RESEARCH ARTICLE

Analyzing Daytime/Nighttime Pedestrian Crash Patterns in Michigan Using Unsupervised Machine Learning Techniques and their Potential as a Decision-making Tool

Baraah Qawasmeh[1,*] , Jun-Seok Oh[1], Valerian Kwigizile[1], Doaa Qawasmeh[2], Arar Al Tawil[2] and Ahmad Aldalqamouni[1]

[1]*Department of Civil and Construction Engineering Western Michigan University Kalamazoo, MI, 49008, United States of America*
[2]*Faculty of Information Technology, Applied Science Private University Amman, Jordan*

Abstract:

Background: Data mining applications are becoming increasingly common across various fields. Numerous data mining methodologies have been utilized in pedestrian crash data analysis. However, association and correspondence analyses have yet to be extensively employed in pedestrian safety literature to support decision-making.

Material and Methods: Road lighting significantly affects pedestrian crashes, highlighting the importance of examining pedestrian crash patterns during daytime and nighttime. This study analyzed pedestrian fatal and injury crashes in Michigan from 2011 to 2021 under varying road lighting conditions. The study identified pedestrian crash patterns using unsupervised machine-learning techniques based on several multidimensional factors. The Association Rules Learning (ARL) technique was used to determine general associations of patterns leading to pedestrian crashes. At the same time, Multiple Correspondence Analysis (MCA) helped understand crash attribute patterns in two notable scenarios: elderly pedestrian crashes during daytime lighting and high-speed midblock crashes during nighttime. These methods revealed key patterns associated with contributing factors to pedestrian crashes in different lighting conditions.

Results: In the analysis of cloud combinations involving elderly pedestrians during daylight, it was found that: 1) elderly pedestrians were significantly involved in severe injury crashes at two-lane midblock locations with a speed limit between 45 and 65 mph; 2) improper actions by elderly drivers were significantly associated with elderly pedestrians on wide roads (>5 lanes); and 3) the complex interaction between city streets with 3-5 lanes was significantly correlated with drivers 'failed to yield' actions. Additionally, in the analysis of cloud combinations involving nighttime pedestrian crashes at high-speed midblock locations, it was found that 1) a specific age group of alcohol- or drug-involved pedestrians (19-30) was significantly associated with young drivers and improper driver actions; 2) during the winter season, young pedestrians were significantly involved in moderate injury crashes at undivided midblock locations; and 3) elderly drivers and 'failed to yield' actions were highly correlated with 3-5 lane roadways.

Discussion: The research findings are expected to raise awareness of Michigan pedestrian crash patterns under daytime and nighttime lighting conditions. They will also recommend safety countermeasures for practitioners to enhance pedestrian safety in walkable cities. The proposed methodologies can uncover relationships that need to be well-documented in the existing literature on pedestrian safety. Additionally, these methods can identify valuable relationships without restricting variables to being either dependent or independent, and they can reveal relationships that might be challenging to detect otherwise. The results will also provide decision rules and visualizations that are easy to understand.

Conclusion: The results showed that the proposed techniques can analyze pedestrian crash data in a way that aligns with understanding pedestrian crash characteristics. Traffic safety administrators could benefit from this methodology as a decision-support tool.

Keywords: Pedestrian, Lighting, ARL, MCA, Darkness, Unsupervised learning, Safety, Nighttime.

*Address correspondence to this author at the Department of Civil and Construction Engineering, Western Michigan University Kalamazoo, MI, 49008, United States of America; E-mail: baraah.qawasmeh@wmich.edu

1. INTRODUCTION

The popularity of walking as a mode of transportation is increasing. In 2021, approximately 3.4 million American workers reported walking to work in the past week, according to the American Community Survey (ACS) by the U.S. Census Bureau [1]. However, pedestrian fatalities and injuries have also risen. In 2021, 7,341 pedestrian fatalities accounted for 18.6 percent of all traffic deaths in the U.S., as reported by the Fatality Analysis Reporting System (FARS) [2]. In Michigan, the situation is more severe, with a pedestrian fatality rate of 1.55 per 100,000 population, ranking it No. 19 among the most dangerous states for pedestrians [3]. In 2022, the National Transportation Safety Board recommended analyzing police crash records at the state level to better assess pedestrian injuries and fatalities [4]. This is a task that requires collaboration among various entities, including transportation authorities, policymakers, and the general public. The Federal Highway Administration (FHWA) has initiated projects to reduce non-motorist fatalities, while the Michigan Department of Transportation (MDOT) has adopted the Toward Zero Deaths (TZD) national strategy to enhance road user safety [5].

One primary driver error leading to crashes is the failure to detect other road users. Notably, pedestrian crashes in dark conditions tend to be more severe [6]. The overrepresentation of pedestrian casualties at night is partly due to reduced visibility and other factors, such as alcohol and drug use by drivers and pedestrians, as well as speeding. In Michigan, 52.0% of suspected severe injury crashes and 75% of fatal pedestrian crashes occurred during nighttime (both with and without streetlights) from 2011 to 2021. Conversely, non-fatal pedestrian crashes are significantly more likely to happen during daylight hours. These distinctions highlight the variation in pedestrian crashes based on illumination conditions, indicating a need for further investigation.

Numerous previous studies and national reports have highlighted that lighting conditions are critical to pedestrian crashes [7-10]. The majority of these studies indicate that fatal pedestrian crashes are more likely to occur at night when hazardous factors, such as reduced visibility, speeding, and intoxicated driving, are heightened. In contrast, nonfatal pedestrian crashes are more likely to occur during the daytime, as drivers can better assess the situation and adjust their movements accordingly. However, this may present a minor safety concern from a visibility perspective. Although less investigated, nonfatal injuries and their contributing factors are more common. This investigation will examine the variation in the associations of contributing factors for pedestrian crashes during daytime and nighttime lighting conditions by correlating pedestrian injury severity with these conditions. It will consider various risk factors, including crash location (intersections and midblock segments), pedestrian age and gender, roadway geometry, and environmental characteristics.

A series of events that are directly or indirectly associated can lead to a singular crash. A crash results from the random interaction of multiple factors, including humans, vehicles, roadways, and environmental conditions. Understanding these interactions and associations is key to pedestrian safety. Machine learning methods, such as the association rules learning (ARL) and multiple correspondence analysis (MCA) methods, can help in finding out patterns affecting crash severity [11-13] and providing valuable insights to support the decision-making process. This investigation will analyze fatal pedestrian crashes and injuries in Michigan over eleven years (2011-2021) under daytime and nighttime road lighting conditions. ARL and MCA will be employed to identify pedestrian crash patterns. Unlike previous studies that treated roadway lighting conditions as a single factor, this research categorizes lighting into daytime and nighttime. These techniques are particularly effective for analyzing observational data, such as pedestrian crash data, collected outside controlled experiments [14]. Prior research linking lighting conditions to pedestrian injury severity often relied on parametric methods that assumed independence among crash-contributing factors. However, understanding the interactions among these factors is crucial for enhancing roadway safety. Therefore, further research using innovative methods is needed to inform the selection of effective countermeasures. While the proposed methods have distinct advantages, they are intended to complement existing statistical techniques for crash data analysis. They will be valuable tools for examining extensive crash characteristics databases from jurisdictions like MDOT and selecting appropriate safety interventions.

2. LITERATURE REVIEW

Safety literature reveals various risk factors

influencing pedestrian crashes and their severity [15]. Some of these variables are associated with pedestrian/ driver demographics (age, gender), roadway features (number of lanes, posted speed limit, presence of intersections, traffic control type, functional classification, and presence of medians), environmental characteristics (weather and lighting condition), and temporal factors, including season and day of the week [16-20]. A study was conducted in San Antonio, Texas, using six-year crash data. They investigated the effects of pedestrian-vehicle crash-related variables on pedestrian injury severity based on the party at fault and identified high-risk locations [21]. The authors concluded that there was a far greater likelihood of fatal and incapacitating injuries when pedestrians were at fault [21].

It is a well-established fact that darkness is a significant risk factor for pedestrian fatalities [22-24]. Wood *et al*. (2005) further explored this issue by analyzing the impact of driver age on nighttime pedestrian recognition. They discovered that elderly drivers have a 60% reduction in the distance at which they can detect pedestrians compared to younger driver age groups [25]. Additionally, alcohol impairment by both drivers and pedestrians is another factor that consistently correlates with crash risk and darkness and pedestrian injury severity [26, 27].

Comprehensive studies that consider lighting conditions and how they relate to the severity of pedestrian injuries are less common in the literature. An earlier study by Siddiqui *et al*. [28] determined how pedestrian injury severity was influenced by lighting conditions at pedestrian crossings. They found that, regardless of lighting conditions, pedestrian fatalities are more likely to occur at segments than at intersections. Machine learning (ML) techniques in data mining have seen explosive growth in recent years and have garnered interest from various research domains, such as safety studies. According to a survey conducted in California using the Multinomial Logit Model (MNL) and the support vector machine method (SVM), contributing factors to pedestrian crash injury severity were determined in both daylight and dark times [29]. The authors investigated how some contributing factors were significant at daytime injury severity (weather condition, roadway type, truck at fault, parked motor vehicle in crash, and driver sobriety), while other factors were significant at nighttime injury severity (rainy weather condition, head-on crash, pedestrian crossing, intersection location, and multilane-undivided-non-freeway roadway) [29]. In both studies, lighting conditions were examined in relation to pedestrian crashes, but there was no investigation into the correlation or association between the contributing factors and the lighting conditions.

Traditional statistical models are characterized by predefined associations and underlying assumptions, which, if disregarded, can result in inaccurate conclusions. These models are limited in their ability to demonstrate the independent impact of a single factor on the severity or probability of a pedestrian crash. Given the frequent occurrence of crashes due to the intricate interplay of numerous factors, it is essential to examine the relationships among multiple crash factors concurrently [30]. To uncover the hidden patterns of association between crash attributes, it is necessary to investigate new and innovative analytical techniques. Also, it has become common practice in traffic safety research to use non-parametric approaches. Using these methods, we can identify hidden associations between various crash factors without restricting their nature. Non-parametric methods enable traffic safety studies to capture the non-linear effects of continuous and discrete variables without the need for prior assumptions. A few recent studies have used ARL to identify significant contributing factors to pedestrian crashes [11-13]. Despite this, none of the studies examined the association between lighting conditions and pedestrian crashes.

Additionally, there needs to be more research that has implemented MCA to analyze crash patterns. For instance, MCA was implemented in Hawaii to investigate the correlations between pedestrian characteristics and circumstances (including at-fault pedestrians, severe injuries, location, and time) [31]. In a previous investigation, the key factors in pedestrian crashes were dissected using MCA in Louisiana, where pedestrian crash data had been collected for eight years [32]. Similarly, in Malta, the correlation between crash variables (demographics, temporal, and geographic) and pedestrian crash characteristics was also investigated using MCA [33]. In other research, MCA was employed to examine contributing factors to pedestrian crashes in specific locations, such as signalized intersections [34]. Nevertheless, none of the mentioned studies have investigated pedestrian crash patterns in locations with varying lighting conditions. They have not been conducted in conjunction with ARL as a decision-support tool to improve the selection of appropriate strategies and safety countermeasures. This gap in the literature emphasizes the necessity of our proposed research.

As ARL and MCA techniques reveal patterns of associations, they might be far more effective tools since no variables are designated dependent or independent. This study will design a novel methodology using two unsupervised learning algorithms, ARL and MCA. This will result in complementarity between them. The ARL algorithm will be applied first to identify the general associations of patterns leading to pedestrian crashes in Michigan. After that, the MCA algorithm will be used to understand pedestrian crash attribute patterns based on the highest lift value rules associated with daytime/ nighttime road lighting conditions, which the ARL algorithm will determine.

So, the primary goals of this study are to (1) identify roadway features, environmental characteristics, and vehicle and driver-related patterns that contribute to the overrepresentation of pedestrian crashes under various lighting conditions, such as daylight, dark-lighted (pedestrian nighttime crashes on a lighted roadway), and dark-unlighted (pedestrian nighttime crashes on an

unlighted roadway), by analyzing eleven years (2011-2021) of pedestrian fatal and injury crashes in Michigan, and (2) to identify safety countermeasures to mitigate the detected pedestrian crash patterns.

The paper is organized as follows: the methods section shows the crash data and associated descriptive statistics and introduces the used methodologies; the result and discussion section discusses the analysis results; the conclusion section summarizes significant study findings and suggests safety countermeasures to mitigate the detected pedestrian crash patterns and describes study limitations.

3. MATERIALS AND METHODS

3.1. Crash Data

Data for the analysis consisted of pedestrian-involved crashes in Michigan from 2011 to 2021. Three crash data files (crash level record, crash location record, and involved-party record) were merged based on the variable "crsh_id", which represents the crash identification number. The final dataset was filtered based on the variable "prty_type, "lit_cd", and injy_svty_cd," which means the type of involved party, the lighting conditions at the time of the crash, and the degree of injury suffered by the involved party as a result of the crash, respectively. By selecting the pedestrian as the involved party type, 25,627 unique pedestrian crashes were extracted from the main database.

The database contains information on the crash lighting condition using a numerical scale (1 = Daylight, 2 = Dawn, 3 = Dusk, 4 = Dark-lighted, 5 = Dark-unlighted, 6 = Unknown), and information on the crash severity levels on a numerical scale (1 = Fatal injury (K), 2 = Severe injury (A), 3 = Moderate injury (B), 4 = Possible injury (C), 5 = No injury (O)). The (KABC) injury classes were filtered, and the no-injury class of 'O' was skipped to focus only on all levels of pedestrian injury severity. After that, lighting conditions (daylight, dark-lighted, and dark-unlighted) were selected, and dawn, dusk, and unknown conditions were skipped because they account for only about 6% of all crashes. Finally, the study came up with 21,158 unique pedestrian injury crashes under (daylight, dark-lighted, and dark-unlighted) lighting conditions using three-step filter criteria. Fig. (**1**) describes the data extraction procedure, and Fig. (**2**) shows crash severity percentages for the three most common lighting conditions (daylight, dark-lighted, and dark-unlighted). In some cases, lighted roadways may not provide sufficient illumination to improve visibility at night, which might be due to the implementation of inappropriate lighting countermeasures. As shown in Fig. (**2**), in Michigan, severe pedestrian crashes are more likely to occur at nighttime, where 68% of the fatal and severe injury pedestrian-involved crashes took place under dark-lighted roadway conditions compared to 58% under dark-unlighted roadway conditions. In comparison, only 34.0% of pedestrian-involved crashes with no injury occurred in either dark-lighted or dark-unlighted conditions.
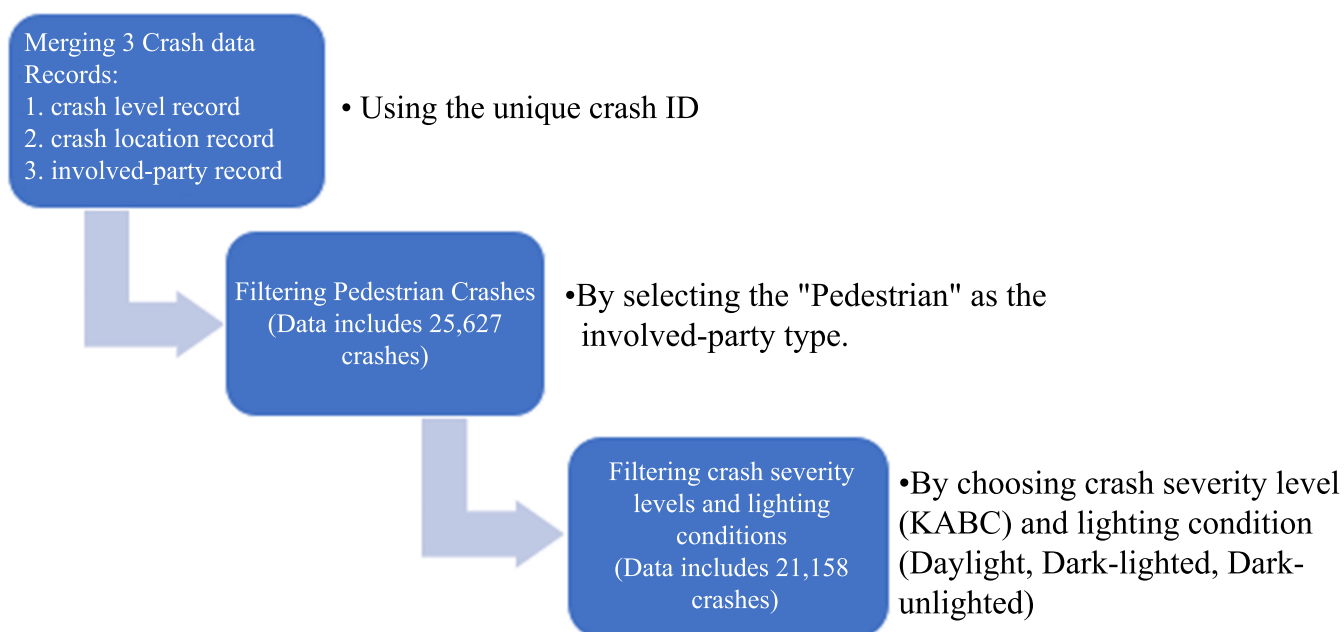

