and started to perform secondary tasks. Meanwhile, highway often has higher speeds, and fewer interruptions and lane change behavior could happen fast without a full investigation of the driving surroundings, especially when two vehicles try to merge to the same lane at the same time. Highways are commonplace where near-crash occurs, and the consequence of the incident normally is more severe due to the high speed. The countermeasure in the second point could also be applied here. In the future, if the automated driving system can detect secondary tasks and automatically increase detection sensitivity and range, the highway is where the level of sensitivity should be set to the highest.

6. Conclusions

This study utilized a naturalistic research dataset to mine the patterns of near-crash events from the viewpoint of secondary task involvement. This dataset contains many unique features related to near-crash events, such as the vehicle maneuver before the near-crash, the evasive maneuver of the vehicle when the near-crash occurs, the driver's behavior before the near-crash, the driver's distraction status before the near-crash, and the hands-on-wheel status of the driver. The association rules mining algorithm has been applied to the categorical datasets regarding the secondary task involvement, and 58 *2-item* association rules, 353 *3-item* association rules, and 255 *4–5-item* rules were mined. The Top twenty-six rules with the highest lift values in each category, near-crash events with or without secondary tasks, are collectively interpreted and summarized. The findings are unique and interesting. For example, the patterns show that the majority of near-crashes that involved performing secondary tasks were rear-end crashes due to the leading vehicle suddenly slowing or stopping. On the contrary, for near-crashes without the involvement of a secondary task, they are less likely to be rear-end crashes and more likely to involve conflicts with adjacent vehicles while changing lanes.

These interesting findings could help transportation researchers and engineers better understand the nature of near-crash events and comprehend the potential risks of performing secondary tasks. This study proves that the drivers who perform secondary tasks are more likely to be involved in rear-end near-crashes, which are preventable. The research also shows that near-crashes can occur during lane-changing situations involving drivers with their full attention on the roadways. Improved blind spot detection may help reduce the occurrence of this type of near-crash. For example, currently, the majority of blind spot detection technologies often detect vehicle movement at adjacent lanes. However, the system may fail when two vehicles from different lanes tend to the same lane at the same time. The findings of this study would encourage future vehicle engineers to enlarge the detection areas and alert the driver when the projected trajectories of the non-adjacent vehicles may lead to potential conflicts.

The current study has several limitations. First, the sample size of the database is small. A larger dataset can help in dividing the databases into training and test data to validate the generated rules. Second, secondary tasks can be categorized into various types. The intensity and duration of the secondary task being performed on roadways with different posted speed limits can produce different outcomes. Future studies can explore the nature of the secondary tasks to justify the outcomes of the generated rules. Third, like any other traffic safety data analysis, the effect of latent variable such as peak-hour may influence the modeling outcomes of this study. Last but not the least, the dataset anonymized participants to avoid the possible misuse of these information. However, if these pieces of information were available, more in-depth analysis could be done and more interesting results can be discovered.

Author Statement

Xiaoqiang Kong developed the initial research idea, performed the analysis, and led the manuscript writing. Subasish Das performed the analysis and co-led the manuscript writing and performed revisions. Yunlong Zhang reviewed the analysis, oversaw the results, provided advice on the results, and contributed to the revisions of the final manuscript. All authors discussed the results and contributed to the final manuscript.

Author disclosure

The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgments

The authors like to thank two anonymous reviewers for their excellent feedback.

References

Arvin, R., Khattak, A.J., 2020. Driving impairments and duration of distractions: assessing crash risk by harnessing microscopic naturalistic driving data. Accid. Anal. Prev. 146, 105733 https://doi.org/10.1016/j.aap.2020.105733.

Ashouri, M.R., Nahvi, A., Azadi, S., 2018. Time delay analysis of vehicle handling variables for near-crash detection of drowsy driving using a bus driving simulator. 2018 6th RSI International Conference on Robotics and Mechatronics (IcRoM). Presented at the 2018 6th RSI International Conference on Robotics and Mechatronics (IcRoM) 243–249. https://doi.org/10.1109/ICRoM.2018.8657519.

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. Transp. Res. Rec. 2672, 245–254. https://doi.org/10.1177/0361198118772703.

