ARTICLE IN PRESS

Accident Analysis and Prevention xxx (xxxx) xxx

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Fatal pedestrian crashes at intersections: Trend mining using association rules

Subasish Das[a,*,1], Reuben Tamakloe[b,2], Hamsa Zubaidi[c], Ihsan Obaid[d,3], Ali Alnedawi[e]

[a] *Texas A&M Transportation Institute, 1111 RELLIS Parkway, Bryan, TX 77807, United States*
[b] *Department of Transportation Engineering, University of Seoul, 163 Seoulsiripdae-ro Dongdaemun-gu, Seoul 02504, South Korea*
[c] *Roads and Transport Department, College of Engineering, University of Al-Qadisiyah, Iraq*
[d] *Oregon State University, 233 Owen Hall, Corvallis, OR 97331-3212, United States*
[e] *School of Engineering, Deakin University, Geelong, Victoria 3220, Australia*

ARTICLE INFO

Keywords:
Pedestrian
Pedestrian safety
Fatal pedestrian crash
Intersection crashes
Data mining
Association rules

ABSTRACT

In 2018, about 6,677 pedestrians were killed on the US roadways. Around one-fourth of these crashes happened at intersections or near intersection locations. This high death toll requires careful investigation. The purpose of this study is to provide an overview of the characteristics and associated crash scenarios resulting in fatal pedestrian crashes in the US. The current study collected five years (2014–2018) of fatal crash data with additional details of pedestrian crash typing. This dataset provides specifics of scenarios associated with fatal pedestrian crashes. This study applied associated rules mining on four sub-groups, which were determined based on the highest frequencies of fatal crash scenarios. This study also developed the top 20 rules for all four sub-groups and used 'a priori' algorithm with 'lift' as a performance measure. Some of the key variable categories such as dark with lighting condition, vehicle going straight, vehicle turning, local municipality streets, pedestrian age range from 45 years and above are frequently presented in the developed rules. The patterns of the rules differ by the pedestrian's position within and outside of crosswalk area. If the pedestrian is outside the crosswalk area, no lighting at dark is associated with high number of crashes. As lift provides quantitative measures in the form of the likelihood, the rules can be transferred into data-driven decision making. The findings of the current study can be used by safety engineers and planners to improve pedestrian safety at intersections.

1. Introduction

As walking and biking have become popular activities in recent years, the number of non-motorist (pedestrian and bicyclist) fatalities have increased significantly. During 2014–2018, 28,842 pedestrian fatalities occurred in the U.S. (National Highway Traffic Safety Administration or NHTSA, 2020). Although intersections represent a small portion of overall roadway miles, around 26% of all fatal pedestrian crashes occurred at intersections or near intersections (NHTSA, 2020). Urban and sub-urban intersection locations are associated with a high number of non-motorist crashes. A pedestrian crash at an interaction can be a product of driver-pedestrian interaction, intersection type and traffic control devices present, and other geometric and scenario-based conditions. To establish countermeasures that can help reduce pedestrian fatalities at intersections, it essential to study and recognize which groups of crash-risk factors are associated with each type of fatal pedestrian crash. Since such studies are currently non-existent in the pedestrian safety field, a rigorous investigation is needed to understand the groups of crash contributory factors associated with pedestrian fatalities considering crash typing information.

Intersection and intersection-related pedestrian fatalities increased by 13% (percentage difference between 1,424 intersection and intersection-related fatalities in 2014 and 1,634 intersection and intersection-related fatalities in 2018) from 2014 to 2018 (NHTSA, 2020). Many state-based traffic crash data do not contain details on pedestrian crashes. The urgency behind pedestrian and bicycle crash

* Corresponding author.
E-mail addresses: s-das@tti.tamu.edu (S. Das), drtamakloe@uos.ackr (R. Tamakloe), hamsa.alzubaidi@qu.edu.iq (H. Zubaidi), obaidi@oregonstate.edu (I. Obaid), a.alnedawi@deakin.edu.au (A. Alnedawi).
[1] ORCID: 0000-0002-1671-2753.
[2] ORCID: 0000-0002-9605-8439.
[3] ORCID: 0000-0001-9048-8342.

https://doi.org/10.1016/j.aap.2021.106306
Received 18 November 2020; Received in revised form 29 June 2021; Accepted 10 July 2021

typing lies in providing a detailed version of the pre-crash events to better illustrate the sequence of events and associated scenarios leading to crash occurrence between vehicles and pedestrians/bicyclists (Harkey et al., 2006). In 2010, NHTSA adopted parts of an application called Pedestrian and Bicycle Crash Analysis Tool (PBCAT) into the Fatality Analysis Reporting System (FARS) (Harkey et al., 2006).

FARS data with pedestrian crash typing provides details of pedestrian possible crossing scenarios. As the dataset provides details of many variables and associated attributes, conventional statistical models are not suitable in identifying the key contributing factors and their interactions. The purpose of this study is to provide an overview of the characteristics and associated crash scenarios caused by fatal pedestrian crashes in the U.S. To perform the analysis, intersection and intersection-related fatal pedestrian crashes were collected using the detailed pedestrian crash typing information for five years (2014–2018) from FARS. This study collected the raw data from FTP website (https://www.nhtsa.gov/file-downloads?p = nhtsa/downloads/FARS/) rather performing queries. It is also important to note that FARS data can be queried via Fatality and Injury Reporting System Tool (FIRST) (https://cdan.dot.gov/query). Data collection from FARS is noted as FIRST in the later part of this paper. In addition, association rules mining was applied to identify the association patterns. As some scenarios are disproportionately higher than the other scenarios, association rules mining was conducted on the sub-group data based on the most frequent pedestrian crash scenarios.

2. Literature review

In this section, the literature review is divided into two major sections: 1) studies on fatal and severe pedestrian injuries in general and 2) studies on intersection-related pedestrian crashes.

2.1. Studies on fatal and severe pedestrian injuries

According to the World Health Organization, vulnerable road users, particularly pedestrians and cyclists, are involved in over half of the 1.35 million traffic-related deaths worldwide each year. To address this issue, transportation researchers have conducted several studies focused on understanding the factors influencing the fatality of crashes involving vulnerable road users. Regarding studies on pedestrian crashes that result in fatal injuries, Anderson et al. (1997) identified that significant travel speed reductions could lead to remarkable reductions in the severity of injuries sustained by the crash victim. The study showed that a 10 km/h reduction in traffic speeds at traffic analysis zones with 60 km/h speed limits reduced pedestrian crash fatalities by 48%. In most areas, traffic collisions involving pedestrians were eliminated. This finding is in line with results from a Maine-based that showed strong associations between speed and pedestrian-involved crash severity (Gårder, 2004).

