CENTER FOR CONNECTED AND AUTOMATED TRANSPORTATION

Final Report #61B
November 2023

Translation of Driver-Pedestrian Behavioral Models at Semi-Controlled Crosswalks into a Quantitative Framework for Practical Self-Driving Vehicle Applications, Part B (Pedestrian Volume Analytics)

Abdullah J. Nafakh
Isaiah C. Mwamba
Sikai Chen
Jon D. Fricker
Samuel Labi

PURDUE UNIVERSITY

CENTER FOR CONNECTED AND AUTOMATED TRANSPORTATION

Report No. 61B **Report Date: November 2023**
Project Start Date: January 1, 2022
Project End Date: December 31, 2022

Translation of Driver-Pedestrian Behavioral Models at Semi-Controlled Crosswalks into a Quantitative Framework for Practical Self-Driving Vehicle Applications – Part B (Pedestrian Volume Analytics)

Abdullah J. Nafakh
Graduate Student

Isaiah C. Mwamba
Graduate Student

Sikai Chen
Visiting Professor

Jon D. Fricker
Professor

Samuel Labi
Professor

Purdue University

ACKNOWLEDGEMENTS AND DISCLAIMER

Funding for this research was provided by the Center for Connected and Automated Transportation under Grant No. 69A3551747105 of the U.S. Department of Transportation, Office of the Assistant Secretary for Research and Technology (OST-R), University Transportation Centers Program. The technical advice provided by Professors Darcy M. Bullock and Sarah M. Hubbard of Purdue University and Professor Sikai Chen of the University of Wisconsin-Madison, are greatly appreciated. The institutions and professors that provided valuable in-kind cost share or other external support for this project, including the Indiana Department of Transportation (Drs. Barry Poturalski and Samy Noureldin), India Institute of Technology Tirupati (Professor Krishna Prapoorna), King Khalid University (Professor Saeed Alqadhi), ChangAn University (Professor Qiang Bai), South East University (Professor Jianchuan Cheng), Hong Kong Polytechnic University (Professor Anthony Chen), and other institutions are also acknowledged. The assistance of graduate students Mohammadhosein Pourgholamali, Jiqian Dong, Richard Ajagu and others are appreciated. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

Suggested APA Format Citation:

Nafakh, A.J., Mwamba, I.C., Chen, S., Fricker, J.D., Labi, S. (2023). Translation of Driver-Pedestrian Behavioral Models at Semi-Controlled Crosswalks into a Quantitative Framework for Practical Self-Driving Vehicle Applications – Part B (Pedestrian Volume Analytics), Technical Report Nr. 61B, Center for Connected and Automated Transportation, Purdue University, West Lafayette, IN.

Contacts

For more information

Samuel Labi, Ph.D.	CCAT
Purdue University	University of Michigan Transportation Research Institute
550 Stadium Mall Drive	2901 Baxter Road
HAMP G175A	Ann Arbor, MI 48152
Phone: (765) 494-5926	uumtri-ccat@umich.edu
Email: labi@purdue.edu	(734) 763-2498

Technical Report Documentation Page

1. Report No. 61B	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle Translation of Driver-Pedestrian Behavioral Models at Semi-Controlled Crosswalks into a Quantitative Framework for Practical Self-Driving Vehicle Applications, Part B (Pedestrian Volume Analytics)		**5. Report Date:** November 2023 **6. Performing Organization Code** N/A
7. Author(s) Abdullah J. Nafakh, Isaiah C. Mwamba, Sikai Chen, Jon D. Fricker, Samuel Labi		**8. Performing Organization Report No. N/A**
9. Performing Organization Name and Address Center for Connected and Automated Transportation, Purdue University, 550 Stadium Mall Drive, W. Lafayette, IN 47907; and University of Michigan Ann Arbor, 2901 Baxter Rd, Ann Arbor, MI 48109		**10. Work Unit No.** **11. Contract or Grant No.** Contract No. 69A3551747105
12. Sponsoring Agency Name and Address U.S. Dept of Transportation, Office of the Assistant Secretary for Research & Technology, 1200 New Jersey Ave. SE, Washington, DC 20590		**13. Type of Report and Period Covered** Final report, Jan 2022 – Dec 2022 **14. Sponsoring Agency Code:** OST-R
15. Supplementary Notes Conducted under the U.S. DOT Office of the Assistant Secretary for Research and Technology's (OST-R) University Transportation Centers (UTC) program.		

16. Abstract

Widespread introduction of cameras installed for monitoring vehicle flow at intersections offers opportunities to leverage this infrastructure to acquire insights into the patterns and trends of pedestrian activities at these locations. This can serve as a valuable data source for both human-driven vehicles (HDVs) and connected and autonomous vehicle (CAV) operations. Data from such equipment help establish pedestrian movement performance and timing thresholds, thereby addressing a gap in the literature. The study leverages data from signalized intersection cameras to (1) prescribe durations for pedestrian walk-interval based on pedestrian volume and geometric features of the intersection, (2) investigate the factors that influence pedestrian demand patterns, and (3) predict pedestrian volumes and tie it to signal timing, to enhance service for all roadway users. The first part of the study provides guidance for walk time interval selection and presents four timing categories ranging from negligible to high volume and prescribes pedestrian walk interval time durations (based on the demand per cycle, storage area for pedestrians, and offset of the pedestrian push-button from the crosswalk). The second part of the study describes scalable techniques for pedestrian movement predictions. Time series correlation and cross-correlation analysis helped provide a comprehensive understanding of the factors affecting pedestrian volumes at intersections. Using the concept of intersection quadrants, the third part of the study presented methods to estimate pedestrian volumes at 15-minute intervals and connected this information to signal timing designs. Machine learning random forest and XGBoost classification models were trained on a large dataset of pedestrian counts. The study results showed that the developed models accurately predict pedestrian volumes per 15-minute intervals for each quadrant of an intersection, with a high degree of precision and a prediction accuracy of 82.3%. Signal timing optimization based on predicted pedestrian volume can significantly improve pedestrian mobility and maximize traffic flow. Overall, the findings of this study provide valuable insights for traffic engineers and planners interested in developing and deploying dynamic pedestrian signal timing systems that are useful in both manual driving autonomous driving environments. The models and data developed from this study can be helpful in efforts to address equitable transportation for the benefit of vulnerable road users specifically, in the current HDV traffic environment, and particularly in the prospective future era of CAVs.

17. Key Words Pedestrians, Intersections, Vulnerable road users, Pedestrian walk interval, Machine learning, Pedestrian data analytics.		**18. Distribution Statement** No restrictions.	
19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 69	**22. Price**

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

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 GENERAL INTRODUCTION

1.1 General Background

Historically, transportation networks have been designed primarily to prioritize vehicular traffic over other modes (pedestrians and cyclists, for example). However, the growing concerns about safety, sustainability, and equity, as highlighted in the recent highway bill (the Infrastructure Investment and Jobs Act), underscores the increasing importance of considering the needs of pedestrians as well. Pedestrians are a vital component of any transportation network, and their mobility needs must be considered to create safe, accessible, and livable cities. Accurate data on pedestrian demand are critical for transportation planning, contributing to improving safety and accessibility. However, it has been difficult to collect data on pedestrians in a systematic manner. Fortunately, infrastructure such as cameras at signalized intersections can be leveraged to collect reliable data on pedestrian demand with minimal effort.

At the intersection level, cameras can provide valuable data on pedestrian activity, including the number of pedestrians crossing at different times of day, the direction of pedestrian movement, and the duration of pedestrian phases. This data can help transportation planners in identifying areas lacking pedestrian infrastructure or where improvements in pedestrian facility would have the greatest impact. It can also inform decisions regarding the placement and timing of pedestrian signals, crosswalks, and other pedestrian facilities.

At the road network level, cameras can be used to track pedestrian movement across multiple intersections, providing insights into how pedestrians navigate the transportation system in general. This information could help assess the impact magnitude of special events on pedestrian activities, identify gaps in the pedestrian network, pinpoint areas with high pedestrian activity but insufficient infrastructure, and to determine areas where physical or policy improvements could be most impactful.

Furthermore, detailed pedestrian information will enhance the accuracy of forecasts for pedestrian demand, even at the most precise geographical granularity levels. Increased reliability will allow for more appropriate timing of pedestrian phases in a dynamic manner, and ultimately, reducing delay for both pedestrian and vehicles. Additionally, reliable forecasts of pedestrian demand can be useful information in the emerging era of connected and automated vehicles (CAVs). CAV awareness of expected pedestrian volumes and patterns at downstream routes can help improve AV route planning and optimization, specifically, reducing travel time by avoiding areas of projected high pedestrian traffic. This could also reduce pedestrian vulnerability and enhance overall road safety.

Overall, leveraging existing monitoring infrastructure, such as cameras at signalized intersections, could provide transportation planners and managers with systematic and accurate data on pedestrian demand. By understanding pedestrian movement patterns at both the intersection and network levels, transportation planners can design and prioritize improvements that will make walking safer, more accessible, and more convenient for all road users. In sum, by collecting data from multiple cameras, transportation planners can assess the overall quality of the pedestrian network, provide data for *ex ante* improvements to pedestrian and pedestrian-related infrastructure and for *ex poste* tracking of the progress of such improvements.