Bakiri, S., Galéra, C., Lagarde, E., Laborey, M., Contrand, B., Ribéreau-Gayon, R., Salmi, L.-R., Gabaude, C., Fort, A., Maury, B., Lemercier, C., Cours, M., Bouvard, M.-P., Orriols, L., 2013. Distraction and driving: results from a case–control responsibility study of traffic crash injured drivers interviewed at the emergency room. Accid. Anal. Prev. 59, 588–592. https://doi.org/10.1016/j.aap.2013.06.004.

Bálint, A., Flannagan, C.A.C., Leslie, A., Klauer, S., Guo, F., Dozza, M., 2020. Multitasking additional-to-driving: prevalence, structure, and associated risk in SHRP2 naturalistic driving data. Accid. Anal. Prev. 137, 105455 https://doi.org/10.1016/j.aap.2020.105455.

Bärgman, J., Lisovskaja, V., Victor, T., Flannagan, C., Dozza, M., 2015. How does glance behavior influence crash and injury risk? A 'what-if' counterfactual simulation using crashes and near-crashes from SHRP2. Transp. Res. Part F Traffic Psychol. Behav. 35, 152–169. https://doi.org/10.1016/j.trf.2015.10.011.

Beck, K.H., Zanjani, F., Allen, H.K., 2019. Social context of drinking among older adults: relationship to alcohol and traffic risk behaviors. Transp. Res. Part F Traffic Psychol. Behav. 64, 161–170. https://doi.org/10.1016/j.trf.2019.05.001.

Chandrapp, A.K., Bhattacharyya, K., Maitra, B., 2016. Estimation of post-encroachment time and threshold wait time for pedestrians on a busy urban corridor in a heterogeneous traffic environment: an experience in Kolkata. Asian Transport Stud. 4, 421–429.

Cheng, B., Lin, Q., Song, T., Cui, Y., Wang, L., Kuzumaki, S., 2011. Analysis of driver brake operation in near-crash situation using naturalistic driving data. Int. J. Automot. Eng. Technol. 2, 87–94. https://doi.org/10.20485/jsaeijae.2.4_87.

Cunningham, M.L., Regan, M.A., 2018. Driver distraction and inattention in the realm of automated driving. Iet Intell. Transp. Syst. 12, 407–413. https://doi.org/10.1049/iet-its.2017.0232.

Das, S., Kong, X., Tsapakis, I., 2019. Hit and run crash analysis using association rules mining. J. Transp. Saf. Secur. 1–20.

Davis, G.A., Hourdos, J., Xiong, H., 2008. Outline of causal theory of traffic conflicts and collisions. In: Presented at the Transportation Research Board 87th Annual Meeting, Transportation Research Board. Washington D.C.

de Rome, L., Brown, J., Baldock, M., Fitzharris, M., 2018. Near-miss crashes and other predictors of motorcycle crashes: findings from a population-based survey. Traffic Inj. Prev. 19, S20–S26. https://doi.org/10.1080/15389588.2018.1536822.

Dingus, T.A., Klauer, S.G., Neale, V.L., Petersen, A., Lee, S.E., Sudweeks, J., Perez, M.A., Hankey, J., Ramsey, D., Gupta, S., 2006. The 100-car Naturalistic Driving Study, Phase II-results of the 100-car Field Experiment. United States Department of Transportation, Washington D.C.

Dingus, T.A., Guo, F., Lee, S., Antin, J.F., Perez, M., Buchanan-King, M., Hankey, J., 2016. Driver crash risk factors and prevalence evaluation using naturalistic driving data. PNAS 113, 2636–2641. https://doi.org/10.1073/pnas.1513271113.

Ersal, T., Fuller, H.J.A., Tsimhoni, O., Stein, J.L., Fathy, H.K., 2010. Model-based analysis and classification of driver distraction under secondary tasks. Ieee Trans. Intell. Transp. Syst. 11, 692–701. https://doi.org/10.1109/TITS.2010.2049741.