To determine critical factors influencing pedestrian fatalities from crash scenes, Dovom et al. (2012) investigated forensic medical data of Mashhad, Iran, from April 2006 to December 2009. The fatal pedestrian crashes comprised almost 40% of all road user fatal crashes. The binary logistic model analysis results depicted that several factors such as pedestrian age, the type of vehicle involved in the crash, and the crash location had a significant effect on death probability at the crash scene. It is important to note that there have been some inconsistencies regarding vehicle type impacts on the injury severity sustained in a pedestrian-vehicle crash. Whereas Henary et al. (2003) established that pedestrian-light truck crashes are more likely to be severe compared to a pedestrian-passenger car crash, Tamakloe et al. (2020) also showed that the severity of pedestrian-bus crashes are more likely to result in reduced severity levels – inconsistent with the findings from Aziz et al. (2013)'s study. A study conducted by Jang et al. (2013), similar to that of Dovom et al. (2012), also showed that alcohol involvement, pedestrian age (<15 years and >65 years), time of day (night), adverse weather, and weekends also contributed to an increase in the likelihood for severe and fatal injuries in pedestrian-vehicle crashes in San Francisco, California. Olszewski et al. (2015) also developed a binary logit model with interaction terms to illustrate the association between pedestrian fatality and influential factors. The model identified several crucial factors, including two-way road, roadway division, non-built-up area, no lighting at night, summer, and mid-block crosswalk location. Uddin and Ahmed (2018) studied Ohio traffic crash data compiled in the Highway Safety Information System (HSIS) during 2009–2013. They estimated both fixed and random parameters ordered probit models of injury severity. The results identified key contributing factors including six-lane roadways, dark-unlighted roadways, struck by a truck, an older pedestrian (65 years old and over), a younger driver (less than 24 yers old), driving under the influence, and speed limits of 40 mph and 50 mph.

To learn more about the patterns of fatal pedestrian crashes, Prato et al. (2012) applied Kohonen neural networks to a database of fatal pedestrian crashes that occurred from 2003 to 2006 to help develop preventive measures and determine how to best allocate resources to mitigate the identified problems. Based on the pattern mapping results, the paper identified that fatal pedestrian crashes are likely to involve young males crossing during the nighttime on both rural and urban roads, teens and children crossing roads in rural communities, pedestrians suddenly appearing and crossing rural roads, the aged crossing at crosswalks located far from intersections, and male pedestrians crossing at night on rural roads. The authors proposed the framing of educational programs and campaigns targeting each of the illustrated patterns and the proper enactment and enforcement of regulations to address the increasing fatality trends.

Of recent, studies have focused on conducting analysis for specific conditions under which pedestrian crashes are most likely to occur. While some researchers focused on analyzing the pedestrian crashes under different weather conditions (Li et al., 2017; Zhai et al., 2019), others focused on the effects of neighbourhood features and the built environment on the severity of pedestrian crashes (Xin et al., 2017). As walking is highly sensitive to changes in weather conditions, researchers diagnosed the relationship between weather condition and pedestrian crash severity outcomes. Li et al. (2017) investigated the influential factors of pedestrian injury severity level under different weather conditions using data from the STATS19 crash database publicly available in (UK). The study implemented a unique technique that combines a classification and regression tree with a random forest approach to identify the features related to the driver, pedestrian, and environment. Results from the study identified that pedestrian age, vehicle maneuver, and speed limits were significant predictors of pedestrian injury severity under both fine and adverse weather conditions. Besides, a similar study conducted by Sun et al. (2019) using a latent class clustering technique showed that both fatal and severe pedestrian crashes are primarily associated with pedestrian age (over 65 years), bad weather conditions, and pedestrian drug or alcohol involvement, light condition (dark), and driving at high speeds.

Regarding the effects of neighborhood and built environment characteristics on pedestrian injury severity, Xin et al. (2017) employed a random parameter generalized ordered probit model with heterogeneity in means and variances approach to quantify the influence of the built environment on pedestrian injury severity outcomes. The study considered 3,867 total pedestrian crashes from the Florida Department of Transportation District 7 from 2011 to 2014. The findings showed that the presence of a school within a range of 0.31 mile from a crash resulted in a greater likelihood of an incapacitating injury rather than a fatal injury. Besides, pedestrian-vehicle crashes occurring at bus stop areas are likely to have a reduced propensity for fatal injuries.

In terms of fatal pedestrian crash fault assessment, Spainhour and Wootton (2007) utilized binary logistic regression to model fault in 318 vehicle–pedestrian crash cases in Florida. From the results, it was identified that inattention, distraction, error in perception, decision-making error, or intoxication increased the likelihood of fault being assigned. In particular, while the mental state of both the driver and pedestrian are critical variables used in fault allocation, average daily traffic (ADT) and driver intoxication were positively significant predictors of driver fault, and the number of lanes a pedestrian tries to cross is a significant positive predictor of pedestrian fault. The study added that taking a pedestrian-centred approach in establishing roadway designs and traffic operation decisions could help mitigate the rising fatality trends.

2.2. Studies on intersection-related pedestrian injuries

Intersection-related pedestrian-vehicle crashes are reported fairly often. The pedestrian injury severity in these crashes is significantly related to vehicle characteristics, driver behavior, pedestrian characteristics, and intersection geometry attributes (ref). The identification of these features related to pedestrian injury severity is critical to future safety improvements and crash reductions. To help reduce the severity of intersection-related pedestrian injuries, Li et al. (2010) investigated field observations in Beijing to explore the safety concerns of pedestrian crossings at signalized intersections. A questionnaire was distributed to pedestrians and motorists at 45 signalized intersections with several questions to complete. The study also explained the traffic conflict analyses and safety implications of pedestrian crossings at wider signalized intersections. The main finding showed that a refuge island in an intersection is acceptable to both pedestrians and drivers. Schneider and Stefanich (2016) developed a new method called the location–movement classification method (LMCM) to classify pedestrian and bicycle crashes. The LMCM was able to provide valuable data that is not included in a well-established NHTSA crash typology. After both typologies were conducted with a sample of 296 pedestrian and 229 bicycle crashes reported in Wisconsin between 2011 and 2013, the LMCM demonstrated that pedestrian crashes are most likely to occur on the farther side of intersections rather than the near side irrespective of injury severity levels. When pedestrian crashes were located on roadway straightaways between intersections, motorists involved in the crash were identified to be traveling straight while pedestrians approached from the motorist's left. These crashes resulted in a significantly greater likelihood of resulting in fatal injuries.

Ma et al. (2018) explored various groups of drivers with different sets of explanatory variables on pedestrian injury severity. They collected 2,614 total crash records (during 2011–2012) from intersections in Cook County, Illinois. The results from the ordered probit modeling approach identified four independent variables that significantly affected pedestrian injury types at intersections for all groups of drivers, including pedestrian age, point of first contact, vehicle type, and weather condition. Xie et al. (2018) investigated pedestrian safety at rural intersections in southwest China and produced an in-depth analysis of 28 fatal pedestrian crashes from 8 low-volume roads as an applicable crash prevention countermeasure. Four different clusters of critical factors negatively affecting pedestrian safety, including lack of pedestrian safety education, inadequate driver training, deficient intersection safety infrastructure, and insufficient traffic law enforcement. A disproportionally high number of fatal pedestrian crashes occur at intersections in correspondence to their small proportion of the entire roadway system.

Further, a study conducted by Xin et al. (2017) showed that about more aged pedestrians (74.3%) involved in a crash at intersection segments are more likely to incur fatal injuries relative to pedestrians of less than 30 years of age in the event that the crash occurred at an intersection compared to those involved in crashes at other road segments in Florida. In a more recent study, Das et al. (2019) employed a three-year (2014 to 2016) FIRST data involving elderly pedestrians to identify the key associations between contributing factors of elderly pedestrian crashes. The study used empirical Bayes (EB) data mining and found several critical risk association patterns. The study showed that variables such as not yielding at intersections and pedestrians crossing at intersections while a vehicle approached from a turn were strongly linked with fatal elderly pedestrians-involved crashes.