1.2 Motivation and Study Objectives

Pedestrian safety and mobility are critical components of urban transportation systems. However, pedestrian data are often scarce or unreliable, making it difficult for transportation agencies to accurately measure pedestrian activity and behavior, and make informed decisions about pedestrian phase timing. To address this issue, accurate and reliable data on pedestrian behavior and activity are needed. Automated camera counting approaches provide a promising solution for collecting large-scale data on pedestrian activity, allowing transportation agencies to better understand pedestrian behavior, measure pedestrian volumes, and optimize pedestrian phase timing to improve safety and efficiency at intersections.

The collection of large-scale data on pedestrian activity using automated camera counting approaches can provide valuable insights into pedestrian behavior and activity, such as pedestrian volumes, pedestrian flow patterns, and pedestrian crossing behavior. This data can help optimize signal timings, improve pedestrian safety, and prioritize pedestrian needs, ultimately leading to a more sustainable and equitable transportation system. Moreover, automated camera counting approaches can provide transportation agencies with a cost-effective and efficient method of collecting pedestrian data. Automating the data collection process enables the transportation agency to collect and collate large amounts of data over extended periods, eliminating the need for manual data collection and reducing labor costs while improving data accuracy and reliability.

Therefore, the objective of this study is to collect large-scale data on pedestrian activities to highlight the importance of pedestrian data in transportation planning and design. This study also seeks to demonstrate the potential benefits of using automated camera-counting approaches to collect this data. By providing transportation agencies with more accurate and reliable data on pedestrian behavior and activity, the safety and mobility of pedestrians can be improved, and transportation systems that are more sustainable, equitable, and efficient for all users, can be developed.

1.3 Scope and Organization

This study developed methodologies that leverage large-scale data from signalized intersection cameras to provide better insights to agencies regarding pedestrian dynamics at both the micro and network levels and overall network performance. The contents of this report are organized in chapters as follows:

- A framework for quantifying the pedestrian walk-interval is established and connected to factors influencing the pedestrian start-up time, such as pedestrian volume and geometric features of the built environment (Chapter 2).
- Large-scale data on pedestrian activities is used to explain variations in volumes in relation to surrounding events using time series and correlation analysis assessing and quantifying the impact certain factors have on pedestrian demand (Chapter 3).
- A data-drive machine learning approach is established to quantify the needed pedestrian walk-interval at any given time during the day per intersection quadrant, potentially enabling a move from fixed to dynamic pedestrian phasing (Chapter 4).
- Chapter 5 presents the overall conclusions from the research in this report and the findings from each chapter.

CHAPTER 2 QUANTIFYING THE PEDESTRIAN WALK INTERVAL

2.1 Introduction

In designing and planning pedestrian facilities, roadway engineers seek to create a safe and comfortable pedestrian environment. To achieve this goal, they consider several factors that influence pedestrian movements. These factors include pedestrian behavior, street design, pedestrian safety, and accessibility. At signalized intersections, the designers of pedestrian timing phases strive to consider pedestrian behavior and the associated dynamics [1].

The pedestrian phase, during which the right-of-way is given solely to pedestrians, consists of two intervals:
(1) *Walk interval* typically begins with the adjacent vehicular through-movement green interval and is designed to permit pedestrians to move from the curb ramp into the crosswalk.
(2) *Pedestrian Clearance*, also referred to as Flashing Don't Walk (FDW) or change interval. follows the walk interval and informs pedestrians to either complete their crossing if already in the intersection or wait till the next cycle is displayed. Finally, the pedestrian phase ends with the solid *Don't Cross* signal.

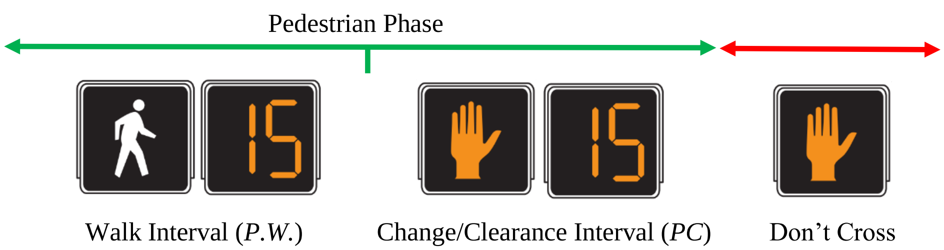


Figure 1. Intervals of the Pedestrian Phase [2]

The duration of the pedestrian phase, seen in Figure 1 (Walk interval + Clearance interval), is calculated using the following equation:

$$G_p = PW + PC$$

Where:
G_p is the green interval duration needed for the pedestrian crossing time.
PW is the walk interval duration. The MUTCD (Manual on Uniform Traffic Control Devices) indicates that the minimum walk duration should be at least 7 seconds but states that a duration as low as 4 seconds may be used if pedestrian volumes are low. The Traffic Signal Operations Handbook suggests using the walk values listed in Figure 2 and Table 1 but does not provide corresponding quantitative values for pedestrian volume.
PC is the clearance/change interval duration. The duration of this interval is computed as the crossing distance divided by the walking speed. The MUTCD recommends a value of 4.0 feet per second (ft/s) walking speed. The Americans with Disabilities Act (ADA) Accessibility Guidelines

for Buildings and Facilities recommended using 3.0 ft/s. Recent work completed by LaPlante and Kaeser has suggested a speed of 3.5 ft/s [2, 3].

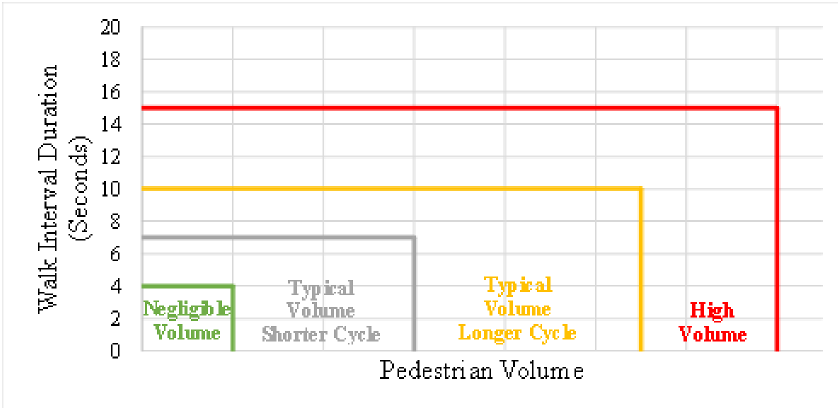


Figure 2. Pedestrian Walk Interval Category Visualization

Table 1. Pedestrian Walk Interval Duration Categories [3]

Conditions	**Walk Interval Duration (*PW*), s**
High pedestrian volume areas	15
Typical pedestrian volume and longer cycle length	10
Typical pedestrian volume and shorter cycle length	7
Negligible pedestrian volume	4

Pedestrian speeds and the clearance interval have been extensively studied in the literature and, consequently, well defined in designers' guidebooks [2, 4-7]. However, little is known about the factors that influence the pedestrian start-up time. As a result, the walk interval guidelines, seen in Table 1, are qualitative rather than quantitative. Studies investigating pedestrian dynamics (i.e., walking speed and start-up time) have considered factors such as pedestrian age and found that, on average, the movement of pedestrians above the age of 65 differs from that of younger pedestrians [8-11]. Other studies considered gender and roadway geometrics such as street width, speed limits, curb height, the number of travel lanes, and traffic cycle length [8, 12, 13]. All of these can be assumed to influence walking speed to a greater extent compared to start-up time.

It has been recommended that the walk interval (P.W.) should be designed to accommodate pedestrians' perception-reaction delay and walking time to the crosswalk. There exist various factors that could result in delaying a pedestrian in accomplishing this task. The social force model is widely used in defining the factors influencing pedestrian dynamics (i.e., avoiding obstacles and keeping a comfort zone away from other pedestrians). Such factors/forces cause a pedestrian to take some time to exit the curb onto the crosswalk just after the walk interval signal is activated [14, 15]. In terms of signal timing, the collective behavior of pedestrians is an important consideration and should be accounted for in walk-interval timing. The walk interval should provide enough time to allow all waiting pedestrians to move onto the crosswalk from the onset of the walk signal.

2.1.1 Motivation and Objectives

There is a gap in the literature regarding quantitative values for pedestrian demand that should be used to select pedestrian walk times. Similarly, the literature does not provide guidance on how other factors such as pedestrian storage areas and distance to pedestrian push-buttons influence the selection of walk times.

This study reports on the observation of pedestrian start-up time and proposes a quantitative model for designers to specify the appropriate walk interval. Specifically, this report seeks to add values to Figure 2 as to determine how many pedestrians are considered "negligible volume" and can be accommodated by the 4-second minimum time; how many pedestrians are considered "typical volume" and require 7 to 10 seconds; and how many pedestrians are considered "high volume" and require 10 to 15 seconds or longer. In addition to examining pedestrian demand, this study examines the impact of storage areas and pedestrian push-button location on the pedestrian start-up time.

With adequate understanding of pedestrian demand and the geometric features that influence the pedestrian start-up time (and consequently, the selected walk-interval), signal designers will be placed in a better position to provide optimal timing decisions that minimize both pedestrian and vehicle delay.

