Accident Analysis and Prevention 194 (2024) 107359

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Crashes involving distracted pedestrians: Identifying risk factors and their relationships to pedestrian severity levels and distraction modes

Md Mahmud Hossain [a,*,1], Huaguo Zhou [a,2], Xiaoduan Sun [b,3], Ahmed Hossain [b,4], Subasish Das [c,5]

[a] Department of Civil and Environmental Engineering, Auburn University, Ramsay Hall, Auburn, AL 36849-5337, USA
[b] Department of Civil Engineering, University of Louisiana at Lafayette, 131 Rex Street, Lafayette, LA 70504, USA
[c] Department of Civil Engineering, Texas State University, 601 University Drive, TX 78666, USA

ARTICLE INFO

Keywords:
Pedestrian distraction
Crash data analysis
Pedestrian severity
Association rule mining
Crash narrative

ABSTRACT

The concept of distracted pedestrians and its impact on highway safety has gained increasing attention in recent years. However, studies focusing exclusively on distracted pedestrian crashes are less pervasive than distracted driving. In addition, most prior studies investigate the harmful effect of cellphone usage while walking, without considering other forms of pedestrian distraction. Also, the existing literature provides limited knowledge on comprehending the affinities between pedestrian distraction and safety consequences. This study aims to reveal the chain of contributing factors involved in distracted pedestrian crashes, considering both pedestrian severity levels and specific distraction-related tasks. Ten years (2010–2019) of related crashes were extracted from the Louisiana Department of Transportation and Development (LADOTD) database, and association rule mining (ARM) was applied to identify the meaningful crash patterns. Different distracting activities of pedestrians were introduced from the narratives of police-investigated crash reports. The study findings exhibit the complex nature of distracted pedestrian crashes by highlighting the intricate relationships between risk factors. On road segments, distracted male pedestrians aged 41–64 were more likely to be fatal/severely injured in dark-not-lighted conditions. Crashes involving pedestrians using electronic devices were often found at intersections. Distractions caused by pets, persons, or objects were strongly associated with crossing segments in rural settings. In-person conversation while standing on roadways in urban residential locations without traffic controls was found to increase vulnerability. Working on vehicles while wearing dark clothes and in dark-not-lighted conditions was identified as an influential factor in crash occurrence. Moreover, careless or inattentive actions of pedestrians while playing on the road segments were associated with a high likelihood of crashes. These study outcomes are crucial in uncovering the coexisting crash characteristics related to distracted pedestrians, which can be helpful in targeting and developing effective educational, design, and enforcement strategies to improve pedestrian safety.

1. Introduction

According to the National Highway Traffic Safety Administration (NHTSA), a pedestrian is a person engaged in activities such as walking, running, jogging, hiking, sitting, or lying on the road during crashes (NHTSA, 2020). This definition excludes individuals traveling on tiny wheels such as scooters, roller skates, baby strollers, and skateboards, as well as motorized and non-motorized wheelchairs. The interaction between pedestrians and motor vehicles poses a serious safety concern due to the heightened risk of pedestrian death and severe injuries. Between

* Corresponding author.
E-mail addresses: mahmud@auburn.edu (M.M. Hossain), hhz0001@auburn.edu (H. Zhou), xsun@louisiana.edu (X. Sun), ahmed.hossain1@louisiana.edu (A. Hossain), subasish@txstate.edu (S. Das).
[1] ORCID: 0000-0002-2737-6951.
[2] ORCID: 0000-0001-5210-036X.
[3] ORCID ID: 0000-0001-7282-1340.
[4] ORCID: 0000-0003-1566-3993.
[5] ORCID: 0000-0002-1671-2753.

https://doi.org/10.1016/j.aap.2023.107359
Received 6 October 2022; Received in revised form 13 June 2023; Accepted 16 October 2023
Available online 1 November 2023

2010 and 2019, pedestrian fatalities in the U.S. increased by about 45 % (NHTSA, 2020), highlighting the need to explore contributing factors associated with pedestrian crashes. Apart from factors like poor road design and hazardous driving, pedestrian conditions such as impairment and distraction can significantly contribute to collision risks (Thompson et al., 2013). The concept of distracted pedestrians and its impact on highway safety has become a growing issue in recent years. Generally, pedestrian distraction can be defined as shifting the attention away from surrounding events critical for safe vehicle interaction (Zhu et al., 2022). Distraction modes for pedestrians include talking or texting on cellphones, listening to music, engaging in in-person conversations while on the street, and more. In 2015, the American Academy of Orthopedic Surgeons (AAOS) reported that approximately 70 % of millennials (aged 18 to 34) and 81 % of adults (aged over 35) considered distracted walking a critical pedestrian safety issue (AAOS, 2015). Furthermore, several observational studies have conveyed the frequent involvement of pedestrians in distraction-related tasks (Schwebel et al., 2022), particularly while crossing the street (Thompson et al., 2013). Although such activities may only last a few seconds (Jiang et al., 2018), pedestrians can lose awareness of their surroundings during critical times, resulting in deadly outcomes or severe injuries (Mwakalonge et al., 2015). The National Safety Council's (NSC) Injury Facts 2015 indicated distracted walking as a significant safety threat subject to the increasing trend of related injuries from 2000 to 2011.

The effect of distraction on pedestrians shares some similarities to those experienced by distracted drivers (Hyman et al., 2010); however, the influence of distraction on safe walking behavior has not been extensively studied. Distracted pedestrians often allocate greater visual attention to the front rather than the periphery of their surroundings (Tapiro et al., 2020). For example, a pedestrian engaged in a video call on a cellphone concentrates less on the surroundings and therefore fails to detect salient objects while crossing the street. In relation to texting during walking, such a reduction in situational awareness could increase the risk of crashes by nearly four times compared to non-distracted walkers (Mwakalonge et al., 2015). It is not only cellphone usage that can distract pedestrians; other distractions like carrying items, dropped objects, or roadside activities can also divert their attention from the roadway environment. From a road safety perspective, crashes can be caused by a combination of multiple factors, including road, environment, vehicle, and human characteristics (Haddon, 1980; Hossain et al., 2021a, 2023a, 2023b). Following the concept, crashes involving distracted pedestrians have not been exclusively investigated. This study aims to reveal the chain of contributing factors involved in distracted pedestrian crashes considering both pedestrian severity levels and specific distraction-related tasks. A dataset of pedestrian crashes spanning ten years (2010–2019) was extracted from Louisiana, and association rule mining (ARM) was applied to identify the associated crash patterns. The narratives of police-investigated crash reports were reviewed to identify the distracting activities of pedestrians prior to the collisions. It is worth noting that Louisiana ranks sixth among the top 20 worst states for pedestrian safety (SGA, 2021). In 2019, Louisiana reported approximately 1,594 pedestrian crashes, marking a 22 % increase from 2010. The jurisdiction is actively working on improving pedestrian safety as part of its Strategic Highway Safety Plan (SHSP) to achieve the goal of 'Destination Zero Deaths'.

2. Literature review

Limited data availability and challenges in identifying modes of pedestrian distraction have contributed to the scarcity of studies solely focused on distracted pedestrian crashes, as compared to the extensive research on distracted driving (Mwakalonge et al., 2015; Stavrinos et al., 2018). Nevertheless, several observational, survey-based, and experimental studies have demonstrated the potential impact of distracted walking on safe interactions between pedestrians and vehicles. This section will discuss potential risk factors associated with pedestrian distraction, which can be identified and derived from police-investigated crash reports.

Children and youths are more likely to engage in distracting activities as pedestrians (Nasar and Troyer, 2013; Tapiro et al., 2020). While distracted, their ability to maintain consistent visual scanning patterns is reduced (Tabibi and Pfeffer, 2003); therefore, they often miss the crucial traffic information necessary for safe road crossings (Meir et al., 2015). In complex road crossing scenarios, older pedestrians (aged over 60) become more vulnerable to dual-task actions due to the slower cognitive processing associated with aging (Mathias et al., 2011). Although male pedestrians demonstrate a higher involvement rate in overall distracted walking incidents (Mwakalonge et al., 2015), the prevalence of specific distracting tasks may vary significantly by gender. For example, females are more likely to talk on cellphones while walking compared to males (Russo et al., 2018). However, multiple studies have reported no gender-specific behavioral variations in the context of crosswalk violations associated with dual-task impairments (Russo et al., 2018; Schwebel et al., 2012). Walkers tend to exhibit greater risky behavior while distracted by electronic devices (Stavrinos et al., 2011; Tapiro and Oron-Gilad, 2016). For instance, manipulating a cellphone could increase intersection crossing time by up to 18 % (Thompson et al., 2013), thereby increasing the risk of intersection crashes, particularly when drivers endeavor to make left or right turns (Russo et al., 2018). Conversely, certain forms of pedestrian distraction are highly coupled with unexpected changes in walking speed, such as chasing pets and running to catch buses or trains (Mohammed, 2021).

Distracted walking is predominately high on weekdays and during business hours (Osborne et al., 2020). However, Schwebel et al. (2022) suggested that further research is needed to investigate changes in distracted pedestrian rates throughout different times of the day, particularly during evening and nighttime hours. Wearing dark clothes at night could be vulnerable for pedestrians (Hossain et al. 2022), directly affecting visibility (Rasouli et al., 2019). In these circumstances, engaging in divided attention tasks can significantly increase the probability of pedestrian-involved collisions. Most observational and simulated studies exploring the impact of distraction on pedestrian safety concentrate on the features of urban environments (Schwebel et al., 2022) and commercial areas with high pedestrian volumes (Miranda-Moreno et al., 2011); therefore, the patterns of pedestrian distraction in residential and rural areas are not well understood. Distracting activities promote risky walking behaviors such as slower walking pace and crossing outside of designated crosswalks. However, the associated vulnerability can be mitigated through the vigilance and effective decision-making of drivers (Russo et al., 2018). The crash risk for distracted pedestrians may vary depending on the presence of traffic controls such as pavement markings, warning signs, and pedestrian signals, as these devices enhance pedestrian visibility and increase driver attention when entering crosswalks (Lu and Noyce, 2009).