In summary, the existing body of literature shows that researchers analyzed pedestrian crashes together regardless of the pedestrian crash typing information, which is a crucial factor in the pedestrian crash safety analysis. Besides, the literature on pedestrian-vehicle crash analysis shows that safety researchers, to a large extent, have explored crash-risk factors impacting the severity of pedestrian-involved crashes using parametric methods. This approach has been effective in characterizing the impact of individual crash-risk factors on the severity of crashes. However, transport safety practitioners who sometimes seek to know the latent trends or groups of crash contributory factors that frequently occur together in fatal pedestrian collisions to make more appropriate and targeted countermeasures to address the safety problem are unable to benefit from these results fully. Besides, parametric methods rely heavily on predetermined assumptions that may affect the accuracy of the results when not handled properly. Recently, studies employing non-parametric approaches, particularly the association rule mining technique, which is a machine learning approach proposed by Agrawal and Srikant (2005), have gained traction in the traffic safety field since they are simple to understand, robust at unearthing patterns in a dataset and do not depend on prespecified assumptions as does the parametric methods (Das et al., 2019; Das and Sun, 2014; Hong et al., 2020; Yu et al., 2019). The association rule mining approach, which applies the a priori algorithm, generates several candidate itemsets (crash-risk factors) using a forward approach. It then applies a robust pruning technique to dramatically remove infrequent itemsets while lessening memory and computing performance requirements. These strengths of the algorithm make it a preferred approach compared to its counterparts, such as the fixed-parameter tractable (FTP) algorithm (Hong et al., 2020). Despite its merits, to the best of our knowledge, no study employed this data mining technique to identify hidden trends and chains of factors influencing fatal pedestrian crash at intersections while considering the detailed pedestrian crash typing information. Therefore, this study aims to bridge this gap by applying association rules mining on the detailed pedestrian crash typing data to help understand the crash scenarios at intersections.

3. Methodology

3.1. Data collection

The research team collected five years (2014–2018) of fatal pedestrian crash data from the FIRST. In the FIRST, pedestrian and bicycle crash typing is conducted by using Ped/Bike Wizard software. The collected data showed that a total of 30,935 pedestrians died in fatal traffic crashes, and intersection and intersection-related (or near to intersection) crashes accounted for 26% of total fatalities (Table1). This study used the variable PB31 (Crash Location) to tag the crashes. It is worth mentioning that the current study is limited to intersections and intersection-related crash scenarios. In this section, exploratory data analysis is provided for additional contexts of the co-occurrence of the key attributes.

3.2. Exploratory data analysis

Understanding pedestrian crashes is crucial for making roadways safer for the pedestrians. Pedestrian exposure in terms of pedestrian volume is often not available. Comparing the status of the states with respect to pedestrian crash and exposure is thus difficult to attain. Each year from 2014 to 2018, California has had the highest number of intersections related fatal pedestrian crashes. Florida, New York, and Texas have also had consistently high numbers of intersection-related fatal pedestrian crashes compared to the other states. Overall, there is a common trend that states with larger populations have a higher

Table 1
Fatal pedestrian crashes during 2014–2018.

Year	At Intersection	Intersection-related	Non-Trafficway	Not at Intersection	Unknown	Total by Year
2014	997	482	92	3,718	36	5,325
2015	1,037	525	98	4,236	29	5,925
2016	1,138	565	99	4,678	26	6,506
2017	1,163	551	102	4,653	33	6,502
2018	1,069	578	82	4,893	55	6,677
Total	**5,404**	**2,701**	**473**	**22,178**	**179**	**30,935**

number of intersections related to fatal pedestrian crashes due to presence of dense urban locations with large number of intersections.

The FIRST data provides a broad range of detailed variables to understand the causation patterns of fatal pedestrian crashes. The pedestrian typing related variables are listed below:

- PEDCTYPE – Crash Type – Pedestrian (55 categories)
- PEDLOC – Crash Location – Pedestrian (5 categories)
- PEDPOS – Pedestrian Position (9 categories)
- PEDDIR – Pedestrian Initial Direction of Travel (5 categories)
- MOTDIR – Motorist Initial Direction of Travel (5 categories)
- MOTMAN – Motorist Maneuver (4 categories)
- PEDLEG – Intersection Leg (3 categories)
- PEDSNR – Pedestrian Scenario (49 categories)
- PEDCGP – Crash Group – Pedestrian (55 categories)
- PBCWALK- Mark Crosswalk Present (0-None, 1-Yes, 9 = Unknown)

Table 2 lists different crash scenarios based on the type of crosswalks. Exposure is a key missing information in this analysis. The presence of a crosswalk may attract more pedestrians to cross the street and the exposure of pedestrians at these sections is expected to be higher than in non-crosswalk locations. The first three scenarios represent 40% (3225/8105*100) of all pedestrian fatalities. For rules mining, these three scenarios are separately examined to understand the hidden patterns. Additionally, rules mining was applied by combing the rest of the data by excluding the first three pedestrian crash scenarios. Thus, the following scenarios were analyzed using rules mining technique:

- Scenario 1: Pedestrian within crosswalk area, traveled from motorist's left (vehicle straight): 15.21 % (1233/8105)
- Scenario 2: Pedestrian within crosswalk area, traveled from motorist's right (vehicle straight): 12.78% (1036/8105)
- Scenario 3: Pedestrian outside crosswalk area, traveled from motorist's left (vehicle straight): 11.79% (956/8105)
- Scenario 4: Other scenarios (scenario 4 to scenario 34 in Table 2): 60.21% (4880/8105)

Although the pedestrian crash group (PEDCGP) has 55 categories, the intersection and intersection-related fatal crash database contains only 14 categories. Like the pedestrian crash scenarios, the presence of crosswalks varies by pedestrian crash groups as shown in Table 3.

As pedestrian crashes are a complex juxtaposition of multiple variables and variable categories, it is also important to have a generalized idea of the co-occurrence. Some of the key observations from Table 2 and Table 3 are below:

- Fatal crashes associated with crossing roadways with vehicle not turning are almost equally distributed at locations with the presence or absence of marked crosswalks. For dash/dart out, this trend is similar.
- Fatal crashes associated with crossing roadways with vehicle turning is more likely to have occurred at locations with the presence of marked crosswalks. For dash/dart out, this trend is similar. Exposure is a key issue in understanding this finding. The presence of marked crosswalks most likely attracts more pedestrians to cross. More careful observations are needed to explore these crashes.
- Fatal crashes with pedestrians within crosswalk areas are more likely to be associated with locations with marked crosswalk.
- Fatal crashes with pedestrians outside of crosswalk areas are more likely to be associated with locations with no marked crosswalks.