Esfahani, H.N., Arvin, R., Song, Z., Sze, N.N., 2019. Prevalence of cell phone use while driving and its impact on driving performance, focusing on near-crash risk: a survey study in Tehran. J. Transp. Saf. Secur. 0, 1–21. https://doi.org/10.1080/19439962.2019.1701166.

Farmer, C.M., Klauer, S.G., McClafferty, J.A., Guo, F., 2015. Relationship of Near-Crash/Crash risk to time spent on a cell phone while driving. Traffic Inj. Prev. 16, 792–800. https://doi.org/10.1080/15389588.2015.1019614.

Gettman, D., Head, L., 2003. Surrogate safety measures from traffic simulation models. Trans. Res. Rec.: J. Trans. Res. Board 1840, 104–115.

Gettman, D., Pu, L., Sayed, T., Shelby, S., Siemens, I.T.S., 2008. Surrogate Safety Assessment Model and Validation. United States. Federal Highway Administration. United States Department of Transportation, Washington D.C.

Guo, F., Fang, Y., 2013. Individual driver risk assessment using naturalistic driving data. Accident Analysis & Prevention. Emerging Res. Methods and Their Appl. Road Safety 61, 3–9. https://doi.org/10.1016/j.aap.2012.06.014.

Guo, F., Klauer, S.G., Hankey, J.M., Dingus, T.A., 2010a. Near crashes as crash surrogate for naturalistic driving studies. Transp. Res. Rec. 2147, 66–74. https://doi.org/10.3141/2147-09.

Guo, F., Klauer, S.G., McGill, M.T., Dingus, T.A., 2010b. Evaluating the Relationship Between Near-crashes and Crashes: Can Near-crashes Serve as a Surrogate Safety Metric for Crashes?.

Guo, F., Simons-Morton, B.G., Klauer, S.E., Ouimet, M.C., Dingus, T.A., Lee, S.E., 2013. Variability in crash and near-crash risk among novice teenage drivers: a naturalistic study. J. Pediatr. 163, 1670–1676. https://doi.org/10.1016/j.jpeds.2013.07.025.

Hamzeie, R., Savolainen, P.T., Gates, T.J., 2017. Driver speed selection and crash risk: insights from the naturalistic driving study. J. Safety Res. 63, 187–194. https://doi.org/10.1016/j.jsr.2017.10.007.

Huisingh, C., Levitan, E.B., Irvin, M.R., MacLennan, P., Wadley, V., Owsley, C., 2017. Visual sensory and visual-cognitive function and rate of crash and near-crash involvement among older drivers using naturalistic driving data. Invest. Ophthalmol. Vis. Sci. 58, 2959–2967. https://doi.org/10.1167/iovs.17-21482.

Huisingh, C., Owsley, C., Levitan, E.B., Irvin, M.R., MacLennan, P., McGwin, G., 2019. Distracted driving and risk of crash or near-crash involvement among older drivers using naturalistic driving data with a case-crossover study design. J. Gerontol. A Biol. Sci. Med. Sci. 74, 550–555. https://doi.org/10.1093/gerona/gly119.

Jiang, X., Wang, W., Bengler, K., 2015. Intercultural analyses of time-to-Collision in vehicle–Pedestrian conflict on an urban midblock crosswalk. Ieee Trans. Intell. Transp. Syst. 16, 1048–1053. https://doi.org/10.1109/TITS.2014.2345555.

Jovanis, P.P., Aguero-Valverde, J., Wu, K.-F., Shankar, V., 2011. Analysis of naturalistic driving event data: omitted-variable Bias and multilevel modeling approaches. Transp. Res. Rec. 2236, 49–57. https://doi.org/10.3141/2236-06.

Kim, H., Song, M., Greatbatch, R., Novotny, A., Doerzaph, Z., 2020. The Virginia Connected Corridor 50 Elite Vehicle Naturalistic Driving Study (VCC50 Elite NDS) (00-029). https://doi.org/10.15787/VTT1/98NBN7.

Klauer, S.G., Dingus, T.A., Neale, V.L., Sudweeks, J.D., Ramsey, D.J., 2006. The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data.