Data mining is an innovative approach that utilizes statistical learning, modeling theories, and effective database management to uncover significant and complex patterns in crash datasets (Fayyad et al., 1996). Traditional statistical parametric models, which establish relationships based on the distributions of independent and dependent variables, have limitations in untangling intricate associations among multiple variables in high-dimensional datasets. As the number of variables increases, these models become less effective and may yield invalid results due to scatteredness. Unlike parametric models, association rule mining (ARM) does not require predetermined assumptions or functional forms (Kumbhare, 2014). Therefore, it has been successfully employed in various transportation safety domains to uncover hidden patterns of interest (Hossain et al., 2021b; Rahman et al., 2023; Das et al., 2022). In relation to crash data analysis, the use of ARM to uncover pedestrian crash patterns has gained much popularity (Das et al., 2021; Hossain A. et al., 2022; Montella et al., 2011; Sivasankaran et al., 2020; Rella Riccardi et al., 2022). For instance, Montella et al. (2011) used ARM to explore the circumstances influencing pedestrian crashes at different pedestrian severity levels. ARM has also been applied to reveal crash scenarios of fatal pedestrian crashes at intersections (Das et al., 2021), as a significant portion of pedestrian crashes occur in or near intersection locations. In addition, ARM has proven useful in analyzing ten years of crash data from Louisiana to identify patterns of pedestrian crashes under different road lighting conditions (Hossain A. et al., 2022). The discovered rules of these studies describe relationships between variables in various circumstances without constraining the nature of the variables. It is noteworthy that ARM has the potential to serve as a decision support tool for road safety agencies, allowing them to target specific crash issues by utilizing its capability to identify coexisting crash characteristics (Pande and Abdel-Aty, 2009). This data mining technique can be applied even when there are only a small number of observations (Ashraf et al., 2021; Hossain et al., 2022).

Enhancing geographical coverage in observational and survey-based studies can often be challenging, while experimental studies conducted in controlled surroundings may only partially reflect real-world walking environments. Consequently, the existing literature provides limited understanding of the relationship between pedestrian distraction and safety outcomes, with a focus on the detrimental effects of cellphone usage during walking rather than considering various sources of distraction. Addressing this research gap, the present study used ten years (2010–2019) of Louisiana crash information 1) to identify the crash scenarios involving distracted pedestrians based on pedestrian severity levels and 2) to explore the association between crash contributing factors and distraction-related tasks, derived from narratives in police-investigated crash reports. As per the authors' knowledge, this study developed the largest dataset on crashes involving distracted pedestrians, encompassing diverse distracting activities performed prior to the crashes. ARM was employed to analyze the crashes. As discussed earlier, this unsupervised data mining method can deal with both large and small datasets with a multitude of factors to interpret the variable interrelations without compromising the actual size of original dataset (Hossain A. et al., 2023). The findings of this research are crucial in revealing the interconnected crash characteristics associated with distracted pedestrians, providing valuable insights for developing effective strategies to enhance pedestrian safety through education, design, and enforcement measures.

3. Methodology

Association rules mining (ARM) identifies frequent itemsets (i.e., collection of variable attributes) that occur together in an event (i.e., individual pedestrian crash) (Agrawal et al., 1993). This data mining technique can readily interpret the association of factors without any predetermined inferences (Pande and Abdel-Aty, 2009). Three commonly used algorithms in ARM are the apriori algorithm, frequent pattern tree algorithm, and constraint-based ARM (Kaur and Madan, 2015). The apriori algorithm is known for its straightforward approach to mining association rules from datasets (Agrawal and Srikant, 1994). This algorithm stands out due to its robust candidate generation method and the use of a pruning technique that enhances efficiency compared to other approaches. Additionally, the apriori algorithm has the advantage of avoiding the unnecessary counting of infrequent candidate itemsets, leading to reduced computation and memory requirements (Hong et al., 2020). In this study, the apriori algorithm was utilized to identify the crash patterns of distracted pedestrians based on their degree of injury severity and sources of distraction. This algorithm of ARM has been increasingly employed by researchers as a decision support tool to discover association rules from multidimensional crash databases, focusing on specific categories of variables (Hossain et al., 2021b, 2022).

Let $K = \{k_1, k_2, k_3, \cdots, k_n\}$ be a set of pedestrian crash database and each observation in K contains a subset of items (a set of variable attributes) in itemset, $J = \{j_1, j_2, j_3, \cdots, j_n\}$. A rule has the form $P \rightarrow Q$ where $P, Q \subseteq J$ and $P \cap Q = \emptyset$. P is the antecedent (left hand side-LHS) and Q is the consequent (right hand side-RHS). In an n-itemset rule, it is possible to have multiple items as the antecedent. For instance, consider a 3-itemset rule such as {lighting condition = dark-lighted, pedestrian action = crossing intersection} → {pedestrian severity = fatal}, where P = {lighting condition = dark-lighted, pedestrian action = crossing intersection} and Q = {pedestrian severity = fatal}. It should be noted that these generated rules imply only interdependencies among factors instead of direct causation. The rules are filtered by three parameters-support (S), confidence (C), and lift (L). The 'support' measures the frequency of occurrence of a specific rule or pattern ($P \rightarrow Q$) in the entire dataset, while 'confidence' represents the proportion of how frequently the pattern $P \rightarrow Q$ occurs together relative to the number of times P occurs in the dataset. The third parameter 'lift' indicates the degree to which items are associated with independent crash events. The equations for these parameters are as follows:

$$Support(P) = \frac{P^{'}}{N}$$

$$Support(Q) = \frac{Q^{'}}{N}$$

$$Support(P \rightarrow Q) = \frac{(P^{'} \cap Q^{'})}{N}$$

$$Confidence(P \rightarrow Q) = \frac{Support(P \rightarrow Q)}{Support(P)}$$

$$Lift(P \rightarrow Q) = \frac{Support(P \rightarrow Q)}{Support(P) \times Support(Q)}$$

Here, N is the number of pedestrian crashes, $P^{'}$ = frequency of occurrences with P, $Q^{'}$ = frequency of occurrences with Q, $(P^{'} \cap Q^{'})$ = frequency of occurrences with both P and Q. Fig. 1 provides a hypothetical example demonstrating how support, confidence, and lift are estimated using the ARM approach.

In the context of association rule discovery, 'lift' plays a crucial role in assessing the strength of a rule by indicating the co-occurrence of antecedent on the conditional likelihood of consequent (Kong et al., 2020). A lift value greater than 1 indicates a positive interrelation between P and Q, whereas a value less than 1 suggests a negative correlation between P and Q. A lift value close to 1 specifies that the occurrence of P is independent in the likelihood of Q. In this study, the analysis was conducted using the 'arules' package in statistical software R (Hahsler et al., 2023).

4. Data

4.1. Data pre-processing

The pedestrian crash data used in this study was obtained from the Louisiana Department of Transportation and Development (LADOTD) crash database for the period between 2010 and 2019. The database contains detailed information about each crash event, including highway, crash, pedestrian, and driver-related characteristics. Following the study objectives, the research team filtered out the crashes in which pedestrians were distracted prior to the collisions. Within the dataset at the pedestrian level, distracted pedestrians are denoted by category C in the 'pedestrian condition' column. A total of 415 crashes involving distracted pedestrians were identified for further analysis. It should be noted that all the extracted crashes were unique and only one pedestrian was found to be distracted in each crash.

To identify the distraction characteristics, the police reports were obtained from the LADOTD database using the unique identification number assigned to each crash. 31 out of the 415 pedestrian crashes did not have a crash report available in the online database. These crashes predominantly occurred on local roads. The crash reports typically include narratives that contain comments from police officers,

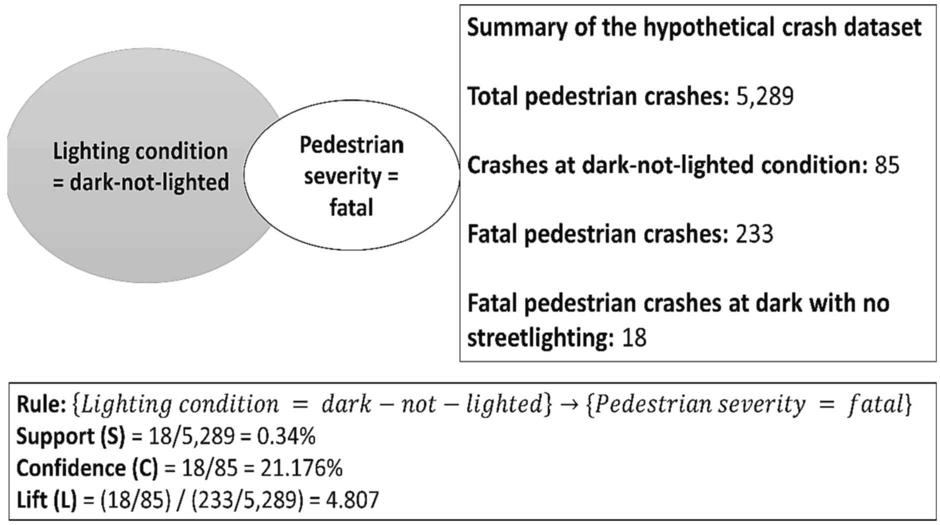

Fig. 1. An example of estimating the three parameters of ARM.

occupants, pedestrians, or other individuals involved in the crash incident. These narratives have proven valuable in previous road safety studies, as they provide additional information beyond the predefined attributes of traditional crash reports (Fitzpatrick et al., 2017; Hossain et al., 2022). In this study, about 80.48 % of distracted pedestrian crashes had police reports with narratives.