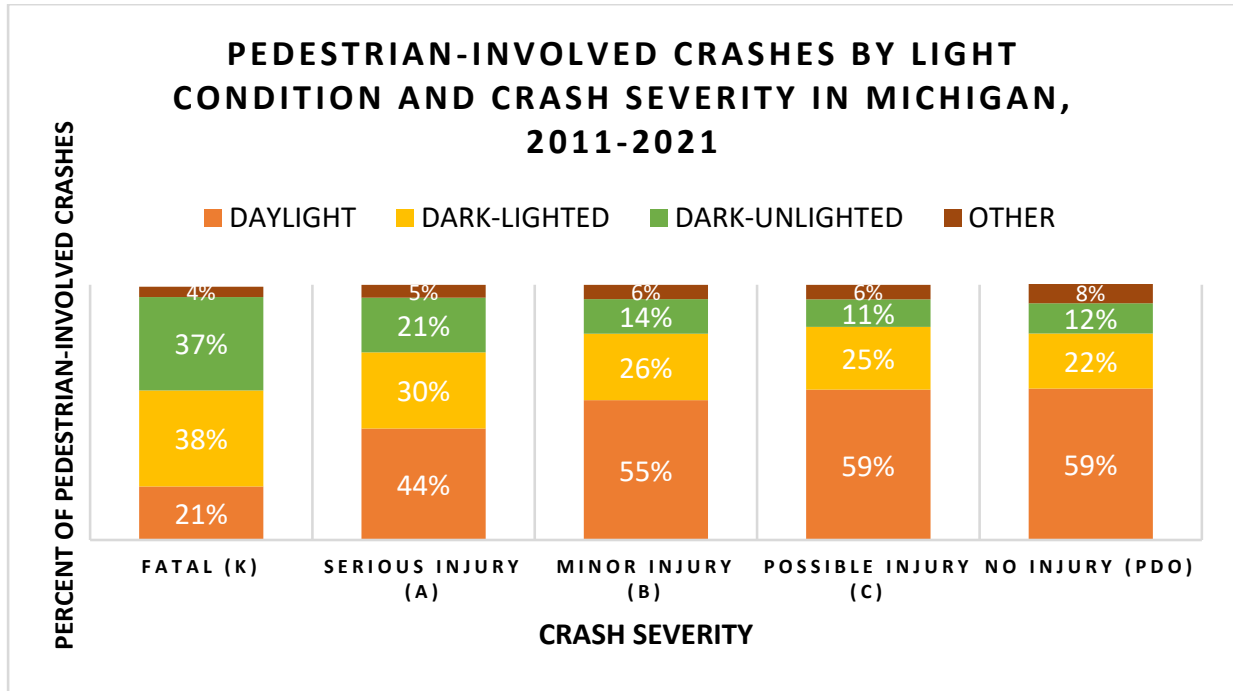**Fig. (1).** Data extraction procedure.

Fig. (2). Pedestrian-involved crashes by light condition and crash severity in Michigan from 2011 to 2021.

Generally, crash databases are classified through multiple variables describing crash characteristics, roadway characteristics, environmental conditions, and characteristics of people involved in the crash (including drivers and pedestrians). Table **1** summarizes the descriptive statistics of all critical variables used in the final dataset. The variable selection method was developed based on previous relevant research, engineering judgment, and findings from previous pedestrian crash analyses (28,31,40–42). Additionally, due to the high skewness of the data, several variables that might have affected the generation of the association rules were excluded. For example, over 50% of pedestrian crashes did not have a traffic control type recorded in the dataset. As a result, the "traffic control type" variable was omitted from the final dataset. For this study, 16 variables were only considered, including 'lighting conditions'. These variables cover several factors contributing to pedestrian crashes: 1. Pedestrian characteristics (age, gender, alcohol/drug involvement); 2. Driver characteristics (age, hazardous action, alcohol/drug involvement); 3. Roadway characteristics (roadway type, presence of intersections, posted speed limit, roadway condition, number of lanes); 4. Environmental conditions (lighting, weather); 5. Temporal factors (day of the week, season); and 6. Crash characteristics (crash severity level).

Table 1. Descriptive statistics of pedestrian crashes that occurred under three different lighting conditions.

Selected Variables	Code	Lighting Conditions		
		Daytime	Nighttime	
		Daylight Count (%)	Dark-lighted Count (%)	Dark-unlighted Count (%)
Severity level	**"Severity"**			
Fatal injury (K)	"Fatal_Injury"	359 (3%)	643 (11%)	627 (18%)
Suspected serious injury (A)	"Serious_Injury"	1,885 (16%)	1,287 (21%)	923 (26%)
Suspected minor injury (B)	"Minor_Injury"	4,049 (35%)	1,923 (32%)	1,002 (28%)
Possible injury (C)	"Possible_Injury"	5,286 (46%)	2,206 (36%)	968 (28%)
Day of week	**"Day_of_week"**	-		
Weekday (M,T,W,T)	"Weekday"	7,371 (64%)	3,280 (54%)	1,905 (54%)
Weekend (F,S,S)	"Weekend"	4,208 (36%)	2,779 (46%)	1,615 (46%)
Season	**"Season"**	-		

(Table 3) contd.....

Selected Variables	Code	Lighting Conditions		
		Daytime	Nighttime	
		Daylight Count (%)	Dark-lighted Count (%)	Dark-unlighted Count (%)
Severity level	**"Severity"**			
Spring (3,4,5)	"Spring"	2,839 (24%)	1,025 (17%)	580 (16%)
Summer (6,7,8)	"Summer"	3,425 (30%)	1,008 (16%)	627 (18%)
Fall (9,10,11)	"Fall"	3,217 (28%)	2,048 (34%)	1,169 (33%)
Winter (12,1,2)	"Winter"	2,098 (18%)	1,978 (33%)	1,144 (33%)
Weather condition	**"Weather_cond"**	-		
Clear	"Clear"	8,079 (70%)	3,613 (60%)	1,991 (57%)
Cloudy	"Cloudy"	2,409 (21%)	972 (16%)	705 (20%)
Rain	"Rain"	650 (5%)	1,085 (18%)	510 (14%)
Snow/Fog/Sleet/Wind	"Severe"	358 (3%)	326 (5%)	261 (7%)
Unknown	"Unknown"	80 (1%)	62 (1%)	57 (2%)
Roadway type	**"Roadway_type"**	-		
Undivided road	"Undivided"	8,985 (78%)	4,669 (77%)	2,851 (80%)
Divided road	"Divided"	912 (8%)	667 (11%)	414 (12%)
One_way road	"One-way"	557 (5%)	255 (4%)	79 (2%)
Other/Unknown	"Other"	1,055 (9%)	488 (8%)	226 (6%)
Highway classification	**"Hwy_clss"**	-		
Interstate	"Interstate"	525 (5%)	359 (6%)	214 (6%)
US Route	"US_route"	588 (5%)	340 (5%)	217 (6%)
Michigan Route	"State_route"	2,052 (18%)	1,361 (22%)	623 (17%)
City Street	"City_str"	7,530 (67%)	3,779 (61%)	2,384 (65%)
Other	"Other"	600 (5%)	377 (6%)	209 (6%)
Crash location	**"Crash_location"**	-		
Intersection	"Intersection"	7,020 (61%)	4,003 (66%)	1,453 (41%)
Midblock	"Midblock"	3,490 (30%)	1,671 (28%)	1,878 (53%)
Other	"other"	1,069 (9%)	385 (6%)	189 (6%)
Posted speed limit	**"Speed_limit"**	-		
<45 MPH	"<45 MPH"	8,789 (78%)	4,700 (76%)	1,584 (42%)
45-65 MPH	"45-65 MPH"	1,940 (17%)	1,076 (18%)	1,674 (45%)
>65 MPH	">65 MPH"	154 (2%)	82 (1%)	168 (5%)
Unknown	"Unknown"	376 (3%)	305 (5%)	310 (8%)
Roadway condition	**"Roadway_cond"**	-		
Dry	"Dry"	9,550 (82%)	3,950 (65%)	2,311 (66%)
Wet	"Wet"	1,277 (11%)	1,682 (28%)	808 (23%)
Slippery	"Slippery"	758 (7%)	424 (7%)	398 (11%)
Number of lanes	**"Num_lanes"**	-		
Single lane	"Single_lane"	1,188 (10%)	290 (5%)	148 (4%)
2 lanes	"2_lanes"	5,414 (47%)	2,093 (35%)	1,964 (56%)
3-5 lanes	"3-5_lanes"	4,430 (38%)	3,237 (53%)	1,280 (36%)
>5 lanes	">5_lanes"	524 (5%)	454 (7%)	136 (4%)
Driver hazardous action	**"driver_hzrd_actn"**	-		
None	"None"	6,755 (58%)	2,770 (46%)	1,361 (38%)
Failed to yield	"Failed_to_yield"	1,089 (9%)	509 (9%)	241 (7%)
Improper action	"Improper_action"	2,767 (24%)	1,996 (33%)	1,486 (42%)
Unknown	"Unknown"	1,006 (9%)	730 (12%)	448 (13%)
Driver age	**"Driver_age"**	-		
Young driver <25	"Young_driver"	1,672 (15%)	922 (15%)	661 (18.3%)
Elderly driver >64	"Elderly_driver"	2193 (19%)	726 (12%)	434 (12%)
Other	"Other_ages"	7,627 (66%)	4,401 (72%)	2,522 (70%)
Alcohol/Drug involvement	**"alch_drug_ind"**	-		
No	"No"	11,029 (95%)	4,773 (79%)	2,598 (74%)
Yes	"Yes"	550 (5%)	1,286 (21%)	922 (26%)
Pedestrian age	**"Ped_age"**	-		

(Table 3) contd.....

Selected Variables	Code	Lighting Conditions		
		Daytime	Nighttime	
		Daylight Count (%)	Dark-lighted Count (%)	Dark-unlighted Count (%)
Severity level	**"Severity"**			
<=18	"<18"	3,380 (29%)	976 (16%)	565 (16%)
19-30	"18-30"	2,285 (20%)	1,704 (28%)	1,013 (28%)
31-64	"31-64"	2,400 (21%)	1,777 (29%)	1,052 (30%)
>64	">64"	3,405 (29%)	1,540 (26%)	875 (25%)
Unknown	"Unknown"	120 (1%)	35 (1%)	31 (1%)
Pedestrian gender	**"ped_gndr"**	-		
Male	"M"	6,641 (57%)	3,810 (63%)	2,423 (68%)
Female	"F"	4,868 (42%)	2,218 (36%)	1,082 (31%)
Unknown	"Unknown"	52 (1%)	35 (1%)	29 (1%)

Table **1** illustrates some of the most notable pedestrian crash features under different lighting conditions, including 1. The likelihood that a crash involving pedestrians occurred at night increases with the severity of the crash. 2. Regardless of lighting conditions, weekday pedestrian crashes or crashes involving males were more likely to occur. 3. There were multiple variables over-represented under 'dark-unlighted' conditions, including midblock crash locations, undivided roadways, roadways with (45-65 mph) speed limits, 2-lane roadways, driver improper actions, young drivers, and middle-aged pedestrians; and under 'dark-lighted' condition (including alcohol/drug involvement=Yes, middle-aged drivers, wet roadway condition, and rain weather condition).

3.2. Association Rules Learning (ARL) Algorithm

ARL is a rule-based machine learning technique that can discover interesting relationships between variables [35]. Also, it is an example of an unsupervised learning algorithm that can be very effective in identifying hidden patterns from a sizeable untagged database and offering significant insights that may be applied to data-driven decision-making [35]. In 1993, this technique was introduced [36], and it has since become widely used. It was utilized to find patterns between products in large-scale transaction data captured by supermarket point-of-sale systems [36]. In addition, ARL is used today in a wide variety of application fields, such as web usage mining, intrusion detection, continuous production, and bio-informatics [35].

Overall, ARL aims to identify item sets that occur together frequently in an event. It looks for any association between the dataset's items based on different rules to discover relevant relations [37]. In this study, 'itemset' is a set of variable categories summarized in Table **1**. At the same time, an 'event' is a pedestrian crash that occurs under a specific lighting condition (daylight, dark-lighted, and dark-unlighted).

3.2.1. Apriori Algorithm

The rules for this study were generated using the "Apriori" algorithm, the most widely used algorithmic framework for applying ARL [38]. A "bottom-up" methodology known as "Candidate Generation" is employed, in which groups of candidates are tested against the data, and frequent subsets are extended one item at a time. The algorithm terminates when no successful extensions are detected [38]. Furthermore, this algorithm employs a Hash tree structure and breadth-first search to effectively count candidate item sets [38].

Let $I = \{i_1, i_2, i_3, ... i_n\}$ be a set of pedestrian crash attributes called items (N) and $D = \{t_1, t_2, t_3, ... t_m\}$ be a pedestrian crash database of transactions where each transaction T contains a set of items $T \in I$. Each transaction has a unique identifier. An association rule is denoted by the form of Antecedent (left-hand-side (L.H.S.) of the rule) → Consequent (right-hand-side (R.H.S.) of the rule) or A → B, where $A \in I$ and $B \in I$. The Antecedent is an item in the data, and the consequent is an item that appears when the Antecedent is combined with another item [36, 39, 40]. The statement A → B is often read as "if A then B", where "if" refers to the Antecedent (A) and "then" refers to the consequent (B). This states that (B) should theoretically occur whenever (A) does in a dataset. It is crucial to note that the antecedent and consequent are disjoint, which means they do not share any elements $(A \cap B = \emptyset)$ [36, 39, 40].

Three traditional measures of significance are associated with each generated rule: Confidence, Support, and Lift [36, 39, 40]. 'Confidence' measures the strength of the rule by estimating the probability P(X|Y), whereas 'Support' shows how frequently the antecedent (A) and consequent (B) of a given rule occur together in the database [36, 39, 40]. Studies in the field of data mining have suggested various measures of significance to generate more insightful rules. The most common measure of performance used in association rules is 'Lift,' which measures the number of times that A and B occur together in comparison to the number of times that one would have predicted if A and B were statistically independent [36, 39, 40]. A and B are considered to be positively dependent on each other if the value of lift is greater than 1, which means that they appear together more frequently in the data; however, if the value of lift is equal to 1, then A and B are independent, and there is no

correlation between them; and if the value of lift is less than 1, then item set A is exclusive to an itemset B [36, 39, 40]. Support, confidence, and lift measures are used to evaluate the importance of an association rule in general instead of a confusion matrix (since there are no class labels present). The equations for Lift (L), Confidence (C), and Support (S) are given below: (Eq: **1**-**5**)

$$S(A) = \frac{n(A)}{N} \tag{1}$$

$$S(B) = \frac{n(B)}{N} \tag{2}$$

$$S(A \rightarrow B) = \frac{n(A \cap B)}{N} \tag{3}$$

$$C(A \rightarrow B) = \frac{S(A \rightarrow B)}{S(A)} \tag{4}$$

$$L(A \rightarrow B) = \frac{C(A \rightarrow B)}{S(B)} = \frac{S(A \rightarrow B)}{S(A) * S(B)} \tag{5}$$

Where; N: number of crashes, n (A): frequency of incidents with A, n(B): frequency of incidents with B, *n (A ∩ B)*: frequency of incidents with both A and B, S $(A \rightarrow B)$ Support of the association rule $(A \rightarrow B)$, C$(A \rightarrow B)$: Confidence of the association rule $(A \rightarrow B)$ and L $(A \rightarrow B)$: Lift of the association rule $(A \rightarrow B)$.

