CPBS
Center for Pedestrian and Bicyclist Safety

A Tier-1 University Transportation Center

Untangling the Growing Pedestrian Safety Problem on Urban Arterials

July 2024

A Report From the Center for Pedestrian and Bicyclist Safety

Christopher R. Cherry
University of Tennessee-Knoxville

Saurav Parajuli
University of Tennessee-Knoxville

About the Center for Pedestrian and Bicyclist Safety (CPBS)

The Center for Pedestrian and Bicyclist Safety (CPBS) is a consortium of universities committed to eliminating pedestrian and bicyclist fatalities and injuries through cutting-edge research, workforce development, technology transfer, and education. Consortium members include: The University of New Mexico; San Diego State University; The University of California Berkeley; The University of Tennessee Knoxville; and The University of Wisconsin Milwaukee. More information can be found at: https://pedbikesafety.org

Disclaimer

Acknowledgments

This study was funded, partially or entirely, by a grant from the Center for Pedestrian and Bicyclist Safety (CPBS), supported by the U.S. Department of Transportation (USDOT) through the University Transportation Centers program. The author would like to thank CPBS and the USDOT for their support of university-based research in transportation, and especially for the funding provided in support of this project. We thank the Tennessee Department of Transportation (TDOT) and the Tennessee Department of Safety and Homeland Security for providing the crash data.

TECHNICAL DOCUMENTATION

1. Project No. 23UTK02	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Untangling the Growing Pedestrian Safety Problem on Urban Arterials		5. Report Date July 2024
		6. Performing Organization Code N/A
7. Author(s) Christopher R. Cherry https://orcid.org/0000-0002-8835-4617 Saurav Parajuli https://orcid.org/0000-0003-4534-2832		8. Performing Organization Report No. N/A
9. Performing Organization Name and Address Center for Pedestrian and Bicyclist Safety Centennial Engineering Center 3020 The University of New Mexico Albuquerque, NM 87131		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. 69A3552348336
12. Sponsoring Agency Name and Address United States of America Department of Transportation Office of Research, Development, and Technology (RD&T)		13. Type of Report and Period Covered Final Report – June 2023 to May 2024
		14. Sponsoring Agency Code USDOT OST-R

15. Supplementary Notes

Report accessible via the CPBS website https://pedbikesafety.org and DOI https://doi.org/10.21949/7qta-3d56

16. Abstract

Pedestrian fatalities in Tennessee are predominantly associated with nighttime, high-speed crashes (35 mph and above), and straight midblock maneuvers, like those observed on urban arterials. The existing literature primarily relies on the functional classification of roads, which can be ambiguous, especially when used for pedestrian safety and its relationship with roadway-related variables. This study introduces the concept of "high-risk crashes" to identify potential hazardous pedestrian crashes beyond state classifications. Using unsupervised learning algorithms (latent class and hierarchical clustering) and supervised learning with manually labeled data, pedestrian crashes in Tennessee were categorized into high-risk crashes and non-high-risk crashes based on road and environmental features with high accuracy. The classification revealed clusters of high-risk crashes along wide, straight streets with higher speed limits, limited pedestrian facilities, and businesses primarily catering to cars, which may not be officially classified as major arterials. Trend analysis of crash involvement and fatalities showed a steep increase in high-risk crashes, suggesting that the rise in pedestrian crash severity in Tennessee can be attributed to these crashes. Logistic regression results indicated that high-risk crashes are more likely to result in fatal injuries in dark conditions, during straight midblock maneuvers, and in non-residential areas. Vehicle size did not significantly impact the likelihood of a fatal crash in high-risk crashes. These findings call for urgent measures focusing primarily on roads, such as lowering speed limits to 35 mph in pedestrian-heavy areas, increasing safe pedestrian crossing opportunities, adopting traffic calming devices, and improving lighting to enhance pedestrian safety.

17. Key Words	18. Distribution Statement
Pedestrians; Safety; Pedestrian safety; Classification; Arterial Highways; High risk locations	No restrictions. This document is available through the National Technical Information Service, Springfield, VA 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	55	

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized.**

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	$candela/m^2$	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	$candela/m^2$	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

Untangling the Growing Pedestrian Safety Problem on Urban Arterials

A Center for Pedestrian and Bicyclist Safety Research Report

July 2024

Christopher R. Cherry
Department of Civil Engineering
University of Tennessee-Knoxville

Saurav Parajuli
Department of Civil Engineering
University of Tennessee-Knoxville

TABLE OF CONTENTS

List of Figures

List of Tables

Acronyms, Abbreviations, and Symbols

GHSA	Governors Highway Safety Association
FHWA	Federal Highway Administration
NHTSA	National Highway Traffic Safety Administration
Ped.	Pedestrian
MCA	Multiple Correspondence Analysis
ARM	Association Rules Mining
LCC	Latent Class Clustering
HC	Hierarchical Clustering
TITAN	Tennessee Integrated Traffic Analysis Network
MMUCC	Model Minimum Uniform Crash Criteria
E-TRIMS	electronic - Tennessee Roadway Information Management System
AIC	Akaike Information Criteria
ANN	Artificial Neural Network
MLP	Multilayer Perceptron
BM	Base Model
IM	Interaction Model
SUV	Sport Utility Vehicle
SE	Standard Error
DUI	Driving Under the Influence

Abstract

Pedestrian fatalities in Tennessee are predominantly associated with nighttime, high-speed crashes (35 mph and above), and straight midblock maneuvers, like those observed on urban arterials. The existing literature primarily relies on the functional classification of roads, which can be ambiguous, especially when used for pedestrian safety and its relationship with roadway-related variables. This study introduces the concept of "high-risk crashes" to identify potential hazardous pedestrian crashes beyond state classifications. Using unsupervised learning algorithms (latent class and hierarchical clustering) and supervised learning with manually labeled data, pedestrian crashes in Tennessee were categorized into high-risk crashes and non-high-risk crashes based on road and environmental features with high accuracy. The classification revealed clusters of high-risk crashes along wide, straight streets with higher speed limits, limited pedestrian facilities, and businesses primarily catering to cars, which may not be officially classified as major arterials. Trend analysis of crash involvement and fatalities showed a steep increase in high-risk crashes, suggesting that the rise in pedestrian crash severity in Tennessee can be attributed to these crashes. Logistic regression results indicated that high-risk crashes are more likely to result in fatal injuries in dark conditions, during straight midblock maneuvers, and in non-residential areas. Vehicle size did not significantly impact the likelihood of a fatal crash in high-risk crashes. These findings call for urgent measures focusing primarily on roads, such as lowering speed limits to 35 mph in pedestrian-heavy areas, increasing safe pedestrian crossing opportunities, adopting traffic calming devices, and improving lighting to enhance pedestrian safety.

Executive Summary

Previous research has identified several factors contributing to fatal pedestrian accidents, including nighttime crashes, high-speed incidents, and midblock crossings. Despite the significant influence of environmental and road characteristics on these crashes, current studies often categorize them using the Federal Highway Administration's (FHWA) functional classification of arterial and non-arterial roads. While not inherently incorrect, FHWA acknowledges significant ambiguity in these guidelines, which can lead to variations in design considerations. Therefore, this study proposes a new approach to identify "high-risk" crashes based on specific road design and environmental features, aiming to enhance the identification of potentially dangerous pedestrian crashes. These crashes are identified using exploratory tools like unsupervised learning and supervised learning algorithms.

We used the Tennessee police crash data for our analyses. The Tennessee Integrated Traffic Analysis Network (TITAN) database compiles comprehensive traffic safety data, including details on persons, crashes, and vehicles involved. We excluded interstate and rural area crashes, focusing on urban environments where most incidents occur. After cleaning and removing missing entries, it contained 17,267 records of pedestrian-involved crashes from January 2009 to September 2019. After a comprehensive literature review to identify the characteristics of pedestrian crashes and their relationship with fatality outcomes, we employ exploratory tools like unsupervised learning and supervised learning algorithms to identify the patterns among the crashes based on the road and environmental variables. We used Latent Class Clustering (LCC) and Hierarchical Clustering (HC) algorithms to categorize crashes into five distinct clusters as listed below:

- Cluster 1: Wet Roads
- Cluster 2: Midblock Locations on Narrow Roads
- Cluster 3: Low-Speed Zones
- Cluster 4: Intersections
- Cluster 5: Midblock Locations on Multilane Roads

When tallying these clusters with fatality data, Cluster 5 was identified as the riskiest cluster, while Cluster 3 was identified as the least risky cluster. The insights from these clusters were used to train the supervised learning model where functional roadway classifications from E-TRIMS were also utilized for determining initial labels for “high-risk” crashes. An Artificial Neural Network (ANN) model was fitted using the Multilayer Perceptron (MLP) classifier with an accuracy of 94.65 percent. This model was used to classify the unlabeled data, with the final categorization including 9611 crashes as “high-risk” and “7656” crashes as low-risk.

Key findings of this study based on the crash categorization are detailed below:

1. High-risk pedestrian crashes predominantly occur on wide, straight roads with higher speed limits, often in non-intersection and dark conditions. These areas frequently lack adequate pedestrian infrastructure such as continuous sidewalks, sufficient signals, and well-spaced crosswalks, particularly in non-residential and business-centric zones.
2. Functional classifications based solely on arterial roads may underrepresent risky pedestrian crash locations, as similar features are found in non-arterial streets common in suburban settings in Tennessee.
3. Demographic analysis reveals that Black pedestrians are disproportionately affected by high-risk crashes compared to White pedestrians, while White drivers are more frequently involved in these incidents, consistent with broader safety literature on racial disparities.
4. Intoxicated pedestrians are overrepresented in high-risk crashes, contrasting with intoxicated drivers who are less likely to be involved, highlighting behavioral differences in crash risk.
5. Despite higher exposure among Black pedestrians, White pedestrians are more likely to die in high-risk crashes, potentially influenced by Tennessee's demographic distributions and vulnerability among certain populations like the unhoused.
6. Children, while less exposed, show lower fatality rates in high-risk crashes, suggesting varying levels of parental vigilance but also raising concerns about non-child-friendly streets in urban areas of Tennessee.
7. Larger vehicles like pickup trucks and SUVs generally increase pedestrian fatality risk, but their differential impact in high-risk crash scenarios may be mitigated by higher travel speeds, underscoring the complexity of vehicle-type interactions in crash outcomes.
8. Trend analysis of "high-risk" and "low-risk" crashes indicates a worsening severity of high-risk crashes over time, contributing to rising pedestrian fatalities in Tennessee, potentially exacerbated by suburbanization trends affecting marginalized communities.

The study emphasizes the urgent need for improved pedestrian safety measures, focusing on enhancing road design and infrastructure to mitigate the growing severity of high-risk pedestrian crashes. Key recommendations based on the study findings are listed below:

1. Consider reducing speeds on streets resembling urban arterials, with a suggested maximum speed limit of 35 mph, to enhance pedestrian safety.
2. Implement road diets on wide arterials by removing two-way turn lanes and strategically placing signalized intersections at regular intervals near businesses to reduce pedestrian crossing distances and encourage speed reduction.
3. Install pedestrian refuge islands at road crossings and enhance lighting and signals in high pedestrian traffic areas to improve visibility and promote safer decision-making.
4. Ensure frequent and well-lit pedestrian crossings, including midblock crossings, are equipped with appropriate lighting and signals to optimize visibility and pedestrian safety.

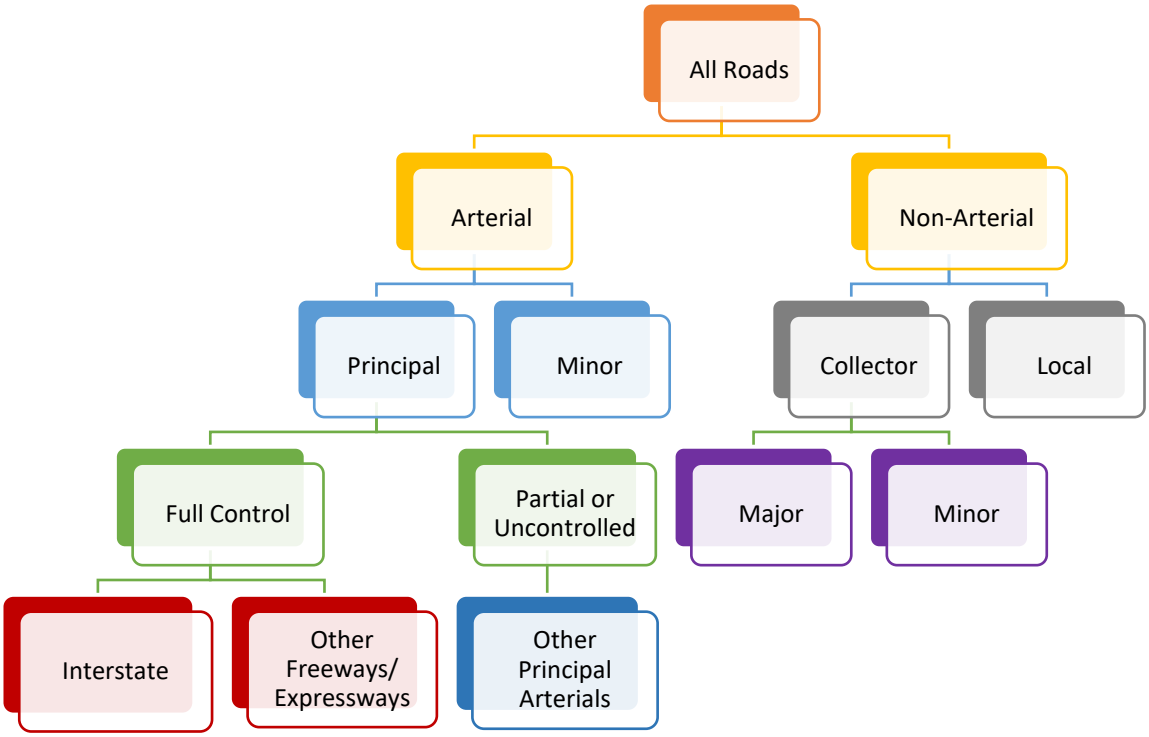
Introduction

The share of pedestrian fatalities in the US has risen consistently, increasing from 13 percent in 2010 to 17.6 percent of all road fatalities in 2021. The Governors Highway Safety Association (GHSA) reports that 7,624 pedestrians were killed in 2021, with 2022 projections pointing to the highest number of pedestrian deaths in 40 years. This reflects a 77 percent increase from 2010 and a 17 percent jump from the 2020 figures. (*Macek*, *2023*). In contrast, other developed countries have achieved advances in pedestrian safety during the same timeframe. (*European Transport Safety Council*, *2020; UK Department for Transport*, *2020*). Safety researchers and news media have identified various factors potentially contributing to this trend, including the growing presence of larger utility vehicles, an aging demographic, and heightened distractions among drivers and pedestrians. (*Schmitt*, *2020; Tyndall*, *2021*). Additionally, research also points to the impact of suburbanization, altering land-use patterns and travel behaviors and ultimately aggravating interactions between pedestrians, road infrastructure, and vehicles. (*Ferenchak & Abadi*, *2021*).

Tennessee ranks among the top five most dangerous states for pedestrians as of 2022 (*Macek*, *2023*). From 2009 to 2019, the state saw its pedestrian fatalities more than double, with a steep increase in the fatality rate from 4.3 to 7.3 deaths per 100 pedestrians involved in crashes, marking a 70 percent increase in fatality rate. Urban areas in Tennessee, particularly mid-blocks on high-speed multilane roads, recorded a disproportionate number of pedestrian deaths, mostly during nighttime hours. Data from this period shows that 74.5 percent of these fatalities occurred in dark conditions, 63.1 percent at mid-block locations, and 77.3 percent on roads where the speed limit was 35 mph or higher. Interestingly, in terms of pedestrian crash involvement, nighttime crashes, midblock crashes, and high-speed pedestrian crashes only accounted for 40 percent of total pedestrian crashes. (*Parajuli et al.*, *2023*). This disparity suggests an overrepresentation of pedestrian fatalities in the abovementioned crash categories in Tennessee. In general, pedestrian safety literature also links severe pedestrian crashes to nighttime, midblock, and high-speed conditions, frequently associated with urban arterials. (*Ferenchak & Abadi*, *2021; Hossain et al.*, *2022; Hu & Cicchino*, *2018*). While current research efforts in identifying risk factors have contributed to our understanding of pedestrian safety, only a few studies have focused on road design characteristics while analyzing pedestrian crashes. Our study examines this relationship, focusing on pedestrian crashes in urban areas in Tennessee.