2.2 Methods

Using video footage from 12 signalized intersection cameras collected between late 2021 and early 2022 in the City of West Lafayette, Indiana, this study examined 1,500 observations of pedestrian start-up time. Figure 3 and Table 2 present the locations of the 12 intersections. The data were extracted from videos recorded using 12 cameras mounted on the traffic light mast arms. Installed cameras recorded the intersections continuously from the day of installation. Video imagery provides a 360-degree view of all intersection approaches and curb ramps (Figure 4). During each cycle, videos were analyzed in terms of start-up time. Start-up time is the duration needed for a waiting pedestrian, or a group of pedestrians, to clear the curb into the crosswalk after the Walk Interval is activated.

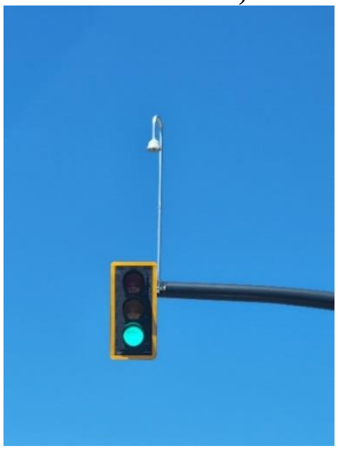
Figure 5 illustrates the visual observation process used in this study to record pedestrian start-up times. In addition, each intersection observation was analyzed in terms of the total number of pedestrians waiting per quadrant, the available storage area for pedestrians per quadrant (curb ramp area), and the distance from the pedestrian push-button to the crosswalk. Figure 6 below presents images depicting some examples of data collected, that potentially represents the explanatory variables for the model. Then a set of statistical regression models was developed to explain the variability in pedestrian start-up time (Y) given the various factors – pedestrian demand in terms of the number of pedestrians per cycle per quadrant (X_1), available storage area (X_2), and distance from the pedestrian push-button to the crosswalk (X_3).

Figure 3. Campus Intersections used to Collect Pedestrian Start-up Time

Table 2. Campus Intersections used to Collect Pedestrian Start-up Time

Intersection		**Location**	
		Longitude	***Latitude***
1	Roebuck Drive and State Street	40.42127	–86.90193
2	State Street and South River Road	40.42180	–86.90425
3	State Street and Chauncey Avenue	40.42333	–86.90693
4	Northwestern Avenue and State Street	40.42400	–86.90820
5	State Street and Andrew Place	40.42403	–86.90925
6	South Grant Street and State Street	40.42399	–86.91034
7	State Street and University Street	40.42424	–86.91689
8	State Street and S. Martin Jischke Drive	40.42423	–86.92170
9	State Street and Airport Road	40.42413	–86.93025
10	South Chauncey Avenue and West Wood Street	40.42197	–86.90766
11	University Street and 3rd Street	40.42724	–86.91664
12	West Stadium Avenue and University Street	40.43132	–86.91680

a) Intersection view with camera location noted



b) Close-up view of camera mounting location

c) Camera view

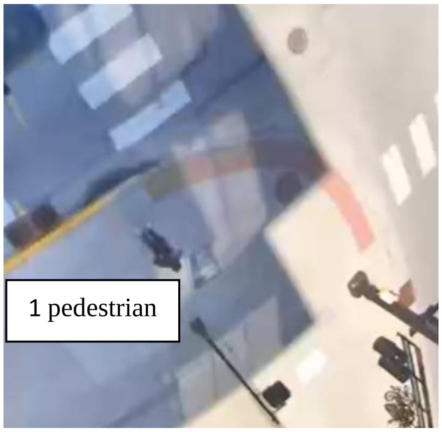
Figure 4. Camera Setup and View at W. Stadium Ave and University St. (#11)

a) Pedestrians waiting for the Walk Interval

b) Pedestrian Walk Interval active (t=0, start the timer)

c) Last waiting pedestrian clears (t=12.5 s, stop the timer)

Figure 5. Pedestrian Start-up Time Observation Process

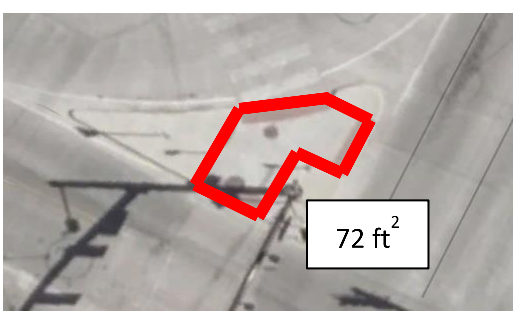


(a) Low ped volume

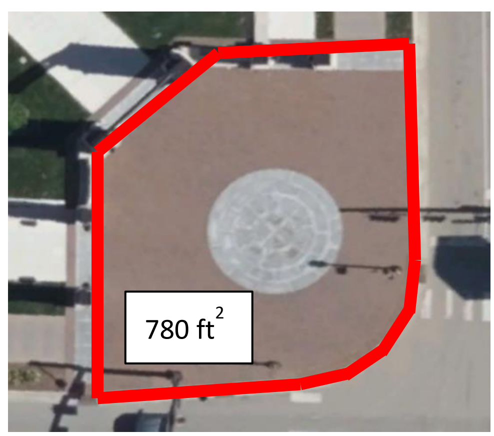
(b) High ped volume

(c) Close pedestrian push-button

(d) Far pedestrian push-button



(e) Small storage area [15]

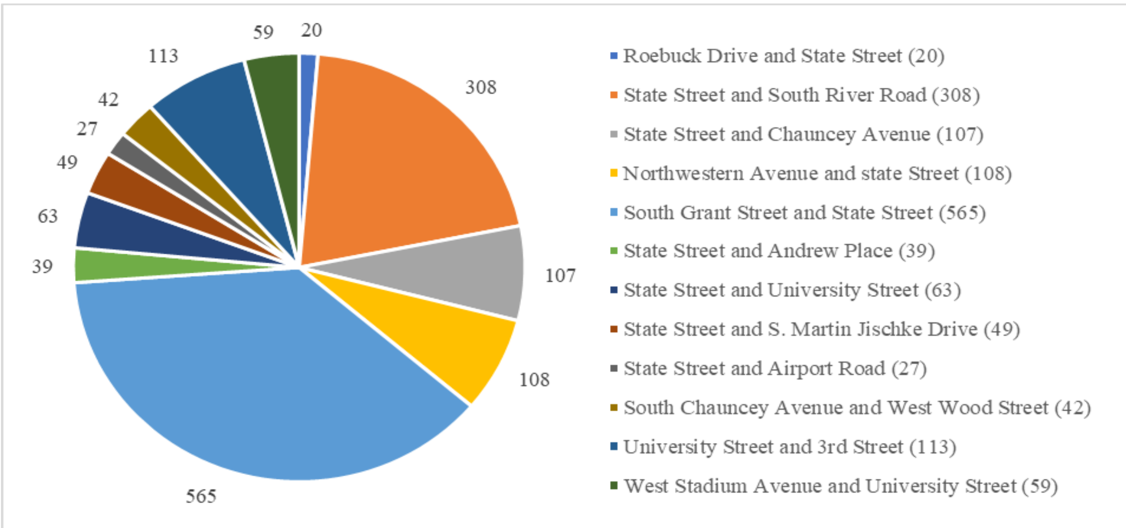


(f) Large storage area [15]

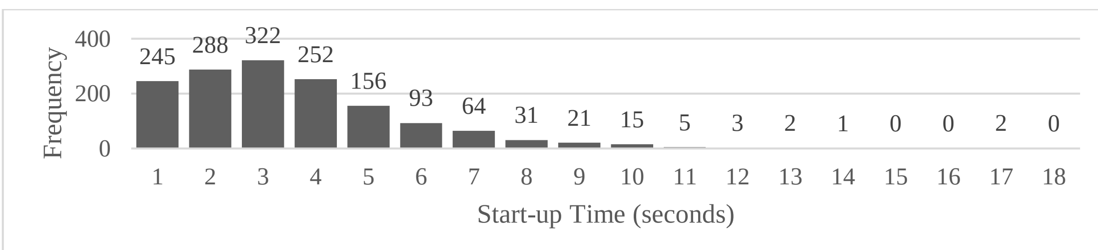
Figure 6. Example Variables that could Potentially Impact the Pedestrian Start-up Time

2.3 Summary Statistics of the Data

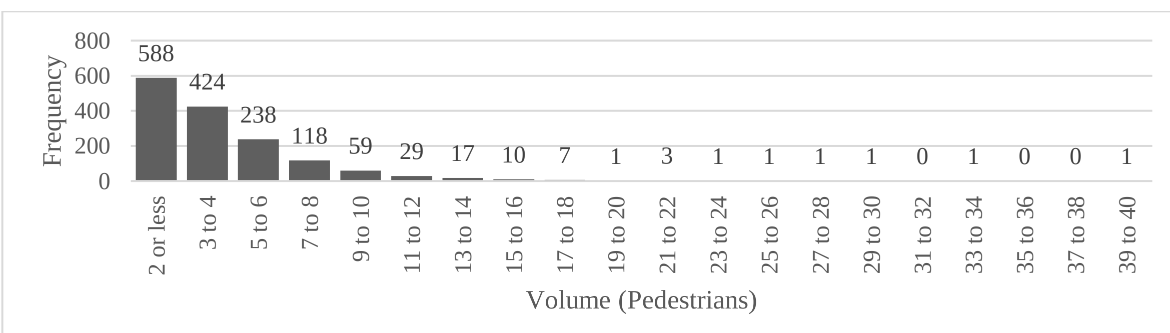
Most of the 1,500 observations were from intersections with heavy pedestrian traffic. Figure 7-a presents the distribution of start-up time observations per intersection. From the data collected, the average pedestrian start-up time was estimated at 4.05 seconds with a standard deviation of 2.17 seconds. The average pedestrian volume was 4.03, with a standard deviation of 3.58. Figures 7-b and 7-c present the observed frequencies of pedestrian start-up time and pedestrian volume.