3.3. Boruta Algorithm for Feature Selection

Since ARL is a technique that introduces noisy factors to the model, it could lead to highly complicated and confusing "decision rules" [41]. Boruta is a feature ranking and selection method based on the Random Forest algorithm [42]. The advantages of Boruta include determining a variable's significance and assisting in statistically selecting significant variables [43]. The algorithm copies features from the original dataset. In this copy, the values in each column are randomly shuffled to create randomization. These shuffled features (Shadow Features) are then joined with the original features [43]. After that, Boruta builds a random forest classifier on the merged dataset and compares the original variables with the randomized variables to measure variable importance. Only variables with greater significance than the randomized variables are considered 'important' [43]. The primary objective of the variable selection procedure was to identify the input variables that had the most significant association with the target variable (*i.e.*, lighting condition) from the list of primary selected variables. These variables were subsequently utilized in the ARL method.

The open-source program R version 4.0.1 with 'Boruta', 'arules', 'arulesViz', and 'arulesCBA' R-packages was used to confirm the feature selection, complete the analysis, and interactively visualize the generated association rules [44, 45].

3.4. Multiple Correspondence Analysis (MCA)

MCA is an extension of correspondence analysis (CA) used to identify association patterns among categorical variables in a large dataset. The data are represented as clouds in a multivariate Euclidean space [46, 47]. It is an unsupervised learning technique that does not require the distinction between dependent (target) and independent variables [48]. The spatial distribution of the points in multiple dimensions can be generated by MCA, with the distance between points as an indicator of the degree of similarity between individual points [12]. Points exhibit a greater degree of remarkable similarity when their distances are reduced [49]. Three key steps determine the spatial arrangement of the points: generating the indicator matrix, calculating individual clouds, and computing category clouds.

Let us consider a scenario where the number of individual records associated with category k is denoted by n_k (with $n_k > 0$), where, $f_k = n_k/n$ is the relative frequency of individuals who are associated with category k. The variables significantly influence the distance between two individual records, each with its own category. The category k is present in a unique record i, and category k' is present in a unique record i' of a given variable z. The equation below defines the squared distance between unique records i and i' (Eq. **6**):

$$d_q^2(i, i') = \frac{1}{f_k} + \frac{1}{f_{k'}} \tag{6}$$

Denoting Q as the number of variables, the overall squared distance between i and i′ is defined by: (Eq. **7**)

$$d_q^2(i, i') = \frac{1}{Q} \sum_{q=Q} d_q^2(i, i') \tag{7}$$

The cloud of categories consists of a weighted set of K points, where each category k is represented by a point M^k with weight n_k. For each variable, the total weight of the category points sums to n, and for the entire set of K categories, the sum is nQ. The relative weight pk of point M^k is calculated as $p^k = n_k/(nQ) = f_k/Q$. For each variable, the sum of the relative weights of the category points equals 1/Q, and for the entire set, the total sum is 1. If $n_{kk'}$ represents the number of records that include both categories k and k′, then the squared distance between M^k and $M^{k'}$ is determined by the following formula: (Eq. **8**)

$$(M^k M^{k'})^2 = \frac{n_k + n_{k'} - 2n_{kk'}}{n_k n_{k'} / n} \tag{8}$$

The subsequent studies provide additional information regarding the fundamental theory of MCA and model development [50, 51]. MCA will be implemented in this investigation through the utilization of the 'FactoMineR' package in the R statistical software version 4.0.1. The methodology proposed is illustrated in Fig. (**3**).

4. RESULTS AND DISCUSSION

4.1. Variable Importance Plot from Boruta Algorithm

Fig. (**4**) shows the variable importance plot resulting from the Boruta algorithm. Blue boxplots correspond to the shadow attribute's minimum, average, and maximum Z scores. Red and green boxplots represent Z scores for

rejected and confirmed attributes. The associated Mean Decrease Accuracy (MDA) for each variable on the x-axis is displayed on the y-axis. MDA measures the accuracy loss when a variable is not included in the model. It is important to note that the rejected attribute has a lower MDA score than Max-shadow. A total of 13 important variables were selected for further analysis after 11 iterations in 9.6 minutes. These variables are alcohol/drug involvement, season, severity, driver hazardous action, pedestrian age, roadway condition, weather condition, speed limit, number of lanes, highway classification, crash location, roadway type, and driver age.

The final database used in this study has 14 columns (variables) and 21,158 rows (each row represents a unique pedestrian crash). There are 53 variables categories, with multiple categories for each variable. The total number of elements in the resulting matrix (21,158 x 53) is 1,121,374 elements. Fig. (5) highlights the relative frequencies of each item (variable category) in the dataset. Notably, the top item 'alch_drug_ind=No' appears 18400 times in the

final database, a significant finding. Therefore, 18400 divided by 21158 equals 0.87, which is the relative frequency of this item. Some of the other commonly appearing variables in the dataset are: "Roadway_type=Undivided" (0.78), "Roadway_cond=Dry" (0.75), "Speed_limit=45 MPH" (0.71), "Driver_age=Other_ages" (0.69), and so on.

4.2. Results from the Apriori Algorithm

The Apriori algorithm effectively identifies interactions among numerous variable categories within a sizeable multidimensional dataset. As mentioned earlier, ARL significance measures, including support, confidence, and lift, are applied to identify frequently occurring items in the crash database. The top patterns can be determined by ranking these frequent items based on their lift values, allowing crash patterns to be organized accordingly. Moreover, the general efficacy of an association rule is assessed using support, confidence, and lift measures rather than a confusion matrix, as the ARL algorithm does not include class labels.
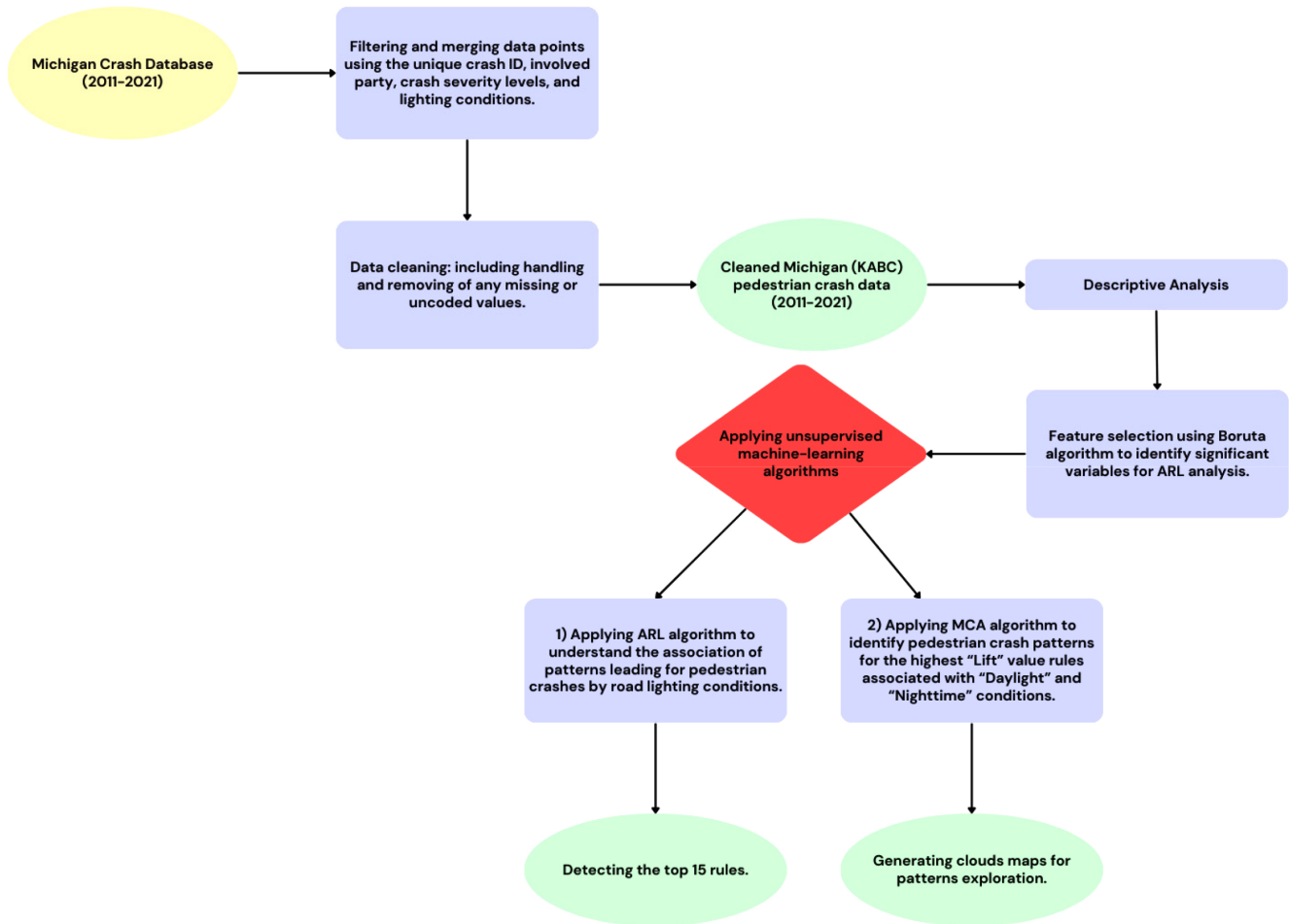
Fig. (3). An analytical framework for the proposed methodology.
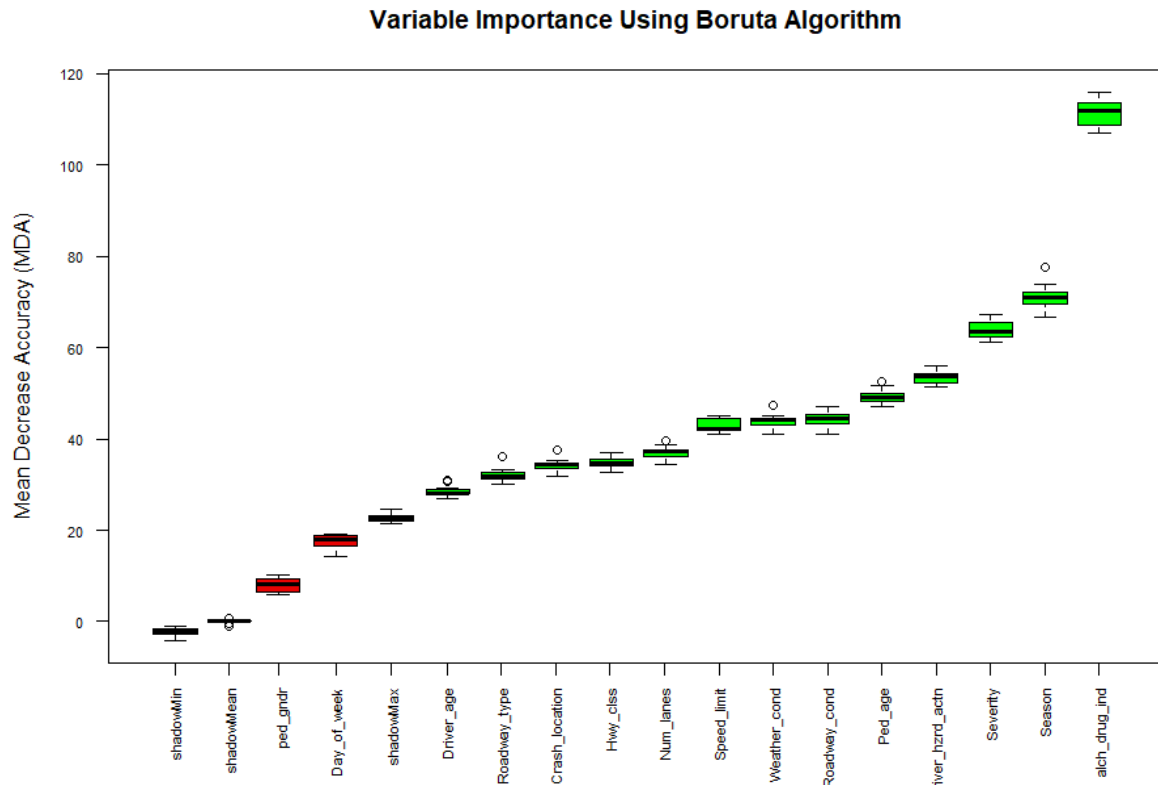
Variable Importance Using Boruta Algorithm

Fig. (4). A variable importance plot produced by the Boruta algorithm.

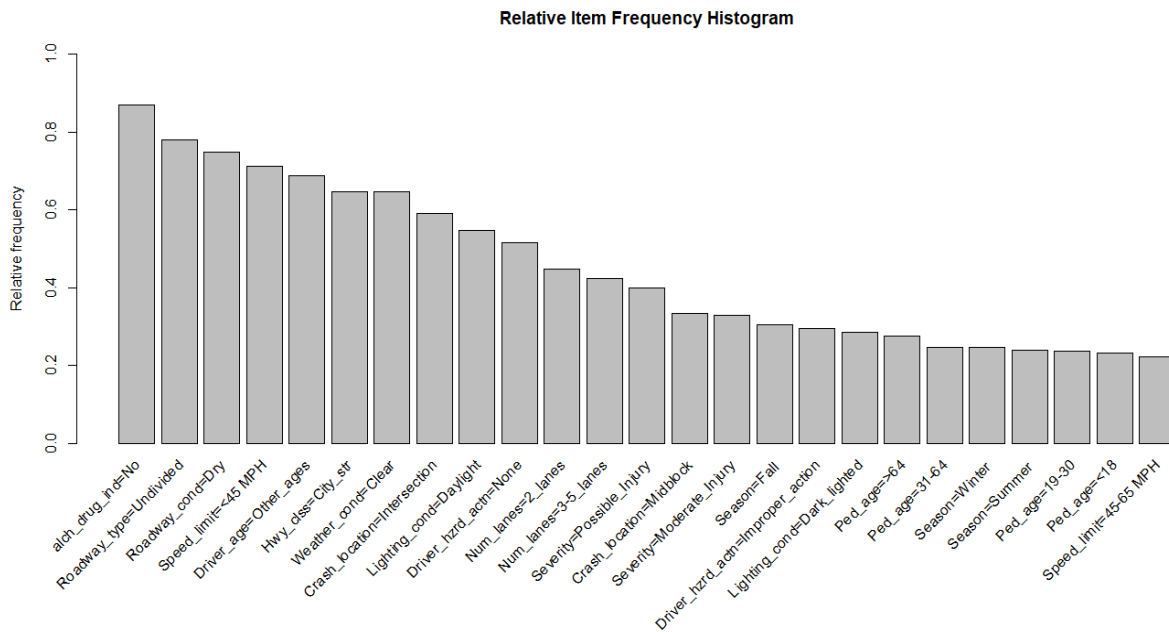**Relative Item Frequency Histogram**

Fig. (5). Relative frequency histogram for the top 25 items.

Table 2. Association rules from apriori algorithm during daytime/nighttime lighting conditions.