Background

The Federal Highway Administration (FHWA) classifies roadways into arterials and non-arterials according to their functions, as shown in Figure 1. Arterials are further classified into principal arterials, including fully controlled roadways such as interstates, freeways, expressways, and minor arterials. While fully controlled arterials do not have pedestrian accesses, other arterials, indicated by “Other Principal Arterials” in Figure 1 generally allow access to pedestrians. In general terms, principal arterial or major arterial refers to this specific type of arterial with partial or uncontrolled access.



(Source: FHWA and CDM Smith)

Figure 1. Federal functional classification of roadways

Major arterials are vital in providing high mobility, connecting metropolitan centers, and granting access through rural areas, with open access to adjacent land uses. Similarly, minor arterials are moderately sized roads serving smaller geographical areas and providing connectivity to the arterial system and are typically smaller than major arterials. Arterials plays an important role in the US road network, accommodating higher traffic volumes and speeds to ensure efficient mobility and accessibility for car drivers, often at the expense of other road user safety. In addition to arterials, collector roads and local roads also allow pedestrians access. In its report, FHWA specifies that this classification of roads comes with some ambiguity and potential overlaps, as it is based on the roadway functionality. For instance, the FHWA classification guidelines have overlaps between arterials and collectors to account for the variation within the functional classes (*Federal Highway Administration*, *2023*).

Arterials are generally broad, multilane roads designed to accommodate high-speed, high-volume traffic with crosswalks that are widely spaced, allowing uncontrolled access to nearby land uses. Due to the higher speeds and potentially heavier vehicles on arterials compared to other local roads, pedestrian activities carry a significant risk, especially in urban arterials where traffic is high, land use is relatively dense, and many arterials are served by transit systems. Furthermore, the wide layout makes arterials challenging to illuminate during nighttime, exacerbating the risks,

especially when pedestrians cross away from designated crosswalks. Therefore, it is crucial to improve pedestrian safety on these roads. It has been a popular practice for state agencies and researchers to use the functional classification of roadways to study the nature and severity of traffic crashes, including pedestrian crashes. However, as indicated by the FHWA's definition, this type of classification has some ambiguity. To that end, this research hypothesizes that the current classification may not fully capture all streets with arterial characteristics, which are often accepted as the most dangerous roads for pedestrians. This could lead to an underrepresentation of crashes and may skew perceived risks associated with urban arterials for both pedestrians and authorities.

In US suburban areas, most roads, including collectors and even some local roads, are designed primarily for vehicle mobility. Thus, streets are characterized by their width and large block lengths, providing ample opportunities for vehicles to exceed speed limits (*Ewing et al.*, *2003*). Furthermore, implementing lower speed limits and adding pedestrian infrastructure, such as regular crossings and adequate lighting, on these streets would be costly. Consequently, most of these roads remain in poor condition for pedestrians, making them as dangerous as major arterials in terms of pedestrian safety, if not more. Therefore, there is a crucial need for research to identify these streets where pedestrian crashes are more likely to be fatal, allowing for targeted policies and engineering interventions to improve safety.

Research Objectives

This study proposes categorizing pedestrian crashes into high-risk and low-risk crashes based solely on roadway and environmental characteristics. Using urban pedestrian crash data in Tennessee, this study employs machine learning techniques like unsupervised and supervised learning to distinguish crashes more likely to have a fatal or severe outcome from crashes with less severe outcomes. By controlling factors such as vehicle size, pedestrian age, and driving or walking under the influence, we aim to understand the impact of road design on pedestrian safety. With proper identification of high-risk crashes, the study will further explore the characteristics of these crashes and attempt to understand the underlying mechanism behind them. Finally, based on the study results, we provide recommendations to mitigate these deadly pedestrian crashes and enhance overall pedestrian safety.

Literature Review

Studies have been tackling traffic safety issues since the invention of the car. However, increased attention to pedestrian safety is a more recent development. As a result, we now have a better understanding of the risk factors affecting pedestrians. The first part of this review focuses on major risk factors associated with pedestrian crashes, with a particular emphasis on roadway design and built-environment factors. The second part will outline exploratory research efforts to understand pedestrian crashes and the methodologies used.

Pedestrian Fatality Risk Factors

The severity of pedestrian crash outcomes is significantly influenced by the transfer of kinetic energy, which is directly linked to the likelihood of fatal outcomes (*Ballesteros et al.*, *2004*). This energy transfer is a product of the vehicle's weight and the square of its speed. Studies have shown that vehicle size, which correlates with weight, and the posted speed limit, which correlates with impact speed (*Elvik et al.*, *2004*), both play crucial roles in determining fatal outcomes. Larger vehicles like SUVs, pickups, and minivans are more likely to result in fatalities compared to smaller cars such as sedans and coupes (*Tyndall*, *2021*). Regarding speed, studies invariably agree that higher posted speed limits on roads are a major cause of pedestrian deaths. (*Hossain et al.*, *2022; Islam*, *2023; Prato et al.*, *2018; Salon & McIntyre*, *2018*). For instance, a study indicates that at an impact speed of about 25 mph, the probability of pedestrian death is 25 percent, whereas at around 55 mph, this probability rises to 90 percent (*Tefft*, *2013*).

In addition to high speeds, wide roads with multiple lanes significantly contribute to the disproportionate number of pedestrian fatalities (*Islam*, *2023; Nabavi Niaki et al.*, *2016; Rab et al.*, *2018*). Schneider, Proulx, et al. examined the top 34 pedestrian fatality hotspots in the US and found that these areas are predominantly multilane roads with high traffic volumes and adjacent commercial land uses. Most of these roads have posted speed limits over 30 mph and are located near low-income neighborhoods (*Schneider et al.*, *2021*). A study notes that the combination of high speeds, poor lighting infrastructure, and crossings at midblock locations are often associated with urban arterials (*Goodman et al.*, *2022*). Multiple studies acknowledge that urban arterials, with infrastructure gaps like discontinuous sidewalks, missing pedestrian signals and signs, and marked pedestrian crossings, are often linked with pedestrian fatality in the US (*Bellis et al.*, *2021; Long Jr & Ferenchak*, *2021; Mansfield et al.*, *2018; Schneider et al.*, *2021*).

The safety literature has also identified other crucial factors that are responsible for causing higher injury severity outcomes in pedestrian crashes. These factors encompass poor visibility during nighttime, intoxicated driving and walking, demographics of pedestrians and drivers, the clothing worn by pedestrians, distractions affecting both parties, surrounding land use, and the influence of advanced vehicle technologies like emergency braking and pedestrian detection systems (*Aziz et al.*, *2013; Keller et al.*, *2011*). Elderly and child pedestrians are more likely to be involved in a fatal traffic crash (*Davis*, *2001; Kim et al.*, *2008*). Similarly, intoxicated pedestrians have a high probability of severe injury during a pedestrian crash (*Dultz & Frangos*, *2013; Zajac & Ivan*,

2003). Studies have also shown that minority populations, such as Black pedestrians, Native Americans, people of color, and low-income populations, bear a disproportionate burden of pedestrian fatalities (*Long Jr & Ferenchak*, *2021; Noland et al.*, *2013; Roll & McNeil*, *2022; Sanders & Schneider*, *2022*). A study found that pedestrian fatality and severe injury hotspots are found in areas with higher percentages of non-white residents, coupled with lower sidewalk coverage (*Long Jr & Ferenchak*, *2021*). The connection between infrastructure gaps and minority populations is attributed to historical neglect in land use development, which has resulted in increased traffic exposure and inadequate pedestrian facilities (*Roll & McNeil*, *2022*). Lastly, inclement weather and walking or driving during the night negatively impact pedestrian injury outcomes. Poor visibility and reduced road friction during harsh weather can create slippery surfaces and increase driver maneuvering errors, thereby heightening the severity of pedestrian injuries (*Li et al.*, *2017*).

Multiple longitudinal studies investigating the steep incline in pedestrian fatality in the US during the last decade also support the findings from severity studies based on cross-sectional data (*Ferenchak & Abadi*, *2021; Hu & Cicchino*, *2018; Schneider*, *2020; Tefft et al.*, *2021*). These studies have identified numerous factors causing the rise in pedestrian fatalities over the years, employing methods such as cross-tabulations, linear regression models, and univariate analyses. Significant contributors identified include higher speeds, arterial roads, nighttime, increased vehicle size, and other related factors. Another longitudinal study in urban Tennessee determined a significant increase in injury severity outcomes for urban arterials during the 2009-2019 period (*Parajuli et al.*, *2023*). Research investigating walking distances notes that Southern states exhibit increasing pedestrian fatality trends due to differences in the built environment, law enforcement practices, and driving culture (*Vellimana & Kockelman*, *2023*).

Some studies have also investigated the relationship between pedestrian behavior and the infrastructural elements of roads. Pedestrian crashes at nighttime are more closely associated with the posted speed limits on roads rather than with speeding as a human factor (*Sanders et al.*, *2022*). This pattern is also evident with pedestrians under the influence of alcohol or drugs. Intoxicated pedestrians are more likely to suffer fatal injuries on roads with infrastructure deficits, such as inadequate lighting (*Hezaveh & Cherry*, *2018*), or in hazardous locations like midblock areas and dark roadways (*Das et al.*, *2020*). Additionally, pedestrian groups often form far from intersections because bus stops are not conveniently located near crossing points, encouraging midblock crossings (*Abaza et al.*, *2018*). Finally, while vehicle technologies such as emergency braking help mitigate pedestrian safety issues, it is still not as effective for poorly lit conditions and high-speed roads, which could prove fatal for pedestrians (*Cicchino*, *2022*).

Exploratory Studies in Pedestrian Safety

Several studies have employed exploratory data analysis methods to identify essential groupings of crash characteristics, thereby allowing for a more comprehensive understanding of the most critical factors involved. Data mining techniques, such as Multiple Correspondence Analysis (MCA) and Association Rules Mining (ARM), have been widely used to group crashes exhibiting

similar features. For instance, a Louisiana study utilized MCA to demonstrate that nighttime crashes with poor lighting conditions were strongly linked to fatal pedestrian outcomes (*Das & Sun*, *2015*). Another study applied ARM to categorize crashes involving children, elderly pedestrians, older drivers and distracted driving behaviors (*Hossain et al.*, *2022*). Moreover, Latent Class Clustering (LCC) has been used to identify distinct pedestrian clusters, revealing unique typologies that are not immediately obvious. For example, one study found a specific cluster of pedestrians crossing roads at non-intersection locations during dark hours, from midnight to 6 am. This study also employed multinomial logistic regression to analyze these clusters and assess injury severity (*Sun et al.*, *2019*). Inspired by this approach, our research begins with unsupervised clustering methods such as LCC and Hierarchical Clustering (HC) and then advances to the use of supervised learning. Through supervised machine learning, we aim to achieve a more precise classification of roadways and to identify factors influencing pedestrian safety on urban roads in Tennessee.

Research Gaps and Contributions

To the best of the authors' knowledge, this is one of very few studies that explore beyond the state agencies' arterial classifications to identify potentially fatal crashes. It brings a novel perspective to the field by exploring the categorization of high-risk crashes according to road and environmental characteristics available in the crash database. Moreover, it performs a comprehensive analysis of these crashes, investigating how they interact with other crash features. This leads to novel insights into the relationships with the existing functional classification of roadways. This study's methodology offers a significant contribution by providing a replicable approach to identifying locations with high pedestrian fatality rates and hazardous road segments for hotspots analysis. Lastly, since pedestrian safety trends in Tennessee closely follow the national trends and has a large suburban population, the study results will not only give a clear picture of why pedestrian crashes are getting more severe each year but also provide insight into the aggravating national pedestrian safety situation during the last decade.

Data and Methodology

Using Tennessee Integrated Traffic Analysis Network (TITAN) police crash data from 2009 to 2019, we categorized crashes as high-risk or low-risk based on road and environmental characteristics through clustering algorithms and supervised learning techniques. We then analyzed these categorized crashes to understand their nature using spatial analysis, trend analysis, and logistic regression models. While Python programming was used for the cleaning and database creation, most of these analyses were performed in the R environment. The following subsections will provide a detailed discussion of the data and methodologies used in this project.

Data

The Tennessee Integrated Traffic Analysis Network (TITAN) database, managed by the Tennessee Department of Safety and Homeland Security, collects all traffic safety-related data, including traffic crashes reported by law enforcement agencies (*Tennessee Highway Safety Office*, *2021*). To ensure uniformity, TITAN adheres to the Model Minimum Uniform Crash Criteria (MMUCC) guidelines (*NHTSA*, *2017*), recording injury outcomes on the KABCO scale, where K denotes a fatal crash and O indicates no injury (*Federal Highway Administration*). TITAN comprises three main datasets: person, crash, and vehicle. The person dataset contains details on all individuals involved in the crash, including demographics, intoxications, actions during the crash, and injury severity outcome. The crash dataset includes specifics such as date, time, location, collision type, lighting conditions, and other infrastructure-related details including if the crash occurred in parking lots and private properties. The vehicle dataset provides information on each vehicle involved, including details about the vehicle's characteristics, maneuvers, and some built-environment information like posted speed associated with the vehicle, road profile, alignment, surface type, number of lanes, travel direction, etc. We stripped off the personally identifiable information from the dataset before proceeding with the analyses.

For this report, we excluded interstate crashes and those identified by police as rural area crashes. Interstate highways, which have fully controlled access, restrict pedestrian presence, and pedestrian crash incidents typically involve emergency stops. Excluding rural area crashes allows us to focus specifically on urban roads in Tennessee, where the majority of crashes and fatalities occur (*Parajuli et al.*, *2023*). We also excluded non-vehicle crashes, such as those involving farm equipment or golf carts. After cleaning the dataset and removing entries with missing values, we were left with 17,267 records of pedestrian-involved crashes from January 2009 to September 2019. Additionally, we utilized electronic - Tennessee Roadway Information Management System (E-TRIMS) data to distinguish roadways that are functionally classified at the state level. This distinction will be important for pre-labeling certain pedestrian crashes for supervised learning, which will be elaborated on later.

Unsupervised Learning Classification

Unsupervised learning is a machine learning technique to group observations into clusters by identifying hidden patterns and similarities. In our study, we employed two methods, Latent Class Clustering (LCC) and Hierarchical Clustering (HC), to cluster pedestrian crash data from urban Tennessee. LCC is popular as a generalized finite modeling technique useful for analyzing categorical data and uncovering hidden typologies, making it relevant in the transportation safety field. For this clustering, we have used an open-source R package called poLCA (*Linzer & Lewis*, *2011*). HC is also a widely used unsupervised learning technique known for its simplicity and intuitive nature, primarily due to its ability to form clusters using dendrograms. For HC, we utilized the agglomerative clustering routine, employing the Jaccard distance matrix (*Jaccard*, *1912*) and Ward's linkage method to compute the hierarchical clusters (*Ward Jr*, *1963*). To determine the optimal number of clusters for LCC, an elbow plot displaying Akaike Information Criteria (AIC) values against the number of clusters will be generated. The same number of clusters will then be used for HC. This approach allows us to analyze each cluster type and assess whether both clustering algorithms consistently identify the most hazardous and least hazardous clusters.

Our study focuses on categorizing crash data based on road and environmental features. Road features include factors such as the number of travel lanes, traffic flow type, traffic control type, surface condition, residential land use, intersection location, parking location, straight maneuver at midblock locations, and posted speed limit. Environmental features, that have a strong bearing on the roadway designs, include lighting conditions, weather conditions, and weekends, offering a comprehensive view of the circumstances surrounding pedestrian crashes. The cross-tabulated clustering algorithm results will help us understand each cluster and identify the most hazardous clusters based on the specified characteristics. It will also aid in making informed decisions for pre-labeling in supervised learning classification.