Two independent reviewers, who were graduate research assistants with over two years of experience working with Louisiana crash data, manually reviewed each crash narrative to classify the distracting activities of pedestrians. The independent feedback of reviewers about the types of distraction for each crash record was matched using unique crash identification numbers. In cases where discrepancies were identified, the narratives were further assessed under the supervision of a highway safety professional. After the review process, the collected crash reports were categorized into three classes that define pedestrian distraction tasks (Table 1). Upon reading the narratives, the reviewers identified approximately 183 police reports as class 1, 123 as class 2, and 78 as class 3. In this study, 'pedestrian distraction' variable has been classified into six categories as follows:

- Electronic devices: talking or listening to a phone, manipulating a cellphone, and engaging with other electronic devices such as iPads or MP3 players.
- In-person conversation: engaging in conversations or verbal altercations with parents, friends, siblings, or other pedestrians.
- pets/persons/objects: attempting to collect any dropped object, trying to catch vehicles, being chased by pets, getting distracted by objects or persons inside the vehicles, and chasing children.
- Roadside activities: pushing vehicles, loading goods, blowing grass, changing tires, cleaning vehicles, cutting grass, helping other vehicles, and other similar activities performed alongside the road.
- Careless/inattention: used when the specific form of distraction cannot be determined, for example, a sudden run or suddenly on the road.
- Unknown: no statement regarding pedestrian or distraction, no crash narrative, and no available police report

Table 1
Crash report classes to define the tasks associated with pedestrian distraction.

Crash report classes	Description
Class 1	Have a precise statement in narratives emphasizing the specific tasks or activities that distracted the pedestrian. For example, "*Driver stated she was traveling east and suddenly a pedestrian started crossing the segment. The pedestrian stated he was talking on the cellphone and did not notice the vehicle on the road*".
Class 2	Have ambiguous statements regarding the distraction activities of pedestrians prior to the crash incident. For instance, "*Driver of vehicle #1 stated that she was heading towards Westbound. Driver of vehicle #1 stated that a girl suddenly started running and came in front of her car. The Driver of vehicle #1 pressed the brake pedal but it was too late to avoid a collision*".
Class 3	Have no crash narrative, or no statement in the narrative regarding pedestrian or their distraction.

Crashes that had class 2 police reports were categorized as 'careless/inattention', whereas crashes with class 3 reports, as well as those with no police reports, were grouped as 'unknown'. Other factors contributing to distracted pedestrian crashes were chosen based on relevant previous studies, the availability of variables in the original database, and engineering judgments. The final dataset contains 415 crashes with fifteen variables, as indicated in Table 2. Among five levels of pedestrian severity (fatal, incapacitating or severe, non-incapacitating or moderate, complaint or possible, and no injury), the fatal and severe crashes were combined. Regarding driver characteristics, the focus was solely on the information of the driver of vehicle 1, as they were usually determined to be the major liability holder for the crash. Since only 3 out of the 415 crashes involved alcohol, the related variable was not included in the analysis. In over 85 % of the crashes, no driver violation or an unknown driver violation was reported; therefore, the variable was not included. Fig. 2 shows the analytical framework of this study with data preprocessing techniques.

4.2. Description of final dataset

The summary statistics of distracted pedestrian crashes for each variable are presented in Table 2. About 44.82 % of crashes occurred while crossing segments or intersections. Kuzel et al. (2008) inferred that distracted pedestrians were more prone to collisions in or around road crossing points. Careless/inattention accounted for 29.64 % of the distracted actions, followed by 18.31 % attributed to pets/persons/objects, 10.84 % to electronic devices, and 9.64 % to roadside activities. Consistent with previous studies (Nasar and Troyer, 2013; Tapiro et al., 2020), a high percentage of distracted pedestrians were young people less than 25 years. Male pedestrians were involved in 63.37 % of the crashes, which aligns with a national study on pedestrian crashes involving headphone use conducted by Lichenstein et al. (2012). The

Table 2
Overview of crashes involving distracted pedestrians.

Categories	Freq.	%	Categories	Freq.	%
Pedestrian action			*Day of the week (DOW)*		
crossing intersection	58	13.98	MTWT	257	61.93
crossing segment	128	30.84	FSS	158	38.07
standing in road	51	12.29	*Lighting condition*		
walking in road	45	10.84	daylight	227	54.7
working on road	21	5.06	dark-lighted	102	24.58
working on vehicle	31	7.47	dark-not-lighted	64	15.42
playing on road	26	6.27	others	22	5.3
others	55	13.25	*Location type*		
Pedestrian distraction			business	81	19.52
electronic devices	45	10.84	residential	189	45.54
in-person conversation	22	5.3	business-residential mixed	102	24.58
pets/persons/objects	76	18.31	open country	26	6.27
roadside activities	40	9.64	others	17	4.1
careless/inattention	123	29.64	*Intersection*		
unknown	109	26.27	no	297	71.57
Pedestrian age			yes	118	28.43
<15y	96	23.13	*Traffic control device*		
15-24y	100	24.1	no control	146	35.18
25-40y	105	25.3	green signal on	29	6.99
41-64y	86	20.72	stop sign	18	4.34
>64y	18	4.34	white dashed line	81	19.52
unknown	10	2.41	yellow dashed line	53	12.77
Pedestrian severity			yellow no passing line	51	12.29
fatal/severe	67	16.14	others	37	8.92
moderate injury	165	39.76	*Area setting*		
possible injury	147	35.42	urban	338	81.45
no injury	36	8.67	rural	77	18.55
Pedestrian gender			*Driver condition*		
female	145	34.94	normal	257	61.93
male	263	63.37	distracted	41	9.88
others	7	1.69	inattentive	50	12.05
Clothing			others	67	16.14
dark	158	38.07	*Movement prior crash*		
light	257	61.93	straight ahead	286	68.92
Crash time			backing	20	4.82
6–11:59am	81	19.52	left/right turn	14	3.37
12–5:59 pm	156	37.59	slowing to stop/left/right	17	4.1
6–11:59 pm	140	33.73	stopped	26	6.27
12–5:59am	38	9.16	others	52	12.53

Note: FSS = Friday, Saturday, Sunday; MTWT = Monday, Tuesday, Wednesday, Thursday.

distribution of crashes across different hours of the day followed a pattern similar to the traditional distribution, with the highest percentage of crashes occurring between 12 pm and 5:59 pm (37.59 %), followed by the period from 6 pm to 11:59 pm (33.73 %). Residential areas experienced the highest proportion of distracted pedestrian crashes (45.54 %) compared to other location classes. Some of the contributing factors exhibited significant skewness towards specific attributes. For example, 61.93 % of crashes happened during weekdays, 71.57 % on segments, 81.45 % in urban areas, and 68.92 % when the driver was proceeding straight.

5. Results and discussion

Determining the appropriate minimum support and confidence values in ARM is critical to obtain interesting and meaningful results. Setting a low support threshold may lead to an abundance of uninteresting rules, whereas a high support value could ignore the significant inherent relationship between categories. This study utilized a trial-and-error approach to set the minimal support and confidence values for each following circumstances, as recommended by previous studies (Hossain et al., 2021b; Pande and Abdel-Aty, 2009; Das et al., 2017). The rules were pruned to remove redundant and repeated associations (Hossain et al., 2021b). Additionally, a maximum of four antecedents was imposed in this study to facilitate a more straightforward interpretation of the rules. In ARM, the lift value is more vital in understanding the strength of an association rule. Rules with a lift value greater than 1 indicate that the co-occurring associations are more frequent than expected (Hossain et al., 2022).

5.1. Rules by pedestrian severity

A minimum support of 2 % and a confidence of 50 % were used to generate rules with three different pedestrian severity levels as consequents: ‘fatal/severe’, ‘moderate injury’, and ‘possible injury’. A total of 1,654 rules were obtained, all of which had a lift value greater than 1. Table 3 exhibits the top 10 rules for each injury severity level in descending order of lift values.

The first 10 rules (A#1–A#10) provide insights into the patterns of fatal/severe injury pedestrian crashes. Rule A#1 states that distracted male pedestrians aged 41–64 have a 3.98 times higher probability of experiencing fatal/severe injuries in dark without streetlighting settings on segments. Jaywalking behavior frequently evolves among pedestrians during dark conditions (Samerei et al., 2021), and any distracting action in such situations could significantly elevate the fatality risk. In similar lighting conditions during the evening and early night hours, fatal/severe injuries are more likely to occur when pedestrians wear dark clothes (A#2). This is because drivers may have difficulty detecting pedestrians wearing dark clothing without sufficient reflective lights (Shinar, 1985). In the generated rules with fatal/severe injuries, ‘pedestrian gender = male’, ‘crash time = 6–11:59 pm’, ‘pedestrian age = 41-64y’, and ‘lighting condition = dark-not-lighted’ show their interactions in almost all associations. Distracted pedestrians with a high degree of severity are more frequently found in rural areas (A#10).