Rule ID	L.H.S.	R.H.S.		S* %	C* %	L*
A1	{Ped_age=>64, Roadway_cond=Dry}	Daytime	Scenario 1: Lighting condition = daylight	*10.9*	*77.8*	*1.42*
A2	{Severity=Possible_Injury, Season=Summer}			13.2	76.1	1.39
A3	{Ped_age=<18, Roadway_cond=Dry}			11.4	76.1	1.39
A4	{Ped_age=<18, Season=Summer}			14.3	75.8	1.38
A5	{Ped_age=<18, Severity=Possible_Injury}			14.3	75.8	1.38
A6	{Season=Summer, Driver_age=Elderly_driver}			10.9	75.3	1.38
A7	{Ped_age=<18, Hwy_clss= City_str}			12.3	74.9	1.37
A8	{Ped_age=<18, Crash_location=Intersection}			10.2	74.6	1.36
A9	{Roadway_cond=Dry, Driver_hzrd_actn=None}			13.4	74.4	1.36
A10	{Ped_age=<18, Speed_limit=<45 MPH}			12.2	74.4	1.36
A11	{Ped_age=>64, Weather_cond=Clear}			11.5	74.1	1.35
A12	{Num_lanes=2_lanes, Season=Summer}			11.0	74.0	1.35
A13	{Weather_cond=Cloudy, Severity=Possible_Injury}			15.3	73.9	1.35
A14	{Crash_location=Intersection, Season=Summer, Severity=Possible_Injury}			10.9	73.5	1.34
A15	{Ped_age>64, Weather_cond=Cloudy, Severity= Moderate _Injury}			12.6	73.0	1.33
B1	{Crash_location=Midblock, Speed_limit=45-65 MPH}	Nighttime (Lighted + Unlighted Roadways)	Scenario 2a: Lighting condition = dark-unlighted Roadway	*1.5*	*73.0*	*4.39*
B2	{Speed_limit=45-65 MPH, Severity=Fatal_Injury}			1.8	64.9	3.90
B3	{Num_lanes=2_lanes, Speed_limit=45-65 MPH}			1.9	62.6	3.77
B4	{Crash_location=Midblock, alch_drug_ind=Yes}			1.5	62.4	3.75
B5	{Crash_location=Midblock, Severity=Fatal_Injury}			1.4	61.8	3.72
B6	{ alch_drug_ind=Yes, Speed_limit=45-65 MPH}			1.0	61.5	3.69
B7	{Severity=Fatal_Injury, Driver_hzrd_actn=Improper_action}			1.4	60.2	3.62
B8	{Speed_limit=45-65 MPH, Driver_hzrd_actn=Improper_action}			1.2	59.7	3.59
B9	{Speed_limit=45-65 MPH, Severity=Severe_Injury}			2.0	59.7	3.59
B10	{Severity=Fatal_Injury, Season=Fall}			1.1	59.3	3.56
B11	{Speed_limit=45-65 MPH, Severity=Fatal_Injury, alch_drug_ind=Yes}			1.4	58.5	3.52
B12	{Crash_location=Midblock, Speed_limit=45-65 MPH, alch_drug_ind=Yes}			1.1	58.0	3.48
B13	{Speed_limit=45-65 MPH, Severity=Fatal_Injury, Ped_age=19-30}			1.1	57.3	3.45
B14	{Speed_limit=45-65 MPH, Roadway_type=Undivided, Severity=Fatal_Injury}			1.7	56.8	3.42
B15	{Crash_location=Midblock, Num_lanes=2_lanes, Severity=Fatal_Injury}			1.3	56.5	3.40
C1	{Weather_cond=Rain, alch_drug_ind=Yes}		Scenario 2b: Lighting condition = dark-lighted roadway	3.1	62.1	2.17
C2	{Severity=Fatal_Injury, alch_drug_ind=Yes}			4.5	58.4	2.04
C3	{Driver_age=Other_ages, alch_drug_ind=Yes}			3.2	57.9	2.02
C4	{Speed_limit=<45 MPH, alch_drug_ind=Yes}			3.7	57.6	2.01
C5	{Crash_location=Intersection, alch_drug_ind=Yes}			3.7	57.0	1.99
C6	{Num_lanes=3-5_lanes, alch_drug_ind=Yes}			3.5	56.7	1.98
C7	{Weather_cond=Rain, Num_lanes=3-5_lanes}			3.0	56.6	1.97
C8	{Weather_cond=Rain, Crash_location=Intersection}			3.0	56.5	1.97
C9	{Weather_cond=Rain, Speed_limit=<45 MPH}			3.1	55.6	1.94
C10	{Weather_cond=Rain, Season=Fall}			3.3	55.2	1.93
C11	{Weather_cond=Rain, Hwy_clss=State_route}			3.0	53.9	1.88
C12	{Num_lanes=3-5_lanes, Severity=Fatal_Injury}			3.6	53.6	1.87
C13	{Weather_cond=Rain, Ped_age=31-64}			3.1	53.5	1.86
C14	{Speed_limit=<45 MPH, alch_drug_ind=Yes, Driver_age=Young_driver}			3.6	53.4	1.86
C15	{Crash_location=Intersection, Ped_age=19-30, alch_drug_ind=Yes}			3.2	53.2	1.85

Note: * S: Support, C: Confidence, and L: Lift.

Three pedestrian crash scenarios were included in the current analysis. The minimum support and confidence values for the daylight, dark-lighted, and dark-unlighted scenarios were 0.1 and 0.7, 0.03 and 0.5, and 0.01 and 0.4, respectively. To provide more insight into improving pedestrian safety despite the difficulty of explaining many rules, this study offers 15 top rules for each of the three scenarios, as mentioned in Table 2. Researchers may require specialized expertise to choose the optimum support and confidence values. The minimum support and confidence value was determined for each scenario after many trials and errors. The values assigned to these parameters (Support and Confidence) may be claimed to be subjective decisions made in every scenario [52]. A high lift value suggests a strong association between L.H.S. and R.H.S [53]. Considering 1.1 as the minimum lift

value, this viewpoint was considered. The study was limited to 3-item set rules to make interpretation more manageable. A rule 'ID' was created for each R.H.S. scenario to identify and describe any specific associated pattern.

4.2.1. Scenario 1: R.H.S. is Lighting Condition = ("Daylight")

To generate the association rules for the daytime lighting scenario, the 'lighting_condition' variable was set to 'daylight' for the R.H.S. The algorithm originally generated 260 rules, but after removing redundant ones [54, 55], only 126 rules remained, which were then ranked based on their lift values. The top 15 rules [A1-A15] for this scenario are listed in Table 2.

According to the highest lift value rule (A1=1.42) associated with daytime lighting, elderly pedestrians (65+ years) are more likely to be involved in crashes in dry roadway conditions. Also, rules (A11, A15) showed that pedestrians aged 65+ in clear weather conditions were more likely to be involved in moderate-injury crashes during daylight hours. This can be explained by the fact that older people typically have impaired vision, hearing, reaction time, and reduced attention, which highly contributes to pedestrian crashes [56]. The generated rules (A3-A5, A7, A8, and A10) also showed other dominant pedestrian age groups. For example, young pedestrians (<18 years) were strongly associated with possible injury in daylight crashes on roads with a speed limit of <45 MPH, on a city street, in the summertime, or at an intersection. This might be because young pedestrians sometimes engage in risky behavior (*e.g.*, crossing intersections without caution) or due to increased pedestrian activities on city roads during summer, such as walking and running. This confirms the findings of a previous study [57], which reported that older pedestrians are more likely to sustain severe injury crashes than younger pedestrians. Therefore, to better understand elderly pedestrian crash attribute patterns, the MCA algorithm based on the highest lift value rule will be applied in the next section.

4.2.2. Scenario 2a: R.H.S. is Lighting Condition = Nighttime ("Dark_unlighted Roadway")

In this nighttime lighting scenario, the association rules were generated using the lighting condition= 'dark-unlighted'. In this scenario, pedestrians are most likely to be involved in crashes. A total of 116 rules were generated, and only 64 remained after pruning redundant rules. Table 2 presents the top 15 rules [B1-B15] sorted by their lift values.

The generated rules (B1-B6, B8-B9, B11-B15) show that pedestrian crash patterns have been noted on roadways with speed limits 45-65 MPH and were strongly associated with midblock locations, pedestrians' or drivers' alcohol/drug involvement, fatal and severe injury crashes, improper driver actions, 2-lane roadways, pedestrian age (19-30), and undivided roadways. It is easily explained by the fact that previous studies have

reported that higher speed limits are intensely associated with higher vehicle speeds and the likelihood that drivers will exceed them in most cases [58, 59]. The driver's improper actions may also increase the risk of pedestrian death according to rule (B7); it can be explained by the fact that dark roads generally decrease pedestrian visibility and increase drivers' response and reaction time. This may lead to motorists and pedestrians misjudging situations frequently on the road during dark conditions. Moreover, (B14) shows a strong association between undivided roadways with speed limits of 45-65 MPH, resulting in fatal pedestrian crashes during dark-unlighted conditions.

Additionally, as shown in rules B1, B4, B11, and B12, alcohol/drug involvement plays a significant role in pedestrian fatalities during dark-unlighted conditions, and this result holds in dark-lighted roads, as explained in scenario 2b. The highest lift value rule for nighttime lighting conditions was B1=4.39, which was associated with pedestrian crashes in dark-unlighted conditions and strongly associated with high-speed limits roads (45-65 MPH) and midblock locations. Therefore, the MCA algorithm based on the highest lift value rule will be applied to better understand pedestrian crash attribute patterns in high-speed midblock locations in the next section.

4.2.3. Scenario 2b: R.H.S. is Lighting Condition = Nighttime ("Dark_lighted Roadway")

In this nighttime lighting scenario, the association rules were generated using the lighting condition= 'dark-lighted'. Only 74 rules remained after pruning redundant rules and were sorted by their lift values. The top 15 rules [C1-C15] for this scenario are listed in Table 2.

There was a clear association between alcohol/drug involvement and pedestrian crashes in dark-lighted conditions (C1-C6, C14, C15). These rules were expected since alcohol and drugs can reduce the driver's vision and impair their ability to evaluate the space, speed, and movement of other vehicles [26, 60]. In addition, alcohol and other sedatives may impair the ability of the driver to process information and respond to critical driving situations (*i.e.*, slow reaction times). Young drivers on roads with speed limits <45 MPH were combined with a factor, such as alcohol/drug involvement, as shown in rule (C14). Rule (C15) shows that middle-aged pedestrians (19-30 years) in intersection locations and with drinking behaviors were strongly associated with crashes in dark-lighted conditions. This could be explained by the lack of driving experience and impairment due to alcohol or drugs among young drivers and middle-aged pedestrians.

Furthermore, rule (C2) indicates that drinking on dark-lighted roads increases the risk of death. This confirms the findings of a previous study [27], which reported that alcohol consumption is associated with a higher mortality rate among pedestrians at night. Moreover, pedestrian crashes occurred more frequently on dark-lighted roads during rainy weather according to rules (C1, C7-C11, C13) and were associated with pedestrians' or drivers' alcohol/

drug involvement, roadways with 3-5 lanes, intersection locations, limited-access roads, roads with speed limits (<45 MPH), pedestrians aged (31- 64 years), and during the fall season. This may be because rainy conditions increase pedestrian crashes caused by jaywalking and risky behaviors [61]. Also, the reduced visibility caused by the rain may make pedestrians and drivers at intersections more likely to be involved in crashes because their response times are longer [61-63]. Researchers have demonstrated in a previous study [27] that pedestrian risk in dark conditions is also associated with posted speed limits. The risk is incredibly high on limited-access roads where the combination of speed and distance may increase pedestrian risk.

4.3. Results from the MCA Algorithm

4.3.1. Identifying the Interaction between the Contributing Factors to Elderly Pedestrian Crashes during the Daytime in Michigan

A higher risk of crash severity has always been associated with pedestrian age. For example, it has been found that pedestrians over the age of 65 are at an increased risk of being killed or severely injured in crashes [64]. According to Eluru *et al.*, pedestrian age is one of the most significant variables in determining fatality risk [65]. There is an increased risk of injury for elderly pedestrians involved in pedestrian-vehicle crashes compared to other age groups [26, 66-72]. In this regard, it is essential to note that, to our knowledge, no prior studies have analyzed crash information regarding pedestrian safety for elderly pedestrians during daytime lighting conditions to identify the interdependence or correlation among crash attributes using unsupervised learning algorithms.

As a result of filtering elderly pedestrians and daytime crashes, 12 categorical variables were considered in this section, as well as 3405 unique pedestrian crash records. Accordingly, the MCA dataset for this section was '3405 x 12'. When considering a table with the dimensions '3405 x 12', the analysis of each crash record in a row can characterize MCA. This scenario is where 12 categorical variables (defined by 12 columns) are categorized into 45 categories. The spatial distribution of points in multiple dimensions can be generated by MCA using these 12 variables. The distance between each point indicates the degree of similarity between the individual points. The crash

data points were initially projected in N dimensions, where N was determined by subtracting the number of variables from the total number of variable categories (45 - 12 = 33). As a result, the total variance was calculated as the ratio of the maximum number of MCA dimensions to the total number of variables (33/12 = 2.75). Using this total variance value (2.75) as the denominator, this step allowed for the analysis of variance in each dimension. For example, the eigenvalue for the first dimension is 0.183. Consequently, dimension 1 accounts for 10.8% of the explained variance, calculated as 0.298 divided by 2.75 (second row in Table 3). The eigenvalue (range 0 to 1) measures the strength of an axis and denotes the extent to which each dimension represents category information. As the eigenvalue rises, the variance among variables in that dimension increases. Low eigenvalues in the database suggest that the variables are diverse, reflecting the complex and unpredictable nature of crash occurrences. The eigenvalues, variance percentages, and cumulative variance percentages for the top 10 dimensions are presented in Table 3.

Additionally, Fig. (6) shows the Screeplot of MCA dimensions and the percentage of variance explained by each dimension. Among the top 10 dimensions, the first dimension explains 10.8% of the total variance, while the second-dimension accounts for 6.5%, bringing the cumulative total to 17.4%. The variance was not accounted for by any of the remaining dimensions over 5.5%. Consequently, only dimensions 1 and 2 were considered for further investigation. Other crash studies have reported comparable variance percentages for the first two dimensions [73].

Each crash variable on the first plane (dim.1 and dim.2) has a coefficient of determination (R^2) and a p-value, as indicated in Table 4. In general, R^2 is a scalar value between 0 and 1, where 0 signifies no correlation, and 1 indicates a strong relationship between the variable and the MCA dimension. The confidence level associated with a variable is denoted by the p-value in the field of statistics. Driver action, crash location, and roadway characteristics (including speed limit, road type, number of lanes, and highway classification) are the most significant variables that are highly correlated with daytime elderly pedestrian crashes in dimension 1 at a 95% confidence level. In contrast, the primary variables in dimension 2 are the posted speed limit, injury severity level, and driver age.

Table 3. Information on the top 10 dimensions.