(a) Observations per Intersection



(b) Start-up Time Frequency



(c) Ped Volume Frequency

Figure 7. Pedestrian Start-up Time Data Observations

2.4 Results

Current guidelines for determining the duration of the pedestrian walk interval, presented in Table 1, categorize the time needed into three categories: (1) “negligible volume” and require 4 seconds, (2) “typical volume” and require 7 to 10 seconds, and (3) “high volume” and require 10 to 15 seconds [2, 3]. Figure 8 and Table 3 below present the descriptive statistics of the study’s observations within these categories. The relation between pedestrian start-up time and the explanatory variables was nearly linear. Therefore, multinomial linear regression was used to explain the variability in the response variable *y* (the start-up time). Three models were built, and their details are listed in Table 4.

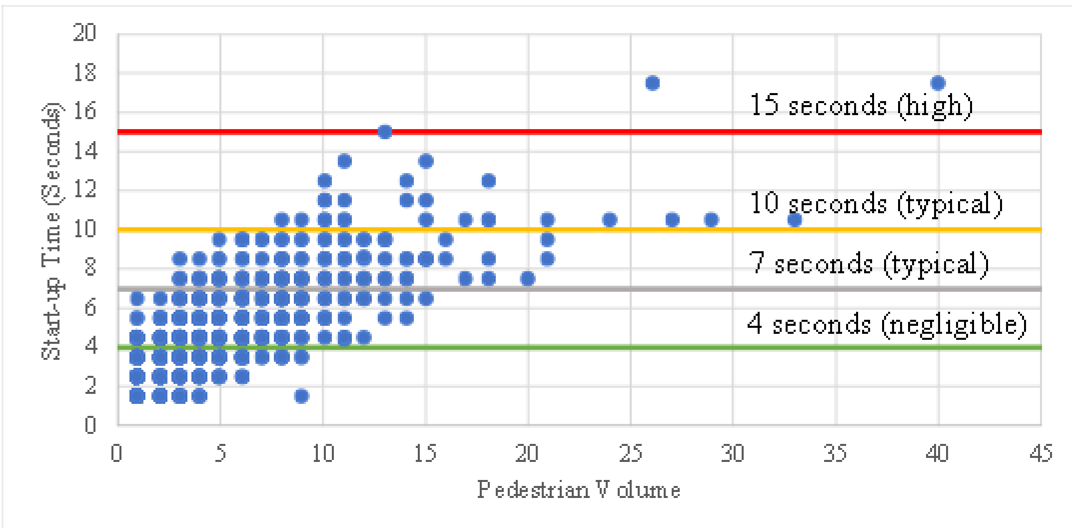


Figure 8. Start-up Time to Pedestrian Volume Relation

Table 3. Pedestrian Walk Interval Start-up Time Observation Statistics

Start-up Time	**Pedestrian Volume**								
	Obs.	**Avg**	**Min**	**Max**	**Std.**	**Percentile**			
						25th	**50th**	**75th**	**90th**
1-4 s	1107	2.75	1	12	1.88	1	2	4	5
4-7 s	313	6.41	1	20	2.87	4	6	8	10
7-10 s	67	11.99	3	33	5.91	8	11	15	19.2
10-15 s	13	15.92	10	40	8.45	11	14	15	24.4

Table 4. Summary of Statistical Models that Predict Start-up Time

(a) Model 1: $\boldsymbol{Startup\ Time = \beta_1(Pedestrian\ Volume)}$

Explanatory Variable Coefficient

	Coefficient
β_1	0.7709

Explanatory Variable Significance

	t-stat	p-value
X_1 (peds)	82.12	0.0000

Goodness-of-Fit

Adj. R2	0.8174
Std. Err.	1.9627
Obs.	1,500

Regression Statistics

	df	SS	MS	F
Regression	1	25979.6336	25979.6336	6744.0812
Residual	1499	5774.4664	3.8522	
Total	1500	31754.1		

(b) Model 2: $\boldsymbol{Startup\ Time = \beta_1(Pedestrian\ Volume) + \beta_2(Storage\ Area) + \beta_3(Push\ Button\ Offset)}$

Explanatory Variable Coefficient

	Coefficient
β_1	0.5460
β_2	0.1933
β_3	-3.4E-06

Explanatory Variable Significance

	t-stat	p-value
X_1 (peds)	55.82	0.0000
X_2 (ft^2)	24.95	3.4E-115
X_3 (ft)	-0.03	0.9739

Goodness-of-Fit

Adj. R^2	0.8963
Std. Err.	1.4770
Obs.	1,500

Regression Statistics

	df	SS	MS	F
Regression	3	28488.2356	9496.0785	4352.7923
Residual	1497	3265.8643	2.1816	
Total	1500	31754.1		

(c) Model 3: $\boldsymbol{Startup\ Time = \beta_1(Pedestrian\ Volume) + \beta_2(Push\ Button\ Offset)}$

Explanatory Variable Coefficient

	Coefficient
β_1	0.5460
β_2	0.1931

Explanatory Variable Significance

	t-stat	p-value
X_1 (peds)	56.37	0.0000
X_2 (ft)	33.92	1.2E-187

Goodness-of-Fit

Adj. R^2	0.8964
Std. Err.	1.4765
Obs.	1,500

Regression Statistics

	df	SS	MS	F
Regression	2	28488.2332	14244.1166	6533.5448
Residual	1498	3265.8667	2.1801	
Total	1500	31754.1		

2.5 Discussion and Recommendations

The data were collected at or near a college campus. Therefore, the authors propose using the 50th percentile values in the pedestrian volume categories listed in Table 5 and seen in Figure 9 as a quantitative guideline for selecting an appropriate pedestrian walk interval duration. It should be noted, however, that the 25th percentile values could provide more conservative values at locations where the pedestrians might have slower start-up times.

Preliminary plots suggest that the relationship between start-up time and the collected explanatory variables is near-linear. Therefore, linear regression models were used to predict start-up time. The statistical models built indicate the significant influence of the variables: (1) pedestrian volume and (2) offset from the push-button to the crosswalk on the pedestrian start-up time. The developed model explains start-up time with a relatively high accuracy (0.8964 R^2).

Table 5. Recommended Walk Interval Duration per Pedestrian Volume

Start-up Time	**Pedestrian Volume (peds/quad/cycle)**			
	Percentile			
	25th	***50th***	**75th**	**90th**
1-4 s	1	*2*	4	5
4-7 s	4	*6*	8	10
7-10 s	8	*11*	15	19.2
10-15 s	11	*14*	15	24.4

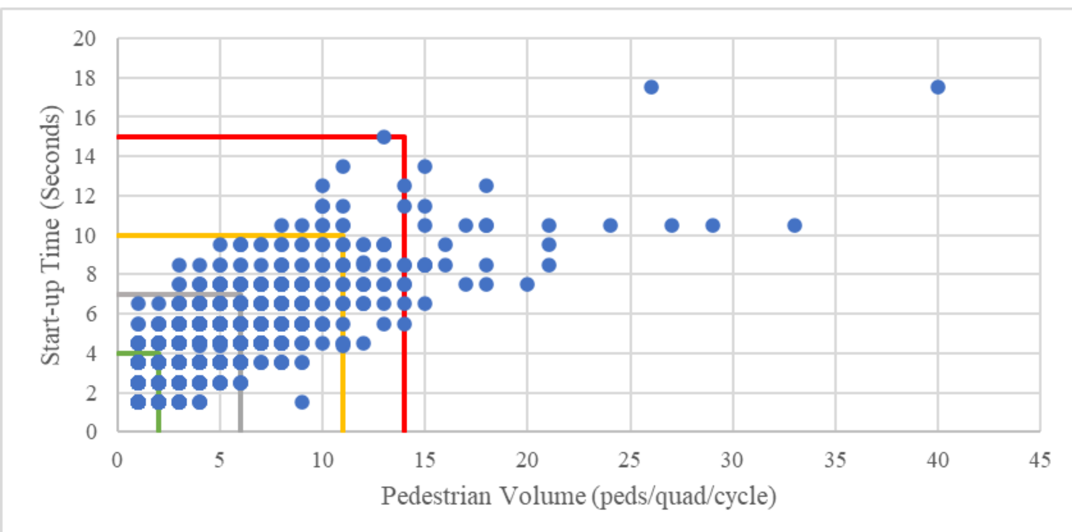


Figure 9. Recommended Pedestrian Walk Interval Duration Using the 50th Percentile

2.6 Conclusion

This study presented a quantitative analysis of the pedestrian walk interval duration based on pedestrian volume observed at 12 signalized intersections at various locations in the City of West Lafayette, Indiana, over a ten-month period. In addition, data on the storage area and offset from the pedestrian push button to the crosswalk were used to explain the variability in pedestrian start-up time. The built statistical model can help traffic signal designers to identify appropriate proper walk interval timing on an intersection-by-intersection basis. In addition, designers are herein provided quantitative data not only to help them in this task but also to support decisions to prioritize close-to-crosswalk push-button locations that could help minimize pedestrian start-up time. Future research should consider examining the impact of different types of pedestrian phasing (i.e., exclusive service and standard concurrent service) on the pedestrian start-up time. Additionally, seasonality can be included in future similar analysis by considering the different seasons (summer, fall, winter, and spring) because pedestrian behavior can be expected to change with inclement weather.