Klauer, S.G., Guo, F., Simons-Morton, B.G., Ouimet, M.C., Lee, S.E., Dingus, T.A., 2014. Distracted driving and risk of road crashes among novice and experienced drivers. N. Engl. J. Med. 370, 54–59. https://doi.org/10.1056/NEJMsa1204142.

Klauer, S.G., Ehsani, J.P., McGehee, D.V., Manser, M., 2015. The effect of secondary task engagement on adolescents' driving performance and crash risk. J. Adolesc. Health, Exploring Teen Driver Safety and Crash Risk: State of the Res. 57, S36–S43. https://doi.org/10.1016/j.jadohealth.2015.03.014.

Kong, X., Das, S., Jha, K., Zhang, Y., 2020. Understanding speeding behavior from naturalistic driving data: applying classification based association rule mining. Accid. Anal. Prev. 144, 105620.

Lee, S.E., Simons-Morton, B.G., Klauer, S.E., Ouimet, M.C., Dingus, T.A., 2011. Naturalistic assessment of novice teenage crash experience. Accid. Anal. Prev. 43, 1472–1479. https://doi.org/10.1016/j.aap.2011.02.026.

Liang, Y., Lee, J.D., Yekhshatyan, L., 2012. How dangerous is looking away from the Road? Algorithms predict crash risk from glance patterns in naturalistic driving. Hum. Factors 54, 1104–1116. https://doi.org/10.1177/0018720812446965.

Liang, Y., Lee, J.D., Horrey, W.J., 2014. A looming crisis: the distribution of off-Road glance duration in moments leading up to crashes/Near-crashes in naturalistic driving. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58, pp. 2102–2106. https://doi.org/10.1177/1541931214581442.

Markkula, G., Engström, J., Lodin, J., Bärgman, J., Victor, T., 2016. A farewell to brake reaction times? Kinematics-dependent brake response in naturalistic rear-end emergencies. Accid. Anal. Prev. 95, 209–226. https://doi.org/10.1016/j.aap.2016.07.007.

Miller, D., Sun, A., Johns, M., Ive, H., Sirkin, D., Aich, S., Ju, W., 2015. Distraction becomes engagement in automated driving. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 59, pp. 1676–1680. https://doi.org/10.1177/1541931215591362.

Mousa, S.R., Bakhit, P.R., Ishak, S., 2019. An extreme gradient boosting method for identifying the factors contributing to crash/near-crash events: a naturalistic driving study. Can. J. Civ. Eng. https://doi.org/10.1139/cjce-2018-0117.

Nadimi, N., Ragland, D.R., Amiri, A.M., 2020. An evaluation of time-to-collision as a surrogate safety measure and a proposal of a new method for its application in safety analysis. Transp. Lett. Int. J. Transp. Res. 12, 491–500. https://doi.org/10.1080/19427867.2019.1650430.

Naji, H.A.H., Xue, Q., Lyu, N., Wu, C., Zheng, K., 2018. Evaluating the driving risk of near-crash events using a mixed-ordered logit model. Sustainability 10, 2868. https://doi.org/10.3390/su10082868.

National Safety Council, 2013. Crashes Involving Cell Phones: Challenges of Collecting and Reporting Reliable Crash Data. National Safety Council Itasca, IL.

NHTSA, 2019. Highway Statistics 2018 [WWW Document]. Persons Fatally Injured In Motor Vehicle Crashes. URL https://www.fhwa.dot.gov/policyinformation/statistics/2018/fi220.cfm (accessed 3.6.20).

Ogeil, R.P., Barger, L.K., Lockley, S.W., O'Brien, C.S., Sullivan, J.P., Qadri, S., Lubman, D.I., Czeisler, C.A., Rajaratnam, S.M.W., 2018. Cross-sectional analysis of sleep-promoting and wake-promoting drug use on health, fatigue-related error, and near-crashes in police officers. BMJ Open 8, e022041. https://doi.org/10.1136/bmjopen-2018-022041.

Osman, O.A., Hajij, M., Bakhit, P.R., Ishak, S., 2019. Prediction of near-crashes from observed vehicle kinematics using machine learning. Transp. Res. Rec. 2673, 463–473. https://doi.org/10.1177/0361198119862629.