Dimension No.	Eigenvalue	% Variance	% Cumulative variance
Dim.1	0.298	10.84	10.84
Dim.2	0.179	6.51	17.35
Dim.3	0.150	5.45	22.8
Dim.4	0.142	5.14	27.94
Dim.5	0.115	4.16	32.10
Dim.6	0.110	3.98	36.08
Dim.7	0.099	3.60	39.68
Dim.8	0.093	3.38	43.06
Dim.9	0.091	3.32	46.38
Dim.10	0.088	3.21	49.59

Scree plot

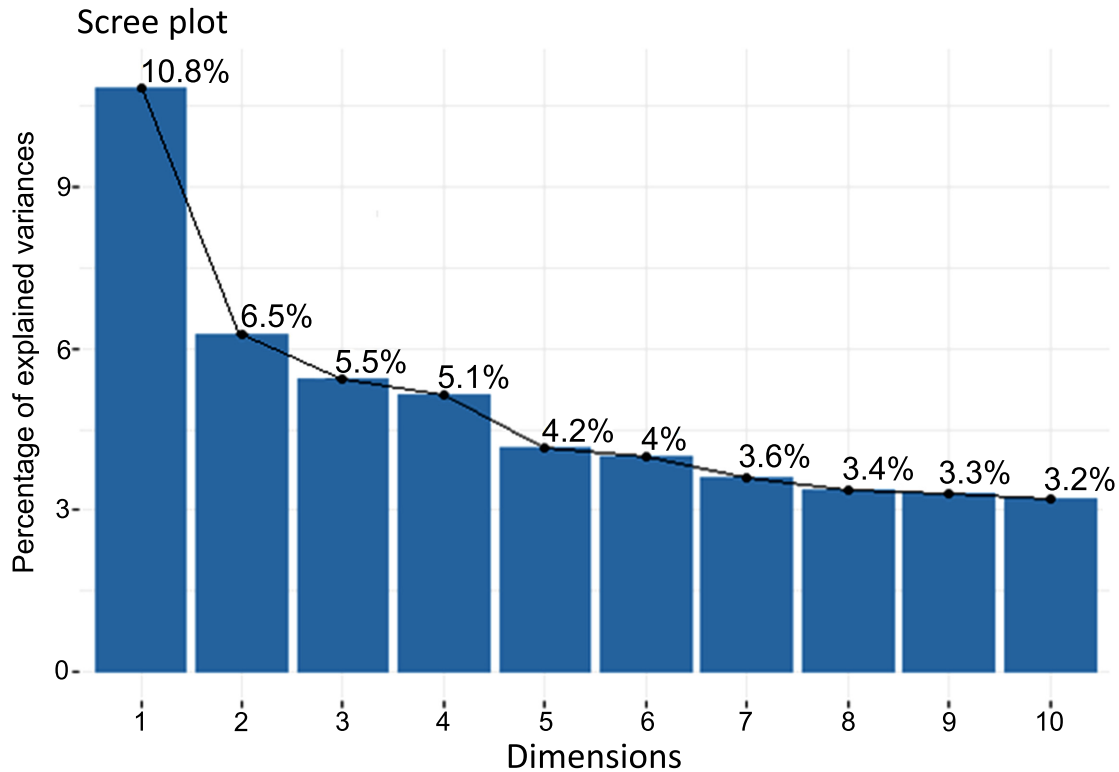**Fig. (6).** Percentage of variances that are explicated for the top ten dimensions.

Table 4. The importance of crash variables on the first plane.

Variables (Dimension 1)	R^2	P-value	Variables (Dimension 2)	R^2	P-value
Crash location	0.811	< 0.001	Driver age	0.633	< 0.001
Number of lanes	0.551	< 0.001	Severity level	0.724	< 0.001
Posted speed limit	0.518	< 0.001	Posted speed limit	0.365	< 0.001
Roadway type	0.743	< 0.001	Number of lanes	0.062	< 0.001
Highway classification	0.858	< 0.001	Crash location	0.039	< 0.001
Driver hazardous action	0.230	< 0.001	Roadway type	0.026	< 0.001
Driver age	0.019	< 0.001	Highway classification	0.016	< 0.001
Roadway condition	0.012	< 0.001	Roadway condition	0.010	< 0.001
Alcohol/drug involvement	0.009	0.0012	Season	0.003	0.0028
Weather condition	0.011	0.0015	-	-	-
Season	0.009	0.0022	-	-	-

Only dimensions 1 and 2 will be considered for data visualizations to facilitate a more straightforward interpretation of the data. Subsequently, the first plane's results in Dim.1 and Dim.2 will be presented and analyzed. All variable categories are represented as points on a "factor map" based on the first two dimensions, as illustrated in Fig. (7). A gradient color scale exemplified the contribution of each variable category. Accordingly, crash attributes, such as severe or rainy weather, slippery pavement, rainy weather, posted speed limits (>65 or 45-65 mph), undivided roads, midblock segments, city street classification, and 3-5 lane roads, were strongly associated with elderly pedestrian crashes during daylight condition, as shown in Fig. (7).

The time gap between a pedestrian approaching the far lane can be challenging for elderly pedestrians. Liu *et al*. reported that pedestrians tend to consider distance rather than the available time gap when approaching a vehicle in the far lane [74]. Furthermore, elderly pedestrians cannot compensate for unsafe crossing decisions by speeding up. Similar considerations were found by Oxley *et al*. when they conducted a simulator-

based study on elderly pedestrians [75]. Furthermore, other studies based on crash data also reported similar results. A survey conducted by Zegeer *et al*. found that elderly pedestrians are more likely to be involved in fatal crashes when crossing wide streets (*i.e*., streets with three to five lanes) [57]. According to a study on pedestrian-vehicle crashes in Florida, the severity of pedestrian injuries is more likely to increase when pedestrians are elderly or intoxicated. Additionally, higher impact speeds, adverse weather conditions, darkness, and the presence of larger vehicles compared to standard passenger cars also contribute to more severe injuries [26].

Clouds are generated in a two-dimensional space by the proximity of variable categories. It is important to recognize that MCA allows for the application of engineering judgment in interpreting co-occurrence patterns. This section will explore the clouds or patterns of co-occurrence that could be linked to elderly pedestrian crashes under daytime lighting conditions. The formation of clusters with redundant variables is typical in many cases. A combination cloud containing redundant information was not chosen despite the proximity of these variables. For example, 'unknown weather condition' or 'uncoded speed limit' do not provide intuitive details, so they were not selected. To elucidate the associations

among variable categories most likely contributing to daytime crashes involving elderly pedestrians, we identified the top three most meaningful combination clouds based on the correlation values of their components (Fig. **8**).

Table **5** shows the results presented in Cloud 1, which indicate that elderly pedestrians have been significantly involved in crashes at two-lane midblock locations having a speed limit between 45 and 65 mph, resulting in severe injury. The relationship between elderly pedestrian severe injury and the location of the crash (midblock) has also been demonstrated in previous research [57]. A high squared loading for a category means that the category is well-represented along that plane. It indicates the importance of the category in explaining the variation captured by that plane. Consequently, Table **5** presents a strong correlation between severe elderly pedestrian crashes and specific road characteristics, such as wide street crossings and high-speed limits. A wide midblock is undoubtedly the most challenging traffic situation for drivers and pedestrians, mainly when no crosswalks exist. Lalika *et al*. (2022) examined factors contributing to fatal/severe injuries in older pedestrians and their possible interdependence. They found that road crossing dangers increased on high-speed roads [70].
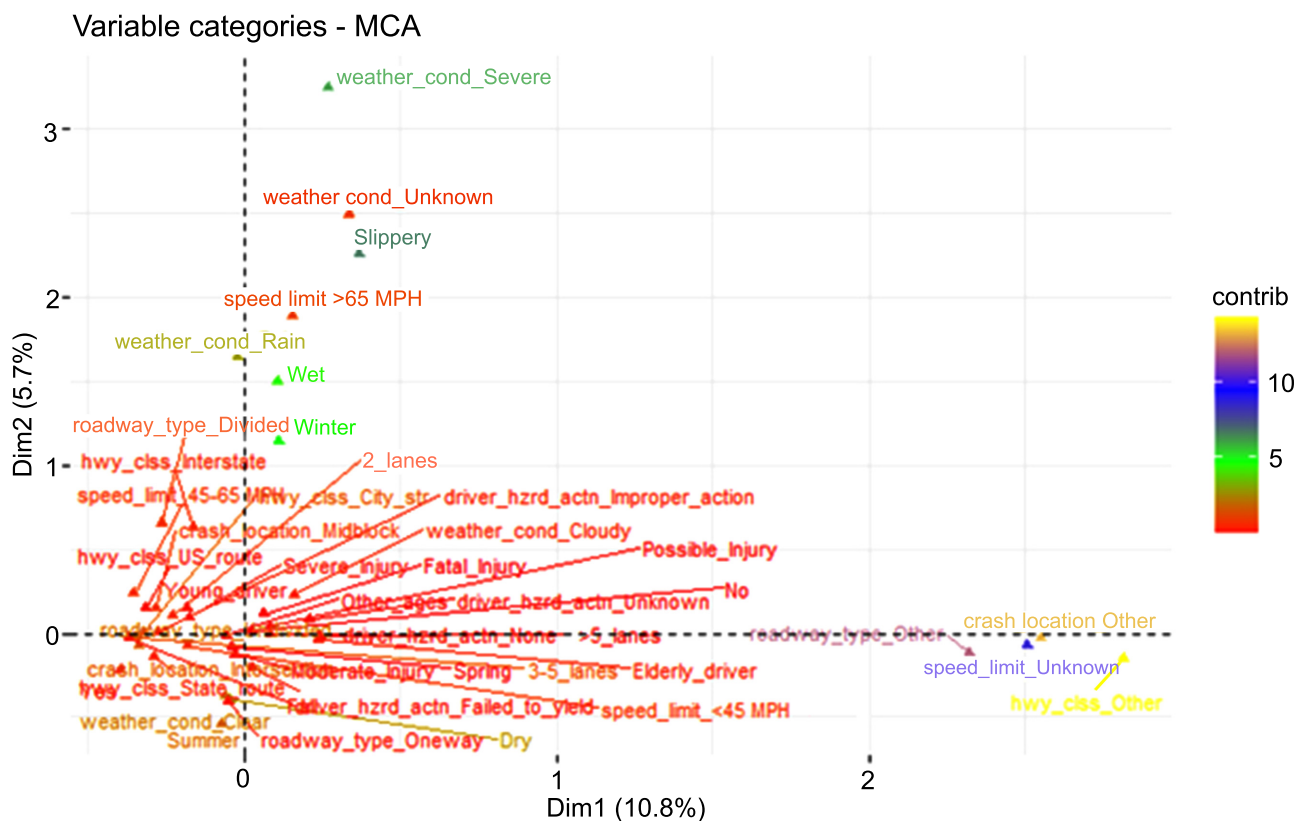
Fig. (7). Factor map of MCA displaying all variable categories on the first plane.
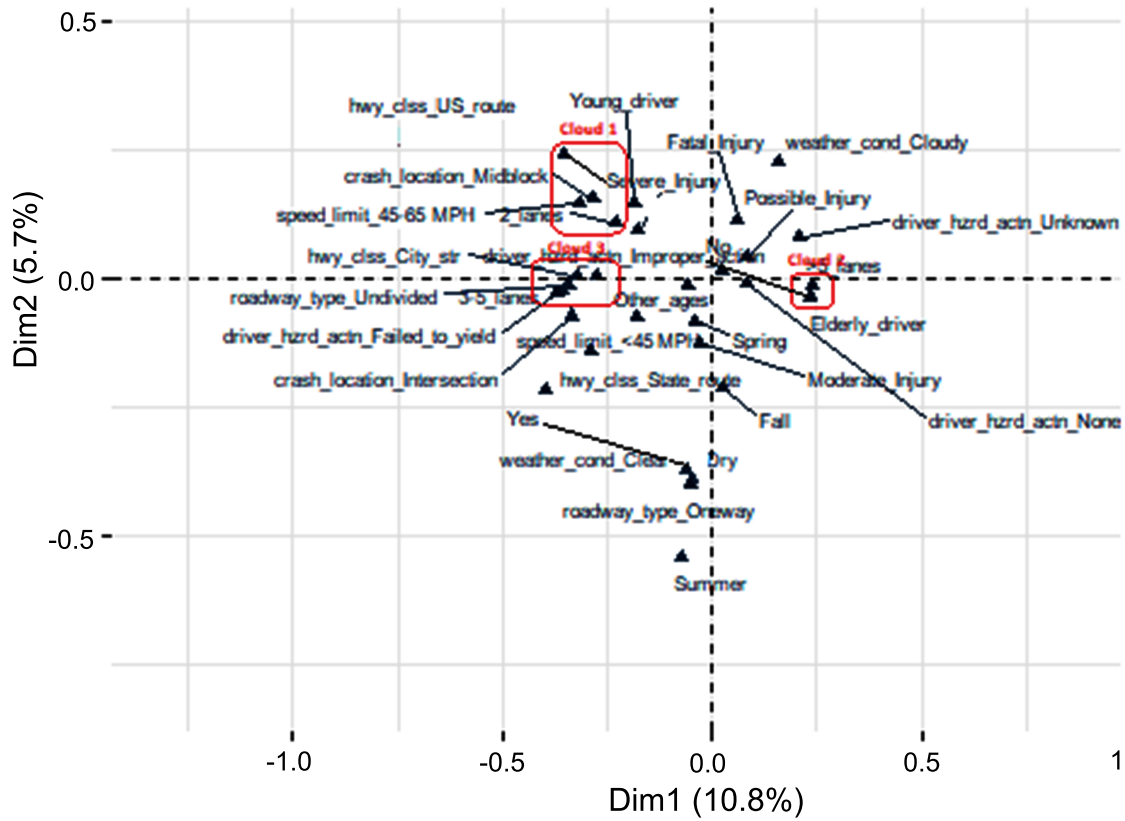
Fig. (8). Close view of the top 3 combination clouds in the factor map of the first plane.

Table 5. Clouds information associated with elderly pedestrian crashes during daylight.

# Cloud	Components	Squared Loading	Correlation	P-value
Cloud#1	Posted speed limit = 45-65 MPH	0.62	0.79	<0.0001
	Crash location = Midblock	0.43	0.65	<0.0001
	Number of lanes = 2 lanes	0.38	0.58	<0.0001
	Severity=Severe Injury	0.32	0.57	<0.0001
Cloud#2	Driver hazardous action = Improper action	0.37	0.61	<0.0001
	Number of lanes >5 lanes	0.28	0.60	<0.0001
	Driver age= Elderly driver	0.19	0.53	<0.0001
Cloud#3	Roadway type = Undivided	0.69	0.83	<0.0001
	Driver hazardous action = Failed to yield	0.56	0.75	<0.0001
	Highway classification = City street	0.46	0.68	<0.0001
	Number of lanes = 3-5 lanes	0.13	0.45	<0.0001

Cloud 2 is strongly associated with the improper action of the elderly driver involving elderly pedestrians on wide roads (> five lanes) during daylight hours. According to previous studies, it has been found that although older drivers behave more cautiously than younger drivers, they misjudge or fail to compensate fully for diminished visual recognition abilities on wide roadways [76].

Cloud 3 illustrates the complex interaction between city streets with 3-5 lanes and driver "failed to yield" action when an elderly pedestrian is involved in a crash

under daylight conditions. Crossing multi-lane traffic is generally accomplished in phases by pedestrians using medians [57]. Consequently, elderly pedestrians may encounter challenges when crossing a multi-lane road without physical separation, as they are compelled to assess traffic in both directions and cannot pause in the center of the road. Additionally, the reason why drivers fail to yield is most likely due to their wrong judgment, distractive driving behaviors, vision/hearing disabilities, or slow reaction time [77].

4.3.2. Identifying the Interaction between the Contributing Factors to Pedestrian Crashes at Midblock Locations with Posted Speed Limit = (45-65 MPH) during Nighttime (Lighted + Unlighted)

Regarding road safety, pedestrian crashes at high-speed locations remain a severe problem. Driving at high speeds will give the driver considerably less time to react and avoid hitting a pedestrian. Additionally, other factors can contribute to a high-speed crash, including humans, vehicles, roadways, and the surrounding environment. It is common for pedestrians to be involved in crashes at midblock locations, especially during nighttime, as mentioned in Table **2**. In 2006, a study was conducted to ascertain which crash variable categories have substantially higher proportions at midblock locations than at intersections by utilizing Kentucky, Florida, and North Carolina databases [78]. They found that midblock location crashes are distributed considerably based on variables, such as lighting conditions and divided *versus* undivided roads. On high-speed roads, environmental conditions have a significant impact on pedestrian crashes as well. For instance, high-speed roadways are hazardous for pedestrians without adequate lighting [27]. Most likely, the primary cause is the lack of visibility and the high speeds of drivers travelling at night, reducing stopping sight distances. Previous studies provided limited information on nighttime pedestrian crashes. They did not explore the complex intersection of contributing factors related to poor lighting conditions in pedestrian crashes at high-speed locations. According to previous literature [79, 80], high-speed roadways are defined as those with posted speed limits greater than 45 mph. We used 45-65 mph as a

threshold value for defining high-speed midblock segments, skipping over freeways (> 70 mph). Although some studies evaluated pedestrian crashes at midblock segments, to our knowledge, research has yet to be conducted addressing the association between key contributing factors related to pedestrian safety at high-speed midblock locations during nighttime (lighted/ unlighted) road conditions using unsupervised learning techniques.