3.3. Association rules mining algorithm

In the transportation safety analysis, the rule-based is a popular modeling approach (Chen et al., 2020; Das et al., 2018; Das et al., 2019b; Kong et al., 2020; Hong et al., 2020; Montella, 2011; Weng and Li, 2019). Association rules are very effective for a large set of unsupervised data by providing key insights that can be used for data-driven decision making. The current study used an 'a priori' algorithm to generate the rules. This algorithm has the applicability of a "bottom up" approach in which frequent subsets are expanded one item at a time by using a breadth-first search following a Hash tree structure (Hahsler et al., 2019; Hahsler, 2020). Consider $I = \{i_1, i_2, i_3, ...i_n,\}$be a set of N distinctive items. Let D be a set of transactions where each transaction T consists of a set of items, such that $T \in I$. Each transaction is associated with only an identifier. An association rule is denoted by a form of Antecedent (left side of the rule) → Consequent (right side of the rule) or A → B, where $A \in I$ and $B \in I$. The conventional measures for association rules are support (Supp or S), confidence (Conf or C), and lift (L). In data mining research studies, many studies proposed new interest measures for rule mining to generate interesting and insightful rules. Lift is the most common performance measure in association rules, and it measures how many more times A and B occur collectively than the amount that would be expected if they were statistically independent. If the value of lift is greater than 1, it indicates that A and B appear more frequently together in the data and are said to be positively dependent on each other.

$$Supp(A) = \frac{n(A)}{n} \quad (1)$$

$$Supp(B) = \frac{n(B)}{n} \quad (2)$$

$$Supp(\mathrm{A}\to\ \mathrm{B}) = \frac{n(AB)}{n} \quad (3)$$

$$Conf(\mathrm{A}\to\mathrm{B}) = \frac{Supp(\mathrm{A}\to\mathrm{B})}{Supp(A)} \quad (4)$$

$$Conf(\mathrm{B}\to\mathrm{A}) = \frac{Supp(\mathrm{B}\to\mathrm{A})}{Supp(B)} \quad (5)$$

$$Lift(\mathrm{A}\to\mathrm{B}) = \frac{Conf(\mathrm{A}\to\mathrm{B})}{Supp(B)} \quad (6)$$

4. Results and discussions

4.1. Scenario-Based rules mining

Key variables are added in the pedestrian crash typing data to develop the final dataset for association rules mining. The final list of

Table 2
Pedestrian fatal-crash scenarios and presence of crosswalks.

No.	Pedestrian Fatal-Crash Scenario (PEDSNR)	Presence of Cross Walk			
		No	Yes	Unknown	Total
1	Pedestrian within crosswalk area, traveled from motorist's left (vehicle straight)	285	925	23	1,233
2	Pedestrian within crosswalk area, traveled from motorist's right (vehicle straight)	229	781	26	1,036
3	Pedestrian outside crosswalk area, traveled from motorist's left (vehicle straight)	674	281	1	956
4	Pedestrian outside crosswalk area, traveled from motorist's right (vehicle straight)	659	243	11	913
5	Pedestrian outside crosswalk area (vehicle turning right)	444	269	63	776
6	Pedestrian outside crosswalk area, approach direction unknown (vehicle straight)	514	178	12	704
7	Unknown/Insufficient Information	293	159	17	469
8	Pedestrian within crosswalk area, approach direction opposite motorist's (vehicle turning left)	47	355	3	405
9	Pedestrian within crosswalk area, approach direction unknown (vehicle straight)	118	258	6	382
10	Pedestrian within crosswalk area, approach direction same as motorist's (vehicle turning left)	21	208	1	230
11	Pedestrian outside crosswalk area, other (vehicle straight)	125	46	2	173
12	Pedestrian within crosswalk area, approach direction unknown (vehicle turning left)	16	118	1	135
13	Pedestrian within crosswalk area, approach direction same as motorist's (vehicle turning right)	9	114	1	124
14	Pedestrian outside crosswalk area, approach direction opposite motorist's (vehicle turning left)	56	42	1	99
15	Pedestrian outside crosswalk area, approach direction unknown (vehicle turning left)	49	28	2	79
16	Pedestrian within crosswalk area, traveled from motorist's right (vehicle turning right)	9	52	2	63
17	Pedestrian within crosswalk area, approach direction unknown (vehicle turning right)	6	49	1	56
18	Pedestrian outside crosswalk area, approach direction same as motorist's (vehicle turning left)	21	25	3	49
19	Pedestrian within crosswalk area, approach direction opposite motorist's (vehicle turning right)	3	33	0	36
20	Pedestrian outside crosswalk area, approach direction unknown (vehicle turning right)	21	13	1	35
21	Pedestrian outside crosswalk area, other (vehicle turning left)	16	6	0	22
22	Pedestrian outside crosswalk area, approach direction opposite motorist's (vehicle turning right)	5	13	0	18
23	Pedestrian within crosswalk area, other (vehicle straight)	12	5	0	17
24	Pedestrian within crosswalk area, traveled from motorist's right (vehicle turning left)	2	15	0	17
25	Pedestrian within crosswalk area, traveled from motorist's left (vehicle turning left)	6	10	0	16
26	Pedestrian outside crosswalk area, approach direction same as motorist's (vehicle turning right)	6	7	1	14
27	Pedestrian within crosswalk area, traveled from motorist's left (vehicle turning right)	8	3	0	11
28	Pedestrian outside crosswalk area, other (vehicle turning right)	4	5	0	9
29	Pedestrian outside crosswalk area, traveled from motorist's right (vehicle turning right)	7	2	0	9
30	Pedestrian within crosswalk area, other (vehicle turning right)	2	3	0	5
31	Pedestrian outside crosswalk area, traveled from motorist's left (vehicle turning left)	3	1	0	4
32	Pedestrian outside crosswalk area, traveled from motorist's left (vehicle turning right)	1	3	0	4
33	Pedestrian outside crosswalk area, traveled from motorist's right (vehicle turning left)	3	1	0	4
34	Pedestrian within crosswalk area, other (vehicle turning left)	0	1	1	2
	Grand Total	3,674	4,252	179	8,105

Table 3
Pedestrian crash scenarios and presence of crosswalks.

Pedestrian Crash Group (PEDCGP)	Presence of Cross Walk			
	No	Yes	Unknown	Total
Crossing Roadway – Vehicle Not Turning	1,789	2,148	73	4,010
Crossing Roadway – Vehicle Turning	256	1,037	17	1,310
Other/Unknown – Insufficient Details	548	385	40	973
Dash/Dart-Out	391	307	20	718
Unusual Circumstances	179	173	11	363
Pedestrian in Roadway – Circumstances Unknown	178	70	8	256
Walking/Running Along Roadway	208	27	6	241
Multiple Threat/Trapped	21	40	0	61
Bus-Related	18	24	1	43
Waiting to Cross	17	24	0	41
Backing Vehicle	26	7	1	34
Working or Playing in Roadway	23	5	0	28
Unique Midblock	10	2	2	14
Driveway Access/ Driveway Access Related	10	3	0	13
Grand Total	3,674	4,252	179	8,105

variables includes pedestrian age, pedestrian gender, presence of crosswalk, pedestrian crash scenario, hour, weather condition, lighting condition, and route category. As explained earlier, the current analysis is limited to four broad pedestrian crash scenarios (scenario 1 to scenario 3 as individual scenario and the rest of the scenarios are considered as scenario 4). For all of these groups, minimum support and minimum confidence are considered as 0.1 and 0.2, respectively. This study used two R packages (arules, arulesViz) to perform the analysis (Hahsler et al., 2019; Hahsler, 2020). Although the explanation of a large set of rules is complicated, this study provides 20 top rules for all four subgroups to provide broader contexts for pedestrian safety improvement.

4.1.1. Rules for scenario 1 [Pedestrians within crosswalk area, traveled from motorist's left (vehicle straight)]

By removing the missing information, this group was based on 1,163 transactions with 44 items (variable categories). By keeping the maximum number of items at 4, a total of 176 rules were produced. The conventional way of explaining data mining results is to select the top rules that can say about the insights in the data. Based on the lift measures, top 20 rules for the selected scenarios (see Table 4 to Table 7) have been explained.