2.7 Chapter Summary

This part of the research analyzed time series data of pedestrian volumes in a campus town and identified factors that influence pedestrian movements. The study assessed the extent to which the time-of-day significantly influences pedestrian volumes. In addition, the study quantified the association between the academic calendar and pedestrian activities. Moreover, the study characterizes the nature of the recurring pattern of pedestrian volumes over time. The study suggests that these findings can inform urban planning and design by highlighting the variation in pedestrian activities and providing insight into forecasting pedestrian volumes. This is beneficial in the current human driven vehicle (HDV) era and is expected to be particularly beneficial in the emerging era of CAVs.

CHAPTER 3 TIME SERIES ANALYSIS OF PEDESTRIAN PATTERNS AT A TYPICAL UNIVERSITY CAMPUS

3.1 Introduction

Funding for the transportation infrastructure is often allocated on the basis of need. However, in allocating resources, not all transportation modes get considered in a manner that reflects the true levels of demand. For example, unlike vehicle-related infrastructure, pedestrian facilities are usually designed based on limited short-duration counts of pedestrian volumes, mainly due to the limitation of existing counting techniques. As such, pedestrian infrastructure is often developed based on underestimates of pedestrian needs (i.e., sidewalks and pedestrian signal timings). Furthermore, misestimation of demand is evidently reflected in the safety of vulnerable roadway users, such as pedestrians and bicyclists. These users constitute almost twenty percent of total traffic fatalities despite making many fewer trips proportionally, compared to motorists [1].

The recent highway bill (Infrastructure Investment and Jobs Act) identified a need for more balanced and equitable service to all transportation modes and users. To do so, accurate data reflecting pedestrian activities is needed. Such data has always been of great interest in fields such as transportation engineering (i.e., pedestrian exposure and safety countermeasure studies), investment planning (i.e., prioritizing non-motorized infrastructure spending), and social and public health planning (i.e., physical health assessments). All these require information on the number of pedestrians at specified the areas of interest [2]. The bill highlights the need to revisit existing methods of traffic count. The NCHRP 797 Guidebook on Pedestrian and Bicycle Volume Data Collection Report provides guidelines for non-motorized data collection methods. Table 6 compares the methods available for pedestrian activity data collection.

Table 6. Comparison of Pedestrian Counting Methods [3]

Characteristic	**Passive Infrared**	**Active Infrared**	**Passive IR + Inductive Loops**	**Radio Beam**	**Automated Video**	**Manual Counts**
Equipment cost	$$	$$$	$$$	$$$	$$	$
Preparation cost	$$	$$	$$$	$$	$$	$
Installation time	🕓	🕓🕓	🕓🕓🕓	🕓🕓	🕓	N/A
Hourly cost	$	$	$	$	$$$	$$$$
Data collector training time	🕓	🕓	🕓	🕓	🕓	🕓🕓🕓
Mobility	+++	++	-	++	+++	+++
Pavement cuts	No	No	Yes	No	No	No

Pedestrian volume data are quite different from vehicle volume data. The former is far more sensitive to external circumstances such as precipitation, temperature, and darkness. The NCHRP 797 report suggests that the volume of bicycles and pedestrians generally exhibits greater variation compared to motor vehicles. Although all kinds of traffic vary over time, pedestrian traffic is significantly more sensitive to the weather on a given day compared to motorized traffic. Additionally, compared to the volumes seen at conventional motorized vehicle count sites, hourly pedestrian and bicycle volumes tend to be relatively modest at most locations.

Long-term factoring approaches used to estimate motorized volumes based on short-term counts (i.e., 24-hour or shorter) may not always be suitable for non-motorized counting due to the increased variability in non-motorized volumes. For example, a 12-hour motorized vehicle count could be transformed into a daily count by dividing the counted volume by the percentage of daily traffic that occurred on average during the count period (based on prior 24-hour or longer counts), and then the result could be adjusted for monthly variations in traffic, as determined from data from a permanent counting station, to arrive at an estimate of average annual daily traffic (AADT). In contrast, estimations of average annual bicycle traffic (AABT) based on 12-hour counts from a midweek day had an average error of 40% off the actual volumes, according to a study of locations in Boulder, Colorado [4]. Research by Niska et al. 2012 and Danish Road Directorate 2004 confirms that it is challenging to estimate AABT on a year-to-year basis or even a 1-week basis. Automated counting methods must be used more frequently because of the extended periods needed for precise non-motorized volume estimates [5].

Studies have used different counting methods for pedestrian demand analysis at the micro and macro levels. At the micro level, pedestrian counts are studied at finer geographical granularity, such as crosswalks and intersections [6-8]. At the macro level, public surveys (i.e., National Household Travel Survey) are often used to collect data to measure pedestrian demand at larger geographical levels such as cities [9,10]. However, macro-level studies often sacrifice accuracy at larger geographical units in lower-cost methods, such as public surveys.

3.1.1 Motivation and Objective

There is a literature gap regarding the analysis of network-wide pedestrian activities over prolonged periods. Similarly, the literature lacks descriptive analysis of the magnitude of impact seasonal trends and special events (i.e., football and basketball games) have on pedestrian activities at both the network and intersection levels simultaneously.

This study leverages data from permanently-installed cameras at signalized intersections to report on the observed pedestrian activities at 19 intersections in West Lafayette, Indiana, between June 2021 and December 2022. The pedestrian activities are recorded using automated video counting methods. The study records pedestrian movements tabulated in 15-minute counts of pedestrian volumes at each intersection. Specifically, this study uses time series analysis to explain the fluctuations in pedestrian volumes overtime at the network level at this campus town. In addition, this research quantifies the magnitude and dynamics of pedestrian volumes on special event days such as football games, basketball games, and graduation ceremonies by introducing a pedestrian activity impact index methodology. The overall study framework is presented in Figure 10.

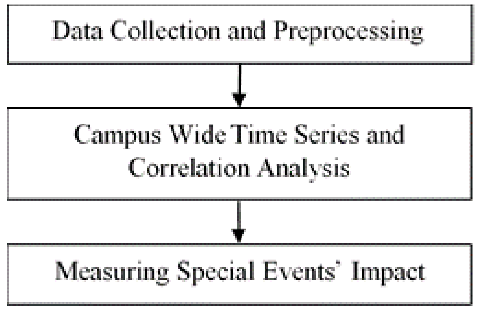


Figure 10. Time-series Study Overall Framework

3.2 Methods

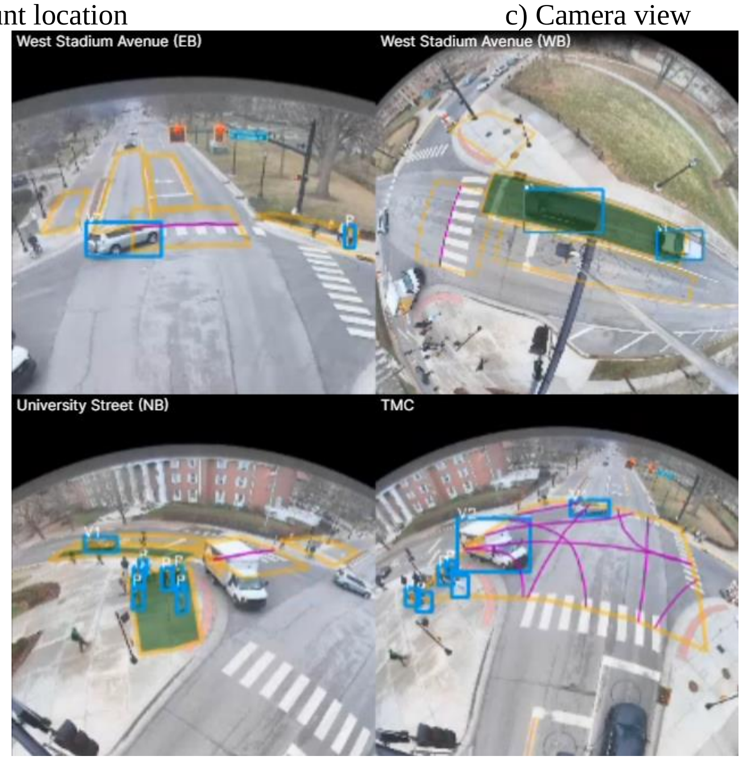
3.2.1 Pedestrian Volume Data Collection and Processing

Pedestrian activity data used in this study were collected from 19 cameras at signalized intersections in the City of West Lafayette, Indiana, between June 2021 and December 2022. As a result, 35,620,413 observations of pedestrian movements were observed and examined in relation to time and seasonal special events. Figure 11 and Table 7 present the camera setup and locations of the 19 intersections used in this study. The cameras used were permanently mounted on the traffic light mast arms, each providing a 360-degree view of the intersection approach and curb ramp. The cameras recorded continuously since the day of installation and allowed for automatic data extraction of volume counts of all roadway users in real time. The conversion of video imagery to meaningful data (traffic counts per direction, in this case) was achieved using computer vision techniques. Computer vision is a field of artificial intelligence (AI) that enables computers and systems to derive meaningful information from digital images, videos, and other visual inputs [11]. In each 15-minute interval, data are aggregated for each direction of travel for vehicles, bicyclists, and pedestrians. Figure 11c presents the detection/tracking view of the cameras used to collect the pedestrian volume data.