The following section examined 11 categorical variables, divided into 43 categories, and 1098 unique pedestrian crash records by filtering nighttime pedestrian crashes at midblock locations with 45-65 mph speed limits. Therefore, the MCA dataset for this section was '1098 x 11'. Consequently, as previously stated, the total variance was determined by dividing the maximum number of MCA dimensions (32) by the total number of variables (11), resulting in a value of 2.91. This allowed for the analysis of variance across each dimension, as mentioned in Table **6**, using the total variance score of 2.91 as the denominator. Table **6** presents the eigenvalues, variance percentages, and cumulative variance percentages for the top 10 dimensions related to pedestrian crashes at high-speed midblock locations. The first dimension explained 6.04% of the total variance among the top 10 dimensions, and the second dimension accounted for 5.76%, bringing the cumulative total to 11.8%. The variance was not accounted for by any of the remaining dimensions over 5.5%. Consequently, only dimensions 1 and 2 were considered for further investigation. Additionally, Fig. (**9**) illustrates the Screeplot of the MCA dimensions and the percentage of variance explained by each dimension.

Each crash variable on the first plane (dim.1 and dim.2) is associated with a coefficient of determination (R^2) and a p-value, as indicated in Table **7**. In dimension 1, the top significant variables that are highly correlated with nighttime pedestrian crashes at high-speed midblock locations at a 95% confidence level are the following: alcohol/drug involvement of pedestrian or driver, number of lanes, crash injury severity level, environmental conditions (season/weather), driver hazardous action, and pedestrian/driver age. In dimension 2, the top variables are roadway type, weather condition, and driver hazardous action.

Table 6. Information on the top 10 dimensions.

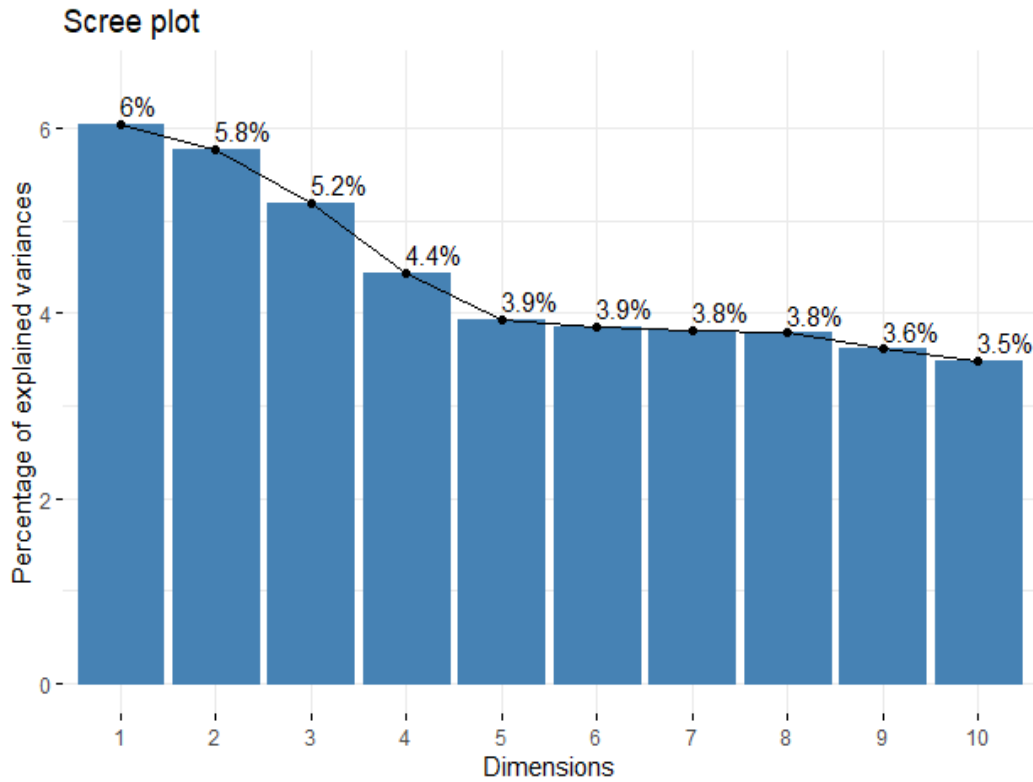| Dimension No.	Eigenvalue	% Variance	% Cumulative Variance
Dim.1	0.176	6.04	6.04
Dim.2	0.168	5.76	11.80
Dim.3	0.151	5.18	16.98
Dim.4	0.129	4.43	21.41
Dim.5	0.114	3.93	25.34
Dim.6	0.112	3.85	29.19
Dim.7	0.111	3.81	33.00
Dim.8	0.110	3.78	36.78
Dim.9	0.105	3.61	40.39
Dim.10	0.101	3.49	43.88

Fig. (9). Percentage of variances that are explicated for the top ten dimensions.

Table 7. The importance of crash variables on the first plane.

Variables (Dimension 1)	R^2	P-value	Variables (Dimension 2)	R^2	P-value
Alcohol/drug involvement	0.500	< 0.001	Roadway type	0.679	< 0.001
Number of lanes	0.418	< 0.001	Weather condition	0.651	< 0.001
Severity	0.221	< 0.001	Driver hazardous action	0.135	< 0.001
Season	0.164	< 0.001	Roadway condition	0.092	< 0.001
Ped age	0.147	< 0.001	Highway classification	0.085	< 0.001
Driver hazardous action	0.131	< 0.001	Season	0.077	< 0.001
Driver age	0.127	< 0.001	Driver age	0.048	< 0.001
Weather condition	0.103	< 0.001	Severity	0.044	< 0.001
Roadway condition	0.055	< 0.001	Driver hazardous action	0.024	< 0.001
Highway classification	0.044	< 0.001	Number of lanes	0.006	0.0026
Roadway type	0.021	< 0.001	-	-	-

Our data visualizations, focusing on dimensions 1 and 2, have led to significant findings. This approach facilitates a more straightforward interpretation of the data, as mentioned earlier. Consequently, we will present and interpret the results of the first plane in Dim.1 and Dim.2. All variable categories are represented as points on a "factor map" based on the first two dimensions, as illustrated in Fig. (**10**). A gradient color scale was used to illustrate the contribution of each variable category. This has led us to identify that crash attributes, such as winter season, pedestrian ages <18/19-30, younger/elderly drivers, undivided roads, improper driver/pedestrian

actions, fatal/moderate injury levels, and 3-5 lane roads, were strongly associated with pedestrian crashes at high-speed midblock locations during nighttime conditions, as indicated in Fig. (**10**).

Previous studies on pedestrian crossing safety have indicated that the speed of an approaching vehicle significantly influences the designated gaps between pedestrians and approaching vehicles [81-83]. The selection of crossing gaps based on distance results in hazardous crossing decisions when approaching vehicles traveling at high speeds, as it overestimates the time

available for crossing [84]. Pedestrians' actions, including darting into oncoming traffic and failing to use crosswalks, are widely considered hazardous, particularly on high-speed roadways. However, only a handful of studies have investigated how pedestrians modify their walking behavior at night [16, 83]. Additionally, higher vehicle speeds are significantly associated with a higher probability of pedestrian crashes and more severe pedestrian injuries [85].

This section examines clouds or co-occurrence patterns in two-dimensional space that may contribute to nighttime pedestrian crashes at high-speed midblock segments. Clusters with redundant variables were found to be common in many cases. A combination cloud containing redundant information was not chosen despite the proximity of these variables, as mentioned earlier. We have pinpointed the three most significant combination clouds to clarify the associations among variable categories that are most likely to influence nighttime pedestrian crashes at high-speed midblock segments, as shown in Fig. (**11**).

One of the most vulnerable groups at risk of being involved in a crash is pedestrians or drivers who are under the influence of alcohol, particularly at night [85]. At high-speed midblock locations during nighttime, cloud 1 is strongly associated with a specific age group of alcohol or drug-involved pedestrians (19-30) as well as young drivers and improper driver actions, as mentioned in Table **8**. Driving under the influence of alcohol may negatively affect an individual's ability to drive. It may affect their vision, hearing, judgment, concentration, coordination, comprehension, reaction time, and other critical skills necessary to operate a vehicle safely and effectively. Moreover, pedestrians are more likely to suffer severe injuries when they are under the influence of alcohol [86]. According to previous studies that examined the combined effect of pedestrian age and alcohol consumption on crash risk, young and middle-aged intoxicated pedestrians are at high risk of being involved in a crash [85]. In addition to impaired cognitive functions and physical skills, impaired pedestrians are likely to make poor judgments and engage in unsafe behavior while walking on the road, such as misjudging the gap between pedestrians and vehicles. Moreover, dangerous driving behaviors by young individuals, such as speeding, recklessness, distraction, or carelessness, are often responsible for nighttime high-speed crashes. Previously, Rahman *et al*. discovered that young drivers exhibit more aggressive/risky driving behaviors, such as speeding, overtaking, and alcohol consumption [87]. This behavior increases the probability of pedestrian crashes at high-speed locations.

Fig. (10). Factor map of MCA displaying all variable categories on the first plane.

Fig. (11). Close view of the top 3 combination clouds in the factor map of the first plane.

Table 8. Clouds information associated with nighttime pedestrian crashes at high-speed midblock locations.

# Cloud	Components	Squared Loading	Correlation	P-value
Cloud#1	Alcohol/Drug involvement = yes	0.52	0.72	<0.0001
	Driver hazardous action = Improper action	0.48	0.69	<0.0001
	Ped age = 19-30	0.24	0.48	<0.0001
	Driver age = Young driver	0.18	0.42	<0.0001
Cloud#2	Roadway type = Undivided	0.39	0.62	<0.0001
	Severity=Moderate Injury	0.25	0.50	<0.0001
	Ped age <18	0.20	0.45	<0.0001
	Season = Winter	0.18	0.42	<0.0001
Cloud#3	Driver hazardous action = Failed to yield	0.53	0.73	<0.0001
	Driver age = Elderly driver	0.33	0.58	<0.0001
	Number of lanes = 3-5 lanes	0.30	0.55	<0.0001

Cloud 2 suggests that during the winter season, young pedestrians were significantly involved in high-speed crashes at undivided midblock locations with moderate injuries during nighttime. The problem may become much more complicated at night since visibility is a significant concern, especially in poor weather conditions. The inexperience and risk-taking behavior of pedestrians under 18 may increase their susceptibility to crashes

because they need to watch for traffic in both directions and cannot pause in the middle of the road. A previous study indicated that young pedestrians perform a less thorough visual search and are more likely to be involved in a traffic crash [88].

Cloud 3 is strongly associated with the elderly driver's "failed to yield" action involving nighttime pedestrians on high-speed segments with 3-5 lanes. Driving becomes

increasingly difficult due to the physical and cognitive changes accompanying aging. Therefore, senior drivers are at a higher risk of making errors due to their deteriorating health, delayed decision-making, and impaired vision [89]. Older drivers may be more likely to be involved in pedestrian crashes in high-speed midblock segments.

CONCLUSION AND STUDY FINDINGS

This study examined pedestrian fatal and injury crashes in Michigan over 11 years (2011-2021) during daytime and nighttime roadway lighting conditions utilizing unsupervised machine learning techniques to identify pedestrian crash patterns based on several multidimensional factors. We analyzed Michigan crash data using the Boruta algorithm for Feature Selection (FS) analysis to assist in the statistical selection of significant variables, the Apriori algorithm for the ARL technique to identify the general associations of patterns leading to pedestrian crashes in Michigan, and the MCA method to understand pedestrian crash attribute patterns for several notable pedestrian crash scenarios based on the highest lift value rules resulting from the ARL algorithm, including elderly pedestrian crashes during daytime lighting conditions and high-speed midblock locations during nighttime lighting conditions.

These methods uncovered key pedestrian crash-related patterns contributing to the daytime and nighttime lighting conditions of Michigan. According to the results, the proposed methodologies can identify associations that need better understanding in the pedestrian safety literature. In addition to identifying valuable relationships without restricting variables by their nature (as dependent or independent), they can also discover relationships that would otherwise be difficult to detect and produce rules and clouds that are easy to understand. It should be noted that, despite the advantages, the proposed techniques are intended to supplement other methods used for statistical crash data analyses. Instead, they can be considered valuable traffic safety decision support tools for analyzing an extensive database of crash characteristics from jurisdictions, such as the MDOT, and selecting appropriate safety countermeasures.

The findings revealed that 75% of fatal pedestrian-involved crashes occurred during nighttime in either dark-lighted or dark-unlighted conditions. The ARL algorithm highlighted several patterns of pedestrian crashes, including the fact that more severe pedestrian crashes are more likely to occur during nighttime. Additionally, in clear weather conditions or during summertime, there is a strong association between daylight pedestrian possible injury crashes and young pedestrians (under 18 years). In contrast, in cloudy weather conditions, daylight pedestrian moderate injury crashes are strongly associated with elderly pedestrians (65 years and older). Furthermore, elderly drivers (65 years and older) were found to be more likely to be associated with pedestrian crashes during daylight. In both dark-lighted and dark-unlighted conditions, alcohol and drug involvement were found to

have a strong association with fatal pedestrian crashes. Pedestrian crashes during dark-lighted conditions also had a strong association with rainy weather, which has been linked to other factors, such as alcohol or drug involvement, roads with 3-5 lanes, intersection locations, limited-access roads, roads with 45-65 MPH speed limits, pedestrians aged 31-64, and the fall season. Additionally, there is a strong association between roadways with a speed limit of 45-65 MPH and fatal/severe pedestrian crashes occurring during dark-unlighted conditions. Midblock locations on roadways with speed limits of 45-65 MPH were strongly associated with pedestrian crashes under dark-unlighted conditions, and driver improper actions were strongly correlated to fatal pedestrian crashes under dark-unlighted conditions.

MCA examined several noteworthy pedestrian crash scenarios, including elderly pedestrian crashes during daytime lighting conditions and high-speed midblock locations during nighttime lighting conditions. In the resulting combination of clouds involving elderly pedestrians during daylight, it was found that elderly pedestrians were significantly involved in severe injury crashes at two-lane midblock locations with a speed limit between 45 and 65 MPH. Additionally, the improper actions of elderly drivers were significantly associated with elderly pedestrians on wide roads (greater than five lanes). Furthermore, in the combination of clouds involving nighttime pedestrian crashes at high-speed midblock locations, it was found that a specific age group of alcohol or drug-involved pedestrians (ages 19-30) was significantly associated with young drivers and improper driver actions. During the winter season, young pedestrians were significantly involved in moderate injury crashes at undivided high-speed midblock locations. Additionally, elderly drivers and "failed to yield" actions were highly correlated with high-speed midblock locations featuring 3-5 lanes.

SUGGESTED COUNTERMEASURES

Although many studies have attempted to reduce fatal and severe pedestrian crashes, this study also aims to raise awareness of pedestrian crash patterns during daytime/nighttime lighting conditions in Michigan and recommend safety countermeasures among practitioners to mitigate them. To achieve walkable cities, we must provide better guidance in planning and designing pedestrian infrastructure that is safe, accessible, and sustainable. The following countermeasures are recommended in response to pedestrian crash risk patterns identified in this study:

- It has been recognized that driver alcohol and drug impairment is a severe safety issue when it comes to pedestrian crashes in dark conditions. According to this study, 203 young pedestrians under 18 were involved in crashes while drinking alcohol (daylight = 67, dark = 136). If the National Minimum Drinking Age Act of 1984 were strictly enforced, especially at night, safety would improve.
- Several evidence-based policy options may reduce alcohol

consumption and alcohol-related crashes, including lowering the blood alcohol concentration limit, utilizing ignition interlock devices in a more significant number, increasing alcohol taxes, limiting the availability of alcohol, and providing treatment support.