Supervised Learning Classification

Although unsupervised models provide valuable insights, they have limitations. One major drawback is their inability to assess the importance of individual features within the model, making it challenging to identify critical crash characteristics. Additionally, since unsupervised learning depends solely on inherent data patterns and structures, noise, and outliers can significantly affect clustering results. Supervised learning classification methods can address these issues, but they require labeling of the data. The TITAN database does not provide comprehensive information for all pedestrian crashes, making it impractical to manually label every crash. However, manual labeling can be confidently applied to certain crashes, such as those near major urban arterials and in low-speed zones like parking lots, allowing us to categorize them as high-risk or low-risk, respectively. For this pre-labeling, we utilized roadways classified as major arterials in the E-TRIMS database and insights from the unsupervised learning clusters to determine the initial labels with high confidence. A neural network model was developed using the predetermined labels after splitting the data into training and testing datasets.

We used the "caret" package in R (*Kuhn*, *2008*) for the supervised learning model training along with the "nnet" package (*Ripley et al.*, *2016*) to fit an Artificial Neural Network (ANN) model. The model uses the Multilayer Perceptron (MLP) classifier utilizing the standardized labeled data and a logistic sigmoid activation function. The MLP can discern complex patterns and relationships within the data with multiple layers, allowing it to make precise and well-informed predictions. As a feedforward ANN model, data in an MLP moves in a single direction, beginning with the input layer, through the hidden layers, and reaching the output layer, with no feedback loops involved (*Gardner & Dorling*, *1998*). The model was fitted using a grid search of hyperparameters for the number of hidden neurons and the regularization parameter. The model was used to classify the unlabeled pedestrian crashes using an appropriate threshold for classifying crashes into two groups.

Logistic Regression

We fitted three logistic regression models to understand the characteristics of high-risk crashes, as identified from the supervised learning classification results. The first model is a binary logit model with risk labels as the dependent variable and crash characteristics as the independent variables. This model will give us a comprehensive understanding of high-risk crashes and their occurrences based on environmental features. The remaining two models include the injury severity model using logistic regression with the fatality outcome as the dependent variable to gain deeper insights into the characteristics of high-risk crashes and other relevant variables concerning fatality. A similar approach was used by Sun et al. where the study used a multinomial logit model for studying the characteristics of unique pedestrian crash clusters (*Sun et al.*, *2019*). Since we are dealing with binary classification, we developed two logistic regression models for injury severity models: the Base Model (BM) and the Interaction Model (IM). BM is a conventional logistic regression model to get a superficial understanding of variables associated with pedestrian crashes. IM is an interaction model, where we use high-risk crashes as the interaction term to identify the relationship between high-risk crashes and other crash variables.

Results and Discussions

This section is divided into multiple sub-sections of results and discussions for coherency.

Unsupervised Learning Classification Results

We used an AIC plot against the number of clusters, to determine the optimal number of clusters for LCC. As seen in Figure 2, the elbow point is approximately at 5 clusters, which we selected as the optimal number of clusters for LCC, and this set was also the easiest to interpret. We also chose the same number of clusters for HC, for consistency and comparative analysis of similar clusters.

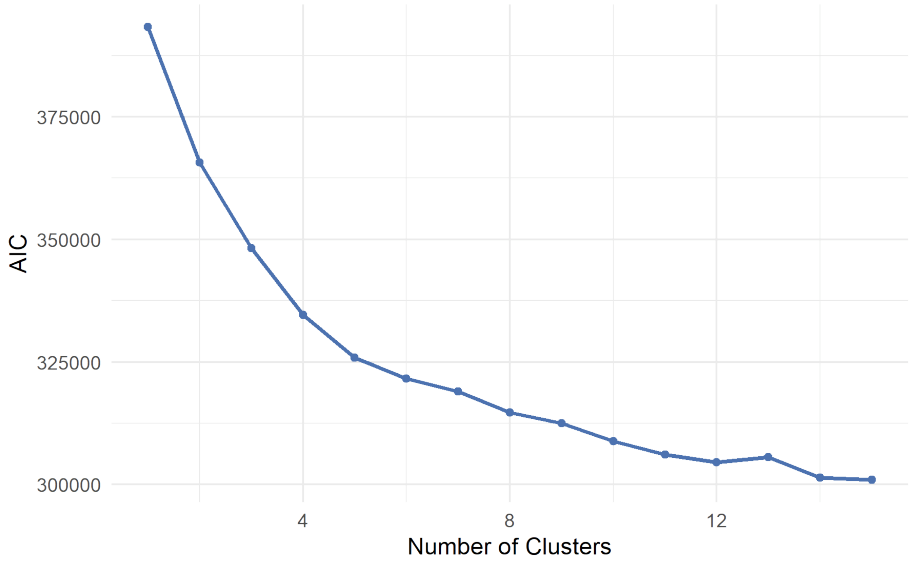


Figure 2. Number of LCC clusters vs. AIC values

Table 1. Cluster sizes for LCC and HC

Clusters	Cluster Size for LCC (%)	Cluster Size for HC (%)
Cluster 1	2,478 (14.4)	2,519 (14.6)
Cluster 2	4,801 (27.8)	5,453 (31.6)
Cluster 3	4,240 (24.6)	3,474 (20.1)
Cluster 4	3,072 (17.8)	3,291 (19.1)
Cluster 5	2,676 (15.5)	2,530 (14.7)
Total	17,267 (100)	17,267 (100)

The cluster labels for both HCC and LC are synced according to their cluster sizes (refer to Table 1) and severity levels (refer to Figure 3). Consequently, the clusters exhibiting similar distributions are labeled with identical names for both algorithm results. Table 1 gives an overview of cluster

sizes for both LCC and HC. In Table 1, we observed that Cluster 2 is the largest cluster and Cluster 1 is the smallest cluster for both algorithms. However, for HC, Clusters 5 and 1 are of similar size. Furthermore, Figure 3 reveals that Cluster 5 and Cluster 2 are among the most hazardous clusters, while Cluster 3 is the safest for clustering both results.

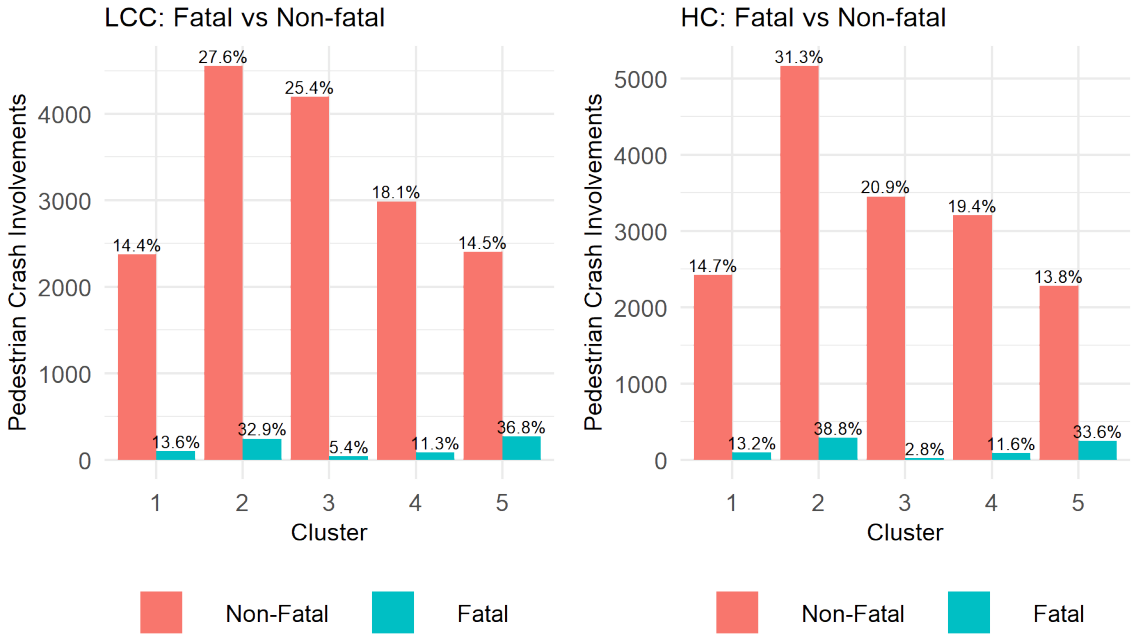


Figure 3. Fatal and non-fatal pedestrian crashes for LCC (left) and HC (right)

Cross-tabulation results for LCC and HC clusters and their relationship with road design and environmental variables are shown by Table 2 and Table 3, respectively. With a succinct description, each of these five clusters from both clustering processes is detailed below.

Cluster 1: Wet Roads

Cluster 1 from both algorithms is associated with pedestrian crashes in inclement weather, mostly rainy conditions with probable visibility issues. Table 2 and Table 3 show that around 73 percent of crashes in this cluster are associated with the rain and more than 95 percent of crashes happening on wet surface roads are classified into this cluster. Furthermore, Figure 4 Suggests that almost all crashes occurring in the rain belong to this cluster. Additionally, more than half of the crashes occurring in dark conditions are included in this cluster. Regarding the fatality rate, this cluster is relatively deadly, with a rate of 4.1 and 3.9 deaths per 100 pedestrians involved in crashes according to LCC and HC results, respectively. The overall fatality rate was 4.3 deaths per 100 pedestrians involved.

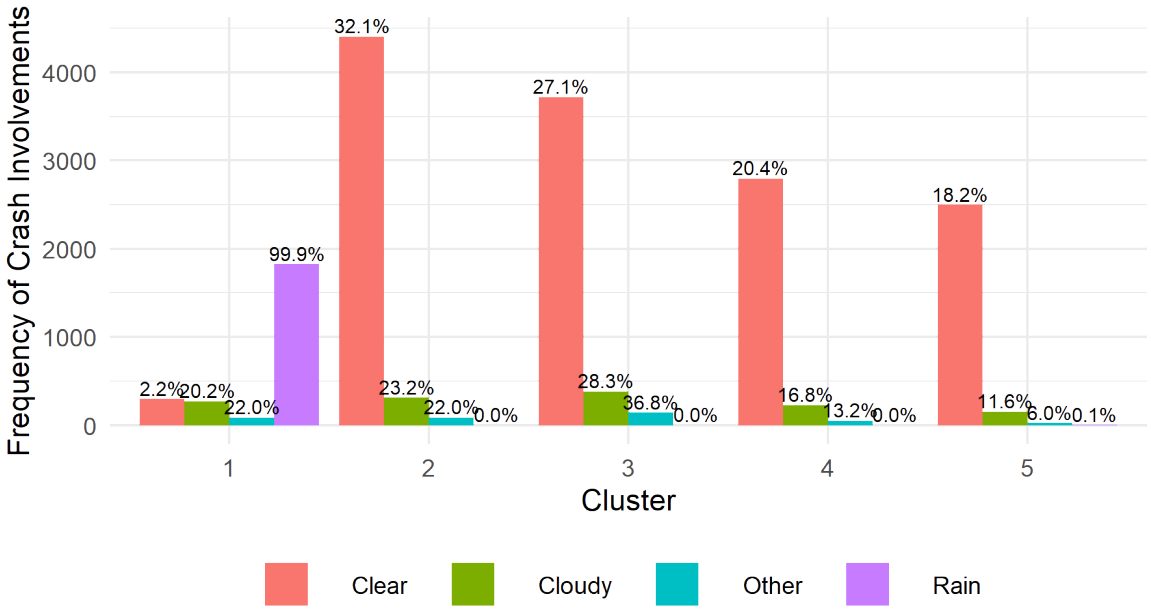


Figure 4. LCC clusters and weather conditions

Cluster 2: Midblock Locations on Narrow Roads

According to the LCC results (Table 2) and as illustrated by Figure 5, Cluster 2 exclusively comprises pedestrian crashes happening on roads with one or two lanes. HC results (Table 3) are similar with almost 90 percent of pedestrian crashes in its cluster happening on these streets. Figure 6 depicts that this cluster category also features more than three-fourths of the crashes happening on roads with speed limits ranging from 20 mph to 40 mph. Moreover, the crashes in this cluster occur mostly on non-intersection locations (~ 84 percent) and roads without traffic controls, suggesting that these crashes are mostly associated with narrow roads and midblock locations. It should be noted that more than 66 percent of crashes in this cluster occur on two-way undivided roadways. This is also the only cluster with the overrepresentation of residential area crashes, further strengthening its association with narrow roads. According to LCC and HC results, the fatality rate for this cluster is 5.1 and 5.3 deaths per 100 pedestrians involved, respectively.

Cluster 3: Low-Speed Zones

With above 85 percent of the crashes (Table 2 and Table 3) associated with parking lots and private properties, this cluster is associated with low-speed zones. Additionally, this cluster primarily consists of crashes occurring at speeds of 15 mph or lower, as shown in Figure 6. Around 75 percent of the crashes belonging to this cluster occur in daylight conditions. This cluster is also the safest cluster, with less than a 1 percent chance of fatal injury for pedestrians involved, according to both LCC and HC results.

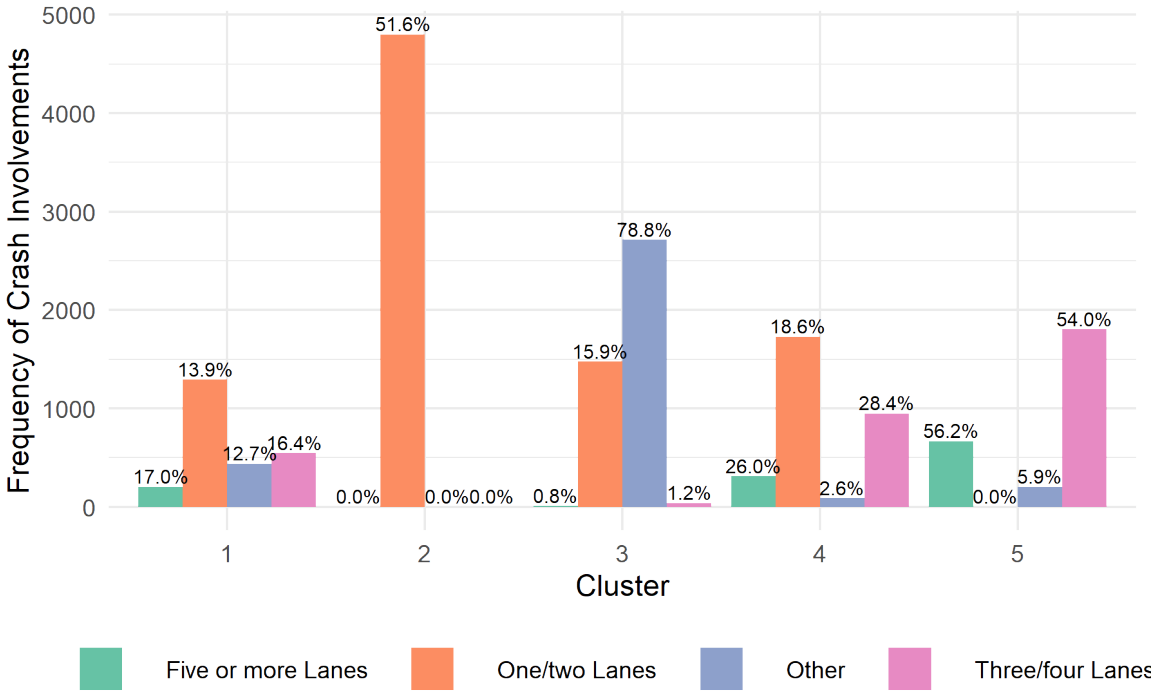


Figure 5. LCC clusters and number of lanes

Cluster 4: Intersections

As seen in Figure 7, this cluster distinguishes itself from other clusters with an overrepresentation of intersection crashes. Almost 85 percent of this cluster is at intersection locations according to the LCC analysis, and 75 percent according to the HC analysis, as shown in Table 2 and Table 3, respectively. This cluster was also primarily associated with locations that have traffic control devices like signage, pedestrian signals, and other traffic control systems. With around 2.7 deaths per 100 pedestrians involved, this cluster is among the less risky clusters after Cluster 3.

Cluster 5: Midblock Locations on Multilane Roads

This group of pedestrian crashes is associated mostly with high-speed wide roads. Both LCC and HC results from Table 2 and Table 3 suggest that this cluster comprises more than 85 percent of crashes occurring on roads with posted speed limits of 35 mph and higher. Additionally, more than 92 percent of the crashes in this cluster happen on multilane roads with at least 3 lanes or more. Most of these crashes happen at non-intersection locations without traffic control systems, and the majority also occur during nighttime. In terms of fatality outcomes, this cluster is the most dangerous among the five with 10.2 deaths per 100 involved according to LCC results (Table 2) and 9.8 deaths according to HC results (Table 3).