Rules B#1-B#10 describe the affiliated attributes that contributed to moderate injury crashes. Rule B#1 implies that distracted pedestrians using electronic devices while walking in residential areas have a 2.52 times higher likelihood of moderate injuries. A similar walking behavior is also apparent in urban areas with identical consequences (B#2). People tend to engage in dual-tasking while walking around their residences during leisure time (Fernández et al., 2020). Rule B#4 indicates a strong association between moderate injuries of distracted pedestrians aged 15–24 and road segments with yellow dashed lines. These lines signify that motorists can overtake slow-moving vehicles using the opposite side of the road. With ‘pedestrian severity = moderate injury’, crashes involving youths are prevalent during afternoon hours (B#6). Additionally, female pedestrians in the same age group are more likely to be distracted on weekends (B#5). The cumulative effect of dark-lighted conditions and FSS increases the risk of moderate injuries among distracted pedestrians (B#7, B#9). The prevalence of late-night walks is expected to be high on weekends (Gulley, 2020); therefore, the chances of distracted walking are higher than usual.

Rules C#1–C#10 explain the co-existing categories influencing possible injury. Distraction behavior in morning hours is found to be strongly associated with male pedestrians (C#1, C#10), daylight (C#2–C#4), youths (C#10), dark clothing (C#4), urban areas (C#2, C#5), and weekdays (C#7). Basch et al. (2015) documented numerous distraction-related tasks of walkers at intersections during the morning commute hours (7:30 to 9:30 am) and recreation hours (9:30 am to 12:30 pm). Rule C#8 indicates that female children engaging in distracting activities in residential areas during afternoon hours have a 2.17 times higher risk of possible injury. To illustrate Rule C#10 (S: 2.17 %, C: 75 %, L: 2.12), it can be interpreted as follows: a) 2.17 % of distracted pedestrian crashes with possible injuries during morning hours involve young male pedestrians, b) out of all distracted pedestrian crashes in a similar environment, 75 % result in possible injuries, c) the proportion of distracted pedestrian crashes with possible injuries among young males in the specified time frame is 2.12 times higher than the same proportion in the complete dataset.

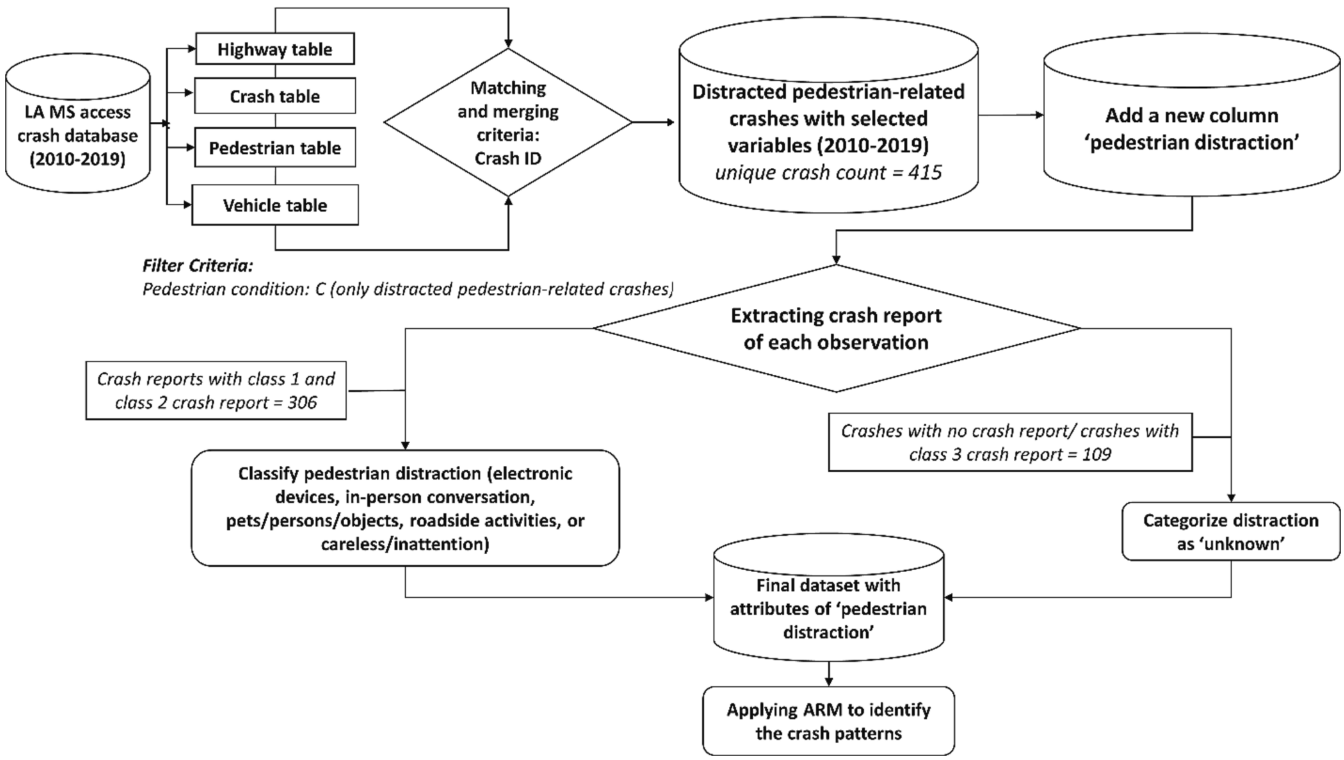

Fig. 2. Study framework.

5.2. Rules by pedestrian distraction

In this study, rules were generated based on different categories of pedestrian distraction, including 'electronic devices', 'in-person conversation', 'pets/persons/objects', 'roadside activities', and 'careless/inattention', as consequent variables. Minimum support and confidence thresholds were determined for each category to generate the rules, taking into account previous studies (Nasim Khan et al., 2020; Rahman et al., 2021).

The support and confidence thresholds for extracting rules with 'pedestrian distraction = electronic devices' as the consequent were set at 1 % and 60 % respectively. A total of 27 rules were generated, with a lift value greater than 1. The top 10 rules based on lift value are presented in Table 4. Among these rules, it is observed that young pedestrians tend to be distracted by electronic devices at intersections on weekdays during the afternoon (#Q1) and at the time of crossing (#D10). While using cellphones, pedestrians are crossed more slowly (Nasar et al., 2008) and less likely to pay attention to traffic (Pešić et al., 2016). Moreover, the risk of moderate injuries is significantly associated with factors such as residential roadways with yellow dashed lines (#D3), urban residential locations (#D5), males wearing dark clothes (#D9), and females wearing light clothes (#D8) when distracted by electronic devices during walking. Environmental attributes such as FSS and the time period of 6–11:59 pm also influence the likelihood of moderate injuries among pedestrians distracted by electronic devices (D#4, D#6). In the case of male pedestrians, similar walking behavior in urban areas during evening and early night hours increases the risk of collision by 6.59 times (D#7). It is worth noting that none of the top associations include 'intersection = no', indicating that electronic device usage is more prevalent at intersections than on road segments.

Association rules were extracted with 'pedestrian distraction = in-person conversation' using the minimum support and confidence of 1 % and 50 %, respectively. A total of 16 rules were generated with a lift value greater than 1. The top 10 rules, ranked by their lift values, are presented in Table 5. Most of these rules include attributes such as 'driver condition = distracted', 'traffic control device = no control', 'DOW = MTWT', 'pedestrian action = standing in road', and 'pedestrian gender = male', and 'location type = residential'. It is expected that people pay less concentration on their surroundings when engaged in conversations with others. Pedestrians aged 25 to 40, while conversing with others on the street, are 11.79 times more likely to be involved in crashes on urban road segments (E#1). The association between 'pedestrian distraction = in-person conversation' and distracted drivers is also observed in specific road and environmental conditions. For example, distracted drivers are involved in related pedestrian crashes during weekdays while traveling on roads with no traffic control (E#4). Both drivers and pedestrians are less likely to notice and yield to other road users when engaged in dual-task activities (Brumfield and Pulugurtha, 2011). Crashes involving in-person conversations are commonly observed among pedestrians standing on urban residential roadways without traffic controls (E#6). This mode of pedestrian distraction is particularly prominent at locations other than intersections (E#1, E#3, E#7, E#10).

To extract rules with 'pedestrian distraction = pets/persons/objects' as the consequent, a minimum support of 1 % and a confidence of 60 % were set. The number of rules generated in this case was 71, with a lift value greater than 1. The top 10 rules by lift value are presented in Table 6. Several rules highlight the combined effect of child pedestrians aged under 15 and business-residential mixed locations (F#1, F#2, F#3, F#5, F#6). According to rule F#10, distractions caused by pets, persons, or objects are more common when crossing segments in rural settings, leading to a 4.55 times higher risk of moderate injuries. Distracting activities related to pets/persons/objects are highly associated with sudden walking speed changes, which can have injury consequences if they occur at or near crosswalks (Alhajyaseen and Iryo-Asano, 2017). With 'pedestrian distraction = pets/persons/objects', the association among multiple pedestrian characteristics has been recognized. For example, rule F#4 specifies that young female pedestrians wearing dark clothes are more likely to be involved in crashes on sites with no traffic control.

Association rules were generated with 'pedestrian distraction = roadside activities' using the minimum support and confidence of 2 %

Table 3
Rules by pedestrian severity.