Oviedo-Trespalacios, O., King, M., Haque, M.M., Washington, S., 2017. Risk factors of mobile phone use while driving in Queensland: prevalence, attitudes, crash risk perception, and task-management strategies. PLoS One 12, e0183361. https://doi.org/10.1371/journal.pone.0183361.

Paul, M., Ghosh, I., 2020. Post encroachment time threshold identification for right-turn related crashes at unsignalized intersections on intercity highways under mixed traffic. Int. J. Inj. Contr. Saf. Promot. 27, 121–135.

Perez, M.A., Sudweeks, J.D., Sears, E., Antin, J., Lee, S., Hankey, J.M., Dingus, T.A., 2017. Performance of basic kinematic thresholds in the identification of crash and

near-crash events within naturalistic driving data. Accid. Anal. Prev. 103, 10–19. https://doi.org/10.1016/j.aap.2017.03.005.
Ranney, T.A., 2008. Driver Distraction: A Review of the Current State-of-Knowledge.
Regan, M.A., Lee, J.D., Young, K., 2008. Driver Distraction: Theory, Effects, and Mitigation. CRC Press.
Regev, S., Rolison, J., Feeney, A., Moutari, S., 2017. Driver distraction is an under-reported cause of road accidents: an examination of discrepancy between police officers' views and road accident reports. DDI2017 E-Proceedings Collection. The Fifth International Conference on Driver Distraction and Inattention.
Sayed, T., Zaki, M.H., Autey, J., 2013. Automated safety diagnosis of vehicle–bicycle interactions using computer vision analysis. Saf. Sci. 59, 163–172. https://doi.org/10.1016/j.ssci.2013.05.009.
Scofield, Larry, 2015. Researcher Dictionary for Safety Critical Event Video Reduction Data. Virginia Tech. Virginia Tech Transportation Institute.
Seacrist, T., Douglas, E.C., Huang, E., Megariotis, J., Prabahar, A., Kashem, A., Elzarka, A., Haber, L., MacKinney, T., Loeb, H., 2018. Analysis of near crashes among teen, young adult, and experienced adult drivers using the SHRP2 naturalistic driving study. Traffic Inj. Prev. 19, S89–S96. https://doi.org/10.1080/15389588.2017.1415433.
Seacrist, T., Douglas, E.C., Hannan, C., Rogers, R., Belwadi, A., Loeb, H., 2020. Near crash characteristics among risky drivers using the SHRP2 naturalistic driving study. J. Safety Res. 73, 263–269. https://doi.org/10.1016/j.jsr.2020.03.012.
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. J. Adolesc. Health, Driver Distraction: A Perennial but Preventable Public Health Threat to Adolesc. 54, S61–S67. https://doi.org/10.1016/j.jadohealth.2013.11.021.
So, J., Lim, I.-K., Kweon, Y.-J., 2015. Exploring traffic conflict-based surrogate approach for safety assessment of highway facilities. Transp. Res. Rec. 2513, 56–62.
Songchitruksa, P., Tarko, A.P., 2006. The extreme value theory approach to safety estimation. Accid. Anal. Prev. 38, 811–822. https://doi.org/10.1016/j.aap.2006.02.003.
Strauss, J., Zangenehpour, S., Miranda-Moreno, L.F., Saunier, N., 2017. Cyclist deceleration rate as surrogate safety measure in Montreal using smartphone GPS data. Accid. Anal. Prev. 99, 287–296. https://doi.org/10.1016/j.aap.2016.11.019.
Stutts, J.C., Reinfurt, D.W., Staplin, L., Rodgman, E., 2001. The Role of Driver Distraction in Traffic Crashes.
Su, J., Chen, J., Wang, H., Chen, W., Wang, K., 2017. Establishment and analysis on typical road traffic near-crash scenarios related to pedestrian in China. Traffic and Trans. 209–214.
Taccari, L., Sambo, F., Bravi, L., Salti, S., Sarti, L., Simoncini, M., Lori, A., 2018. Classification of crash and near-crash events from dashcam videos and telematics. 2018 21st International Conference on Intelligent Transportation Systems (ITSC). Presented at the 2018 21st International Conference on Intelligent Transportation Systems (ITSC) 2460–2465. https://doi.org/10.1109/ITSC.2018.8569952.