Table 4
Association rules for scenario 1 [pedestrian within crosswalk area, traveled from motorist's left (vehicle straight)]

Rule	Antecedent	Consequent	S	C	L	N
1	PBSEX = Male + PBCWALK = Yes + ROUTE = Local Street Municipality	LGT_COND = Dark Lighted	0.151	0.739	1.171	176
2	PBSEX = Male + PBCWALK = Yes	PBAGE = 55–64	0.108	0.234	1.168	126
3	PBSEX = Male + PBCWALK = Yes + PEDCGP = Crossing Roadway – Vehicle Not Turning + ROUTE = Local Street Municipality	LGT_COND = Dark Lighted	0.123	0.733	1.161	143
4	PBSEX = Male + PBCWALK = Yes + PEDCGP = Crossing Roadway – Vehicle Not Turning	LGT_COND = Dark Lighted	0.269	0.728	1.153	313
5	PBSEX = Male + PEDCGP = Crossing Roadway – Vehicle Not Turning + ROUTE = Local Street Municipality	LGT_COND = Dark Lighted	0.154	0.728	1.152	179
6	PBSEX = Male + PBCWALK = Yes	LGT_COND = Dark Lighted	0.334	0.722	1.143	389
7	PBSEX = Female + PEDCGP = Crossing Roadway – Vehicle Not Turning + LGT_COND = Dark Lighted	PBCWALK = Yes	0.133	0.852	1.139	155
8	PBSEX = Female + PBCWALK = Yes + PEDCGP = Crossing Roadway – Vehicle Not Turning	ROUTE = Local Street Municipality	0.108	0.500	1.139	126
9	PBAGE = 55–64 + PBCWALK = Yes	LGT_COND = Dark Lighted	0.109	0.718	1.136	127
10	PBSEX = Female + LGT_COND = Dark Lighted	ROUTE = Local Street Municipality	0.102	0.498	1.134	119
11	PBSEX = Female + PEDCGP = Crossing Roadway – Vehicle Not Turning	ROUTE = Local Street Municipality	0.134	0.495	1.128	156
12	PBCWALK = Yes + ROUTE = Local Street Municipality	LGT_COND = Dark Lighted	0.241	0.711	1.125	280
13	PBSEX = Male + ROUTE = Local Street Municipality	LGT_COND = Dark Lighted	0.193	0.708	1.121	225
14	PBSEX = Female + LGT_COND = Dark Lighted	PBCWALK = Yes	0.171	0.833	1.114	199
15	PBCWALK = Yes + PEDCGP = Crossing Roadway – Vehicle Not Turning + ROUTE = Local Street Municipality	LGT_COND = Dark Lighted	0.193	0.699	1.107	225
16	PBSEX = Male + PEDCGP = Crossing Roadway – Vehicle Not Turning	PBAGE = 55–64	0.107	0.221	1.103	125
17	PBAGE = 45–54	LGT_COND = Dark Lighted	0.118	0.695	1.101	137
18	PBAGE = 55–64	PBSEX = Male	0.140	0.700	1.097	163
19	PEDCGP = Crossing Roadway – Vehicle Not Turning + LGT_COND = Dark Lighted	PBCWALK = Yes	0.404	0.820	1.097	470
20	PBCWALK = Yes + ROUTE = Local Street Municipality	PBSEX = Female	0.133	0.393	1.090	155

Note: S = Support, C = Confidence, L = Lift, N = Count of Events.

Table 4 shows the top 20 rules generated from the subgroup data based on fatal pedestrian crashes associated with the scenario, 'pedestrian within crosswalk area, traveled from motorist's left (vehicle straight).' The top rule is *'PBSEX = Male + PBCWALK = Yes + ROUTE = Local Street Municipality → LGT_COND = Dark Lighted'* with the measures (Support = 0.151, Confidence = 0.739, Lift = 1.171, and Frequency = 176). It indicates that the proportion of male pedestrian fatalities on local municipality roadways (in the presence of crosswalks) at dark (with lighting) condition is 1.171 times the proportion of all-male pedestrian fatalities on local municipality roadways (in the presence of crosswalks) for the complete data. This finding is in line with Hanson et al. (2013). Some frequent attributes include 'LGT_COND = Dark Lighted,' (14 rules), 'PBCWALK = Yes' (13 rules) 'ROUTE = Local Street Municipality,' (10 rules) 'PBSEX = Male' (8 rules). Female pedestrians are present in six rules, and elder pedestrians (age 55–64 yrs) are present in 4 rules. The rule with the maximum count in Table 4 is *'PEDCGP = Crossing Roadway – Vehicle Not Turning + LGT_COND = Dark Lighted → PBCWALK = Yes'* with rules measures (Support = 0.404, Confidence = 0.820, Lift = 1.097, and Frequency = 470).

4.1.2. Rules for scenario 2 [Pedestrians within crosswalk area, traveled from motorist's right (vehicle straight)]

By removing the missing information, this group was based on 972 transactions with 44 items (variable categories). By keeping the maximum number of items at 4, a total of 183 rules were produced.

Table 5 shows the top 20 rules generated from the subgroup data based on fatal pedestrian crashes associated with the scenario, 'pedestrian within crosswalk area, traveled from motorist's right (vehicle straight).' The top rule is *'PBAGE = 45–54 → LGT_COND = Dark Lighted'* with the measures (Support = 0.117, Confidence = 0.659, Lift = 1.181, and Frequency = 114). It indicates that the fatalities of 45–54 years old pedestrians in the dark (with lighting) condition is 1.181 times the proportion of all fatalities of 45–54 years old pedestrians for the complete data. The results not in line with Hanson et al. (2013). Some frequent attributes include: 'PBSEX = Male' (11 rules), 'ROUTE = Local Street Municipality' (10 rules), 'LGT_COND = Dark Lighted' (10 rules), and 'PBCWALK = Yes' (8 rules). Daylight conditions are present in 4 rules, and female pedestrians are only present in 1 rule. The rule with the maximum count in Table 5 is *'PBSEX = Male + PEDCGP = Crossing Roadway – Vehicle Not Turning → LGT_COND = Dark Lighted'* with rule measures (Support = 0.284, Confidence = 0.613, Lift = 1.099, and Frequency = 276).

4.1.3. Rules for scenario 3 [Pedestrians outside crosswalk area, traveled from motorist's left (vehicle straight)]

By removing the missing information, this group was based on 920 transactions with 44 items (variable categories). By keeping the maximum number of items at 4, a total of 181 rules were produced.

Table 6 shows the top 20 rules generated from the subgroup data based on fatal pedestrian crashes associated with the scenario, 'pedestrian outside crosswalk area, traveled from motorist's left (vehicle straight).' The top rule is *'PBSEX = Male + PEDCGP = Crossing Roadway – Vehicle Not Turning + ROUTE = Local Street Municipality → LGT_COND = Dark Lighted'* with the measures (Support = 0.112, Confidence = 0.757, Lift = 1.298, and Frequency = 103). It indicates that the proportion of male pedestrian fatalities (vehicle not turning related) on local municipality roadways at dark (with lighting) condition is 1.298 times the proportion of all-male pedestrian fatalities (vehicle not turning

Table 5
Association rules for scenario 2 [pedestrian within crosswalk area, traveled from motorist's right (vehicle straight)]