Rule	LHS	S (%)	C (%)	L
A#1	pedestrian age = 41-64y, pedestrian gender = male, lighting condition = dark-not-lighted, intersection = no	2.17	64.29	3.98
A#2	crash time = 6–11:59 pm, lighting condition = dark-not-lighted, clothing = dark, driver condition = normal	2.17	64.29	3.98
A#3	pedestrian gender = male, lighting condition = dark-not-lighted, driver condition = normal, movement prior crash = straight ahead	2.89	63.16	3.91
A#4	pedestrian age = 41-64y, lighting condition = dark-not-lighted, intersection = no	2.17	60.00	3.72
A#5	pedestrian gender = male, crash time = 6–11:59 pm, lighting condition = dark-not-lighted, driver condition = normal	2.17	60.00	3.72
A#6	lighting condition = dark-not-lighted, clothing = dark, driver condition = normal, movement prior crash = straight ahead	2.17	60.00	3.72
A#7	crash time = 6–11:59 pm, lighting condition = dark-not-lighted, driver condition = normal, movement prior crash = straight ahead	2.65	57.89	3.59
A#8	pedestrian age = 41-64y, pedestrian gender = male, lighting condition = dark-not-lighted	2.17	56.25	3.48
A#9	crash time = 6–11:59 pm, lighting condition = dark-not-lighted, driver condition = normal	2.89	54.55	3.38
A#10	pedestrian age = 41-64y, pedestrian gender = male, intersection = no, area setting = rural	2.17	50.00	3.10
B#1	pedestrian action = walking in road, pedestrian distraction = electronic devices, location type = residential	2.41	100.00	2.52
B#2	pedestrian action = walking in road, pedestrian distraction = electronic devices, area setting = urban	2.41	100.00	2.52
B#3	pedestrian action = walking in road, pedestrian distraction = electronic devices	2.89	92.31	2.32
B#4	pedestrian age = 15-24y, intersection = no, traffic control device = yellow dashed line	2.17	90.00	2.26
B#5	pedestrian age = 15-24y, pedestrian gender = female, DOW = FSS, intersection = no	2.17	90.00	2.26
B#6	pedestrian age = 15-24y, crash time = 12–5:59 pm, traffic control device = no control, movement prior crash = straight ahead	2.17	90.00	2.26
B#7	DOW = FSS, lighting condition = dark-lighted, clothing = light, movement prior crash = straight ahead	2.65	84.62	2.13
B#8	pedestrian action = walking in road, pedestrian age = 15-24y, intersection = no, movement prior crash = straight ahead	2.41	83.33	2.10
B#9	pedestrian distraction = careless/inattention, DOW = FSS, lighting condition = dark-lighted, movement prior crash = straight ahead	2.41	83.33	2.10
B#10	pedestrian age = 15-24y, traffic control device = yellow dashed line	2.17	81.82	2.06
C#1	pedestrian gender = male, crash time = 6–11:59am, driver condition = others	2.17	90.00	2.54
C#2	crash time = 6–11:59am, lighting condition = daylight, driver condition = others, area setting = urban	2.17	90.00	2.54
C#3	crash time = 6–11:59am, lighting condition = daylight, driver condition = others	2.17	81.82	2.31
C#4	crash time = 6–11:59am, lighting condition = daylight, intersection = no, clothing = dark	2.89	80.00	2.26
C#5	crash time = 6–11:59am, driver condition = others, area setting = urban	2.65	78.57	2.22
C#6	lighting condition = daylight, intersection = no, traffic control device = others	2.41	76.92	2.17
C#7	pedestrian distraction = unknown, crash time = 6–11:59am, DOW = MTWT, lighting condition = daylight	2.41	76.92	2.17
C#8	pedestrian age = <15y, pedestrian gender = female, crash time = 12–5:59 pm, location type = residential	2.41	76.92	2.17
C#9	crash time = 6–11:59 am, movement prior crash = others	2.17	75.00	2.12
C#10	pedestrian age = 15–24 y, pedestrian gender = male, crash time = 6–11:59am	2.17	75.00	2.12

Note: A#1–A#10: fatal/severe as consequent, B#1–B#10: moderate injury as consequent, C#1–C#10: possible injury as consequent.

Table 4
Rules for 'pedestrian distraction = electronic devices'.

Rule	LHS	S (%)	C (%)	L
D#1	pedestrian age = 15–24 y, crash time = 12–5:59 pm, DOW = MTWT, intersection = yes	1.20	83.33	7.69
D#2	pedestrian age = 15-24y, crash time = 12–5:59 pm, intersection = yes	1.45	75.00	6.92
D#3	pedestrian action = walking in road, pedestrian severity = moderate injury, location type = residential, traffic control device = yellow dashed line	1.45	75.00	6.92
D#4	pedestrian action = walking in road, pedestrian severity = moderate injury, DOW = FSS, area setting = urban	1.45	75.00	6.92
D#5	pedestrian action = walking in road, pedestrian severity = moderate injury, location type = residential, area setting = urban	1.93	72.73	6.71
D#6	pedestrian action = walking in road, pedestrian severity = moderate injury, pedestrian gender = male, crash time = 6–11:59 pm	1.20	71.43	6.59
D#7	pedestrian action = walking in road, pedestrian gender = male, crash time = 6–11:59 pm, area setting = urban	1.20	71.43	6.59
D#8	pedestrian action = walking in road, pedestrian severity = moderate injury, pedestrian gender = female, clothing = light	1.20	71.43	6.59
D#9	pedestrian action = walking in road, pedestrian severity = moderate injury, pedestrian gender = male, clothing = dark	1.20	71.43	6.59
D#10	pedestrian action = crossing intersection, pedestrian age = 15-24y, DOW = MTWT, intersection = yes	1.20	71.43	6.59

Table 5
Rules for 'pedestrian distraction = in-person conversation'.

Rule	LHS	S (%)	C (%)	L
E#1	pedestrian action = others, pedestrian age = 25-40y, intersection = no, area setting = urban	1.20	62.50	11.79
E#2	pedestrian gender = male, DOW = MTWT, driver condition = distracted, traffic control device = no control	1.45	60.00	11.32
E#3	DOW = MTWT, intersection = no, driver condition = distracted, traffic control device = no control	1.45	60.00	11.32
E#4	DOW = MTWT, driver condition = distracted, traffic control device = no control	1.69	58.33	11.00
E#5	pedestrian action = standing in road, location type = residential, clothing = light, traffic control device = no control	1.20	55.56	10.48
E#6	pedestrian action = standing in road, location type = residential, traffic control device = no control, area setting = urban	1.20	55.56	10.48
E#7	pedestrian action = standing in road, pedestrian gender = male, intersection = no, traffic control device = no control	1.20	55.56	10.48
E#8	pedestrian gender = male, driver condition = distracted, traffic control device = no control	1.69	50.00	9.43
E#9	pedestrian action = standing in road, pedestrian gender = male, traffic control device = no control	1.20	50.00	9.43
E#10	location type = residential, intersection = no, driver condition = distracted, traffic control device = no control	1.20	50.00	9.43

and 60 %, respectively. A total of 15 rules were formed with a lift value greater than 1, and the top 10 rules in the descending order of lift values are exhibited in Table 7. It is observed that working on a vehicle is

Table 6
Rules for 'pedestrian distraction = pets/persons/objects'.

Rule	LHS	S (%)	C (%)	L
F#1	pedestrian age = <15 y, DOW = FSS, location type = business-residential mixed, traffic control device = no control	1.20	100.00	5.46
F#2	pedestrian age = <15 y, pedestrian severity = moderate injury, location type = business-residential mixed, traffic control device = no control	1.20	100.00	5.46
F#3	pedestrian age = <15 y, location type = business-residential mixed, intersection = no, traffic control device = no control	1.20	100.00	5.46
F#4	pedestrian age = 15-24y, pedestrian gender = female, clothing = dark, traffic control device = no control	1.45	100.00	5.46
F#5	pedestrian age = <15 y, pedestrian severity = moderate injury, location type = business-residential mixed, intersection = no	1.69	87.50	4.78
F#6	pedestrian age=<15y, location type = business-residential mixed, traffic control device = no control	1.45	85.71	4.68
F#7	pedestrian action = others, pedestrian severity = moderate injury, pedestrian gender = female, movement prior crash = straight ahead	1.45	85.71	4.68
F#8	pedestrian action = others, pedestrian severity = moderate injury, location type = business-residential mixed	1.20	83.33	4.55
F#9	pedestrian action = crossing segment, pedestrian severity = moderate injury, area setting = rural	1.20	83.33	4.55
F#10	pedestrian severity = moderate injury, crash time = 6–11:59am, traffic control device = no control	1.20	83.33	4.55

Table 7
Rules for 'pedestrian distraction = roadside activities'.

Rule	LHS	S (%)	C (%)	L
G#1	pedestrian action = working on vehicle, pedestrian severity = possible injury	2.41	100.00	10.38
G#2	pedestrian action = working on vehicle, lighting condition = dark-not-lighted, clothing = dark	2.17	100.00	10.38
G#3	pedestrian action = working on vehicle, clothing = dark	2.89	92.31	9.58
G#4	pedestrian action = working on vehicle, location type = residential	2.17	90.00	9.34
G#5	pedestrian action = working on vehicle, area setting = urban	3.61	83.33	8.65
G#6	pedestrian action = working on vehicle, crash time = 6–11:59 pm, lighting condition = dark-not-lighted	2.41	83.33	8.65
G#7	pedestrian action = working on vehicle, lighting condition = dark-not-lighted	3.37	77.78	8.07
G#8	pedestrian action = working on road, lighting condition = daylight	2.41	76.92	7.98
G#9	pedestrian action = working on vehicle, crash time = 6–11:59 pm	2.41	76.92	7.98
G#10	pedestrian action = working on vehicle, DOW = MTWT	3.61	68.18	7.07

associated with 9 out of the 10 generated rules. Engaging in such activities while wearing dark clothes under dark-not-lighted conditions increases the risk of crashes by 10.38 times (G#2). Crashes involving working on a vehicle are frequent a) in urban areas (G#5), b) on weekdays (G#10), c) in residential locations (G#4), and d) from 6 to 11:59 pm (G#6). Previous studies have also highlighted the vulnerability of roadside activities, such as working on a vehicle and performing any activity in the shared road space (Baltes, 1998; Spainhour et al., 2006).