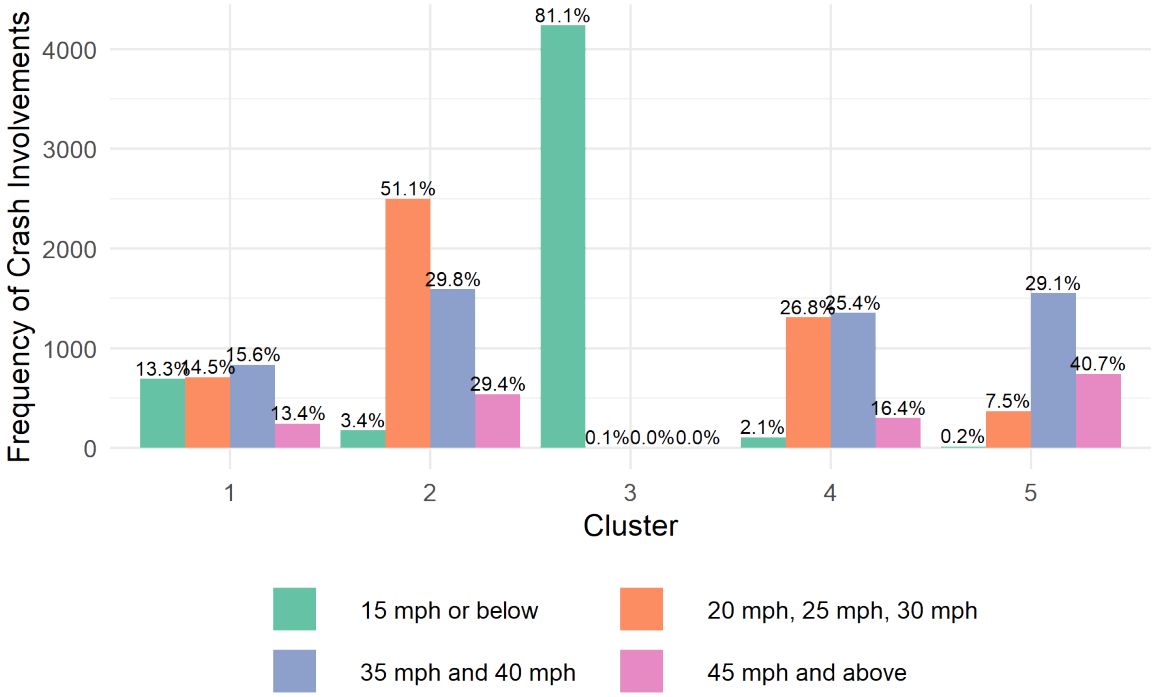


Figure 6. LCC clusters and posted speed limits

Figure 7. LCC clusters and intersections

Table 2. Cross-tabulation of Latent Class Clustering (LCC) results

Road & Environmental features	LCC1 (%)	LCC2 (%)	LCC3 (%)	LCC4 (%)	LCC5 (%)	Total (%)
Outcome						
Non-fatal	2,377 (95.9)	4,557 (94.9)	4,200 (99.1)	2,988 (97.3)	2,403 (89.8)	16,525 (95.7)
Fatal	101 (4.1)	244 (5.1)	40 (0.9)	84 (2.7)	273 (10.2)	742 (4.3)
At Intersection						
No	1,728 (69.7)	4,033 (84.0)	4,199 (99.0)	489 (15.9)	2,198 (82.1)	12,647 (73.2)
Yes	750 (30.3)	768 (16.0)	41 (1.0)	2,583 (84.1)	478 (17.9)	4,620 (26.8)
Light Condition						
Dark-Lighted	1,055 (42.6)	1,276 (26.6)	768 (18.1)	778 (25.3)	1,164 (43.5)	5,041 (29.2)
Dark-Not Lighted	286 (11.5)	690 (14.4)	124 (2.9)	128 (4.2)	218 (8.1)	1,446 (8.4)
Dawn/Dusk	96 (3.9)	194 (4.0)	107 (2.5)	103 (3.4)	100 (3.7)	600 (3.5)
Daylight	1,008 (40.7)	2,586 (53.9)	3,114 (73.4)	2,035 (66.2)	1,166 (43.6)	9,909 (57.4)
Other	33 (1.3)	55 (1.1)	127 (3.0)	28 (0.9)	28 (1.0)	271 (1.6)
Straight maneuver at midblock						
No	1,240 (50.0)	1,702 (35.5)	3,171 (74.8)	1,989 (64.7)	922 (34.5)	9,024 (52.3)
Yes	1,238 (50.0)	3,099 (64.5)	1,069 (25.2)	1,083 (35.3)	1,754 (65.5)	8,243 (47.7)
Land Use						
Residential	669 (27.0)	2,633 (54.8)	867 (20.4)	691 (22.5)	428 (16.0)	5,288 (30.6)
Non-residential	1,809 (73.0)	2,168 (45.2)	3,373 (79.6)	2,381 (77.5)	2,248 (84.0)	11,979 (69.4)
Posted Speed Limits						
15 mph or below	693 (28.0)	176 (3.7)	4,236 (99.9)	108 (3.5)	13 (0.5)	5,226 (30.3)
20 mph, 25 mph, 30 mph	708 (28.6)	2,498 (52.0)	4 (0.1)	1,309 (42.6)	369 (13.8)	4,888 (28.3)
35 mph and 40 mph	832 (33.6)	1,590 (33.1)	0 (0.0)	1,355 (44.1)	1,552 (58.0)	5,329 (30.9)
45 mph and above	245 (9.9)	537 (11.2)	0 (0.0)	300 (9.8)	742 (27.7)	1,824 (10.6)
Traffic Control System (TCS)						
No Control	1,777 (71.7)	4,801 (100)	3,802 (89.7)	0 (0.0)	2,676 (100)	13,056 (75.6)
Stop sign/ Yield sign	110 (4.4)	0 (0.0)	48 (1.1)	584 (19.0)	0 (0.0)	742 (4.3)
TCS-With Ped Signal	240 (9.7)	0 (0.0)	0 (0.0)	1,177 (38.3)	0 (0.0)	1,417 (8.2)
TCS-Without Ped Signal	183 (7.4)	0 (0.0)	0 (0.0)	884 (28.8)	0 (0.0)	1,067 (6.2)
Others	168 (6.8)	0 (0.0)	390 (9.2)	427 (13.9)	0 (0.0)	985 (5.7)
Traffic Flow						
1-Way Trafficway	62 (2.5)	79 (1.6)	238 (5.6)	134 (4.4)	58 (2.2)	571 (3.3)
2-Way Divided w/ Barrier	85 (3.4)	159 (3.3)	15 (0.4)	145 (4.7)	149 (5.6)	553 (3.2)
2-Way Divided w/o Barrier	477 (19.2)	1,043 (21.7)	67 (1.6)	773 (25.2)	863 (32.2)	3,223 (18.7)
2-Way Not Divided	1,324 (53.4)	3,280 (68.3)	1,341 (31.6)	1,824 (59.4)	1,249 (46.7)	9,018 (52.2)
Other	530 (21.4)	240 (5.0)	2,579 (60.8)	196 (6.4)	357 (13.3)	3,902 (22.6)
Trafficway Type						
Parking Lot	456 (18.4)	69 (1.4)	2,692 (63.5)	14 (0.5)	36 (1.3)	3,267 (18.9)
Private Property/Road	140 (5.6)	81 (1.7)	985 (23.2)	12 (0.4)	17 (0.6)	1,235 (7.2)
Trafficway	1,882 (75.9)	4,651 (96.9)	563 (13.3)	3,046 (99.2)	2,623 (98.0)	12,765 (73.9)

Road & Environmental features	LCC1 (%)	LCC2 (%)	LCC3 (%)	LCC4 (%)	LCC5 (%)	Total (%)
Number of Lanes						
Five or more Lanes	203 (8.2)	0 (0.0)	9 (0.2)	310 (10.1)	669 (25.0)	1,191 (6.9)
Three/four Lanes	546 (22.0)	0 (0.0)	40 (0.9)	948 (30.9)	1,804 (67.4)	3,338 (19.3)
One/two Lanes	1,292 (52.1)	4,801 (100)	1,477 (34.8)	1,726 (56.2)	0 (0.0)	9,296 (53.8)
Other	437 (17.6)	0 (0.0)	2,714 (64.0)	88 (2.9)	203 (7.6)	3,442 (19.9)
Surface Condition						
Dry	0 (0.0)	4,725 (98.4)	4,016 (94.7)	3,012 (98.0)	2,654 (99.2)	14,407 (83.4)
Others	26 (1.0)	76 (1.6)	213 (5.0)	60 (2.0)	22 (0.8)	397 (2.3)
Wet	2,452 (99.0)	0 (0.0)	11 (0.3)	0 (0.0)	0 (0.0)	2,463 (14.3)
Weather						
Clear	297 (12.0)	4,405 (91.8)	3,719 (87.7)	2,796 (91.0)	2,497 (93.3)	13,714 (79.4)
Cloudy	271 (10.9)	311 (6.5)	379 (8.9)	225 (7.3)	155 (5.8)	1,341 (7.8)
Other	85 (3.4)	85 (1.8)	142 (3.3)	51 (1.7)	23 (0.9)	386 (2.2)
Rain	1,825 (73.6)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.0)	1,826 (10.6)
Weekend						
No	1,854 (74.8)	3,565 (74.3)	3,192 (75.3)	2,523 (82.1)	2,010 (75.1)	13,144 (76.1)
Yes	624 (25.2)	1,236 (25.7)	1,048 (24.7)	549 (17.9)	666 (24.9)	4,123 (23.9)
Number of Involvements	2,478	4,801	4,240	3,072	2,676	17,267

Table 3. Cross-tabulation of Hierarchical Clustering (HC) results

Road & Environmental features	HC1 (%)	HC2 (%)	HC3 (%)	HC4 (%)	HC5 (%)	Total (%)
Outcome						
Non-fatal	2,421 (96.1)	5,165 (94.7)	3,453 (99.4)	3,205 (97.4)	2,281 (90.2)	16,525 (95.7)
Fatal	98 (3.9)	288 (5.3)	21 (0.6)	86 (2.6)	249 (9.8)	742 (4.3)
At Intersection						
No	1,793 (71.2)	4,546 (83.4)	3,407 (98.1)	844 (25.6)	2,057 (81.3)	12,647 (73.2)
Yes	726 (28.8)	907 (16.6)	67 (1.9)	2,447 (74.4)	473 (18.7)	4,620 (26.8)
Light Condition						
Dark-Lighted	1,044 (41.4)	1,404 (25.7)	672 (19.3)	781 (23.7)	1,140 (45.1)	5,041 (29.2)
Dark-Not Lighted	293 (11.6)	681 (12.5)	139 (4.0)	124 (3.8)	209 (8.3)	1,446 (8.4)
Dawn/Dusk	96 (3.8)	220 (4.0)	94 (2.7)	102 (3.1)	88 (3.5)	600 (3.5)
Daylight	1,039 (41.2)	3,089 (56.6)	2,534 (72.9)	2,178 (66.2)	1,069 (42.3)	9,909 (57.4)
Other	47 (1.9)	59 (1.1)	35 (1.0)	106 (3.2)	24 (0.9)	271 (1.6)
Straight maneuver at midblock						
No	1,293 (51.3)	2,254 (41.3)	2,497 (71.9)	2,120 (64.4)	860 (34.0)	9,024 (52.3)
Yes	1,226 (48.7)	3,199 (58.7)	977 (28.1)	1,171 (35.6)	1,670 (66.0)	8,243 (47.7)
Land Use						
Residential	707 (28.1)	2,798 (51.3)	545 (15.7)	819 (24.9)	419 (16.6)	5,288 (30.6)
Non-residential	1,812 (71.9)	2,655 (48.7)	2,929 (84.3)	2,472 (75.1)	2,111 (83.4)	11,979 (69.4)

Road & Environmental features	HC1 (%)	HC2 (%)	HC3 (%)	HC4 (%)	HC5 (%)	Total (%)
Posted Speed Limits						
15 mph or below	729 (28.9)	620 (11.4)	3,330 (95.9)	509 (15.5)	38 (1.5)	5,226 (30.3)
20 mph, 25 mph, 30 mph	724 (28.7)	2,556 (46.9)	131 (3.8)	1,197 (36.4)	280 (11.1)	4,888 (28.3)
35 mph and 40 mph	841 (33.4)	1,653 (30.3)	10 (0.3)	1,300 (39.5)	1,525 (60.3)	5,329 (30.9)
45 mph and above	225 (8.9)	624 (11.4)	3 (0.1)	285 (8.7)	687 (27.2)	1,824 (10.6)
Traffic Control System (TCS)						
No Control	1,831 (72.7)	5,233 (96.0)	3,472 (99.9)	3 (0.1)	2,517 (99.5)	13,056 (75.6)
Stop sign/ Yield sign	111 (4.4)	53 (1.0)	0 (0.0)	575 (17.5)	3 (0.1)	742 (4.3)
TCS-With Ped Signal	245 (9.7)	78 (1.4)	1 (0.0)	1,092 (33.2)	1 (0.0)	1,417 (8.2)
TCS-Without Ped Signal	159 (6.3)	35 (0.6)	1 (0.0)	871 (26.5)	1 (0.0)	1,067 (6.2)
Others	173 (6.9)	54 (1.0)	0 (0.0)	750 (22.8)	8 (0.3)	985 (5.7)
Traffic Flow						
1-Way Trafficway	65 (2.6)	66 (1.2)	232 (6.7)	155 (4.7)	53 (2.1)	571 (3.3)
2-Way Divided w/ Barrier	75 (3.0)	82 (1.5)	18 (0.5)	146 (4.4)	232 (9.2)	553 (3.2)
2-Way Divided w/o Barrier	476 (18.9)	1,111 (20.4)	75 (2.2)	757 (23.0)	804 (31.8)	3,223 (18.7)
2-Way Not Divided	1,352 (53.7)	3,599 (66.0)	1,099 (31.6)	1,821 (55.3)	1,147 (45.3)	9,018 (52.2)
Other	551 (21.9)	595 (10.9)	2,050 (59.0)	412 (12.5)	294 (11.6)	3,902 (22.6)
Trafficway Type						
Parking Lot	478 (19.0)	353 (6.5)	2,222 (64.0)	189 (5.7)	25 (1.0)	3,267 (18.9)
Private Property/Road	151 (6.0)	229 (4.2)	748 (21.5)	91 (2.8)	16 (0.6)	1,235 (7.2)
Trafficway	1,890 (75.0)	4,871 (89.3)	504 (14.5)	3,011 (91.5)	2,489 (98.4)	12,765 (73.9)
Number of Lanes						
Five or more Lanes	186 (7.4)	50 (0.9)	6 (0.2)	310 (9.4)	639 (25.3)	1,191 (6.9)
Three/four Lanes	545 (21.6)	179 (3.3)	15 (0.4)	889 (27.0)	1,710 (67.6)	3,338 (19.3)
One/two Lanes	1,324 (52.6)	4,846 (88.9)	1,251 (36.0)	1,768 (53.7)	107 (4.2)	9,296 (53.8)
Other	464 (18.4)	378 (6.9)	2,202 (63.4)	324 (9.8)	74 (2.9)	3,442 (19.9)
Surface Condition						
Dry	4 (0.2)	5,381 (98.7)	3,412 (98.2)	3,106 (94.4)	2,504 (99.0)	14,407 (83.4)
Others	122 (4.8)	42 (0.8)	62 (1.8)	159 (4.8)	12 (0.5)	397 (2.3)
Wet	2,393 (95.0)	30 (0.6)	0 (0.0)	26 (0.8)	14 (0.6)	2,463 (14.3)
Weather						
Clear	285 (11.3)	4,354 (79.8)	3,430 (98.7)	3,127 (95.0)	2,518 (99.5)	13,714 (79.4)
Cloudy	253 (10.0)	1,035 (19.0)	8 (0.2)	45 (1.4)	0 (0.0)	1,341 (7.8)
Other	160 (6.4)	64 (1.2)	36 (1.0)	114 (3.5)	12 (0.5)	386 (2.2)
Rain	1,821 (72.3)	0 (0.0)	0 (0.0)	5 (0.2)	0 (0.0)	1,826 (10.6)
Weekend						
No	1,901 (75.5)	4,090 (75.0)	2,555 (73.5)	2,701 (82.1)	1,897 (75.0)	13,144 (76.1)
Yes	618 (24.5)	1,363 (25.0)	919 (26.5)	590 (17.9)	633 (25.0)	4,123 (23.9)
Number of Involvements	2,519	5,453	3,474	3,291	2,530	17,267

Unsupervised Learning Classification Discussions

The clustering results provide a comprehensive understanding of five types of crashes occurring in Tennessee according to the road and environment characteristics: crashes in wet roads, narrow roads at midblock locations, low-speed zones, intersections, and wide roads at midblock locations. The results were consistent across both clustering algorithms. These results were also consistent with the spatial distribution of crashes for each cluster.