Rule	Antecedent	Consequent	S	C	L	N
1	PBAGE = 45–54	LGT_COND = Dark Lighted	0.117	0.659	1.181	114
2	PEDCGP = Crossing Roadway – Vehicle Not Turning + LGT_COND = Daylight	ROUTE = Local Street Municipality	0.107	0.556	1.174	104
3	PBCWALK = Yes + LGT_COND = Daylight	ROUTE = Local Street Municipality	0.113	0.556	1.173	110
4	LGT_COND = Daylight	ROUTE = Local Street Municipality	0.142	0.554	1.170	138
5	PBSEX = Male + PBCWALK = Yes + PEDCGP = Crossing Roadway – Vehicle Not Turning	LGT_COND = Dark Lighted	0.222	0.645	1.155	216
6	PBSEX = Male + PBCWALK = Yes	LGT_COND = Dark Lighted	0.276	0.626	1.122	268
7	PBSEX = Female + LGT_COND = Daylight	PBCWALK = Yes	0.102	0.839	1.117	99
8	PBSEX = Male + PBCWALK = Yes + ROUTE = Local Street Municipality	LGT_COND = Dark Lighted	0.136	0.623	1.115	132
9	PBAGE = 75–90	PEDCGP = Crossing Roadway – Vehicle Not Turning	0.131	0.852	1.114	127
10	PBSEX = Male + ROUTE = Local Street Municipality	LGT_COND = Dark Lighted	0.179	0.621	1.113	174
11	PBSEX = Male + PBCWALK = Yes + LGT_COND = Dark Lighted + ROUTE = Local Street Municipality	PEDCGP = Crossing Roadway – Vehicle Not Turning	0.115	0.848	1.109	112
12	PBSEX = Male + PEDCGP = Crossing Roadway – Vehicle Not Turning	PBAGE = 55–64	0.106	0.229	1.100	103
13	PBSEX = Male + PEDCGP = Crossing Roadway – Vehicle Not Turning	LGT_COND = Dark Lighted	0.284	0.613	1.099	276
14	PBCWALK = Yes + LGT_COND = Dark Lighted + ROUTE = Local Street Municipality	PEDCGP = Crossing Roadway – Vehicle Not Turning	0.174	0.837	1.093	169
15	PBSEX = Male + LGT_COND = Dark Lighted + ROUTE = Local Street Municipality	PEDCGP = Crossing Roadway – Vehicle Not Turning	0.149	0.833	1.089	145
16	PEDCGP = Crossing Roadway – Vehicle Not Turning + LGT_COND = Dark Lighted	ROUTE = Local Street Municipality	0.223	0.514	1.085	217
17	PBSEX = Male + PBCWALK = Yes + PEDCGP = Crossing Roadway – Vehicle Not Turning	ROUTE = Local Street Municipality	0.177	0.513	1.084	172
18	PEDCGP = Crossing Roadway – Vehicle Not Turning + ROUTE = State Highway	PBCWALK = Yes	0.133	0.811	1.081	129
19	PBSEX = Male	PBCWALK = None	0.146	0.243	1.076	142
20	PBCWALK = None + PEDCGP = Crossing Roadway – Vehicle Not Turning	PBSEX = Male	0.109	0.646	1.073	106

related) on local municipality roadways for the complete data. Some of the frequent attributes are: 'LGT_COND = Dark Lighted' (11 rules), 'ROUTE = Local Street Municipality' (8 rules), 'PBSEX = Male' (8 rules), and 'PEDCGP = Crossing Roadway' (8 rules). In 3 rules with 'LGT_COND = Dark Not Lighted', the common factor is the absence of marked crosswalks. The rule with the maximum count in Table 6 is *'ROUTE = Local Street Municipality → LGT_COND = Dark Lighted'* with the measures (Support = 0.226, Confidence = 0.658, Lift = 1.128., and Frequency = 208).

4.1.4. *Rules for scenario 4 (Other Scenarios)*

By removing the missing information, this group was based on 4,612 transactions with 44 items (variable categories). By keeping the maximum number of items at 4, a total of 168 rules were produced. Table 7 shows the top 20 rules generated from the subgroup data based on fatal pedestrian crashes associated with the rest cases. The top rule is *'LGT_COND = Daylight → PEDCGP = Crossing Roadway – Vehicle Turning'* with the measures (Support = 0.108, Confidence = 0.384, Lift = 2.373, and Frequency = 828). It indicates that the pedestrian fatalities at daylight condition when the vehicle is turning is 2.373 times the proportion of all fatalities at daylight condition for the complete data. Some frequent attributes include: 'PBCWALK = Yes' (10 rules), 'LGT_COND = Dark Lighted' (9 rules), 'PEDCGP = Crossing Roadway – Vehicle Not Turning' (9 rules), and 'ROUTE = Local Street Municipality' (7 rules). Male pedestrians are present in 6 rules. The rule with the maximum count in Table 7 is *'PEDCGP = Crossing Roadway – Vehicle Not Turning → LGT_COND = Dark Lighted'* with rules measures (Support = 0.298, Confidence = 0.595, Lift = 1.181, and Frequency = 2,281).

5. Conclusions

Intersection and intersection-related pedestrian fatalities contribute nearly one-fourth of all pedestrian fatalities in the US. To clearly understand the factors influencing fatal pedestrian crashes for the development of targeted countermeasures geared toward dealing with this menace, it is imperative to investigate the crash contributory factors affecting pedestrian crashes. To this end, this study contributed to the literature on pedestrian safety by investigating and providing practical knowledge on the patterns and associations between crash contributory factors influencing fatal pedestrian crashes occurring at intersections using a robust data mining technique while considering the unique pedestrian crash typing. To achieve the objective of this study, the FIRST data with pedestrian typing information collected over a five-year period was used in the analysis. Although this data contains a detailed list of pedestrian crash scenarios and pedestrian crash groups, intersection and intersection-related pedestrian fatalities are limited to a handful of crash scenarios and pedestrian crash groups. As the collected

Table 6
Association rules for scenario 3 [Pedestrian outside crosswalk area, traveled from motorist's left (vehicle straight)]

Rule	Antecedent	Consequent	S	C	L	N
1	PBSEX = Male + PEDCGP = Crossing Roadway – Vehicle Not Turning + ROUTE = Local Street Municipality	LGT_COND = Dark Lighted	0.112	0.757	1.298	103
2	PBSEX = Male + PBCWALK = None + ROUTE = Local Street Municipality	LGT_COND = Dark Lighted	0.112	0.730	1.252	103
3	PBSEX = Male + ROUTE = Local Street Municipality	LGT_COND = Dark Lighted	0.161	0.715	1.225	148
4	PBCWALK = None + PEDCGP = Crossing Roadway – Vehicle Not Turning	LGT_COND = Dark Not Lighted	0.141	0.287	1.211	130
5	PBSEX = Male + PBCWALK = None	LGT_COND = Dark Not Lighted	0.137	0.284	1.198	126
6	LGT_COND = Dark Not Lighted	PBCWALK = None	0.197	0.830	1.188	181
7	PBCWALK = None + PEDCGP = Crossing Roadway – Vehicle Not Turning + LGT_COND = Dark Lighted	ROUTE = Local Street Municipality	0.113	0.406	1.183	104
8	PBCWALK = None + LGT_COND = Dark Lighted	ROUTE = Local Street Municipality	0.158	0.406	1.182	145
9	PBSEX = Male + PBCWALK = None	PBAGE = 45–54	0.115	0.239	1.181	106
10	PEDCGP = Crossing Roadway – Vehicle Not Turning + ROUTE = Local Street Municipality	LGT_COND = Dark Lighted	0.159	0.682	1.169	146
11	PBCWALK = None + PEDCGP = Crossing Roadway – Vehicle Not Turning	PBAGE = 45–54	0.115	0.234	1.157	106
12	PBSEX = Male + PBCWALK = Yes	LGT_COND = Dark Lighted	0.141	0.674	1.154	130
13	PBSEX = Male + PEDCGP = Crossing Roadway – Vehicle Not Turning	PBAGE = 45–54	0.110	0.232	1.146	101
14	PBSEX = Female	ROUTE = Local Street Municipality	0.118	0.389	1.133	109
15	ROUTE = Local Street Municipality	LGT_COND = Dark Lighted	0.226	0.658	1.128	208
16	PBCWALK = Yes + PEDCGP = Crossing Roadway – Vehicle Not Turning	LGT_COND = Dark Lighted	0.130	0.649	1.111	120
17	PBCWALK = Yes	LGT_COND = Dark Lighted	0.195	0.649	1.111	179
18	LGT_COND = Dark Lighted + ROUTE = State Highway	PBSEX = Male	0.113	0.765	1.103	104
19	PBAGE = 65–74	PBCWALK = None	0.102	0.758	1.085	94
20	PBSEX = Male + PEDCGP = Crossing Roadway – Vehicle Not Turning	ROUTE = State Highway	0.138	0.291	1.072	127