The minimum support and confidence were set at 2 % and 70 %, respectively, to generate rules with 'pedestrian distraction = careless/inattention' as consequent. A total of 51 rules were generated, and the top 10 rules by lift value are shown in Table 8. The majority of crashes involving careless/inattentive actions of pedestrians occur while playing on the road (7 out of 10 rules). Activities like chasing balls during play can easily distract pedestrians from their surroundings (Schofer et al., 1995), thereby increasing the risk of collisions (Christie et al., 2007). With 'pedestrian distraction = careless/inattention', these pedestrian actions are particularly linked to urban areas, road segments, and residential locations. For instance, crashes involving careless/inattentive pedestrians playing on the street are more likely to occur on road segments within urban residential areas (H#7). Female child pedestrians are often found to have a lack of situational awareness while on road segments (H#2) and in residential locations (H#10). According to rule H#9, the probability of careless/inattentive pedestrians being involved in crashes is 2.65 times higher during afternoon hours on segments within business areas where vehicles are moving straight.

5.3. Key findings

The results obtained from ARM analysis reveal the correlated characteristics related to crashes involving distracted pedestrians, categorized according to the pedestrian severity and types of distracting activities. To provide a concise overview of the study, the main findings are outlined as follows:

- Pedestrians engaged in divided attention tasks in dark-not-lighted conditions are at a higher risk of fatal/severe injuries. This behavior is particularly prevalent among male and older pedestrians. Conversely, pedestrians with moderate injuries are strongly associated with electronic device usage while walking in urban residential areas. In relation to possible injuries, distracting actions are prevalent during morning hours.
- On weekdays, crashes involving pedestrians distracted by electronic devices are often evident at intersections and while crossing them.

Table 8
Rules for 'pedestrian distraction = careless/inattention'.

Rule	LHS	S (%)	C (%)	L
H#1	pedestrian action = playing on road, DOW = MTWT, traffic control device = no control	2.17	90.00	3.04
H#2	pedestrian age = <15 y, pedestrian gender = female, lighting condition = daylight, intersection = no	2.89	85.71	2.89
H#3	pedestrian action = playing on road, lighting condition = daylight, intersection = no, movement prior crash = straight ahead	2.65	84.62	2.85
H#4	pedestrian action = playing on road, lighting condition = daylight, intersection = no, rurality = urban	2.65	84.62	2.85
H#5	pedestrian action = playing on road, location type = residential, intersection = no, traffic control device = no control	2.41	83.33	2.81
H#6	pedestrian action = playing on road, location type = residential, intersection = no, movement prior crash = straight ahead	2.17	81.82	2.76
H#7	pedestrian action = playing on road, location type = residential, intersection = no, rurality = urban	2.17	81.82	2.76
H#8	pedestrian action = playing on road, intersection = no, traffic control device = no control, movement prior crash = straight ahead	2.65	78.57	2.65
H#9	crash time = 12–5:59 pm, location type = business, intersection = no, movement prior crash = straight ahead	2.65	78.57	2.65
H#10	pedestrian age = <15y, pedestrian gender = female, lighting condition = daylight, location type = residential	2.65	78.57	2.65

On the contrary, weekend crashes related to electronic devices while walking are found to be frequent in urban settings.

- Drivers are found to be distracted in crashes involving pedestrians talking with other individuals on the road. These crashes are often observed on road segments without traffic control devices.
- Business-residential mixed locations are found to be influential for crashes involving child pedestrians distracted by pets/persons/objects. Such distractions are more prevalent in rural areas during crossing segments.
- Working on vehicles in low-light conditions while wearing dark clothing is found to be vulnerable.
- Careless/inattentive actions of pedestrians while playing on the street in urban residential areas are strongly associated with crash occurrences.

6. Conclusions

Due to the emerging distraction resources, identifying the chain of risk factors influencing distracted walking behavior has become crucial. However, limited research specifically focuses on the traits associated with distracted pedestrian crashes. This study addresses this gap by analyzing ten years of crash data from Louisiana (2010–2019) to determine the grouping of factors related to pedestrian severity outcomes and types of distraction. A major contribution of this study is the inclusion of various distraction-related tasks reported in police-investigated crash narratives. Moreover, this study is the first to reveal meaningful crash patterns involving distracted pedestrians. Using the apriori algorithm enabled the development of rules that highlight the most common combinations of factors in these crashes. Compared to traditional parametric modeling, ARM can handle datasets more effectively without relying on predefined hypotheses, such as the independence of covariates. The study results can guide in making effective countermeasures to reduce the distraction behavior of pedestrians. The identified associations provide valuable insights into the factors that can be targeted to promote safe interactions between pedestrians and vehicles.

The findings of this study highlight the complex nature of distracted pedestrian crashes and shed light on the interrelationships between various risk factors. The generated rules indicate that senior adults aged 41–64 are particularly vulnerable to fatal/severe injuries when distracted while walking. Developing educational and media campaigns with relevant content, layout, exposure, and communication channels is essential to target these demographics (Levi et al., 2013). For instance, promoting safe walking knowledge through medical brochures, which are more likely to be read by older adults, can effectively raise awareness among this specific group (Downing and Jones, 2008). In this study, children and youths were found to be distracted by electronic devices and pets/persons/objects. Implementing community-level safe-walking programs, such as Safe Routes to School (SRTS), with the active involvement of parents, school administrators, teachers, and childcare providers, can help modify the distracted walking behavior of adolescents. Among pedestrians, electronic device usage was more frequent while crossing intersections, whereas in-person conversations were highly associated with places other than intersections. Won et al. (2020) introduced a prototype mobile system called SaferCross, designed to warn pedestrians using cellphones when engaged in critical actions such as crossing segments or intersections at a slow speed. However, the effectiveness of such technologies in reducing distracted walking is still under investigation. In this study, the outcome of ARM showed the cumulative effect of dark clothing and dark-not-lighted condition in multiple crash scenarios with high severity of distracted pedestrians. Retro-reflective stripes on clothing have been recognized for enhancing pedestrian visibility and safety in low-lighting conditions (Black et al., 2021). However, these concepts require more publicity to increase social acceptance. Transportation planners have recently focused on incorporating modern technologies to improve road illumination. Adaptive lighting, LED lighting, and Bollard lighting have been found to be effective in creating a safer walking environment at night (Markowitz and Smith, 2017).

While initial descriptive statistics suggested that distracted pedestrian crashes were more prevalent in urban areas, the generated rules revealed that pedestrians were at a higher risk of experiencing fatal or severe injuries in rural locations. Also, pedestrians were found to be distracted by pets/persons/objects while crossing segments in rural areas. In most cases, sidewalks do not exist in rural areas; therefore, pedestrians are advised to walk in the direction of traffic. Hall et al. (2004) recommended the implementation of advanced warning and pedestrian detection systems, dynamic signs, and midblock crossing signals in rural locations prone to pedestrian collisions. Installing such traffic control devices is crucial, as road segments with no traffic control significantly contribute to pedestrian distraction. However, further research is needed to determine the most effective control devices for different walking scenarios. In this study, driver distraction was associated with crashes in which pedestrians were talking with others on the street. Therefore, there is a scope to reduce distracted pedestrian crashes by preventing drivers from being distracted. Further investigation can provide insights into the specific types of driver distractions associated with these collisions. Distracted pedestrians with injury outcomes were repeatedly visible in urban residential locations and roadways marked with yellow dashed lines. Conducting road safety assessments that consider geometric features and functional elements can help prioritize interventions such as speed reduction and roadway redesign to improve pedestrian safety in these areas. Existing pedestrian safety curricula primarily focus on the dangers of cellphone usage while walking, neglecting other distracting tasks such as being distracted by pets/persons/objects, working on a vehicle, or working on the road. The hazardous consequences of these distractions, as identified in this study, should be incorporated into intervention programs to raise awareness and promote safer pedestrian behaviors. Modern vehicle models are equipped with advanced pedestrian safety features like pedestrian airbags and pedestrian detection systems, which have the potential to mitigate injury severity (Fleming, 2012). However, the high cost and increased insurance rates associated with these vehicles may pose challenges to their widespread adoption. Therefore, a successful intervention strategy should encompass a combination of environmental modifications, technological advancements, educational initiatives, and enforcement efforts to enhance overall pedestrian safety.