Figure 8 features a spatial distribution of pedestrian crashes for all five LCC clusters around Nashville, Tennessee. Crashes in Cluster 1 do not show a discernible pattern; they are clustered around some roads but also scattered across various locations on the map. It is understandable as weather-related crashes are spread roughly over the datasets. Cluster 2 is interesting because the crashes appear randomly scattered across the map. A closer look reveals that most of the pedestrian crashes in this cluster occur on residential streets, giving a more distributed appearance. This appearance also conforms with the clustering results stating the relationship of these crashes with narrow roads. Cluster 3 crashes occur in random patches, often in areas with large parking lots, shopping centers, grocery stores, and malls. Cluster 4 crashes seem to be clustered along certain roads, but on closer inspection, most of these crashes occur at intersections with other streets. The spatial arrangement of Cluster 5 crashes is quite intriguing; unlike other clusters, these crashes are not concentrated in the Downtown area but are instead neatly concentrated along specific streets. These streets are multilane, high-speed roads, mostly classified as major arterials according to the E-TRIMS dataset.

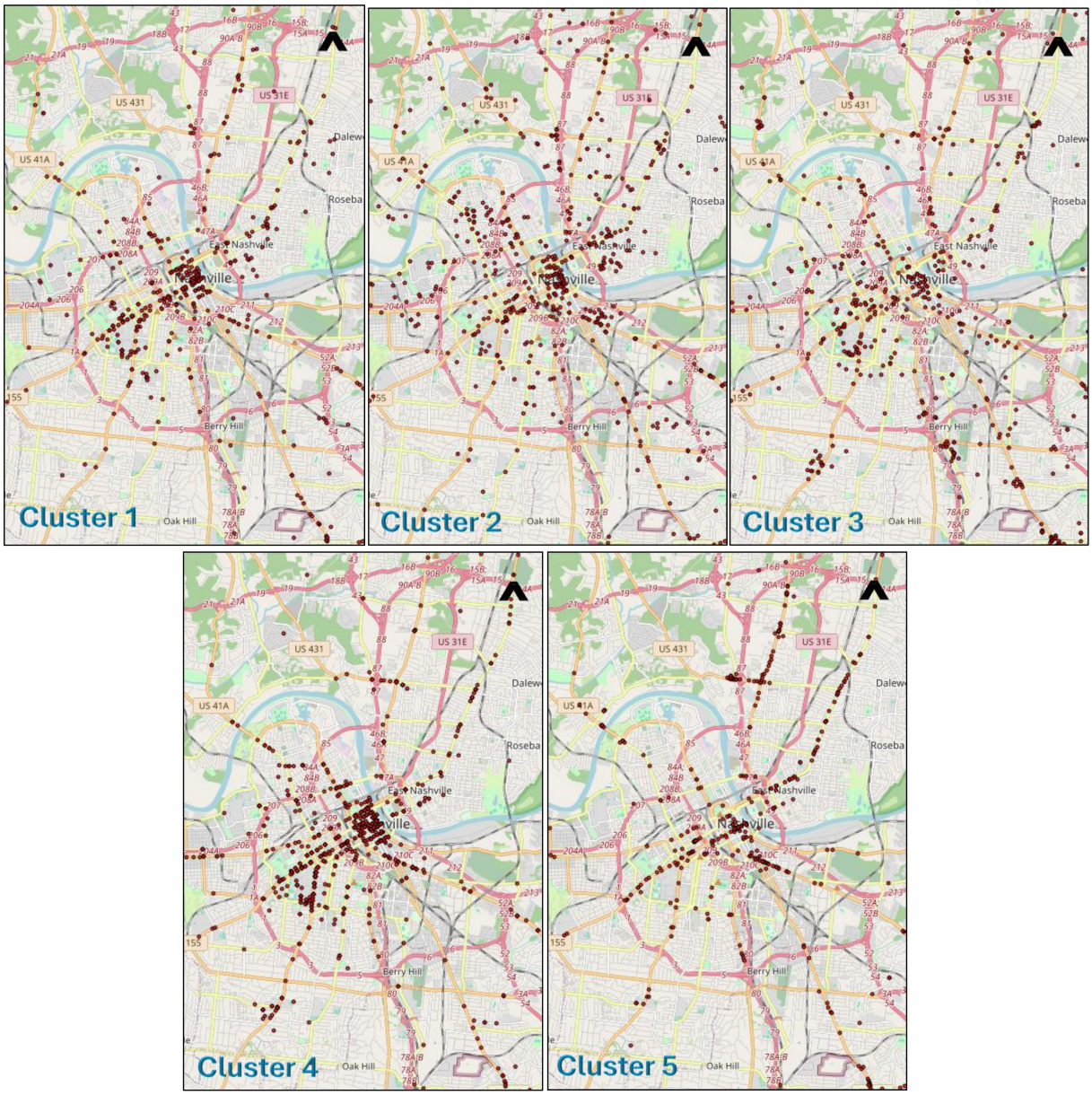


Figure 8. Spatial visualization of LCC clusters in Nashville

It is safe to assume that crashes in parking lots and on private roads are among the least hazardous, with a low likelihood of fatality. In contrast, crashes at non-intersection locations on high-speed roads are among the most dangerous for pedestrians. This information is crucial for the next steps, as it allows us to confidently use these crash pre-labels for supervised learning classification. Supervised learning offers better control in distinguishing between high-risk and low-risk pedestrian crashes, unlike unsupervised learning, where the clusters are less flexible.

Supervised Learning Classification Results

Streets classified as "major arterials" in the E-TRIMS dataset are often the roads with higher speeds and multiple lanes. Through the literature review, we have established that pedestrians face a higher risk of death on these roads. This finding is further supported by unsupervised learning results, which show that pedestrian crashes on these roads have about a 10 percent chance of resulting in a fatal injury. To identify the pre-labels for high-risk crashes, we overlaid the existing crash data with the E-TRIMS dataset. We then labeled non-intersection pedestrian crashes occurring within 100 feet of major arterials, as defined by E-TRIMS, as high-risk crashes. Similarly, as supported by the pedestrian safety literature and LCC and HC clustering results, we labeled the non-intersection-related parking lot as low-risk crashes.

The pre-processing step resulted in 2,092 crash points associated with major arterials in the E-TRIMS dataset, labeled as high-risk crashes. Additionally, 3,823 pedestrian crashes were identified in parking lots and categorized as low-risk crashes. The remaining crashes were left unlabeled. To balance the dataset for future training, 1,800 low-risk crash points were randomly selected, and their labels were changed to unlabeled. This adjustment provided us with balanced pre-labels: 2,092 high-risk crashes with 237 fatalities, indicating a fatality rate of 11.3 deaths per 100 pedestrians involved, and 2,023 low-risk crash points with 24 fatalities, indicating a fatality rate of 1.2 deaths per 100 pedestrians involved. Figure 9 Offers a visualization of these pre-labels, illustrating the distribution of fatality outcomes.

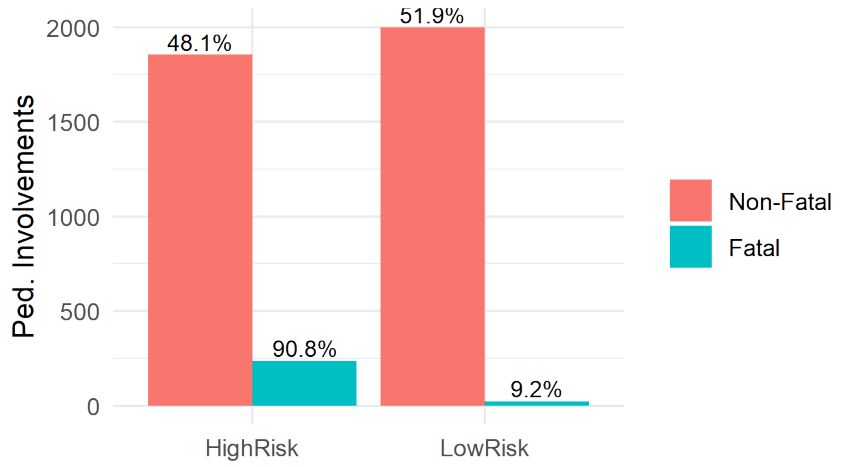


Figure 9. Classification pre-labels vs. fatality outcomes

We then proceeded with the training of the classification model based on these pre-labels. An ANN model was fitted with the 70-30 train-test split using the MLP classifier with 10-fold cross-validation. Like the clustering process, only road design and environmental variables were input into the model to determine high-risk and low-risk crashes. These variables include lighting condition type, weather type, traffic control presence/type, traffic way flow type, travel lanes,

surface condition, posted speed limit, weekend indicator, non-residential land use indicator, intersection indicator, and midblock indicator. To optimize the neural network model, we performed a hyperparameter search using a grid search approach. Specifically, we tuned two key parameters: the number of hidden layers and the weight decay parameter. We considered a range from 1 to 7 units in the hidden layer and explored a sequence of decay values from 10^{-4} to 10^{-1} on a logarithmic scale, resulting in 10 values. We identified the best-fitting model with 3 units in the hidden layer and a decay of 0.0015, achieving a cross-validation score of 0.941. The accuracy of the model on the test dataset was 94.65 percent, suggesting a robust model based on the pre-labels. With this model, we then proceeded to classify the unlabeled pedestrian crashes. For a balanced classification of low-risk and high-risk crashes, we defined the cut-off probability for classification as 0.9 instead of the default 0.5.

Cross-tabulation

Table 4 presents a detailed two-way crosstab analysis of classification results and fatality outcomes against crash variables, including road design and environmental variables, pedestrian and driver characteristics, and vehicle categories. The neural network classification results show similarities with the clustering algorithm results. For instance, high-risk crashes predominantly occur on roads with speeds of 35 mph or more, comprising 68 percent of these crashes. This category also includes crashes primarily happening in the dark and other low-light conditions such as dawn and dusk, accounting for over 50 percent of crashes in these conditions. High-risk crashes frequently occur on multilane roads, and two-way divided roads without traffic barriers, and involve straight maneuvers at midblock locations. Conversely, low-risk crashes are mainly found in low-speed areas like parking lots, non-midblock locations, and daylight conditions. In terms of pedestrian demographics, high-risk crashes are more common among males (63 percent male within the high-risk category compared to 53 percent in the low-risk), Black pedestrians (35 percent in the high-risk compared to 26 percent in the low-risk), and individuals aged 16 to 54 years (80 percent compared to 69 percent.

Regarding fatalities within high-risk categories, these are overrepresented in dark conditions, at speeds above 45 mph, in non-residential areas, in midblock locations, among pedestrians above the age of 50, intoxicated pedestrians, male pedestrians, White pedestrians, and in crashes involving male or intoxicated drivers. For low-risk categories, fatalities are overrepresented in residential areas, crashes involving pickups and SUVs, male pedestrians, pedestrians walking under the influence, pedestrians living far from home, and drivers under the influence. There are no significant differences among driver characteristics and vehicle types. For a deeper understanding of these categories, we have employed statistical modeling, discussed later in this section.

Table 4. Cross-tabulation of supervised learning classification with fatality outcome vs. crash features

Variables	High-Risk		Low-Risk		Total
	Non-fatal	Fatal	Non-fatal	Fatal	
Road Design and Environmental Variables					
Light Condition					
Dark-Lighted	3,081 (34.3)	380 (60.9)	1,545 (20.5)	35 (29.7)	5,041
Dark-Not Lighted	986 (11.0)	125 (20.0)	328 (4.4)	7 (5.9)	1,446
Dawn/Dusk	310 (3.4)	16 (2.6)	265 (3.5)	9 (7.6)	600
Daylight	4,479 (49.8)	97 (15.5)	5,267 (69.9)	66 (55.9)	9,909
Other	131 (1.5)	6 (1.0)	133 (1.8)	1 (0.8)	271
At Intersection					
No	5,230 (58.2)	487 (78.0)	6,824 (90.5)	106 (89.8)	12,647
Yes	3,757 (41.8)	137 (22.0)	714 (9.5)	12 (10.2)	4,620
Posted Speed Limit					
15 mph or below	224 (2.5)	2 (0.3)	4,953 (65.7)	47 (39.8)	5,226
20 mph, 25 mph, 30 mph	2,782 (31.0)	69 (11.1)	1,989 (26.4)	48 (40.7)	4,888
35 mph and 40 mph	4,447 (49.5)	309 (49.5)	554 (7.3)	19 (16.1)	5,329
45 mph and above	1,534 (17.1)	244 (39.1)	42 (0.6)	4 (3.4)	1,824
Land Use					
Residential	2,175 (24.2)	96 (15.4)	2,937 (39.0)	80 (67.8)	5,288
Non-residential	6,812 (75.8)	528 (84.6)	4,601 (61.0)	38 (32.2)	11,979
Straight Maneuver at Midblock Locations					
No	3,976 (44.2)	90 (14.4)	4,890 (64.9)	68 (57.6)	9,024
Yes	5,011 (55.8)	534 (85.6)	2,648 (35.1)	50 (42.4)	8,243
Number of Lanes					
Five or more Lanes	1,075 (12.0)	100 (16.0)	16 (0.2)	0 (0.0)	1,191
One/two Lanes	4,679 (52.1)	257 (41.2)	4,276 (56.7)	84 (71.2)	9,296
Other	290 (3.2)	26 (4.2)	3,096 (41.1)	30 (25.4)	3,442
Three/four Lanes	2,943 (32.7)	241 (38.6)	150 (2.0)	4 (3.4)	3,338
Trafficway Type					
Parking Lot	76 (0.8)	1 (0.2)	3,168 (42.0)	22 (18.6)	3,267
Private Property or Road	25 (0.3)	0 (0.0)	1,193 (15.8)	17 (14.4)	1,235
Trafficway	8,886 (98.9)	623 (99.8)	3,177 (42.1)	79 (66.9)	12,765
Trafficway Flow					
1-Way Trafficway	247 (2.7)	7 (1.1)	313 (4.2)	4 (3.4)	571
2-Way Divided w/ Traffic Barrier	413 (4.6)	51 (8.2)	89 (1.2)	0 (0.0)	553
2-Way Divided w/o Traffic Barrier	2,909 (32.4)	173 (27.7)	141 (1.9)	0 (0.0)	3,223
2-Way Not Divided	4,715 (52.5)	291 (46.6)	3,931 (52.1)	81 (68.6)	9,018
Other	703 (7.8)	102 (16.3)	3,064 (40.6)	33 (28.0)	3,902