(a)Intersection view showing camera location

b) Camera mount location

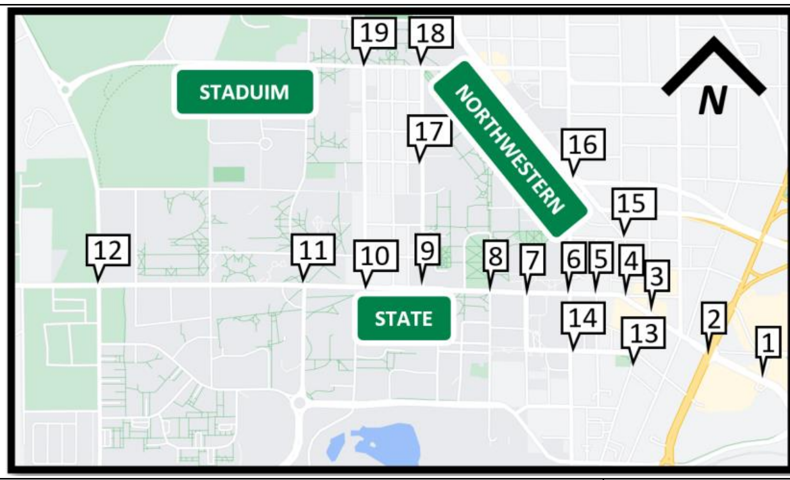
c) Camera view



d) Camera detection/tracking view

Figure 11. Camera Installed at W. Stadium Ave and University St. (Intersection #18)

Table 7. Map and List of Intersections where Cameras are Mounted



Intersection		Location	
		Lat	*Long*
1	Roebuck Drive and State Street	40.4212	–86.9019
2	State Street and South River Road	40.4218	–86.9042
3	State Street and Chauncey Avenue	40.4233	–86.9069
4	Northwestern Avenue and State Street	40.4240	–86.9082
5	State Street and Andrew Place	40.4240	–86.9092
6	South Grant Street and State Street	40.4239	–86.9103
7	State Street and Sheetz Street	40.4240	–86.9122
8	State Street and Marsteller Street	40.4241	–86.9138
9	State Street and University Street	40.4242	–86.9168
10	State Street and Russell Street	40.4242	–86.9191
11	State Street and S. Martin Jischke Drive	40.4242	–86.9217
12	State Street and Airport Road	40.424	–86.9302
13	South Chauncey Avenue and West Wood Street	40.4219,	–86.9076
14	West Wood Street and South Grant Street	40.4221	–86.9103
15	North Street and Northwestern Ave	40.4258	–86.9080
16	North Grant Street and Northwestern Ave	40.4280	–86.9104
17	University Street and 3rd Street	40.4272	–86.9166
18	West Stadium Avenue and University Street	40.4313	–86.9168
19	Russell Street and West Stadium Avenue	40.4242	–86.9191

3.2.2 Times Series, Correlation Analysis, and Special Events

First, the data were split into four main periods consistent with the university academic calendar: (1) summer semester and break, (2) fall semester, (3) Christmas break, and (4) spring semester. Then, student t-tests at a 95% confidence interval were used to determine if the means of selected time frames were statistically different. The tests were conducted to examine if pedestrian activities on different days of the week and the different academic semesters are statistically different. Student's t-test were used to test the hypotheses about the mean of a sample drawn from a normally distributed population when the population standard deviation is unknown [12].

$H_o: \mu_1 = \mu_2$
$H_\alpha: \mu_1 \neq \mu_2$

The following five scenarios were tested:

1. Whether (1) Fall vs. (2) Spring activities are statistically different.
2. Whether (1) Summer vs. (2) Christmas Break activities are statistically different.
3. Whether (1) Monday, Wednesday, and Friday vs. (2) Tuesday and Thursday activities are statistically different.
4. Whether (1) Saturday vs. (2) Sunday activities are statistically different.
5. Whether (1) Weekdays vs. (2) Weekends activities are statistically different.

Pedestrian activity data collected during the analysis is examined using time series analysis to help understand the underlying causes of trends or systemic patterns over time. Furthermore, to explore the relationship between activities and special events within the geospatial vicinity of each intersection by conducting the following steps:

1. Observed data: a plot of the aggregated data per 15-minute intervals of all 19 intersections.
2. Trend data: a plot of the aggregated data per 15-minute intervals using a rolling average to detect the overall trend over time.
3. Seasonality data: a plot of the aggregated data per 15-minute intervals after de-trending and averaging to detect day-to-day and week-to-week patterns.
4. Residual data: a plot of the residuals of each time interval.

In addition, autocorrelation plots are derived to establish the relations between lagged time series and the present versions. In other words, they determine the correlation between the past and the present or future. Cross-correlation plots were also developed to determine the relationship between pedestrian volumes of two different intersections (i.e., intersection A volume at time step *i* to determine intersection B volume at time step i+1).

The autocorrelation function at lag k with average pedestrian volume α is defined as:

$$r_k = \frac{\sum_{i=1}^{n-k}(y_i - \alpha)(y_{i+k} - \alpha)}{\sum_{i=1}^{n}(y_i - \alpha)^2}$$

Finally, the impact of football games, basketball games, and graduation ceremonies was examined by visualizing pedestrian activities on the day of the event and by calculating a special-event impact index. The special event impact index is calculated using the following equation:

$$Impact\ Index = \frac{\left(\frac{Vol_x - Vol_{x-7}}{\frac{(Vol_x + Vol_{x-7})}{2}}\right) * 100\% + \left(\frac{Vol_x - Vol_{x+7}}{\frac{(Vol_x + Vol_{x+7})}{2}}\right) * 100}{2}$$

Where:
Vol_x is the pedestrian volume of the day of event
Vol_{x-7} is the pedestrian volume of the same day of event but for the previous week
Vol_{x+7} is the pedestrian volume of the same day of event but for the following week

The special events impact index was calculated for 12 football games, 26 men's basketball games, and nine graduation ceremonies. This index is calculated for each intersection by measuring the total day activities in relation to the same day of the previous week and the following week, as long as those days were during the same academic season (i.e., all during Fall 2021) and did not have special events. In other words, if the total pedestrian volume on Friday (game day) for intersection #1 is 1200 and was 600 for the past Friday and 550 for the following Friday, with all three days being during the Fall 2022 semester and the previous and following Fridays did not have special events, then the special event impact index is +103.81%.

3.3 Summary Statistics of the Data

Between June 1st, 2021 and December 31st, 2022, a total of 35,620,413 pedestrian movements were recorded, with an average campus 15-minute pedestrian volume of 640.84, a minimum volume of 0, and a maximum volume of 7509. The standard deviation was 852.77, and 80% of observed intervals were below 500 pedestrian movements per 15 minutes. Figure 12 presents an overview of pedestrian volumes through the analysis period.

Of the 19 intersections, University St., and 3rd St. (#17) had the highest average pedestrian activities at 85.61 pedestrians per 15-minute interval, the interval with the maximum number of pedestrians throughout the analysis period in 1780, and the highest volume standard deviation at 170.54. Table 8 presents the descriptive statistics of the data from the 19 intersections.