There are some limitations in this study that could be addressed in future research. Pedestrian distraction determination may be based on the testimony of a biased participant in the event (driver) while the other party (pedestrian following a collision with a motor vehicle) is likely incapacitated. Another limitation is the potential for underreporting and overreporting issues in police-investigated crashes, which may have influenced the findings of this study. To improve the reliability of the data, future research can explore alternative data sources or employ more rigorous data collection methods. It is worth noting that more than 85 % of the recorded speed information in the original crash database was either miscoded or unknown, and therefore, not included in the analysis. Investigating the association between vehicle speed and pedestrian distraction crashes could provide valuable insights and can be considered in future studies. Additionally, the original dataset had no information on traffic control devices for pedestrians at the crash locations, and over 90 % of pedestrian violation information was recorded as 'unknown'. Future research could incorporate real-world observations and laboratory experiments to gather more comprehensive data on these variables. In this study, the analysis is limited to 5-itemsets, and further research can be conducted to determine the long patterns from a multitude of factors. It would be beneficial to conduct research that specifically examines crash mechanisms, such as the sequence of events, in the identified crash scenarios involving distracted pedestrians.

Author contributions

The authors confirm contribution to the paper as follows: study conception and design: Hossain M.M.; data preparation: Hossain M.M., Hossain A., Sun X., Zhou H.; analysis and interpretation of results: Hossain M.M.; draft manuscript preparation: Hossain M.M., Zhou H., Hossain A., Sun X., Das S. All authors reviewed the results and approved the final version of the manuscript.

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.

Data availability

The authors do not have permission to share data.

References

AAOS, 2015. Distracted Walking Study. American Academy of Orthopaedic Surgeons (AAOS). https://orthoinfo.aaos.org/en/staying-healthy/distracted-walking/.

Agrawal, R., Imieliński, T., Swami, A., 1993. Mining association rules between sets of items in large databases 207–216. doi:10.1145/170035.170072.

Agrawal, R., Srikant, R., 1994. Fast algorithms for mining association rules. Proc. Conf. Very Large Data Bases VLDB'94, 487–499.

Alhajyaseen, W.K.M., Iryo-Asano, M., 2017. Studying critical pedestrian behavioral changes for the safety assessment at signalized crosswalks. Saf. Sci. 91, 351–360. https://doi.org/10.1016/j.ssci.2016.09.002.

Ashraf, M.T., Dey, K., Mishra, S., Rahman, M.T., 2021. Extracting rules from autonomous-vehicle-involved crashes by applying decision tree and association rule methods, 036119812110184 Transp. Res. Rec. J. Transp. Res. Board. https://doi.org/10.1177/03611981211018461.

Baltes, M.R., 1998. Descriptive analysis of crashes involving pedestrians in Florida, 1990–1994. Transp. Res. Rec. 1636, 138–145. https://doi.org/10.3141/1636-22.

Basch, C.H., Ethan, D., Zybert, P., Basch, C.E., 2015. Pedestrian behavior at five dangerous and busy manhattan intersections. J. Community Health 40 (4), 789–792. https://doi.org/10.1007/s10900-015-0001-9.

Black, A.A., Bui, V., Henry, E., Ho, K., Pham, D., Tran, T., Wood, J.M., 2021. Using retro-reflective cloth to enhance drivers' judgment of pedestrian walking direction at night-time. J. Saf. Res. 77, 196–201. https://doi.org/10.1016/j.jsr.2021.03.002.

Brumfield, R., Pulugurtha, S.S.. When distracted road users cross paths. Public Roads. http://www.fhwa.dot.gov/publications/publicroads/11novdec/01.cfm.

Christie, N., Ward, H., Kimberlee, R., Towner, E., Sleney, J., 2007. Understanding high traffic injury risks for children in low socioeconomic areas: A qualitative study of parents' views. Inj. Prev. 13 (6), 394–397. https://doi.org/10.1136/ip.2007.016659.

Das, S., Avelar, R.E., Dixon, K.K., Sun, X., 2017. Pedestrian crash analysis using association rules mining. Transportation Research Board 96th Annual Meeting. https://trid.trb.org/view/1437503.

Das, X., Kong "Jack", S., Hossain, M.M., Exploration on prior driving modes for automated vehicle collisions 2022 J. Urban Sci Int 10.1080/12265934.2022.2142650.

Das, S., Tamakloe, R., Zubaidi, H., Obaid, I., Alnedawi, A., 2021. Fatal pedestrian crashes at intersections: Trend mining using association rules. Accid. Anal. Prev. 160 https://doi.org/10.1016/j.aap.2021.106306.

Downing, K., Jones, L., 2008. Designing an educational strategy for increasing organ donor registration among older adults. Prog. Transplant. 18 (4), 290–296. https://doi.org/10.7182/prtr.18.4.0030q7j47t400464.

Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., Fayyad, U.M., Smith, P., 1996. From data mining to knowledge discovery: An overview. Adv. Knowl. Discov. Data Min. 1, 1–34.

Fernández, C., Vicente, M.A., Carrillo, I., Guilabert, M., Mira, J.J., 2020. Factors influencing the smartphone usage behavior of pedestrians: Observational study on "Spanish smombies". J. Med. Internet Res. 22, 8. https://doi.org/10.2196/19350.

Fitzpatrick, C.D., Rakasi, S., Knodler, M.A., 2017. An investigation of the speeding-related crash designation through crash narrative reviews sampled via logistic regression. Accid. Anal. Prev. 98, 57–63. https://doi.org/10.1016/j.aap.2016.09.017.

Fleming, B., 2012. New automotive electronics technologies [automotive electronics]. IEEE Veh. Technol. Mag. 7 (4), 4–12. https://doi.org/10.1109/MVT.2012.2218144.

Gulley, D., 2020. Curating the Campus Nightlife Experience: Documenting a Program Model for Late Night and Weekend Programming. University of Nebraska, Lincoln.

Haddon, W., 1980. Advances in the epidemiology of injuries as a basis for public policy. Public Health Rep. 95 (5), 411–421.

Hahsler, A.M., Buchta, C., Gruen, B., Hornik, K., Hahsler, M.M., 2023. Package 'arules'. https://cran.uib.no/web/packages/arules/arules.pdf.

Hall, J.W., Brogan, J.D., Kondreddi, M., 2004. Pedestrian Safety on Rural Highways. Federal Highway Administration Report #FHWA-SA-04-008. http://www.pedbikeinfo.org/cms/downloads/Ped_Safety_RuralHighways.pdf.

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

Hossain, M.M., Rahman, M.A., Sun, X., Mitran, E., 2021a. Investigating underage alcohol-intoxicated driver crash patterns in Louisiana. Transp. Res. Rec. 2675 (11), 769–782. https://doi.org/10.1177/03611981211019742.

Hossain, M.M., Sun, X., Mitran, E., Rahman, M.A., 2021b. Investigating fatal and injury crash patterns of teen drivers with unsupervised learning algorithms. IATSS Res. 45 (4), 561–573. https://doi.org/10.1016/j.iatssr.2021.07.002.

Hossain, A., Sun, X., Thapa, R., Codjoe, J., 2022. Applying association rules mining to investigate pedestrian fatal and injury crash patterns under different lighting conditions. Transp. Res. Rec. 2676 (6), 659–672. https://doi.org/10.1177/03611981221076120.

Hossain, A., Sun, X., Islam, S., Alam, S., Mahmud Hossain, M., 2023. Identifying roadway departure crash patterns on rural two-lane highways under different lighting conditions: Association knowledge using data mining approach. J. Saf. Res. https://doi.org/10.1016/j.jsr.2023.01.006.

Hossain, M.M., Zhou, H., Das, S., Sun, X., Hossain, A., 2023a. Young drivers and cellphone distraction: Pattern recognition from fatal crashes. J. Transp. Saf. Secur. 1–26 https://doi.org/10.1080/19439962.2022.2048763.

Hossain, M.M., Zhou, H., Rahman, M.A., Das, S., Sun, X., 2022. Cellphone-distracted crashes of novice teen drivers: Understanding associations of contributing factors for crash severity levels and cellphone usage types. Traffic Inj. Prev. 23 (7), 390–397. https://doi.org/10.1080/15389588.2022.2097667.

Hossain, M.M., Zhou, H., Sun, X., 2023b. A clustering regression approach to explore the heterogeneous effects of risk factors associated with teen driver crash severity. Transp. Res. Rec. 2677 (7), 1–21. https://doi.org/10.1177/03611981221150927.

Hyman, I.E., Boss, S.M., Wise, B.M., McKenzie, K.E., Caggiano, J.M., 2010. Did you see the unicycling clown? Inattentional blindness while walking and talking on a cell phone. Appl. Cogn. Psychol. 24 (5), 597–607. https://doi.org/10.1002/acp.1638.

Jiang, K., Ling, F., Feng, Z., Ma, C., Kumfer, W., Shao, C., Wang, K., 2018. Effects of mobile phone distraction on pedestrians' crossing behavior and visual attention allocation at a signalized intersection: An outdoor experimental study. Accid. Anal. Prev. 115, 170–177. https://doi.org/10.1016/j.aap.2018.03.019.

Kaur, J., Madan, N., 2015. Association Rule Mining: A Survey. Int. J. Hybrid Inf. Technol. 8 (7), 239–242. https://doi.org/10.14257/ijhit.2015.8.7.22.

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

Kumbhare Santosh, V., Chobe, T.A., 2014. An overview of association rule mining algorithms. Int. J. Comput. Sci. Inf. Technol. Int. J. Comput. Sci. Inf. Technol. 5 (1), 927–930.

Kuzel, M.J., Heller, M.F., Sala, J.B., Ciccarelli, L., Gray, R., 2008. A review of real-world collisions involving distracted pedestrians. Proc. XX Annu. Int. Occup. Ergon. Saf. Conf. 115–121.

Levi, S., De Leonardis M, D., Antin, J., Angel, L., 2013. Identifying countermeasure strategies to increase safety of older pedestrians. National highway traffic safety administration, U.S. Department of Transportation.