- Michigan could significantly reduce daylight elderly pedestrian-related crashes by imposing more restrictions on senior age-related regulations for drivers [57].
- It is important to implement safety lighting counter-measures that could significantly reduce pedestrian crashes, especially on high-speed roads and midblock locations [90-93], such as rectangular rapid flashing beacons, pedestrian hybrid beacons, high-visibility crosswalks, and pedestrian visibility enhancements at crosswalks.
- A more extensive awareness campaign regarding speeding's consequences is needed to change the culture's attitude toward speeding.
- There are several strategies for improving pedestrian safety in the built environment, including pedestrian safety zones, marked crosswalks, sidewalks, pedestrian overpasses, and fencing [94].

STUDY LIMITATIONS AND FUTURE WORK

This study employed an apriori algorithm to identify association rules in crash data, incorporating fundamental adjustments to the thresholds for evaluating these rules. These modifications aimed to enhance the suitability of the ARL algorithm for crash data analysis. Given that the dataset included rare events of interest, the thresholds for minimum support were lowered, allowing analysts to uncover associations between such rare events.

However, several limitations were identified. First, multiple variables contained 'unknown' or 'other' categories, such as driver hazardous actions, age, and roadway type. This issue arises primarily from the limited information collected in police reports, highlighting the need for improved data collection methods in future studies, as noted in a previous study [95]. Second, further investigation is needed to determine the optimal values for the support and confidence parameters, which could enhance the robustness of the findings. Third, this study restricted the rules to 3-item sets, producing 15 rules for each scenario. Future research could increase the number of rules to reveal more interesting patterns and associations, thereby enriching the analysis of pedestrian crash data.

Future research should also explore the use of microscopic simulation modeling [96]. Such simulations can offer valuable insights into individual pedestrian and vehicle interactions, facilitating a deeper understanding of how various factors influence crash dynamics at a granular level. By simulating different scenarios and interventions, researchers can effectively evaluate the potential impact of countermeasures under real-world conditions.

Furthermore, future research should consider the inclusion of more comprehensive datasets that cover diverse geographical locations, time frames, and demo-

graphic factors. This could lead to more generalizable findings and a deeper understanding of the variations in pedestrian crash patterns across different contexts. Importantly, the potential of advanced machine learning techniques, such as deep learning or ensemble methods, to enhance the identification of complex interactions among contributing factors should not be overlooked.

Investigating the temporal dynamics of pedestrian crashes, *i.e.*, examining how patterns evolve over time, could also provide valuable insights for traffic safety interventions. However, it is crucial to note that a multidisciplinary approach involving collaboration with traffic engineers, urban planners, and public health officials is essential. This approach can facilitate the development of effective countermeasures aimed at reducing pedestrian fatalities and injuries, underscoring the need for a holistic approach to traffic safety.

AUTHOR CONTRIBUTIONS

B.Q.: Conceptualization; B.Q.: methodology; B.Q.: software; B.Q.: validation; B.Q.: formal analysis; B.Q.: investigation; B.Q. and D.Q.: resources; B.Q. and A.D.: data curation; B.Q., D.Q., and A.T.: writing—original draft preparation; B.Q., J.O. and V.K.: writing—review and editing; B.Q.: visualization; J.O. and V.K.: supervision; J.O. and V.K.: project administration; B.Q.: funding acquisition.

LIST OF ABBREVIATIONS

FARS	=	Fatality Analysis Reporting System
FHWA	=	Federal Highway Administration
MDOT	=	Michigan Department of Transportation
TZD	=	Toward Zero Deaths
ARL	=	Association Rules Learning
MCA	=	Multiple Correspondence Analysis
ML	=	Machine Learning
MNL	=	Multinomial Logit Model
L.H.S	=	Left-Hand-Side
R.H.S	=	Right-Hand-Side
CA	=	Correspondence Analysis
MDA	=	Mean Decrease Accuracy
FS	=	Feature Selection

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

The data supporting the findings of the article is available publicly in the [Michigan Traffic Crash Facts (MTCF)] at : https://www.michigantrafficcrashfacts.org/legacy/querytool#q1;0;2023;
Reference Year: [2011-2021].

CONFLICT OF INTEREST

The author declares no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

[1] "The League of American Bicyclists, Topic I: Rates of biking and walking", https://data.bikeleague.org/show-your-data/national-data/rates-ofbiking-and-walking/

[2] "NHTSA, Fatality and Injury Reporting System (FIRS)", https://www.nhtsa.gov/file-downloads?p=nhtsa/downloads/FARS/

[3] "Smart Growth America, Dangerous By Design 2021", https://smartgrowthamerica.org/dangerous-by-design-2021-2/

[4] "NTSB, 2022 Annual Report to Congress", https://www.ntsb.gov/about/reports/Documents/2022 Annual Report to Congress.pdf

[5] "MDOT, Moving-Michigan-Toward-Zero-Deaths", https://www.michigan.gov/-/media/Project/Websites/MDOT/Travel/Safety/Efforts/TZD/Moving-Michigan-Toward-Zero-Deaths.pdf?rev=c5eb12c0da254d8290cddaa55d53534a

[6] J. Parks, P. Bowman, C. Peterson, and C. Flannagan, "Analysis Report: Pedestrian-involved crashes in Michigan (2017-2021)", https://smarter-usa.org/wp-content/uploads/2024/04/2021-Analysis-Report-Motorcycle-Involved-Crashes-in-Michigan-2017-2021.pdf

[7] R.L. Sanders, R.J. Schneider, and F.R. Proulx, "Pedestrian fatalities in darkness: What do we know, and what can be done?", *Transp. Policy,* vol. 120, pp. 23-39, 2022.
[http://dx.doi.org/10.1016/j.tranpol.2022.02.010]

[8] W.A. Harrell, "The impact of pedestrian visibility and assertiveness on motorist yielding", *J. Soc. Psychol.,* vol. 133, no. 3, pp. 353-360, 1993.
[http://dx.doi.org/10.1080/00224545.1993.9712153]

[9] I. Kwan, and J. Mapstone, "Interventions for increasing pedestrian and cyclist visibility for the prevention of death and injuries", *Cochrane Libr.,* vol. 2009, no. 4, p. CD003438, 2006.
[http://dx.doi.org/10.1002/14651858.CD003438.pub2] [PMID: 17054171]

[10] "NHTSA, National Pedestrian Safety Month", https://www.trafficsafetymarketing.gov/sites/tsm.gov/files/2021-10/THE FINAL NHTSA Ped Toolkit Layout_508_9.30.pdf

[11] S. Jung, X. Qin, and C. Oh, "Improving strategic policies for pedestrian safety enhancement using classification tree modeling", *Transp. Res. Part A Policy Pract.,* vol. 85, pp. 53-64, 2016.
[http://dx.doi.org/10.1016/j.tra.2016.01.002]

[12] S. Das, R. Avelar, K. Dixon, and X. Sun, "Investigation on the wrong way driving crash patterns using multiple correspondence analysis", *Accid. Anal. Prev.,* vol. 111, pp. 43-55, 2018.
[http://dx.doi.org/10.1016/j.aap.2017.11.016] [PMID: 29172044]

[13] M.M. Besharati, and A. Tavakoli Kashani, "Which set of factors contribute to increase the likelihood of pedestrian fatality in road crashes?", *Int. J. Inj. Contr. Saf. Promot.,* vol. 25, no. 3, pp. 247-256, 2018.
[http://dx.doi.org/10.1080/17457300.2017.1363781] [PMID: 28853326]

[14] D.J. Hand, "Principles of data mining", *Drug Saf.,* vol. 30, no. 7, pp. 621-622, 2007.
[http://dx.doi.org/10.2165/00002018-200730070-00010] [PMID: 17604416]

[15] B. Qawasmeh, S. Qawasmeh, A. Al Tawil, and D. Qawasmeh, "Estimation of Trip-based Generation Models and Calibration of Mode Choice Models for the American Travel Behavior", *Open Transplant. J.,* vol. 18, 2024.e26671212348473
[http://dx.doi.org/10.2174/0126671212348473241010122638]

[16] D. Dai, "Identifying clusters and risk factors of injuries in pedestrian–vehicle crashes in a GIS environment", *J. Transp. Geogr.,* vol. 24, pp. 206-214, 2012.
[http://dx.doi.org/10.1016/j.jtrangeo.2012.02.005]

[17] M. Pour-Rouholamin, and H. Zhou, "Investigating the risk factors associated with pedestrian injury severity in Illinois", *J. Safety Res.,* vol. 57, pp. 9-17, 2016.
[http://dx.doi.org/10.1016/j.jsr.2016.03.004] [PMID: 27178074]

[18] P. Stoker, A. Garfinkel-Castro, M. Khayesi, W. Odero, M.N. Mwangi, M. Peden, and R. Ewing, "Pedestrian safety and the built environment: a review of the risk factors", *J. Plann. Lit.,* vol. 30, no. 4, pp. 377-392, 2015.
[http://dx.doi.org/10.1177/0885412215595438]

[19] D. Eustace, K. Dey, M.T. Rahman, and B. Qawasmeh, *Emerging Delivery and Mobility Services.,* 2022.https://trid.trb.org/View/1941867
[http://dx.doi.org/10.1061/9780784415986.ch4]

[20] C.R. Kemnitzer, C.N. Pope, A. Nwosu, S. Zhao, L. Wei, and M. Zhu, "An investigation of driver, pedestrian, and environmental characteristics and resulting pedestrian injury", *Traffic Inj. Prev.,* vol. 20, no. 5, pp. 510-514, 2019.
[http://dx.doi.org/10.1080/15389588.2019.1612886] [PMID: 31180735]

[21] K. Billah, H.O. Sharif, and S. Dessouky, "Analysis of pedestrian–motor vehicle crashes in San Antonio, Texas", *Sustainability,* vol. 13, no. 12, p. 6610, 2021.
[http://dx.doi.org/10.3390/su13126610]

[22] C.V. Zegeer, and M. Bushell, "Pedestrian crash trends and potential countermeasures from around the world", *Accid. Anal. Prev.,* vol. 44, no. 1, pp. 3-11, 2012.
[http://dx.doi.org/10.1016/j.aap.2010.12.007] [PMID: 22062330]

[23] J.M. Sullivan, and M.J. Flannagan, "The role of ambient light level in fatal crashes: inferences from daylight saving time transitions", *Accid. Anal. Prev.,* vol. 34, no. 4, pp. 487-498, 2002.
[http://dx.doi.org/10.1016/S0001-4575(01)00046-X] [PMID: 12067111]

[24] J. Griswold, B. Fishbain, S. Washington, and D.R. Ragland, "Visual assessment of pedestrian crashes", *Accid. Anal. Prev.,* vol. 43, no. 1, pp. 301-306, 2011.
[http://dx.doi.org/10.1016/j.aap.2010.08.028] [PMID: 21094328]

[25] J.M. Wood, R.A. Tyrrell, and T.P. Carberry, "Limitations in drivers' ability to recognize pedestrians at night", *Hum. Factors,* vol. 47, no. 3, pp. 644-653, 2005.
[http://dx.doi.org/10.1518/001872005774859980] [PMID: 16435703]

[26] C. Lee, and M. Abdel-Aty, "Comprehensive analysis of vehicle–pedestrian crashes at intersections in Florida", *Accid. Anal. Prev.,* vol. 37, no. 4, pp. 775-786, 2005.
[http://dx.doi.org/10.1016/j.aap.2005.03.019] [PMID: 15869737]

[27] J.M. Sullivan, "*Characteristics of pedestrian risk in darkness*", University of Michigan, Ann Arbor, Transportation Research Institute, 2001. Available from: https://deepblue.lib.umich.edu/bitstream/handle/2027.42/49450/UMTRI-2001-33.pdf%3Bjsessionid%3D8D89987166EEBF3CD96FBC7E280B5050?sequence%3D1

[28] N.A. Siddiqui, X. Chu, and M. Guttenplan, "Crossing locations, light conditions, and pedestrian injury severity", *Transp. Res. Rec.,* vol. 1982, no. 1, pp. 141-149, 2006.
[http://dx.doi.org/10.1177/0361198106198200118]

[29] S. Mokhtarimousavi, "A time of day analysis of pedestrian-involved crashes in California: Investigation of injury severity, a logistic regression and machine learning approach using HSIS data", *Inst. Transp. Eng. ITE J.,* vol. 89, no. 10, pp. 25-33, 2019.
https://www.hsisinfo.org/pdf/2019_ITEJ_HSIS_Award.pdf

[30] A. Montella, M. Aria, A. D'Ambrosio, and F. Mauriello, "Data-mining techniques for exploratory analysis of pedestrian crashes", *Transp. Res. Rec.,* vol. 2237, no. 1, pp. 107-116, 2011.
[http://dx.doi.org/10.3141/2237-12]

[31] K. Kim, and E. Yamashita, "Corresponding characteristics and circumstances of collision-involved pedestrians in Hawaii",

Transp. Res. Rec., vol. 2073, no. 1, pp. 18-24, 2008.
[http://dx.doi.org/10.3141/2073-03]

[32]   S. Das, and X. Sun, "Factor association with multiple correspondence analysis in vehicle–pedestrian crashes", *Transp. Res. Rec.,* vol. 2519, no. 1, pp. 95-103, 2015.
[http://dx.doi.org/10.3141/2519-11]

[33]   T. Bajada, and M. Attard, "A typological and spatial analysis of pedestrian fatalities and injuries in Malta", *Res. Transp. Econ.,* vol. 86, p. 101023, 2021.
[http://dx.doi.org/10.1016/j.retrec.2020.101023]

[34]   X. Xu, S.C. Wong, and K. Choi, "Quasi-induced exposure method for pedestrian safety at signalized intersections", *J. Transp. Saf. Secur.,* vol. 8, no. 2, pp. 129-147, 2016.
[http://dx.doi.org/10.1080/19439962.2014.1003630]

[35]   T. Chen, V. Sampath, M.C. May, S. Shan, O.J. Jorg, J.J. Aguilar Martín, F. Stamer, G. Fantoni, G. Tosello, and M. Calaon, "Machine Learning in Manufacturing towards Industry 4.0: From 'For Now' to 'Four-Know'", *Appl. Sci.,* vol. 13, no. 3, p. 1903, 2023.
[http://dx.doi.org/10.3390/app13031903]

[36]   R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases", *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data,* 1993pp. 207-216
[http://dx.doi.org/10.1145/170035.170072]

[37]   "Javatpoint, Association Rule Learning", https://www.javatpoint.com/association-rule-learning

[38]   P. Tanna, and Y. Ghodasara, "Using Apriori with WEKA for frequent pattern mining", *arXiv Prepr arXiv1406.7371,* 2014.
https://arxiv.org/pdf/1406.7371

[39]   I. Ntoutsi, N. Pelekis, and Y. Theodoridis, "Pattern comparison in data mining: A survey", In: *Research and Trends in Data Mining Technologies and Applications.,* IGI Global, 2007, pp. 86-120.
[http://dx.doi.org/10.4018/978-1-59904-271-8.ch004]

[40]   D. Demir, and T. Bucak, "The Place and Importance of Earth Markets Within Scope of Eco Gastronomy The Example of Foça Earth Market", *Int. J. Appl. Sci. Technol.,* vol. 4, no. 2, 2014.
https://avesis.deu.edu.tr/yayin/86afe5be-e931-49c8-8744-f137ffc9 56a8/the-place-and-importance-of-earth-markets-within-scope-of- eco-gastronomy-the-example-of-foca-earth-market

[41]   A. Ait-Mlouk, T. Agouti, and F. Gharnati, "Mining and prioritization of association rules for big data: Multi-criteria decision analysis approach", *J. Big Data,* vol. 4, no. 1, p. 42, 2017.
[http://dx.doi.org/10.1186/s40537-017-0105-4]

[42]   R.C. Chen, C. Dewi, S.W. Huang, and R.E. Caraka, "Selecting critical features for data classification based on machine learning methods", *J. Big Data,* vol. 7, no. 1, p. 52, 2020.
[http://dx.doi.org/10.1186/s40537-020-00327-4]

[43]   M.B. Kursa, and W.R. Rudnicki, "Feature selection with the Boruta package", *J. Stat. Softw.,* vol. 36, no. 11, pp. 1-13, 2010.
[http://dx.doi.org/10.18637/jss.v036.i11]

[44]   R.C. Team, "R: A language and environment for statistical computing", https://cir.nii.ac.jp/crid/1370298755636824325

[45]   M. Hahsler, S. Chelluboina, K. Hornik, and C. Buchta, "The arules R-package ecosystem: analyzing interesting patterns from large transaction data sets", *J. Mach. Learn. Res.,* vol. 12, pp. 2021-2025, 2011.
https://www.jmlr.org/papers/volume12/hahsler11a/hahsler11a.pdf

[46]   H. Abdi, "Discriminant correspondence analysis", *Encycl. Meas. Stat,* vol. 2007, pp. 284-290, 2007.
https://personal.utdallas.edu/~herve/Abdi-DCA2007-pretty.pdf

[47]   N. Kalayci, and M.A. Basaran, "A combined approach using multiple correspondence analysis and log-linear models for student perception in quality in higher education", *Procedia Econ. Finance,* vol. 17, pp. 55-62, 2014.
[http://dx.doi.org/10.1016/S2212-5671(14)00878-8]

[48]   C. Spychala, J. Armand, C. Dombry, and C. Goga, "Multivariate statistical analysis for exploring road crash-related factors in the Franche-Comté region of France", *Commun. Stat. Case Stud. Data Anal. Appl.,* vol. 7, no. 3, pp. 442-474, 2021.