Variables	High-Risk		Low-Risk		Total
	Non-fatal	Fatal	Non-fatal	Fatal	
Traffic Control System					
No Control	5,535 (61.6)	529 (84.8)	6,884 (91.3)	108 (91.5)	13,056
Stop sign/ Yield sign	592 (6.6)	3 (0.5)	145 (1.9)	2 (1.7)	742
TCS-With Ped Signal	1,264 (14.1)	39 (6.2)	113 (1.5)	1 (0.8)	1,417
TCS-Without Ped Signal	1,021 (11.4)	42 (6.7)	4 (0.1)	0 (0.0)	1,067
Others	575 (6.4)	11 (1.8)	392 (5.2)	7 (5.9)	985
Surface Condition					
Dry	7,282 (81.0)	526 (84.3)	6,494 (86.2)	105 (89.0)	14,407
Others	200 (2.2)	7 (1.1)	187 (2.5)	3 (2.5)	397
Wet	1,505 (16.7)	91 (14.6)	857 (11.4)	10 (8.5)	2,463
Weather					
Clear	7,047 (78.4)	488 (78.2)	6,079 (80.6)	100 (84.7)	13,714
Cloudy	564 (6.3)	48 (7.7)	717 (9.5)	12 (10.2)	1,341
Other	237 (2.6)	11 (1.8)	138 (1.8)	0 (0.0)	386
Rain	1,139 (12.7)	77 (12.3)	604 (8.0)	6 (5.1)	1,826
Weekend					
No	7,042 (78.4)	446 (71.5)	5,568 (73.9)	88 (74.6)	13,144
Yes	1,945 (21.6)	178 (28.5)	1,970 (26.1)	30 (25.4)	4,123
Vehicle Characteristics					
Vehicle Category					
Heavy vehicles	146 (1.6)	27 (4.3)	118 (1.6)	10 (8.5)	301
Medium-heavy vehicles	147 (1.6)	15 (2.4)	157 (2.1)	4 (3.4)	323
Minivan	346 (3.9)	26 (4.2)	349 (4.6)	4 (3.4)	725
Others	196 (2.2)	16 (2.6)	133 (1.8)	0 (0.0)	345
Passenger cars	4,448 (49.5)	265 (42.5)	3,583 (47.5)	36 (30.5)	8,332
Pickups	1,334 (14.8)	116 (18.6)	1,172 (15.5)	26 (22.0)	2,648
SUV	1,510 (16.8)	115 (18.4)	1,448 (19.2)	31 (26.3)	3,104
Unknown	860 (9.6)	44 (7.1)	578 (7.7)	7 (5.9)	1,489
Pedestrian Characteristics					
Pedestrian age					
15 and younger	1,180 (13.1)	13 (2.1)	1,389 (18.4)	23 (19.5)	2,605
16 - 34	3,359 (37.4)	134 (21.5)	2,330 (30.9)	14 (11.9)	5,837
35 - 49	1,913 (21.3)	147 (23.6)	1,458 (19.3)	20 (16.9)	3,538
50 - 64	1,882 (20.9)	236 (37.8)	1,431 (19.0)	29 (24.6)	3,578
65 and above	637 (7.1)	94 (15.1)	929 (12.3)	32 (27.1)	1,692
unknown	16 (0.2)	0 (0.0)	1 (0.0)	0 (0.0)	17
Pedestrian Sex					
Female	3,398 (37.8)	174 (27.9)	3,560 (47.2)	37 (31.4)	7,169
Male	5,589 (62.2)	450 (72.1)	3,978 (52.8)	81 (68.6)	10,098

Variables	High-Risk		Low-Risk		Total
	Non-fatal	Fatal	Non-fatal	Fatal	
Pedestrian Race					
Black	3,189 (35.5)	197 (31.6)	1,983 (26.3)	35 (29.7)	5,404
White	3,876 (43.1)	353 (56.6)	3,779 (50.1)	61 (51.7)	8,069
Other	1,922 (21.4)	74 (11.9)	1,776 (23.6)	22 (18.6)	3,794
Walking under Influence					
No or Untested	8,345 (92.9)	480 (76.9)	7,312 (97.0)	103 (87.3)	16,240
Yes	642 (7.1)	144 (23.1)	226 (3.0)	15 (12.7)	1,027
Crash Location from Pedestrian's Home					
More than 2 miles	4,325 (48.1)	281 (45.0)	3,279 (43.5)	71 (60.2)	7,956
Less than 2 miles	3,973 (44.2)	302 (48.4)	3,808 (50.5)	41 (34.7)	8,124
Unknown	689 (7.7)	41 (6.6)	451 (6.0)	6 (5.1)	1,187
Driver Characteristics					
Driver Age					
15 - 24	1,268 (14.1)	103 (16.5)	1,121 (14.9)	15 (12.7)	2,507
25 - 54	3,738 (41.6)	309 (49.5)	2,775 (36.8)	60 (50.8)	6,882
55 and above	1,820 (20.3)	129 (20.7)	1,675 (22.2)	26 (22.0)	3,650
Unknown	2,161 (24.0)	83 (13.3)	1,967 (26.1)	17 (14.4)	4,228
Driver Sex					
Female	4,849 (54.0)	240 (38.5)	4,234 (56.2)	46 (39.0)	9,369
Male	4,138 (46.0)	384 (61.5)	3,304 (43.8)	72 (61.0)	7,898
Driver Race					
Black	2,595 (28.9)	198 (31.7)	1,838 (24.4)	40 (33.9)	4,671
White	4,198 (46.7)	336 (53.8)	3,898 (51.7)	59 (50.0)	8,491
Other	2,194 (24.4)	90 (14.4)	1,802 (23.9)	19 (16.1)	4,105
Driving Under Influence					
No or Untested	8,794 (97.9)	557 (89.3)	7,330 (97.2)	101 (85.6)	16,782
Yes	193 (2.1)	67 (10.7)	208 (2.8)	17 (14.4)	485
Number of Involvements	8,987	624	7,538	118	17,267

Table 5. Binary logit model fitted on pedestrian crash characteristics with classification labels from supervised learning

Variables	Coef.	Std. Error (SE)	Statistic
Dependent Variable: Risk Label (High Risk = 1)			
Lighting (Base: Daylight)			
Dark - lighted	1.209***	0.058	20.695
Dark - not lighted	0.394***	0.087	4.525
Dawn/dusk	-0.655***	0.125	-5.225
Intersection vs otherwise	0.705***	0.054	12.979
Straight maneuver at midblock	1.171***	0.053	22.262
Residential area vs. Otherwise	-1.811***	0.054	-33.661
Posted Speed Limit (Base: 20mph - 30 mph)			
15 mph and below	-3.078***	0.088	-35.020
35 mph and 40 mph	1.662***	0.053	31.572
45 mph and above	3.945***	0.151	26.105
Vehicle Category (Base: Passenger cars)			
Heavy Vehicles	0.171	0.181	0.945
Medium Heavy (e.g. Delivery vans)	-0.169	0.174	-0.974
Minivans	0.215 ·	0.123	1.749
Pickup trucks	-0.039	0.073	-0.543
SUVs	0.022	0.066	0.338
Other/ Unknown	-0.194**	0.092	-2.112
Pedestrian age (Base: 16-34)			
15 and below	-0.209**	0.073	-2.870
35-49	-0.057	0.067	-0.848
50-64	0.132 ·	0.067	1.957
65 and above	0.117	0.092	1.272
Pedestrian sex: Male vs female	0.007	0.049	0.153
Pedestrian race: White vs otherwise	-0.256***	0.051	-5.009
Pedestrian walking under the influence	0.195 ·	0.110	1.765
Driver Age (Base: 25-54)			
15-24	0.107	0.073	1.452
55 and above	-0.035	0.065	-0.542
Other/ Unknown	-0.482***	0.106	-4.556
Driver sex: male vs female	-0.003	0.054	-0.051
Driver Race (Base: White)			
Black	-0.24***	0.061	-3.956
Other/ Unknown	0.129	0.091	1.414
Driving Under the Influence (DUI)	-0.326**	0.149	-2.192
Intercept	-0.646***	0.093	-6.958
Degrees of Freedom	Total: 17266	Residual: 17235	
Deviance	Null: 23890	Residual: 11520	
AIC value	11580		

***p-value < 0.001, **p-value < 0.01, * p-value < 0.05, · p-value < 0.1

Logistic Regression Models

Table 5 presents the results of the logistic regression model where the dependent variable is the risk label for pedestrian crashes, with the high-risk category coded as 1. The model indicates that, compared to daylight conditions, crashes occurring in dark but lighted conditions are significantly more likely to be high-risk, with a log-odds increase of 1.209. Similarly, crashes in unlighted dark conditions have a log-odds increase of 0.394 for being high-risk. Crashes at intersections and those involving straight maneuvers at midblock locations are also more likely to be high-risk.

Conversely, crashes in residential areas are significantly more likely to be low-risk, with a log-odds decrease of -1.811 ($p < 0.001$). Vehicle types do not show significant indicators of high-risk crashes, suggesting no distinct relationship between vehicle type and crash risk. Similarly, pedestrian age and sex generally do not significantly predict high-risk crashes, except for child pedestrians (aged 15 and below), who are more associated with low-risk crashes, with a log-odds decrease of -0.209 ($p < 0.01$).

Pedestrian intoxication shows a weak relationship with high-risk crashes (coef. = 0.195, $p < 0.1$), while driver intoxication is strongly associated with low-risk crashes (coef. = -0.326, $p < 0.01$). Non-White pedestrians are more likely to be involved in high-risk crashes, with a log-odds increase of 0.256 ($p < 0.001$), while Black drivers are more likely to be associated with low-risk crashes, with a log-odds decrease of -0.240 ($p < 0.001$).

We fitted two injury severity models, BM and IM, with fatal outcomes as the dependent variable using logistic regression. The results are presented in Table 6. The Base Model (BM) identifies the crucial variables that dictate the probability of a pedestrian fatality given a crash. This model serves as a reference for the Interaction Model (IM), which explores the relationship between high-risk crash categories and fatality outcomes using interaction terms.

According to the BM model, crashes occurring in dark conditions are significantly associated with fatal outcomes compared to daylight conditions. Specifically, the log odds of a fatal outcome increase by 1.25 in dark-lighted conditions and by 1.08 in dark conditions without lighting. Limited lighting during dawn and dusk also raises the log odds by 0.784, with all these conditions being significant at the 0.001 level. Additionally, intersections are strongly associated with non-fatal outcomes (coef. = -0.467), whereas straight maneuvers at midblock locations are linked to fatal outcomes (coef. = 0.939).

Speed limits also play a significant role. Compared to a speed limit range of 20-30 mph, speeds of 15 mph and below are associated with non-fatal outcomes, with a log odds decrease of -0.846. Higher speeds are strongly linked to fatal outcomes, with log odds increases of 0.792 for 35-40 mph and 1.494 for 45 mph and above, all significant at a 99.9 percent confidence level.

Regarding vehicle characteristics, heavy vehicles like trucks and tractor-trailers are the most dangerous for pedestrians, with a log odds increase of 1.795 for fatality. Medium-heavy vehicles, such as large delivery vans, also pose a significant risk, with a log odds increase of 0.858 compared to passenger cars, significant at a 99.9 percent confidence level. Private vehicles, including

minivans (coef. = 0.424), pickup trucks (0.499), and SUVs (coef. = 0.474), are also significantly more hazardous to pedestrians than passenger cars.

Pedestrian age is another critical factor. Compared to the 16-34 age group, older adults are more likely to be involved in fatal crashes, with elderly pedestrians aged 65 and above showing a log odds increase of 1.88 (p-value < 0.001). There is no clear relationship between pedestrian sex and fatality outcomes, but White pedestrians (coef. = 0.301, p-value < 0.001) are more likely to be involved in fatal crashes compared to pedestrians of other races. Additionally, intoxicated walking (coef. = 0.643) and driving (coef. = 1.373) significantly affect fatality outcomes (p-value < 0.001).

The estimates for the IM model in Table 6 are similar for the most part, with some notable differences. The main-effect variables for lighting conditions show lower magnitudes of log odds for fatal outcomes, with an increase of 0.537 (p-value < 0.01) for dark-lighted conditions and 0.637 for dark-not-lighted conditions. In contrast, the interaction term for high-risk crashes in dark-lighted conditions has a significant log-odds increase of 0.896 (p-value < 0.001), while the interaction with dark-not-lighted conditions shows a smaller increase of 0.624. These interaction results indicate that high-risk pedestrian crashes in dark-lighted conditions are more prone to fatal outcomes than other lighting conditions. Similarly, although the main-effect coefficient for straight maneuvers at midblock locations loses significance, its interaction with high-risk crashes is significant at a 99.9 percent confidence level with a log-odds increase of 1.081.

Interestingly, while the residential area did not have a significant relationship with fatal outcomes in the BM model, in the IM model, the main effect shows a positive relationship (coefficient = 0.950), and the interaction term with high-risk crashes shows a negative relationship (coefficient = -1.262), both significant at the 0.001 level. This suggests that pedestrians have higher chances of fatal crashes in non-residential areas than in residential areas when the crash belongs to the high-risk category.

Additionally, the coefficients of main effects in the IM model for vehicle categories such as heavy vehicles (2.385 from 1.795), pickup trucks (0.677 from 0.499), and SUVs (0.701 from 0.474) have significant (p-value < 0.01) increased in magnitude, while there is no clear relationship between these vehicles when interacted with high-risk crashes, except for heavy vehicles which have a log odds decrease of -0.804 with a lower significance level of 0.1. This implies that these vehicle categories significantly impact fatal outcomes in low-risk crashes but not in high-risk crashes compared to passenger cars.

The significant log-odds increase of 0.462 for the interaction term for white pedestrians compared to non-white pedestrians suggests that white pedestrians are mostly associated with high-risk crashes. Lastly, child pedestrians, although not significantly related to fatal outcomes in the BM model, now have a positive coefficient of 0.776 (p-value < 0.01) for the main effect and a negative coefficient of -1.56 (p-value < 0.01) for the interaction term with high-risk crashes. This indicates that child pedestrians are more likely to experience fatal outcomes in low-risk crashes and less likely in high-risk crashes, compared to the base category of pedestrians aged 16-34.

Table 6. Injury severity modeling with the fatal outcome as the dependent variable

	BM		IM	
Variables	**Coef.**	**SE**	**Coef.**	**SE**
Dependent Variable: Fatal Outcome = 1				
Lighting (Base: Daylight)				
Dark - lighted	1.25***	0.105	0.537**	0.224
Dark - not lighted	1.084***	0.137	0.637**	0.289
Dawn/dusk	0.784***	0.231	0.637·	0.346
Intersection vs otherwise	-0.467***	0.103	-0.379	0.255
Straight maneuver at midblock	0.939***	0.101	0.164	0.186
Residential area vs. Otherwise	0.024	0.100	0.950***	0.208
Posted Speed Limit (Base: 20mph - 30 mph)				
15 mph and below	-0.846***	0.182	-0.593**	0.231
35 mph and 40 mph	0.792***	0.118	0.713**	0.220
45 mph and above	1.494***	0.127	0.485	0.778
Vehicle Category (Base: Passenger cars)				
Heavy Vehicles	1.795***	0.218	2.385***	0.361
Medium Heavy (e.g. Delivery vans)	0.858**	0.270	1.299**	0.436
Minivans	0.424**	0.214	-0.214	0.609
Pickup trucks	0.499***	0.120	0.677**	0.242
SUVs	0.474***	0.113	0.701**	0.239
Other/ Unknown	0.643***	0.182	0.649·	0.342
Pedestrian age (Base: 16-34)				
15 and below	-0.044	0.196	0.776**	0.326
35-49	0.597***	0.122	1.013**	0.310
50-64	1.276***	0.114	1.603***	0.293
65 and above	1.88***	0.141	2.325***	0.318
Pedestrian sex: Male vs female	0.128	0.091	0.102	0.091
Pedestrian race: White vs otherwise	0.301***	0.089	-0.071	0.186
Pedestrian walking under the influence	0.643***	0.109	0.673**	0.309
Driving Under the Influence (DUI)	1.373***	0.148	1.762***	0.278
Risk Label: High-risk vs low-risk			-0.046	0.457
High-risk × **Lighting (Base: Daylight)**				
High-risk × Dark - lighted			0.896***	0.260
High-risk × Dark - not lighted			0.624·	0.332
High-risk × Dawn/dusk			0.186	0.468
High-risk × **Intersection vs otherwise**			-0.045	0.279
High-risk × **Straight maneuver at midblock**			1.081***	0.231
High-risk × **Residential area vs. Otherwise**			-1.262***	0.248
High-risk × **Posted Speed Limit (Base: 20mph - 30 mph)**				
High-risk × *15 mph and below*			0.278	0.783
High-risk × *35 mph and 40 mph*			0.141	0.268

	BM		IM	
Variables	**Coef.**	**SE**	**Coef.**	**SE**
High-risk × 45 mph and above			1.067	0.794
High-risk × **Vehicle Category (Base: Passenger cars)**				
High-risk × Heavy Vehicles			-0.804 ·	0.450
High-risk × Medium Heavy (e.g. Delivery vans)			-0.627	0.553
High-risk × Minivans			0.768	0.652
High-risk × Pickup trucks			-0.211	0.274
High-risk × SUVs			-0.282	0.271
High-risk × Other/ Unknown			0.039	0.373
High-risk × **Ped. age (Base: 16-34)**				
High-risk × 15 and below			-1.56***	0.459
High-risk × 35-49			-0.507	0.338
High-risk × 50-64			-0.400	0.318
High-risk × 65 and above			-0.556	0.356
High-risk × **Ped. Race: White vs otherwise**			0.462**	0.207
High-risk × **Ped. Walking under influence**			-0.095	0.331
High-risk × **Driving Under the Influence (DUI)**			-0.473	0.328
Driver age, sex, and race	controlled		controlled	
Intercept	-6.347***	0.218	-6.624***	0.397
Total Degrees of Freedom	17,266		17,266	
Residual Degrees of Freedom	17,235		17,235	
Null Deviance	6,122		6,122	
Residual Deviance	4,610		4,517	
AIC value	4,674		4,631	

***p-value < 0.001, **p-value < 0.01, * p-value < 0.05, · p-value < 0.1

Supervised Learning Classification Discussions

Using insights from clustering algorithms and unsupervised learning results, we applied supervised learning to classify pedestrian crashes in Tennessee into two groups: high-risk and low-risk crashes, based on road and environmental characteristics. Through cross-tabulation classification and various logistic regression models, we identified the characteristics of high-risk crashes. The following sub-section will elaborate on these high-risk crash characteristics, while subsequent sub-sections will dive deeper into the spatial and temporal visualization of these crashes.