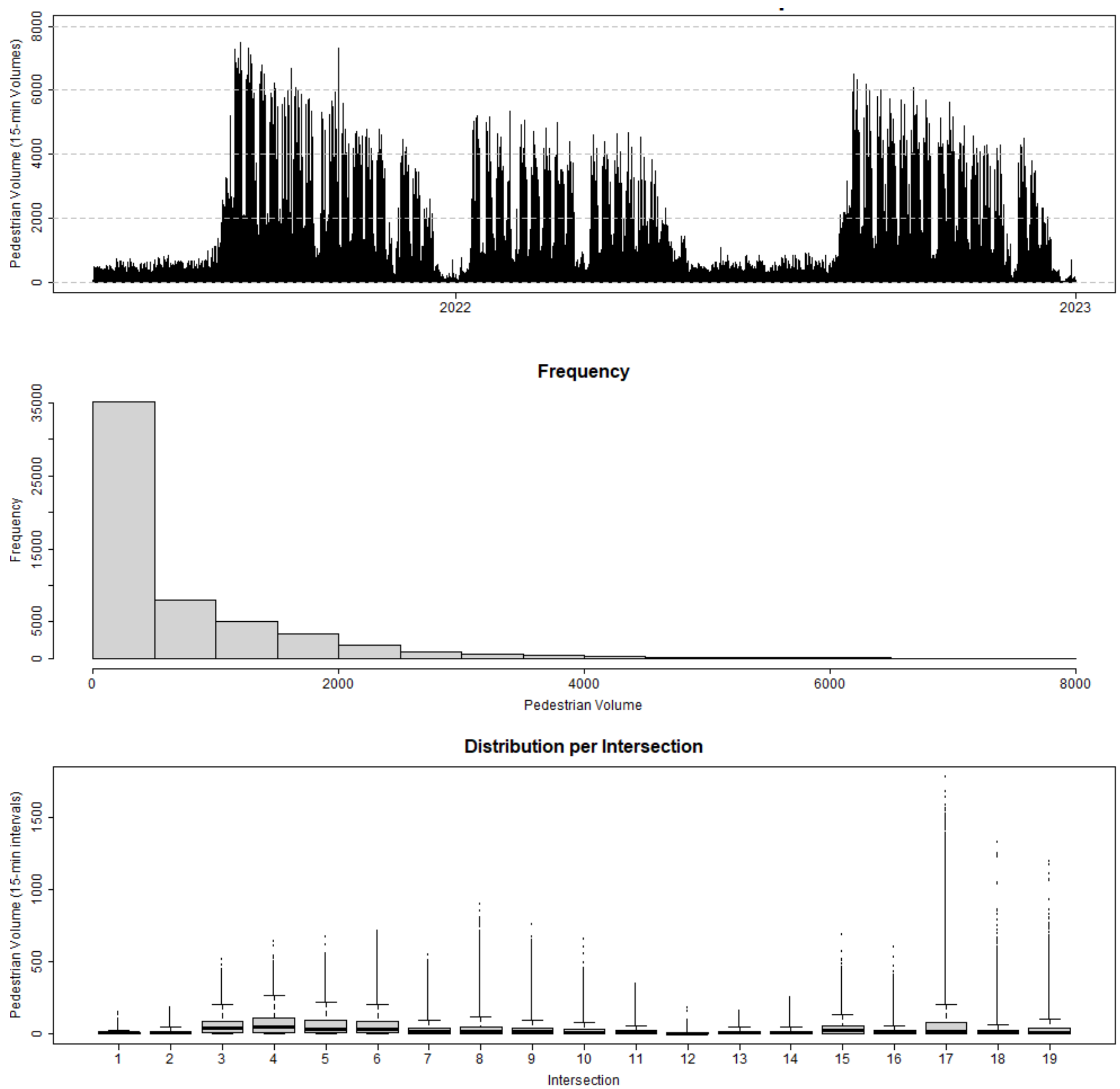


Figure 12. Campus Pedestrian Activity Volume, Volume Distribution, and Intersection Volume Distribution

Table 1**.** Intersection Descriptive Statistics for Pedestrian Volumes

Int.	**Min**	**1st Q**	**Median**	**Mean**	**3rd Q**	**Max**	**Std. Dev.**
1	0.00	1.00	5.00	6.78	10.00	148.00	6.88
2	0.00	1.00	7.00	11.67	18.00	178.00	14.07
3	0.00	7.00	36.00	53.44	84.00	517.00	55.84
4	0.00	8.00	47.00	67.59	110.00	638.00	69.66
5	0.00	5.00	30.00	61.10	91.00	668.00	76.73
6	0.00	5.00	28.00	60.62	84.00	714.00	82.97
7	0.00	2.00	12.00	31.56	39.00	544.00	50.92
8	0.00	3.00	15.00	39.65	47.00	900.00	69.95
9	0.00	2.00	13.00	32.94	39.00	758.00	56.94
10	0.00	2.00	10.00	24.79	31.00	653.00	39.74
11	0.00	2.00	8.00	17.43	23.00	341.00	24.99
12	0.00	0.00	1.00	2.99	4.00	179.00	4.60
13	0.00	1.00	6.00	12.41	18.00	154.00	15.89
14	0.00	1.00	6.00	16.40	18.00	251.00	26.80
15	0.00	3.00	19.00	36.83	55.00	691.00	46.06
16	0.00	1.00	4.00	21.47	21.00	598.00	41.83
17	0.00	2.00	16.00	85.61	81.00	1780.00	170.54
18	0.00	1.00	6.00	26.96	25.00	1330.00	54.59
19	0.00	1.00	8.00	30.60	42.00	1193.00	51.17

3.4 Results

The overall analysis period is decomposed into periods based on whether the school was in session or not. This yielded four main periods: (1) Summer semester and break, (2) Fall semester, (3) Christmas break, and (4) Spring semester. The Fall semester saw the highest average activities at 1022 for 2021 and 879.8 for 2022, followed by the Spring semester at 738 for 2022, Summer at 322.9 for 2021 and 273.1 for 2022, and then Christmas break at 127.1 for 2021-22 and 70.67 for 2022-23. Table 9 and Figure 13 provide the activity levels at each analysis period. In addition, Figure 14 presents the data decomposition into a trend, seasonality, and random plots.

Table 9. Descriptive Statistics for Pedestrian Volumes for the Different School Sessions

Season of Year	Dates (from-to)	Activities (volume per 15-minute intervals)				
		Min.	1st Q	Mean	3rd Q	Max
Summer 2021	6/1/2021 – 8/22/2021	0	55	322.9	401	5215
Fall 2021	8/23/2021 – 12/18/2022	0	147	1022	1510	7509
Christmas 2021-2022	12/19/2022 – 1/9/2022	0	13	127.1	176	1141
Spring 2022	1/10/2022 – 5/7/2022	0	97	738	1133.2	5349
Summer 2022	5/8/2022 – 8/21/2022	0	38	273.1	376	3172
Fall 2022	8/22/2022 – 12/17/2022	0	97	879.8	1353	6510
Christmas 2022-2023	12/18/2022 – 12/31/2022	0	5	70.67	100	725

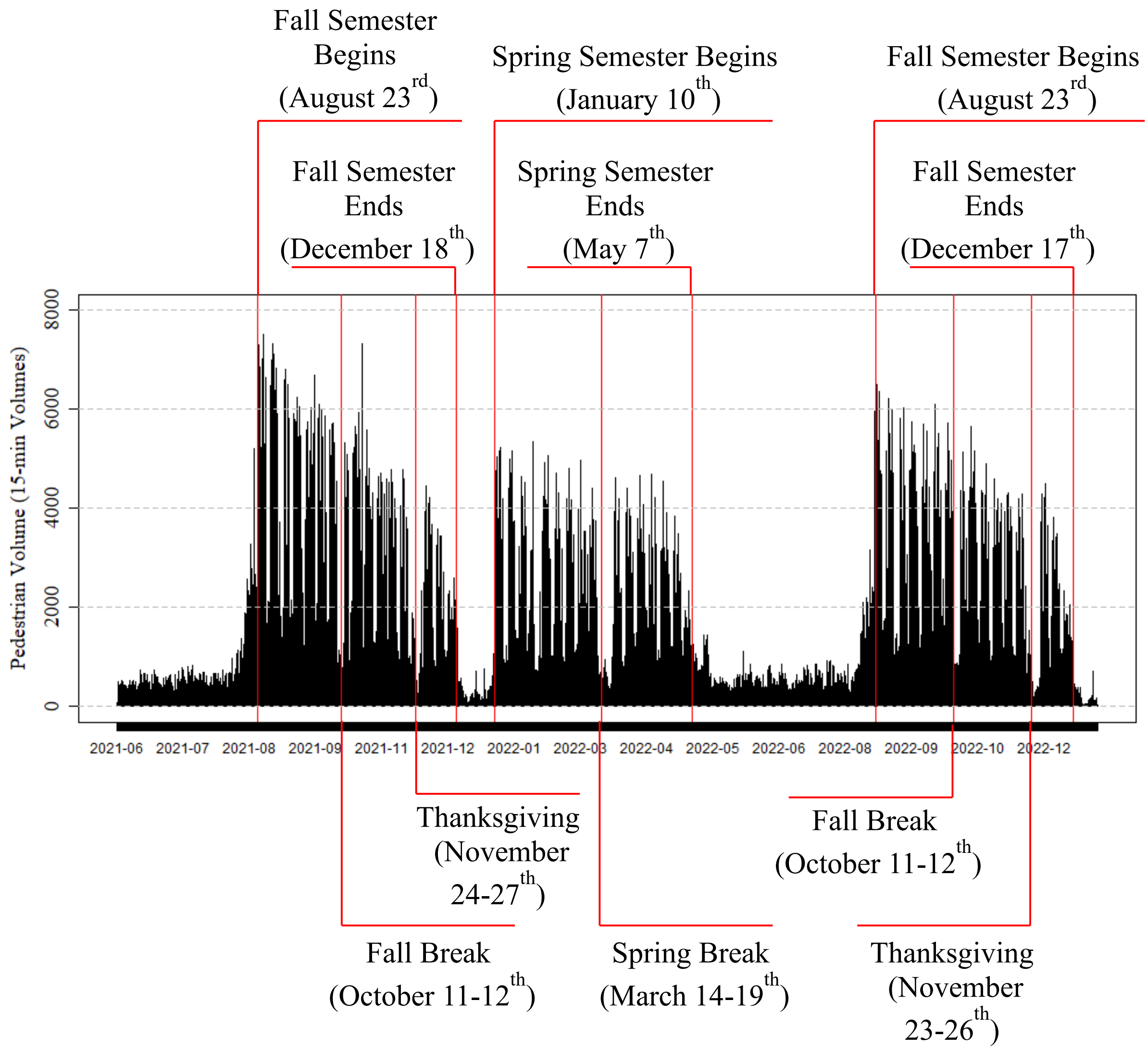


Figure 13. Pedestrian Activity Visualization During on/off School Sessions

The results of the Student t-test indicate the statistical difference between the four analysis periods considered in this study: (1) Summer semester and break, (2) Fall semester, (3) Christmas break, and (4) Spring semester. In addition, weekdays, weekends, Saturdays, and Sundays were found to be statistically different. Table 10 presents a summary of the statistical tests. The time series decomposition plot (Figure 14) presents the observed aggregate campus values, the trend using a moving average, the seasonal plot, and the random plot. Aggregate campus autocorrelation values measuring the correlation of values with their lag (former time step values) are above 0.600 for the first 13 lags, where each lag is a 15-minute interval. In addition, a clear descending pattern is evident across the 50 lags, moving from a positive to a negative correlation around the 30th lag. Figure 15 presents a plot of the first 50 lags with values for the first ten lags.