[http://dx.doi.org/10.1080/23737484.2021.1929562]

[49]   A. Kaminska, A. Ickowicz, P. Plouin, M.F. Bru, G. Dellatolas, and O. Dulac, "Delineation of cryptogenic Lennox–Gastaut syndrome and myoclonic astatic epilepsy using multiple correspondence analysis", *Epilepsy Res.,* vol. 36, no. 1, pp. 15-29, 1999.
[http://dx.doi.org/10.1016/S0920-1211(99)00021-2]   [PMID: 10463847]

[50]   M. Greenacre, and J. Blasius, *Multiple Correspondence Analysis and Related Methods,* CRC Press, 2006.
[http://dx.doi.org/10.1201/9781420011319]

[51]   S. Lê, J. Josse, and F. Husson, "FactoMineR: an R package for multivariate analysis", *J. Stat. Softw.,* vol. 25, no. 1, pp. 1-18, 2008.
[http://dx.doi.org/10.18637/jss.v025.i01]

[52]   S. Das, A. Dutta, R. Avelar, K. Dixon, X. Sun, and M. Jalayer, "Supervised association rules mining on pedestrian crashes in urban areas: identifying patterns for appropriate countermeasures", *IJUS,* vol. 23, no. 1, pp. 30-48, 2019.
[http://dx.doi.org/10.1080/12265934.2018.1431146]

[53]   M. Hahsler, B. Grün, and K. Hornik, "arules-A computational environment for mining association rules and frequent item sets", *J. Stat. Softw.,* vol. 14, no. 15, pp. 1-25, 2005.
[http://dx.doi.org/10.18637/jss.v014.i15]

[54]   F. Thabtah, "Rules pruning in associative classification mining", *Proceedings of the IBIMA Conference,* 2005pp. 7-15 https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi =9dbc5e593453c3509358b6c234825bbd44dad0a5

[55]   B. Liu, W. Hsu, and Y. Ma, "Pruning and Summarizing the Discovered Associations", *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* 1999pp. 125-134 https://dl.acm.org/doi/pdf/10.1145/312129.312216

[56]   A.J. McKnight, "Driver and Pedestrian Training", *Spec. Rep. 218 Transp. an Aging Soc. Improv. Mobil. Saf. Older Pers.,* vol. 2, pp. 101-133, 1988.
https://onlinepubs.trb.org/onlinepubs/sr/sr218v2.pdf#page=114

[57]   C.V. Zegeer, J.C. Stutts, H. Huang, M. Zhou, and E. Rodgman, "Analysis of elderly pedestrian accidents and recommended countermeasures", *Transp. Res. Rec.,* no. 1405, pp. 56-63, 1993.
https://www.safetylit.org/citations/index.php?fuseaction=citations .viewdetails&citationIds[]=citjournalarticle_603261_38

[58]   L.S. Friedman, D. Hedeker, and E.D. Richter, "Long-term effects of repealing the national maximum speed limit in the United States", *Am. J. Public Health,* vol. 99, no. 9, pp. 1626-1631, 2009.
[http://dx.doi.org/10.2105/AJPH.2008.153726] [PMID: 19608943]

[59]   A.T. McCartt, and W. Hu, "Effects of vehicle power on passenger vehicle speeds", *Traffic Inj. Prev.,* vol. 18, no. 5, pp. 500-507, 2017.
[http://dx.doi.org/10.1080/15389588.2016.1241994]   [PMID: 27753503]

[60]   B. Qawasmeh, and D. Eustace, Effectiveness of Cable Median Barriers in Preventing Cross Median Crashes and Related Casualties in the United States-A Systematic Review.*International Road Federation World Meeting & Exhibition.,* Springer: Cham, 2022, pp. 345-354.https://link.springer.com/chapter/10.1007/978-3-030-798 01-7_24
[http://dx.doi.org/10.1007/978-3-030-79801-7_24]

[61]   X. Zhai, H. Huang, N.N. Sze, Z. Song, and K.K. Hon, "Diagnostic analysis of the effects of weather condition on pedestrian crash severity", *Accid. Anal. Prev.,* vol. 122, pp. 318-324, 2019.
[http://dx.doi.org/10.1016/j.aap.2018.10.017] [PMID: 30412822]

[62]   S. Mondal, N. Roy, R.A. Laskar, I. Sk, S. Basu, D. Mandal, and N.A. Begum, "Biogenic synthesis of Ag, Au and bimetallic Au/Ag alloy nanoparticles using aqueous extract of mahogany (Swietenia mahogani JACQ.) leaves", *Colloids Surf. B Biointerfaces,* vol. 82, no. 2, pp. 497-504, 2011.
[http://dx.doi.org/10.1016/j.colsurfb.2010.10.007]   [PMID: 21030220]

[63]   A. Theofilatos, and G. Yannis, "A review of the effect of traffic and

weather characteristics on road safety", *Accid. Anal. Prev.,* vol. 72, pp. 244-256, 2014.
[http://dx.doi.org/10.1016/j.aap.2014.06.017] [PMID: 25086442]

[64] S.M. Siram, V. Sonaike, O.B. Bolorunduro, W.R. Greene, S.Z. Gerald, D.C. Chang, E.E. Cornwell III, and T.A. Oyetunji, "Does the pattern of injury in elderly pedestrian trauma mirror that of the younger pedestrian?", *J. Surg. Res.,* vol. 167, no. 1, pp. 14-18, 2011.
[http://dx.doi.org/10.1016/j.jss.2010.10.007] [PMID: 21109262]

[65] N. Eluru, C.R. Bhat, and D.A. Hensher, "A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes", *Accid. Anal. Prev.,* vol. 40, no. 3, pp. 1033-1054, 2008.
[http://dx.doi.org/10.1016/j.aap.2007.11.010] [PMID: 18460372]

[66] S. Das, K. Jha, K. Fitzpatrick, M. Brewer, and T.H. Shimu, "Pattern identification from older bicyclist fatal crashes", *Transp. Res. Rec.,* vol. 2673, no. 6, pp. 638-649, 2019.
[http://dx.doi.org/10.1177/0361198119841863]

[67] J.D. Febres, M.Á. Mariscal, S. Herrera, and S. García-Herrero, "Pedestrians' injury severity in traffic accidents in Spain: A pedestrian actions approach", *Sustainability,* vol. 13, no. 11, p. 6439, 2021.
[http://dx.doi.org/10.3390/su13116439]

[68] B.Y. Henary, J. Ivarsson, and J.R. Crandall, "The influence of age on the morbidity and mortality of pedestrian victims", *Traffic Inj. Prev.,* vol. 7, no. 2, pp. 182-190, 2006.
[http://dx.doi.org/10.1080/15389580500516414] [PMID: 16854713]

[69] L. Hu, X. Wu, J. Huang, Y. Peng, and W. Liu, "Investigation of clusters and injuries in pedestrian crashes using GIS in Changsha, China", *Saf. Sci.,* vol. 127, p. 104710, 2020.
[http://dx.doi.org/10.1016/j.ssci.2020.104710]

[70] L. Lalika, A.E. Kitali, H.J. Haule, E. Kidando, T. Sando, and P. Alluri, "What are the leading causes of fatal and severe injury crashes involving older pedestrian? Evidence from Bayesian network model", *J. Safety Res.,* vol. 80, pp. 281-292, 2022.
[http://dx.doi.org/10.1016/j.jsr.2021.12.011] [PMID: 35249608]

[71] B.P.Y. Loo, and K.L. Tsui, "Pedestrian injuries in an ageing society: insights from hospital trauma registry", *J. Trauma,* vol. 66, no. 4, pp. 1196-1201, 2009.
[http://dx.doi.org/10.1097/TA.0b013e31817fdef6] [PMID: 19359937]

[72] M.G. Mohamed, N. Saunier, L.F. Miranda-Moreno, and S.V. Ukkusuri, "A clustering regression approach: A comprehensive injury severity analysis of pedestrian–vehicle crashes in New York, US and Montreal, Canada", *Saf. Sci.,* vol. 54, pp. 27-37, 2013.
[http://dx.doi.org/10.1016/j.ssci.2012.11.001]

[73] Q. Yuan, X. Li, C. Wang, Y. Li, and Y. Gao, "Cluster and factor analysis on data of fatal traffic crashes in China", *2017 4th International Conference on Transportation Information and Safety (ICTIS),* Banff, AB, Canada, 2017, pp. 211-224.
[http://dx.doi.org/10.1109/ICTIS.2017.8047768]

[74] Y.C. Liu, and Y.C. Tung, "Risk analysis of pedestrians' road-crossing decisions: Effects of age, time gap, time of day, and vehicle speed", *Saf. Sci.,* vol. 63, pp. 77-82, 2014.
[http://dx.doi.org/10.1016/j.ssci.2013.11.002]

[75] J.A. Oxley, E. Ihsen, B.N. Fildes, J.L. Charlton, and R.H. Day, "Crossing roads safely: An experimental study of age differences in gap selection by pedestrians", *Accid. Anal. Prev.,* vol. 37, no. 5, pp. 962-971, 2005.
[http://dx.doi.org/10.1016/j.aap.2005.04.017] [PMID: 15993827]

[76] D.A. Owens, J.M. Wood, and J.M. Owens, "Effects of age and illumination on night driving: a road test", *Hum. Factors,* vol. 49, no. 6, pp. 1115-1131, 2007.
[http://dx.doi.org/10.1518/001872007X249974] [PMID: 18074710]

[77] X. Zhao, X. Zhang, and J. Rong, "Study of the effects of alcohol on drivers and driving performance on straight road", *Math. Probl. Eng.,* vol. 2014, 2014.
https://onlinelibrary.wiley.com/doi/pdf/10.1155/2014/607652

[78] L. Sandt, and C.V. Zegeer, "Characteristics Related to Midblock Pedestrian–Vehicle Crashes and Potential Treatments", *Transp. Res. Rec.,* vol. 1982, no. 1, pp. 113-121, 2006.
[http://dx.doi.org/10.1177/0361198106198200115]

[79] S. Stapleton, T. Kirsch, T.J. Gates, and P.T. Savolainen, "Factors affecting driver yielding compliance at uncontrolled midblock crosswalks on low-speed roadways", *Transp. Res. Rec.,* vol. 2661, no. 1, pp. 95-102, 2017.
[http://dx.doi.org/10.3141/2661-11]

[80] B. Naik, J. Appiah, and L.R. Rilett, "Are dilemma zone protection systems useful on high speed arterials with signal coordination? A case study", *JITSR,* vol. 18, no. 2, pp. 219-229, 2020.
[http://dx.doi.org/10.1007/s13177-019-00193-2]

[81] R. Lobjois, and V. Cavallo, "Age-related differences in street-crossing decisions: The effects of vehicle speed and time constraints on gap selection in an estimation task", *Accid. Anal. Prev.,* vol. 39, no. 5, pp. 934-943, 2007.
[http://dx.doi.org/10.1016/j.aap.2006.12.013] [PMID: 17275774]

[82] T. Petzoldt, "On the relationship between pedestrian gap acceptance and time to arrival estimates", *Accid. Anal. Prev.,* vol. 72, pp. 127-133, 2014.
[http://dx.doi.org/10.1016/j.aap.2014.06.019] [PMID: 25035969]

[83] B. Qawasmeh, V. Kwigizile, and J.S. Oh, "Performance and safety effectiveness evaluation of mini-roundabouts in Michigan", *J. Eng. Appl. Sci. (Asian Res. Publ. Netw.),* vol. 70, no. 1, p. 36, 2023.https://jeas.springeropen.com/articles/10.1186/s44147-023-00208-6

[84] Y.S. Chung, and C.Y. Chang, "Farther and safer: An illusion engendered by incapability?", *Transp. Res., Part F Traffic Psychol. Behav.,* vol. 54, pp. 110-123, 2018.
[http://dx.doi.org/10.1016/j.trf.2018.01.020]

[85] W. A. Leaf, "Literature review on vehicle travel speeds and pedestrian injuries",
[http://dx.doi.org/10.1037/e446312008-001]

[86] R. Miles-Doan, "Alcohol use among pedestrians and the odds of surviving an injury: Evidence from Florida law enforcement data", *Accid. Anal. Prev.,* vol. 28, no. 1, pp. 23-31, 1996.
[http://dx.doi.org/10.1016/0001-4575(95)00030-5] [PMID: 8924182]

[87] M. Rahman, K.M. Kockelman, and K.A. Perrine, "Investigating risk factors associated with pedestrian crash occurrence and injury severity in Texas", *Traffic Inj. Prev.,* vol. 23, no. 5, pp. 283-289, 2022.
[http://dx.doi.org/10.1080/15389588.2022.2059474] [PMID: 35584352]

[88] D. Lord, A. Smiley, and A. Haroun, "Pedestrian accidents with left-turning traffic at signalized intersections: Characteristics, human factors, and unconsidered issues", https://www.pedbikeinfo.org/cms/downloads/PedAccidents_LeftTurn.pdf

[89] S. Koppel, M. Bohensky, J. Langford, and D. Taranto, "Older drivers, crashes and injuries", *Traffic Inj. Prev.,* vol. 12, no. 5, pp. 459-467, 2011.
[http://dx.doi.org/10.1080/15389588.2011.580802] [PMID: 21972856]

[90] L.E. Dougald, "Effectiveness of a rectangular rapid-flashing beacon at a midblock crosswalk on a high-speed urban collector", *Transp. Res. Rec.,* vol. 2562, no. 1, pp. 36-44, 2016.
[http://dx.doi.org/10.3141/2562-05]

[91] A. Maxwell, J. Kennedy, I. Routledge, P. Knight, and K. Wood, "Puffin Pedestrian Crossing Accident Study", https://www.safetylit.org/citations/index.php?fuseaction=citations.viewdetails&citationIds[]=citjournalarticle_334820_19

[92] K. Fitzpatrick, and E.S. Park, "Safety Effectiveness of the HAWK Pedestrian Crossing Treatment", https://rosap.ntl.bts.gov/view/dot/41661

[93] L. Chen, C. Chen, and R. Ewing, "The relative effectiveness of pedestrian safety countermeasures at urban intersections—Lessons from a New York City experience", https://nacto.org/wp-content/uploads/2015/04/relative_effectiveness_of_pedestrian_safety_counter_measures_chen.pdf

[94]   B.S. Qawasmeh, *Safety Assessment for Vulnerable Road Users Using Automated Data Extraction with Machine-Learning Techniques,* 2024. https://scholarworks.wmich.edu/dissertations/4112

[95]   B.S. Qawasmeh, "Safety Assessment for Vulnerable Road Users Using Automated Data Extraction with Machine-Learning Techniques",

[96]   B. Qawasmeh, "Enhancing Work Zone Safety: Evaluating Static Merge Strategies Through Microscopic Traffic Simulation", *Open Transp. J.,* vol. 18, 2024.
[http://dx.doi.org/10.2174/0118744478330254240801062605]