Characteristics of High-risk Crashes

High-risk crashes are defined as the most dangerous group of crashes for pedestrians, with the highest chance of fatality in a pedestrian crash, based on road design and environmental factors. These crashes are strongly associated with non-intersection or midblock locations on straight roads. They frequently occur in non-residential areas and during dark conditions, often in places

with some form of lighting infrastructure. They are also common on roads with multiple lanes and higher speed limits, typically above 35 mph. As these crashes usually happen at midblock locations, they generally do not involve traffic control systems. They mostly occur on two-way divided roads without traffic barriers or on two-way undivided roads and rarely involve parking lots or private properties. These road characteristics align strongly with the road characteristics of arterials in the US, as depicted by several safety studies. We did not observe an association between higher posted speed limits and fatality in high-risk crashes. This might be because crashes with higher posted speed limits are predominantly classified as high-risk crashes, leading to an imbalance in the data.

While men and women are equally represented in high-risk crashes, they predominantly involve non-white pedestrians, including Black pedestrians. The overrepresentation of racial minorities and its link with hazardous road design aligns with the findings of past studies (*Haddad et al.*, *2023; Roll & McNeil*, *2022*). Impaired pedestrians are also disproportionately represented in this type of crash. On the other hand, White drivers are more likely to be behind the steering wheel in these crashes. Conversely, we find that children are less represented in this category. A possible reason for their lower exposure is that these crash locations appear unsafe, leading parents to keep their children away from these areas. The type of vehicle involved typically is not associated with this type of crash.

Regarding pedestrian fatality, high-risk crashes are mostly associated with dark-lighted conditions, straight maneuvers of vehicles at midblock locations, and non-residential areas. Although Black pedestrians are predominantly involved in these types of crashes, fatalities are more frequently associated with White pedestrians. On the other hand, child pedestrians are not only less represented in this type of crash but also less likely to suffer fatal outcomes. This may be due to the extra precautionary measures taken to protect children, even when they are exposed to such hazardous situations. The impact of intoxication on fatality outcomes for both pedestrians and drivers shows no significant difference between high-risk and low-risk crashes.

The relationship between vehicle types and high-risk crashes presents interesting findings. When looking at overall crashes, the probability of fatality is approximately 65 percent higher when a pedestrian is struck by a pickup truck and about 61 percent higher when struck by an SUV, compared to a passenger car. However, when considering high-risk crashes and their interaction with vehicle types, the main effect reveals that the chances of fatality are 97 percent higher for pickup trucks and 101 percent higher for SUVs in low-risk crash scenarios, although such scenarios are rare. Conversely, the interaction term between vehicle type and high-risk crashes shows no clear relationship between vehicle size and fatality outcomes. One possible explanation is that vehicles tend to travel at higher speeds in areas prone to high-risk crashes. This increased speed may offset the differences in vehicle size, leading to an equal or greater transfer of kinetic energy during collision (*Ballesteros et al.*, *2004*). This confirms findings from Parajuli et al. that showed that most of increase in fatality risk in Tennessee was associated with increased fatalities involving passenger cars (*Parajuli et al. 2023).*

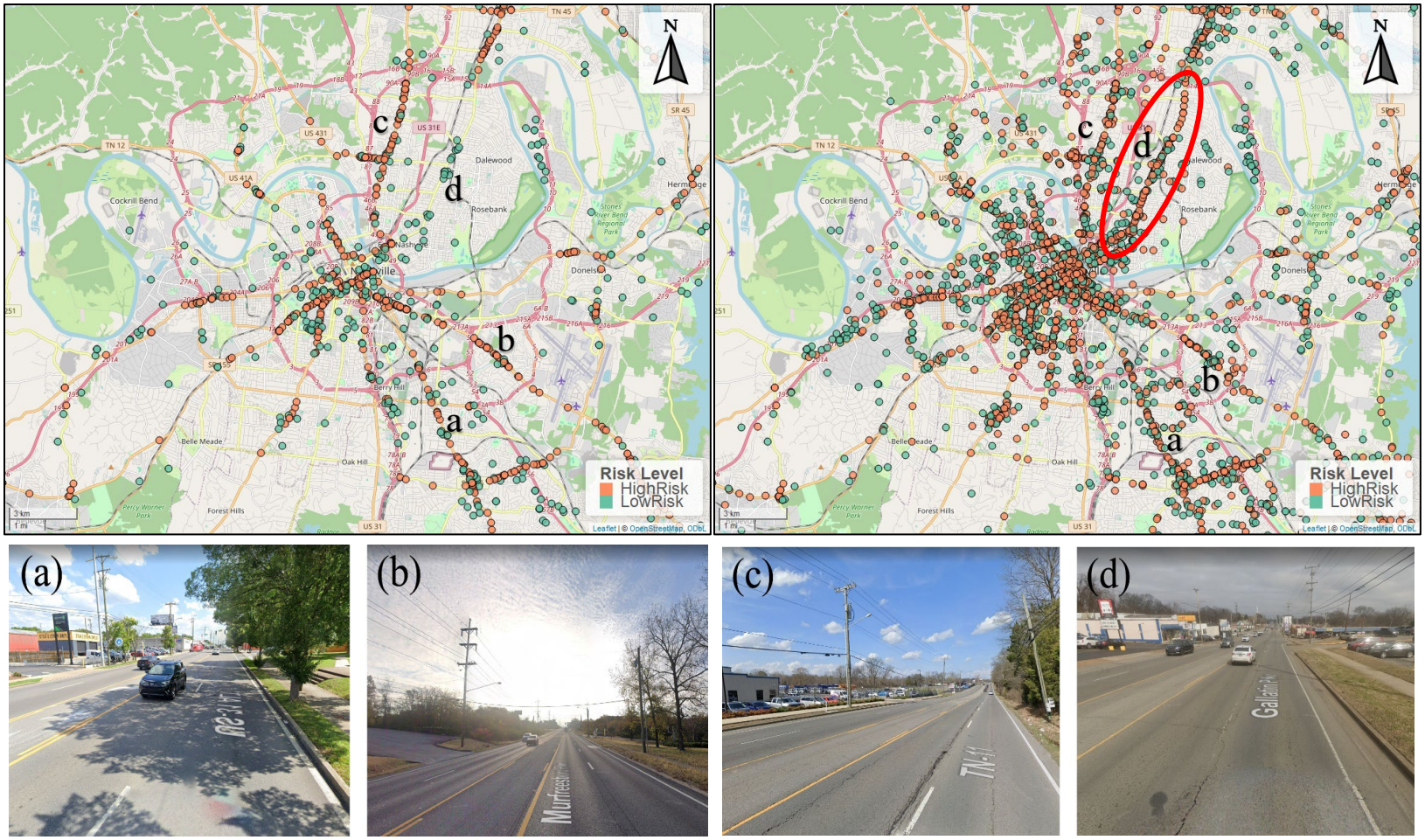


Figure 10. [Top] Pedestrian crashes in Nashville: Initial labels (top-left) and final classification labels (top-right) [Bottom] Comparison of visible road features various streets – a) Nolensville Pike, b) Murfreesboro Pike, c) Dickerson Pike, and d) Gallatin Pike, Nashville, TN

Identification of High-Risk Streets

After the identification of high-risk crashes, we can connect the dots and see the whole picture of streets across the cities that are associated with high-risk crashes. Among the 17,267 total filtered crashes in urban Tennessee from 2009 to 2019, 13,152 crashes were initially unlabeled while 4,115 crashes had initial labels. Among the crashes that had pre-labels, 2092 crashes, which were identified as high-risk crashes based on extant literature and unsupervised learning results of this study, were derived from the proximity of crash locations to streets with "major arterials" labels (within 100 ft) as depicted in the E-TRIMS map. The E-TRIMS map relies on state agencies' classification of arterials. To understand the relationship between these pre-labels and finalized classification results, we performed a spatial visualization of pedestrian crashes across the city of Nashville.

Figure 10 (top-left) is a visualization of pre-labels, with orange dots representing the "high-risk" crashes, while the green dots represent "low-risk" crashes. We can observe that these initial labels are heavily concentrated along the major urban arterials of Nashville, such as Nolensville Pike, Murfreesboro Pike, Dickerson Pike, and other major arterials. Figure 10 (top-right) illustrates the final classification of high-risk and low-risk crashes throughout the city. The new classification also visibly clusters high-risk crashes along the major arterials. However, there is a noticeable cluster of crashes along newer roads, particularly a newly formed cluster along Gallatin Pike, highlighted in red for clarity. This indicates that our methodology extends beyond traditional arterial definitions, identifying several new streets as pedestrian fatality hotspots in Nashville. Consequently, we can infer that these streets, especially Gallatin Pike, possess some arterial characteristics that are hazardous to pedestrians.

We analyzed Google Street View for these streets to investigate further, as shown in Figure 10 [Bottom]. Figure 10a, Figure 10b, and Figure 10c represent state-defined major arterials in Nashville: Nolensville Pike, Murfreesboro Pike, and Dickerson Pike, respectively. We also included Gallatin Pike in Figure 10d for a side-by-side comparison of visible road features. Referencing these four figures, we can identify the characteristics of high-risk streets for pedestrians. It should be noted that these streets are classified solely based on road design and environmental factors, excluding pedestrian, driver, and vehicle characteristics. They typically feature long, straight sections, multiple lanes, and wide roads with two-way turn lanes. The figures also reveal a lack of adequate pedestrian infrastructure, such as pedestrian crossings, signals, continuous sidewalks, and proper bus stop facilities for transit riders. Furthermore, the adjacent land use is primarily car-focused, with multiple driveways and businesses catering primarily to drivers. Gallatin Pike serves as a prominent example of a street with arterial characteristics, making it a "high-risk" street for pedestrians.

As cities experience significant population growth and suburbanization (*Tennessee State Data Center*, *2023*), we hypothesize that more streets will evolve into high-risk areas, prioritizing cars and potentially worsening pedestrian safety concerns in the state. Overall, the current functional definition of arterials is inadequate for identifying streets that are potentially dangerous for

pedestrians. Relying solely on state agencies' classification of arterials could result in the underrepresentation of high-risk crashes and overlook potential pedestrian fatality hotspots.

Trend Visualization of High-Risk Crashes

Over the years, pedestrian fatalities in Tennessee have increased significantly, while pedestrian involvement in crashes has remained relatively constant. Having identified the high-risk crashes, it would be interesting to examine the trend of these high-risk crashes over time and their relationship to the overall trend in pedestrian fatalities. To ensure a comprehensive representation, we normalized the 2019 figures based on historical trends, as our observations for 2019 only extended until the end of September. Using data from Table 7. Historical crash involvement and fatalities in Tennessee, we calculated that both crash involvement (28.9 percent) and pedestrian fatalities (29.5 percent) were slightly overrepresented from the start of October to the end of December. Crash involvement shows lower variance compared to crash fatalities. Therefore, we used 1/ (1 - 0.29) = 1.41 as the normalization parameter to roughly estimate the crash involvement and fatalities for the entirety of 2019. Figure 11 Shows the adjusted pedestrian fatalities in the state from 2009 to 2019.

Table 7. Historical crash involvement and fatalities in Tennessee

	Crash Involvement			Fatal Outcomes		
Year	**Total**	**Oct to Dec**	**%**	**Total**	**Oct to Dec**	**%**
2009	1,206	361	29.9	40	16	40.0
2010	1,252	369	29.5	52	11	21.2
2011	1,460	416	28.5	57	17	29.8
2012	1,632	435	26.7	49	13	26.5
2013	1,531	427	27.9	63	15	23.8
2014	1,557	445	28.6	63	14	22.2
2015	1,761	547	31.1	82	37	45.1
2016	1,827	491	26.9	68	19	27.9
2017	1,842	541	29.4	88	24	27.3
2018	1,825	551	30.2	101	31	30.7
2019*	1,374	*Missing*	*Missing*	79	*Missing*	*Missing*
Average			*28.9*			*29.5*
2019**	1,936	-	-	111	-	-

* Missing Data for October to December 2019; ** Estimated Figures for 2019

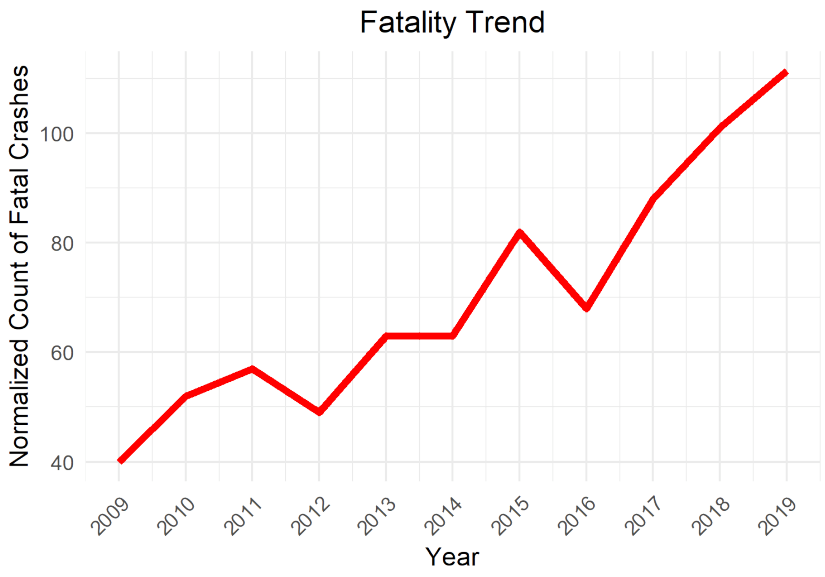


Figure 11. Pedestrian fatality trend (normalized) in Tennessee

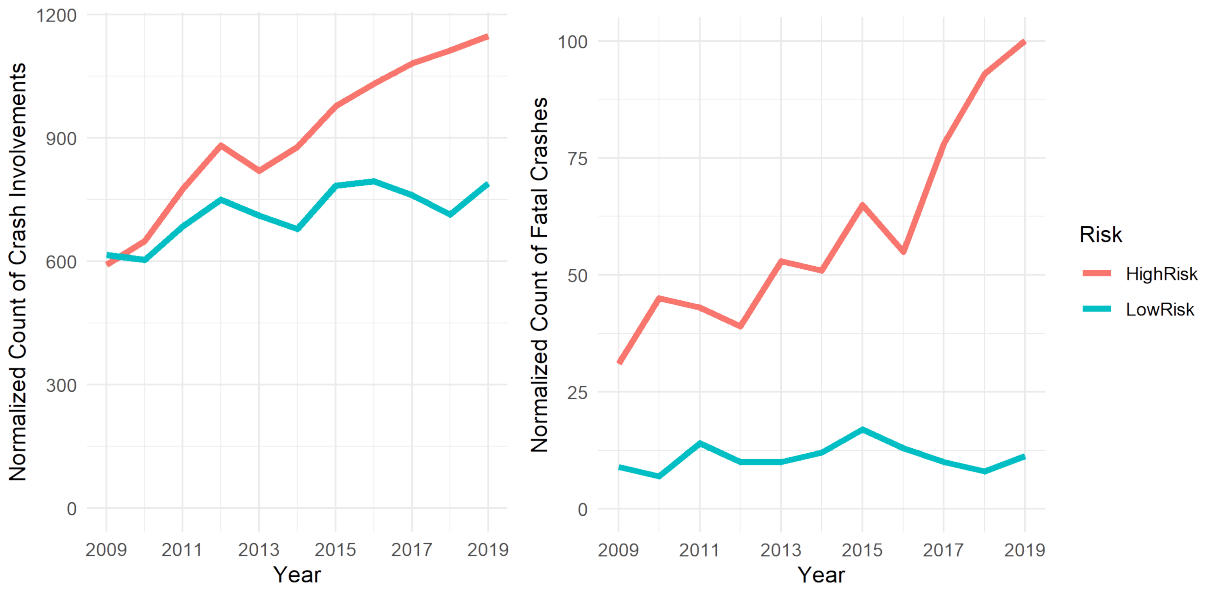


Figure 12. Pedestrian trends for a) involvements (right) and b) fatality (left)

From Figure 12, we see a clear distinction in trends between high-risk and low-risk crashes. High-risk crashes exhibit a steep increase in fatality counts (Figure 12a) and a relatively moderate rise in involvement counts (Figure 12b). These observations suggest that high-risk crashes, which are inherently more likely to result in fatalities, have been increasing over the years, leading to a significant rise in pedestrian fatalities. Consequently, high-risk crashes substantially contribute to the overall increase in the severity of pedestrian crashes in Tennessee. This finding supports the preliminary univariate findings by Parajuli et al., which emphasize road design features as a key factor in the increase of pedestrian injury severity (*Parajuli et al.*, *2023*).