Table 7
Association rules for scenario 4 (other scenarios).

Rule	Antecedent	Consequent	S	C	L	N
1	LGT_COND = Daylight	PEDCGP = Crossing Roadway – Vehicle Turning	0.108	0.384	2.373	828
2	LGT_COND = Dark Not Lighted	PBCWALK = None	0.116	0.719	1.589	891
3	PBCWALK = Yes	PEDCGP = Crossing Roadway – Vehicle Turning	0.128	0.244	1.509	985
4	PBCWALK = Yes + ROUTE = Local Street Municipality	LGT_COND = Daylight	0.103	0.402	1.430	792
5	PBSEX = Male + PBCWALK = Yes + PEDCGP = Crossing Roadway – Vehicle Not Turning	LGT_COND = Dark Lighted	0.115	0.686	1.361	883
6	PBCWALK = Yes + PEDCGP = Crossing Roadway – Vehicle Not Turning	LGT_COND = Dark Lighted	0.171	0.642	1.275	1313
7	PEDCGP = Crossing Roadway – Vehicle Not Turning + ROUTE = Local Street Municipality	LGT_COND = Dark Lighted	0.132	0.631	1.252	1015
8	PBSEX = Male + PEDCGP = Crossing Roadway – Vehicle Not Turning	LGT_COND = Dark Lighted	0.205	0.628	1.246	1570
9	PBSEX = Female + ROUTE = Local Street Municipality	PBCWALK = Yes	0.112	0.648	1.231	858
10	LGT_COND = Daylight	ROUTE = Local Street Municipality	0.153	0.544	1.228	1172
11	LGT_COND = Daylight	PBCWALK = Yes	0.178	0.634	1.206	1367
12	PEDCGP = Crossing Roadway – Vehicle Not Turning	LGT_COND = Dark Lighted	0.298	0.595	1.181	2281
13	PBSEX = Male + PBCWALK = Yes	LGT_COND = Dark Lighted	0.183	0.593	1.178	1401
14	PBSEX = Male + ROUTE = Local Street Municipality	LGT_COND = Dark Lighted	0.155	0.573	1.138	1185
15	PBCWALK = None + LGT_COND = Dark Lighted	PEDCGP = Crossing Roadway – Vehicle Not Turning	0.121	0.561	1.122	930
16	PBCWALK = None + LGT_COND = Dark Lighted	PBSEX = Male	0.151	0.698	1.111	1157
17	PEDCGP = Crossing	PBCWALK = Yes	0.122	0.582	1.106	936

(continued on next page)

Table 7 (*continued*)

Rule	Antecedent	Consequent	S	C	L	N
	Roadway – Vehicle Not Turning + ROUTE = Local Street Municipality					
18	ROUTE = Local Street Municipality	PBCWALK = Yes	0.257	0.580	1.103	1969
19	PBSEX = Male + PBCWALK = Yes	PEDCGP = Crossing Roadway – Vehicle Not Turning	0.168	0.545	1.091	1288
20	ROUTE = State Highway	PEDCGP = Crossing Roadway – Vehicle Not Turning	0.128	0.539	1.078	980

data format is categorical in nature, it is useful to utilize data mining techniques to identify insight rules. Instead of conducting rules mining on the complete dataset, this study grouped the data into four major groups (based on the frequency of pedestrian crash scenarios) to develop association rules for individual groups. Applying ‘a priori’ algorithm and ‘lift’ as the performance measures, the top 20 rules (based on lift values) for all four groups are listed and explained. Some of the key variable categories such as dark with lighting at intersections, vehicle going straight, vehicle turning, local municipality streets, pedestrian age range from 45 years and above, male pedestrians, intersections at local municipality roadways are frequently present in the generated rules. It is also found that intersections without lighting at dark segments associated with fatal crashes occurred in locations without marked crosswalks. As association rules mining provide quantitative measures in the form of lift or likelihood, the risk factors can be transferred into data-driven decision making.

The study has some limitations. First, this study is limited to pedestrian crash typing variables with the addition of a few more person-level variables. Additional contributing factors can be added to generate a more robust study. Second, the current study did not apply any optimization technique to determine the minimum threshold of support and confidence for any of the groups. This was done by performing trial and error. Third, the current study is focused on rule generation and association pattern determination. The findings are not addressed with respect to different interventions and countermeasures. Fourth, the study is limited to fatal crashes only. There are no available data sources that can provide state level all crash information as well as there is no viable source that can provide exposure metrics of pedestrians. Availability of these critical data can make this study more robust. Searching ‘pedestrian intersection’ in crash modification factor or CMF Clearinghouse data (https://rpubs.com/subasish/506776) generates 169 relevant CMFs for different countermeasures. Recommendation of countermeasures requires careful observation of the scenarios before recommendation, which is currently out of the scope of this study. Future studies can address the limitations of the current study.

CRediT authorship contribution statement

Subasish Das: Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. **Reuben Tamakloe:** Methodology, Writing – original draft, Writing – review & editing. **Hamsa Zubaidi:** Methodology, Writing – review & editing. **Ihsan Obaid:** Methodology, Writing – review & editing. **Ali Alnedawi:** Methodology, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We like to thank three anonymous reviewers for their excellent feedback.

References

Agrawal, R., Srikant, R., 2005. Fast Algorithms for Mining Association Rules. Proc. 20th int. conf. very large data bases, VLDB 487–499.

Anderson, R.W., McLean, A., Farmer, M.J., Lee, B., Brooks, C., 1997. Vehicle travel speeds and the incidence of fatal pedestrian crashes. Accid. Anal. Prev. 29 (5), 667–674. https://doi.org/10.1016/S0001-4575(97)00036-5.

Aziz, H.M.A., Ukkusuri, S.V., Hasan, S., 2013. Exploring the determinants of pedestrian–vehicle crash severity in New York City. Accid. Anal. Prev. 50, 1298–1309. https://doi.org/10.1016/j.aap.2012.09.034.

Chen, L.u., Huang, S., Yang, C., Chen, Q., 2020. Analyzing factors that influence expressway traffic crashes based on association rules: using the shaoyang-xinhuang section of the shanghai-kunming expressway as an example. J. Transp. Eng., Part A: Systems 146 (9), 05020007. https://doi.org/10.1061/JTEPBS.0000425.