Lichenstein, R., Smith, D.C., Ambrose, J.L., Moody, L.A., 2012. Headphone use and pedestrian injury and death in the United States: 2004e2011. Inj. Prev. 18 (5), 287–290. https://doi.org/10.1136/injuryprev-2011-040161.

Lu, G. Xiao-Z., Noyce, D.A., 2009. Pedestrian Crosswalks at Midblock Locations. Transp. Res. Rec. J. Transp. Res. Board 2140, 63–78. doi:10.3141/2140-07.

Markowitz, F., Smith, A., 2017. Innovations in Lighting for Pedestrian Safety and Walkability. ITE Western District Annual Meeting, San Diego.

Mathias, J.R., Dodd, M.E., Walters, K.B., Yoo, S.K., Erik, A., Huttenlocher, A., 2011. Walking & talking: Dual-task effects on street crossing behavior in older adults. Psychol. Aging 33 (11), 1212–1217.

Meir, A., Oron-Gilad, T., Parmet, Y., 2015. Are child-pedestrians able to identify hazardous traffic situations? Measuring their abilities in a virtual reality environment. Saf. Sci. 80, 33–40. https://doi.org/10.1016/j.ssci.2015.07.007.

Miranda-Moreno, L.F., Morency, P., El-Geneidy, A.M., 2011. The link between built environment, pedestrian activity and pedestrian-vehicle collision occurrence at signalized intersections. Accid. Anal. Prev. 43 (5), 1624–1634. https://doi.org/10.1016/j.aap.2011.02.005.

Mohammed, H.A., 2021. Assessment of distracted pedestrian crossing behavior at midblock crosswalks. IATSS Res. 45 (4), 584–593. https://doi.org/10.1016/j.iatssr.2021.07.003.

Montella, A., Aria, M., D'Ambrosio, A., Mauriello, F., 2011. Data-mining techniques for exploratory analysis of pedestrian crashes. Transp. Res. Rec. 2237, 107–116. https://doi.org/10.3141/2237-12.

Mwakalonge, J., Siuhi, S., White, J., 2015. Distracted walking: Examining the extent to pedestrian safety problems. J. Traffic Transp. Eng. (english Ed. 2 5, 327–337. https://doi.org/10.1016/j.jtte.2015.08.004.

Nasar, J., Hecht, P., Wener, R., 2008. Mobile telephones, distracted attention, and pedestrian safety. Accid. Anal. Prev. 40 (1), 69–75. https://doi.org/10.1016/j.aap.2007.04.005.

Nasar, J.L., Troyer, D., 2013. Pedestrian injuries due to mobile phone use in public places. Accid. Anal. Prev. 57, 91–95. https://doi.org/10.1016/j.aap.2013.03.021.

Nasim Khan, M., Das, A., Ahmed, M.M., 2020. Non-parametric association rules mining and parametric ordinal logistic regression for an in-depth investigation of driver speed selection behavior in adverse weather using SHRP2 naturalistic driving study data. Transp. Res. Rec. 2674 (11), 101–119. https://doi.org/10.1177/0361198120941509.

NHTSA, 2020. Pedestrains: Traffic Safety Facts 2019. National Highway Traffic Safety Administration. U.S. Department of Transportation.

Osborne, R., Horberry, T., Young, K.L., 2020. Pedestrian distraction from Smartphones: An end-user perspective on current and future countermeasures. Transp. Res. Part F Traffic Psychol. Behav. 73, 348–361. https://doi.org/10.1016/j.trf.2020.07.007.

Pande, A., Abdel-Aty, M., 2009. Market basket analysis of crash data from large jurisdictions and its potential as a decision support tool. Saf. Sci. 47 (1), 145–154. https://doi.org/10.1016/j.ssci.2007.12.001.

Pešić, D., Antić, B., Glavić, D., Milenković, M., 2016. The effects of mobile phone use on pedestrian crossing behaviour at unsignalized intersections - Models for predicting unsafe pedestrians behaviour. Saf. Sci. 82, 1–8. https://doi.org/10.1016/j.ssci.2015.08.016.

Rahman, M.A., Das, S., Sun, X., Sun, M., Hossain, M.M., 2023. Using unsupervised learning to investigate injury-associated factors of animal-vehicle crashes. Int. J. Inj. Contr. Saf. Promot. 30 (2), 210–219. https://doi.org/10.1080/17457300.2022.2125532.

Rahman, M.A., Sun, X., Sun, M., Shan, D., 2021. Investigating characteristics of cellphone distraction with significance tests and association rule mining. IATSS Res 45 (2), 198–209. https://doi.org/10.1016/j.iatssr.2020.09.001.

Rasouli, A., Kotseruba, I., Tsotsos, J.K., 2019. It's not all about size: On the role of data properties in pedestrian detection. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 11129 LNCS, 210–225. https://doi.org/10.1007/978-3-030-11009-3_12.

Rella Riccardi, M., Mauriello, F., Scarano, A., Montella, A., 2022. Analysis of contributory factors of fatal pedestrian crashes by mixed logit model and association rules. Int. J. Inj. Contr. Saf. Promot. doi:10.1080/17457300.2022.2116647.

Russo, B.J., James, E., Aguilar, C.Y., Smaglik, E.J., 2018. Walking on the wild side: distracted pedestrians and traffic safety. Transfers 13–18.

Samerei, S.A., Aghabayk, K., Shiwakoti, N., Karimi, S., 2021. Modelling bus-pedestrian crash severity in the state of Victoria, Australia. Int. J. Inj. Contr. Saf. Promot. 28 (2), 233–242. https://doi.org/10.1080/17457300.2021.1907597.

Schofer, J.L., Christoffel, K.K., Donovan, M., Lavigne, J.V., Tanz, R.R., Wills, K.E., 1995. Child pedestrian injury taxonomy based on visibility and action. Accid. Anal. Prev. 27 (3), 317–333. https://doi.org/10.1016/0001-4575(94)00074-V.

Schwebel, D.C., Canter, M.F., Hasan, R., Griffin, R., White, T.R., Johnston, A., 2022. Distracted pedestrian behavior: An observational study of risk by situational environments. Traffic Inj. Prev. doi:10.1080/15389588.2022.2076282.

Schwebel, D.C., Stavrinos, D., Byington, K.W., Davis, T., O'Neal, E.E., De Jong, D., 2012. Distraction and pedestrian safety: How talking on the phone, texting, and listening to music impact crossing the street. Accid. Anal. Prev. 45, 266–271. https://doi.org/10.1016/j.aap.2011.07.011.

SGA, 2021. Dangerous by Design. Smart Growth America. Smart Growth America (SGA), Transportation for America.

Shinar, D., 1985. The effects of expectancy, clothing reflectance, and detection criterion on nighttime pedestrian visibility. Hum. Factors 27 (3), 327–333. https://doi.org/10.1177/001872088502700308.

Sivasankaran, S.K., Natarajan, P., Balasubramanian, V., 2020. Identifying patterns of pedestrian crashes in urban metropolitan roads in India using association rule mining. Transp. Res. Procedia 48, 3496–3507. https://doi.org/10.1016/j.trpro.2020.08.102.

Spainhour, L.K., Wootton, I.A., Sobanjo, J.O., Brady, P.A., 2006. Causative factors and trends in Florida pedestrian crashes. Transp. Res. Rec. J. Transp. Res. Board 1982, 1, 90–98. doi:10.1177/0361198106198200112.

Stavrinos, D., Byington, K.W., Schwebel, D.C., 2011. Distracted walking: cell phones increase injury risk for college pedestrians. J. Saf. Res. 42 (2), 101–107. https://doi.org/10.1016/j.jsr.2011.01.004.

Stavrinos, D., Pope, C.N., Shen, J., Schwebel, D.C., 2018. Distracted walking, bicycling, and driving: systematic review and meta-analysis of mobile technology and youth crash risk. Child Dev. 89 (1), 118–128. https://doi.org/10.1111/cdev.12827.

Tabibi, Z., Pfeffer, K., 2003. Choosing a safe place to cross the road: The relationship between attention adn identification of safe and dangerous road-crossing sites. Child Care Health Dev. 29 (4), 237–244. https://doi.org/10.1046/j.1365-2214.2003.00336.x.

Tapiro, H., Oron-Gilad, T., 2016. Validation of the Semi-immersive dome shaped pedestrians' virtual reality simulator. Ben-Gurion University of the Negev.

Tapiro, H., Oron-Gilad, T., Parmet, Y., 2020. Pedestrian distraction: The effects of road environment complexity and age on pedestrian's visual attention and crossing behavior. J. Saf. Res. 72, 101–109. https://doi.org/10.1016/j.jsr.2019.12.003.

Thompson, L.L., Rivara, F.P., Ayyagari, R.C., Ebel, B.E., 2013. Impact of social and technological distraction on pedestrian crossing behaviour: An observational study. Inj. Prev. 19 (4), 232–237. https://doi.org/10.1136/injuryprev-2012-040601.

Won, M., Shrestha, A., Park, K.J., Eun, Y., 2020. SaferCross: enhancing pedestrian safety using embedded sensors of smartphone. IEEE Access 8, 49657–49670. https://doi.org/10.1109/ACCESS.2020.2980085.

Zhu, H., Han, T., Alhajyaseen, W.K.M., Iryo-Asano, M., Nakamura, H., 2022. Can automated driving prevent crashes with distracted Pedestrians? An exploration of motion planning at unsignalized Mid-block crosswalks. Accid. Anal. Prev. 173 https://doi.org/10.1016/j.aap.2022.106711.