Tang, K., Kuwahara, M., 2011. Implementing the concept of critical post-encroachment time for all-red clearance interval design at signalized intersections. In: Proceedings of the Eastern Asia Society for Transportation Studies Vol. 8 (The 9th International Conference of Eastern Asia Society for Transportation Studies, 2011). Eastern Asia Society for Transportation Studies, pp. 299–299.
Tian, R., Li, L., Chen, M., Chen, Y., Witt, G.J., 2013. Studying the Effects of Driver Distraction and Traffic Density on the Probability of Crash and Near-Crash Events in Naturalistic Driving Environment. IEEE Trans. Intell. Transp. Syst. 14, 1547–1555. https://doi.org/10.1109/TITS.2013.2261988.
Tivesten, E., Dozza, M., 2014. Driving context and visual-manual phone tasks influence glance behavior in naturalistic driving. Transp. Res. Part F Traffic Psychol. Behav. 26, 258–272. https://doi.org/10.1016/j.trf.2014.08.004.
TXDOT, 2020. CRIS Query [WWW Document]. URL https://cris.dot.state.tx.us/public/Query/app/welcome (accessed 1.25.21).
Uzondu, C., Jamson, S., Lai, F., 2018. Exploratory study involving observation of traffic behaviour and conflicts in Nigeria using the Traffic Conflict Technique. Saf. Sci. 110, 273–284. https://doi.org/10.1016/j.ssci.2018.08.029.
Vogel, K., 2003. A comparison of headway and time to collision as safety indicators. Accid. Anal. Prev. 35, 427–433. https://doi.org/10.1016/S0001-4575(02)00022-2.
Wandtner, B., Schömig, N., Schmidt, G., 2018. Effects of non-driving related task modalities on takeover performance in highly automated driving. Hum. Factors 60, 870–881. https://doi.org/10.1177/0018720818768199.
Wang, J., Zheng, Y., Li, X., Yu, C., Kodaka, K., Li, K., 2015. Driving risk assessment using near-crash database through data mining of tree-based model. Accid. Anal. Prev. 84, 54–64. https://doi.org/10.1016/j.aap.2015.07.007.
Wang, C., Xu, C., Xia, J., Qian, Z., Lu, L., 2018. A combined use of microscopic traffic simulation and extreme value methods for traffic safety evaluation. Transp. Res. Part C Emerg. Technol. 90, 281–291. https://doi.org/10.1016/j.trc.2018.03.011.
Wu, K.-F., Jovanis, P.P., 2012. Crashes and crash-surrogate events: exploratory modeling with naturalistic driving data. Accid. Anal. Prev. 45, 507–516. https://doi.org/10.1016/j.aap.2011.09.002.
Wu, K.-F., Aguero-Valverde, J., Jovanis, P.P., 2014. Using naturalistic driving data to explore the association between traffic safety-related events and crash risk at driver level. Accid. Anal. Prev. 72, 210–218. https://doi.org/10.1016/j.aap.2014.07.005.
Wynne, R.A., Beanland, V., Salmon, P.M., 2019. Systematic review of driving simulator validation studies. Saf. Sci. 117, 138–151. https://doi.org/10.1016/j.ssci.2019.04.004.
Xu, R., Luo, F., 2020. Risk prediction and early warning for air traffic controllers' unsafe acts using association rule mining and random forest. Saf. Sci. 135, 105125.
Ye, M., Osman, O.A., Ishak, S., Hashemi, B., 2017. Detection of driver engagement in secondary tasks from observed naturalistic driving behavior. Accid. Anal. Prev. 106, 385–391. https://doi.org/10.1016/j.aap.2017.07.010.
Zheng, L., Ismail, K., Meng, X., 2014a. Traffic conflict techniques for road safety analysis: open questions and some insights. Can. J. Civ. Eng. 41, 633–641.
Zheng, L., Ismail, K., Meng, X., 2014b. Freeway safety estimation using extreme value theory approaches: a comparative study. Accid. Anal. Prev. 62, 32–41. https://doi.org/10.1016/j.aap.2013.09.006.