Das, S., Dutta, A., Avelar, R., Dixon, K., Sun, X., Jalayer, M., 2019. Supervised association rules mining on pedestrian crashes in urban areas: identifying patterns for appropriate countermeasures. Int. J. Urban Sci. 23 (1), 30–48. https://doi.org/10.1080/12265934.2018.1431146.

Das, S., Kyle, L., Avelar, R., Dixon, K., Bommanayakanahalli B. Improper Passing-Related Crashes on Rural Roadways: Using Association Rules Negative Binomial Miner. Presented at the TRB Annual Meeting, Washington DC, 2018.

Das, S., Bibeka, A., Sun, X., "Tracy" Zhou H., Jalayer, M. Elderly. Pedestrian Fatal Crash-Related Contributing Factors: Applying Empirical Bayes Geometric Mean Method. Transportation Research Record: Journal of the Transportation Research Board, Vol. 2673, No. 8, 2019a, p. pp 254-263.

Das, S., Dutta, A., Sun, X., 2020. Patterns of rainy weather crashes: applying rules mining. J. Trans. Safety Secur. 12 (9), 1083–1105.

Das, S., Sun, X., 2014. Investigating the Pattern of Traffic Crashes under Rainy Weather by Association Rules in Data Mining. Transportation Research Board 93rd Annual Meeting. Transportation Research Board, Washington DC.

Dovom, H.Z., Saffarzadeh, M., Dovom, M.Z., Nadimi, N., 2012. An analysis of pedestrian fatal accident severity using a binary logistic regression model. ITE J. 82 (4), 38–43.

Gårder, P.E., 2004. The impact of speed and other variables on pedestrian safety in Maine. Accid. Anal. Prev. 36 (4), 533–542. https://doi.org/10.1016/S0001-4575(03)00059-9.

Hahsler, M., Buchta, C., Gruen, B., Hornik K. (2019). arules: Mining Association Rules and Frequent Itemsets. R package version 1.6-4. https://CRAN.R-project.org/package=arules Accessed: August 1, 2020.

Hahsler, M. arulesViz: Visualizing Association Rules and Frequent Itemsets. R package version 1.3-3. https://CRAN.R-project.org/package=arulesViz Accessed: August 1, 2020.

Harkey, D., Tsai, S., Thomas, L., Hunter, W. Pedestrian and Bicycle Crash Analysis Tool (PBCAT): Version 2.0 Application Manual. Report No. FHWA-HRT-06-089, 2006.

Henary, B.Y., Crandall, J., Bhalla, K., Mock, C.N., Roudsari, B.S., 2003. Child and adult pedestrian impact: the influence of vehicle type on injury severity, in: 47th Annual Proceedings of the Association for the Advancement of Automotive Medicine. Lisbon, Portugal, pp. 105–126.

Hong, J., Tamakloe, R., Park, D., 2020a. Discovering Insightful Rules among Truck Crash Characteristics using Apriori Algorithm. J. Adv. Transp. 2020, 1–16. https://doi.org/10.1155/2020/4323816.

Hong, J., Tamakloe, R., Park, D., 2020b. Application of association rules mining algorithm for hazardous materials transportation crashes on expressway. Accid. Anal. Prev. 142, 105497. https://doi.org/10.1016/j.aap.2020.105497.

Jang, K., Park, S.H., Kang, S., Song, K.H., Kang, S., Chung, S., 2013. Evaluation of Pedestrian Safety. Transp. Res. Rec. J. Transp. Res. Board 2393 (1), 104–116. https://doi.org/10.3141/2393-12.

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. https://doi.org/10.1016/j.aap.2020.105620.

Li, D., Ranjitkar, P., Zhao, Y., Yi, H., Rashidi, S., 2017. Analyzing pedestrian crash injury severity under different weather conditions. Traffic Inj. Prev. 18 (4), 427–430.

Li, L., Yang, X., Yin, L., 2010. Exploration of Pedestrian Refuge Effect on Safety Crossing at Signalized Intersection. Transp. Res. Rec.: Journal of the Transportation Research Board, No. 2193, 44–50.

Ma, Z., Lu, X., Chien, S.I.-J., Hu, D. Investigating Factors Influencing Pedestrian Injury Severity at Intersections. Traffic Injury Prevention, Vol. 19, No. 2, 2018, p. pp 159-164.

Montella, A., 2011. Identifying crash contributory factors at urban roundabouts and using association rules to explore their relationships to different crash types. Accid. Anal. Prev. 43 (4), 1451–1463.

NHTSA. Traffic Safety Facts: Pedestrian 2018 Data. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812850 Accessed: November 16, 2020.

Olszewski, P., Szagala, P., Wolanski, M., Zielinska, A. Pedestrian Fatality Risk in Accidents at Unsignalized Zebra Crosswalks in Poland. Accident Analysis & Prevention, Vol. 84, 2015, p. pp 83-91.

Prato, C.G., Gitelman, V., Bekhor, S. Mapping Patterns of Pedestrian Fatal Accidents in Israel. Accident Analysis & Prevention, Vol. 44, No. 1, 2012, p. pp 56-62.

Schneider, R.J., Stefanich, J. Application of the Location–Movement Classification Method for Pedestrian and Bicycle Crash Typing. Transportation Research Record: Journal of the Transportation Research Board, No. 2601, 2016, p. pp 72–83.

Spainhour, L.K., Wootton, I.A. Modeling Fault in Fatal Pedestrian Crashes by Using Various Data Sources. Transportation Research Record: Journal of the Transportation Research Board, No. 2002, 2007, p. pp 64-71.

Sun, M., Sun, X., Shan, D., 2019. Pedestrian crash analysis with latent class clustering method. Accid. Anal. Prev. 124, 50–57. https://doi.org/10.1016/j.aap.2018.12.016.

Tamakloe, R., Hong, J., Park, D., 2020. A copula-based approach for jointly modeling crash severity and number of vehicles involved in express bus crashes on expressways considering temporal stability of data. Accid. Anal. Prev. doi:10.1016/j.aap.2020.105736.

Uddin, M., Ahmed, F. Pedestrian Injury Severity Analysis in Motor Vehicle Crashes in Ohio. Safety, Vol. 4, No. 2, 2018.

Weng, J., Li, G., 2019. Exploring shipping accident contributory factors using association rules. J. Transp. Safety Secur. 11 (1), 36–57.

Xie, X., Nikitas, A., Liu, H. A Study of Fatal Pedestrian Crashes at Rural Low-Volume Road Intersections in Southwest China. Traffic Injury Prevention, Vol. 19, No. 3, 2018, p. pp 298-304.

Xin, C., Guo, R., Wang, Z., Lu, Q., Lin, P.-S. The Effects of Neighborhood Characteristics and the Built Environment on Pedestrian Injury Severity: A Random Parameters Generalized Ordered Probability Model with Heterogeneity in Means and Variances. Analytic Methods in Accident Research, Vol. 16, 2017, p. pp 117-132.

Yu, S., Jia, Y., Sun, D., 2019. Identifying factors that influence the patterns of road crashes using association rules: A case study from Wisconsin. United States. Sustain. 11 (7), 1–14. https://doi.org/10.3390/su11071925.

Zhai, X., Huang, H., Sze, N.N., Song, Z., Hon, K.K., 2019. Diagnostic analysis of the effects of weather condition on pedestrian crash severity. Accid. Anal. Prev. 122, 318–324. https://doi.org/10.1016/j.aap.2018.10.017.