Conclusions and Recommendations

Using unsupervised and supervised learning, the study identified hazardous pedestrian crash patterns in urban Tennessee based on road and situational factors. Clustering results indicated that pedestrian crashes in midblock locations on higher-speed roads were the most dangerous, while parking lot crashes were the safest. This insight led to the training and classification of pedestrian crashes into low-risk and high-risk categories using a supervised learning model, which provided valuable insights into the identification and characteristics of high-risk crashes. The study methodology could be replicated in determining fatality hotspot locations and risky road sections for pedestrians.

High-risk crashes were found to occur on wide, straight roads with higher posted speed limits, typically in non-intersection locations and during dark conditions. Further inspection revealed that these streets often lack adequate pedestrian infrastructure, such as continuous sidewalks, sufficient pedestrian signals, and closely spaced crosswalks. These streets are usually in non-residential areas and have businesses primarily catering to cars, with multiple driveways. While these streets share characteristics with arterials, our investigation shows that non-arterial streets with similar features are common in predominantly suburban areas in states like Tennessee. Consequently, relying solely on the functional classification of arterials to identify risky crashes may lead to the underrepresentation of high-risk pedestrian crashes. Moreover, design and intervention policies that are targeted at (or avoid) functionally classified arterials (e.g., traffic calming strategies) could miss important roads with elevated risk.

The understanding of high-risk crashes extended beyond road design variables to include the relationship between demographics and these risky crashes. The classification revealed that Black pedestrians are more frequently exposed to high-risk crashes compared to White pedestrians, while White drivers are more often associated with these crashes. These findings of racial disparity are consistent with existing safety literature. Additionally, intoxicated pedestrians are overrepresented in high-risk crashes, whereas intoxicated drivers are less likely to be involved. When examining injury severity, we discovered some counterintuitive results. Despite higher exposure among Black pedestrians, White pedestrians are more likely to die in high-risk crashes. This may be due to the majority demographic in Tennessee being White, which is reflected in more vulnerable populations such as the unhoused. Furthermore, child pedestrians are both less exposed and less likely to die in these crashes, suggesting significant parental awareness of the issue, particularly on risky infrastructure. However, this also highlights a concerning issue, indicating that urban areas in Tennessee may not be safe for children. The study also found that while larger vehicles like pickup trucks and SUVs significantly increase pedestrian fatality risk in general, their impact is less clear in high-risk crash scenarios, possibly due to higher travel speeds in such areas neutralizing vehicle size effects. Future research should delve deeper into these nuances to better understand the interactions between demographics, vehicle types, and high-risk crash scenarios.

Finally, trend analyses indicate that high-risk crashes have become more severe over the years, contributing to the overall increase in pedestrian fatalities in Tennessee. Suburbanization has likely

led to marginalized communities and minorities being increasingly overexposed to higher-speed suburban roads. These roads often lack adequate pedestrian infrastructure, such as continuous sidewalks and appropriately spaced crossings, which are costly to implement in sprawling suburban areas. The revelation that high-risk crashes are driving the increase in pedestrian crash severity in Tennessee underscores the urgent need for immediate pedestrian safety improvements, focusing on road design and environmental factors.

Building upon our findings, we propose actionable recommendations: Begin by identifying streets that resemble urban arterials and consider reducing speeds, while establishing a maximum speed limit of 35 mph for important arterials. Often speed limit reductions do not proportionately increase travel times (when including signal delay). Implement road diets on wide arterials to remove two-way turn lanes and strategically place signalized intersections at regular intervals near prominent businesses or transit stops. This approach shortens pedestrian crossing distances, encourages speed reduction, and provides additional U-turn opportunities. Install pedestrian refuge islands at road crossings and improve lighting and signals in high pedestrian traffic areas to enhance visibility and encourage safer decision-making. Finally, ensure the presence of frequent and well-lit pedestrian crossings, including midblock crossings, using appropriate signals to optimize visibility and pedestrian safety.

A notable limitation of this study is its reliance solely on police crash data for road and environmental characteristics. Future research should broaden its scope to include additional data sources such as satellite imagery, detailed roadway management systems, and other comprehensive data sources. This would provide more detailed information on pedestrian infrastructure and other road characteristics not captured in the crash data. Instead of automating the pre-labeling process, future studies could benefit from manually labeling crashes after thoroughly examining them, including the crash narratives, to improve initial crash labeling. Additionally, future research could explore more advanced machine learning techniques, which have the potential to yield enhanced and robust results.

References

Abaza, O. A., Arafat, M., & Chowdhury, T. D. (2018). Study on pedestrian road crossing compliance at high pedestrian crash locations of Anchorage, Alaska. International Conference on Transportation and Development 2018.

Aziz, H. A., Ukkusuri, S. V., & Hasan, S. (2013). Exploring the determinants of pedestrian–vehicle crash severity in New York City. *Accident Analysis & Prevention*, *50*, 1298-1309.

Ballesteros, M. F., Dischinger, P. C., & Langenberg, P. (2004). Pedestrian injuries and vehicle type in Maryland, 1995–1999. *Accident Analysis & Prevention*, *36*(1), 73-81.

Bellis, R., Buthe, B., Guglielmone, M., Rahman, B., & Davis, S. L. (2021). Dangerous by design 2021.

Cicchino, J. B. (2022). Effects of automatic emergency braking systems on pedestrian crash risk. *Accident Analysis & Prevention*, *172*, 106686.

Das, S., Ashraf, S., Dutta, A., & Tran, L.-N. (2020). Pedestrians under influence (PUI) crashes: Patterns from correspondence regression analysis. *Journal of safety research*, *75*, 14-23.

Das, S., & Sun, X. (2015). Factor association with multiple correspondence analysis in vehicle–pedestrian crashes. *Transportation Research Record*, *2519*(1), 95-103.

Davis, G. A. (2001). Relating severity of pedestrian injury to impact speed in vehicle-pedestrian crashes: Simple threshold model. *Transportation Research Record*, *1773*(1), 108-113.

Dultz, L. A., & Frangos, S. G. (2013). The impact of alcohol in pedestrian trauma. *Trauma*, *15*(1), 64-75.

Elvik, R., Christensen, P., & Amundsen, A. H. (2004). *Speed and road accidents: an evaluation of the Power Model*. Transportøkonomisk Institutt.

European Transport Safety Council. (2020). *Urgent action needed to tackle deaths of pedestrians and cyclists* https://etsc.eu/urgent-action-needed-to-tackle-deaths-of-pedestrians-and-cyclists/

Ewing, R., Schieber, R. A., & Zegeer, C. V. (2003). Urban sprawl as a risk factor in motor vehicle occupant and pedestrian fatalities. *American journal of public health*, *93*(9), 1541-1545.

Federal Highway Administration. KABCO Injury Classification Scale and Definitions. In.

Federal Highway Administration. (2023). *Highway Functional Classification Concepts, Criteria and Procedures*. https://www.fhwa.dot.gov/planning/processes/statewide/related/hwy-functional-classification-2023.pdf

Ferenchak, N. N., & Abadi, M. G. (2021). Nighttime pedestrian fatalities: A comprehensive examination of infrastructure, user, vehicle, and situational factors. *Journal of safety research*, *79*, 14-25.

Gardner, M. W., & Dorling, S. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric environment*, *32*(14-15), 2627-2636.

Goodman, D., Hillman, T., Ciabotti, J., & Gelinne, D. (2022). *Arterial roads and pedestrian safety: a resource for state, metropolitan planning organization, and local planners and engineers*. https://www.pedbikeinfo.org/resources/resources_details.cfm?id=5358

Haddad, A. J., Mondal, A., Bhat, C. R., Zhang, A., Liao, M. C., Macias, L. J., Lee, M. K., & Watkins, S. C. (2023). Pedestrian crash frequency: Unpacking the effects of contributing factors and racial disparities. *Accident Analysis & Prevention*, *182*, 106954.

Hezaveh, A. M., & Cherry, C. R. (2018). Walking under the influence of the alcohol: A case study of pedestrian crashes in Tennessee. *Accident Analysis & Prevention*, *121*, 64-70.

Hossain, A., Sun, X., Thapa, R., & Codjoe, J. (2022). Applying association rules mining to investigate pedestrian fatal and injury crash patterns under different lighting conditions. *Transportation Research Record*, *2676*(6), 659-672.

Hu, W., & Cicchino, J. B. (2018). An examination of the increases in pedestrian motor-vehicle crash fatalities during 2009–2016. *Journal of safety research*, *67*, 37-44.

Islam, M. (2023). An exploratory analysis of the effects of speed limits on pedestrian injury severities in vehicle-pedestrian crashes. *Journal of Transport & Health*, *28*, 101561.

Jaccard, P. (1912). The distribution of the flora in the alpine zone. 1. *New phytologist*, *11*(2), 37-50.

Keller, C. G., Dang, T., Fritz, H., Joos, A., Rabe, C., & Gavrila, D. M. (2011). Active pedestrian safety by automatic braking and evasive steering. *IEEE Transactions on Intelligent Transportation Systems*, *12*(4), 1292-1304.

Kim, J.-K., Ulfarsson, G. F., Shankar, V. N., & Kim, S. (2008). Age and pedestrian injury severity in motor-vehicle crashes: A heteroskedastic logit analysis. *Accident Analysis & Prevention*, *40*(5), 1695-1702.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of statistical software*, *28*, 1-26.

Li, D., Ranjitkar, P., Zhao, Y., Yi, H., & Rashidi, S. (2017). Analyzing pedestrian crash injury severity under different weather conditions. *Traffic injury prevention*, *18*(4), 427-430. https://doi.org/10.1080/15389588.2016.1207762

Linzer, D. A., & Lewis, J. B. (2011). poLCA: An R package for polytomous variable latent class analysis. *Journal of statistical software*, *42*, 1-29.

Long Jr, B., & Ferenchak, N. N. (2021). Spatial equity analysis of nighttime pedestrian safety: role of land use and alcohol establishments in Albuquerque, NM. *Transportation research record*, *2675*(12), 622-634.

Macek, K. (2023). *Pedestrian Traffic Fatalities by State: 2022 Preliminary Data* (Spotlight on Highway Safety, Issue.

Mansfield, T. J., Peck, D., Morgan, D., McCann, B., & Teicher, P. (2018). The effects of roadway and built environment characteristics on pedestrian fatality risk: A national assessment at the neighborhood scale. *Accident Analysis & Prevention*, *121*, 166-176.

Nabavi Niaki, M. S., Fu, T., Saunier, N., Miranda-Moreno, L. F., Amador, L., & Bruneau, J.-F. (2016). Road lighting effects on bicycle and pedestrian accident frequency: Case study in Montreal, Quebec, Canada. *Transportation research record*, *2555*(1), 86-94.

NHTSA. (2017). MMUCC Guideline: Model Minimum Uniform Crash Criteria. In: National Highway Traffic Safety Administration, US Department of Transportation.

Noland, R. B., Klein, N. J., & Tulach, N. K. (2013). Do lower income areas have more pedestrian casualties? *Accident Analysis & Prevention*, *59*, 337-345. /https://doi.org/10.1016/j.aap.2013.06.009

Parajuli, S., Cherry, C. R., Zavisca, E., & Rogers III, W. (2023). Are Pedestrian Crashes Becoming More Severe? A Breakdown of Pedestrian Crashes in Urban Tennessee. *Transportation Research Record*, 03611981231198475.

Prato, C. G., Kaplan, S., Patrier, A., & Rasmussen, T. K. (2018). Considering built environment and spatial correlation in modeling pedestrian injury severity. *Traffic injury prevention*, *19*(1), 88-93.

Rab, M. A., Qi, Y., & Fries, R. N. (2018). *Comparison of Contributing Factors to Pedestrian Crossing Crash Severity at Locations with Different Controls in Illinois*.

Ripley, B., Venables, W., & Ripley, M. B. (2016). Package 'nnet'. *R package version*, *7*(3-12), 700.

Roll, J., & McNeil, N. (2022). Race and income disparities in pedestrian injuries: Factors influencing pedestrian safety inequity. *Transportation research part D: transport and environment*, *107*, 103294.

Salon, D., & McIntyre, A. (2018). Determinants of pedestrian and bicyclist crash severity by party at fault in San Francisco, CA. *Accident Analysis & Prevention*, *110*, 149-160.

Sanders, R. L., & Schneider, R. J. (2022). An exploration of pedestrian fatalities by race in the United States. *Transportation research part D: transport and environment*, *107*, 103298.

Sanders, R. L., Schneider, R. J., & Proulx, F. R. (2022). Pedestrian fatalities in darkness: What do we know, and what can be done? *Transport policy*, *120*, 23-39.

Schmitt, A. (2020). *Right of Way: Race, Class, and the Silent Epidemic of Pedestrian Deaths in America*. Island Press.

Schneider, R. J. (2020). United States pedestrian fatality trends, 1977 to 2016. *Transportation Research Record*, *2674*(9), 1069-1083.

Schneider, R. J., Proulx, F. R., Sanders, R. L., & Moayyed, H. (2021). United States fatal pedestrian crash hot spot locations and characteristics. *Journal of transport and land use*, *14*(1), 1-23.

Sun, M., Sun, X., & Shan, D. (2019). Pedestrian crash analysis with latent class clustering method. *Accident Analysis & Prevention*, *124*, 50-57.

Tefft, B. C. (2013). Impact speed and a pedestrian's risk of severe injury or death. *Accident Analysis & Prevention*, *50*, 871-878.

Tefft, B. C., Arnold, L. S., & Horrey, W. J. (2021). *Examining the Increase in Pedestrian Fatalities in the United States, 2009-2018 (Research Brief)* [Research Brief].

Tennessee Highway Safety Office. (2021). *Tennessee Integrated Traffic Analysis Network*.

Tennessee State Data Center. (2023). *2022 Population Estimates Show Big Cities on the Move in Tennessee*. https://tnsdc.utk.edu/2023/05/18/2022-population-estimates-show-big-cities-on-the-move-in-tennessee/

Tyndall, J. (2021). Pedestrian deaths and large vehicles. *Economics of Transportation*, *26-27*, 100219. https://doi.org/10.1016/j.ecotra.2021.100219

UK Department for Transport. (2020). *Reported road casualties in Great Britain: 2019 annual report*.

Vellimana, M., & Kockelman, K. (2023). Darkness and Death in the US: Walking Distances Across the Nation by Time of Day and Time of Year. *Findings*.

Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, *58*(301), 236-244.

Zajac, S. S., & Ivan, J. N. (2003). Factors influencing injury severity of motor vehicle–crossing pedestrian crashes in rural Connecticut. *Accident Analysis & Prevention*, *35*(3), 369-379. https://doi.org/10.1016/S0001-4575(02)00013-1