Table 10. Student t-test Results Summary

	Mean	LCL (95%)	UCL (95%)	t-value	p-value	Statistically Different?
(1) Are Fall vs. Spring activities significantly different?						
Fall Semester	951.071	192.960	233.161	20.779	3.57e-95	Yes
Spring Semester	738.011					
(2) Is Summer vs. Christmas Break activities significantly different?						
Summer break	294.959	182.471	197.150	50.695	0.000	Yes
Christmas Break	105.149					
(3) Are Monday, Wednesday, and Friday vs. Tuesday and Thursday activities significantly different?						
MWF	707.227	22.578	-15.183	0.384	0.701	No
T Th	710.924					
(4) Are Saturday vs. Sunday activities significantly different?						
Sa	530.361	103.488	136.883	14.109	6.72e-45	Yes
Su	410.176					
(5) Are Weekday vs. Weekend activities significantly different?						
Weekdays	708.710	225.558	250.566	37.316	1.60e-300	Yes
Weekend	470.648					

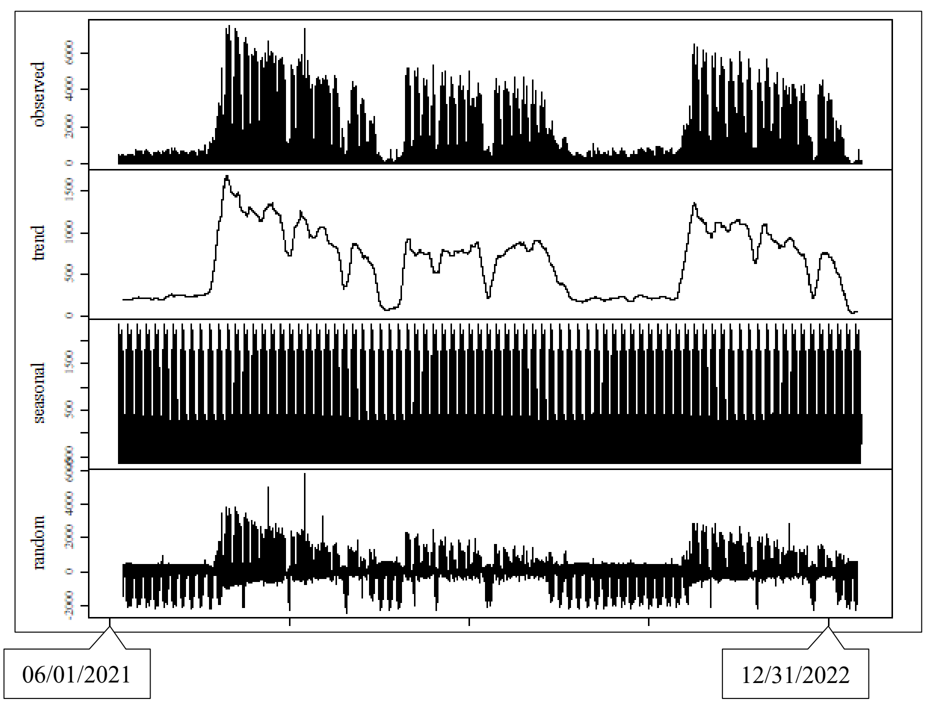
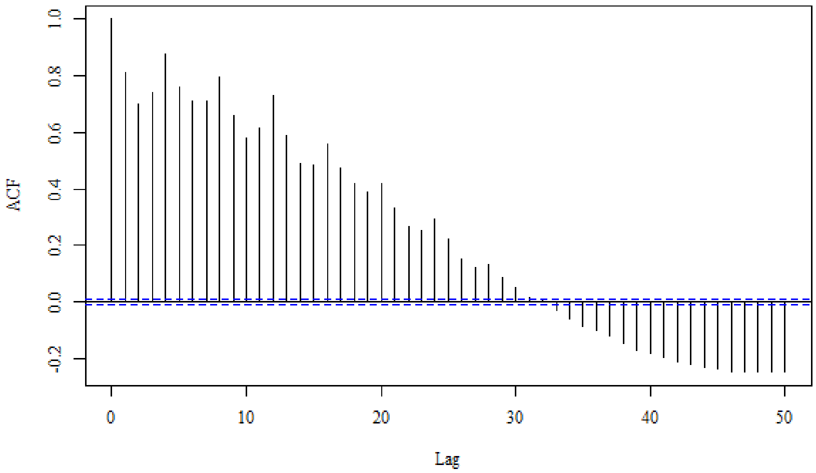


Figure 14. Decomposition of Campus-Wide Additive Time Series



Lag	ACF
0	1.000
1	0.815
2	0.709
3	0.748
4	0.881
5	0.766
6	0.717
7	0.716
8	0.802
9	0.667
10	0.588

Figure 15. Aggregate Campus Pedestrian Activity Autocorrelation

Cross-correlation values (indicating the correlation of an intersection value with the lags of other intersections) show that State St. and Chauncey Ave. (#4) is the intersection that is most correlated to other intersections (with an average 10-lag cross-correlation of 0.693). On the contrary, the Roebuck Dr. and State St. intersection (#1) is the least correlated (with average 10-lag cross-correlation of 0.454). Table 6 presents the average 10-lag cross-correlation values for all the 19 intersections. The cross-correlation values for the 19 intersections are provided in Table 11.

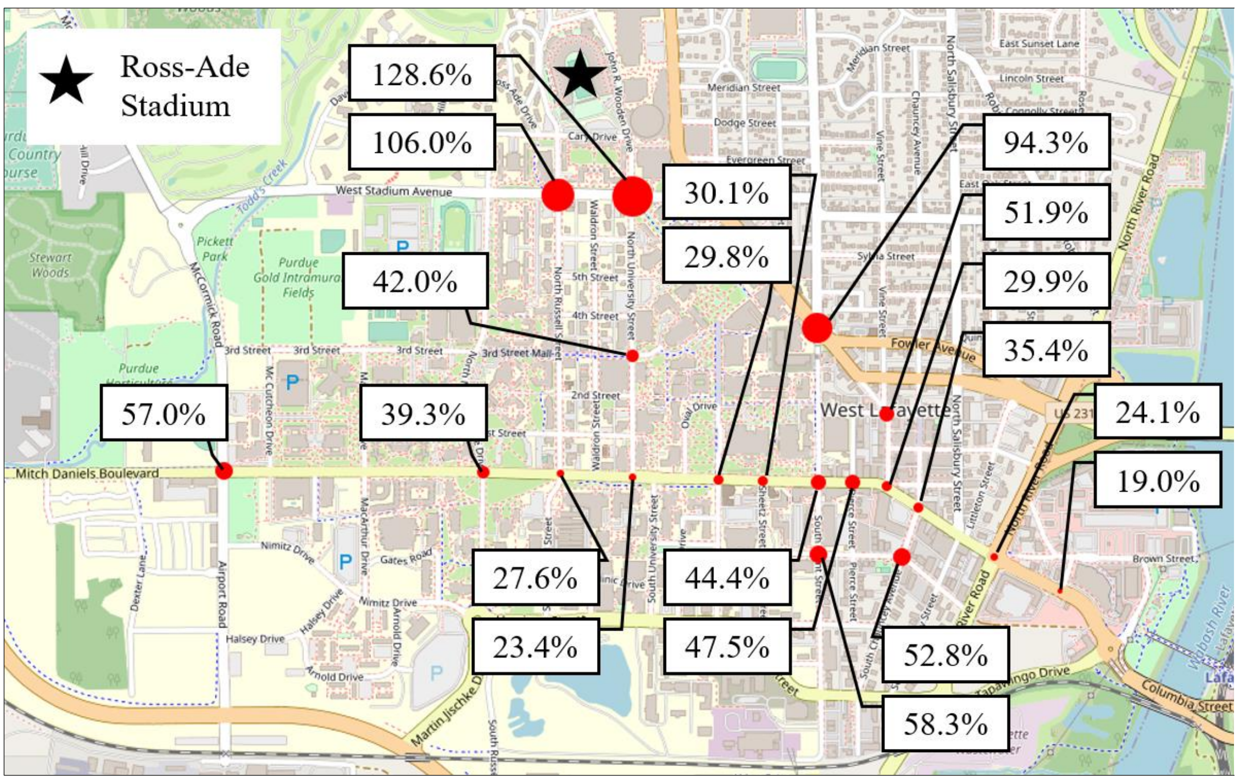
Table 11. Intersection Cross Correlation (average 10-lag values)

	Intersection																		
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**
1	0.64	0.63	0.60	0.61	0.49	0.49	0.36	0.36	0.35	0.38	0.44	0.45	0.54	0.35	0.52	0.32	0.31	0.35	0.43
2	0.61	0.77	0.71	0.75	0.68	0.68	0.55	0.55	0.54	0.57	0.64	0.58	0.71	0.56	0.69	0.51	0.50	0.53	0.61
3	0.66	0.79	0.84	0.84	0.73	0.71	0.55	0.55	0.55	0.57	0.65	0.59	0.74	0.54	0.76	0.49	0.48	0.53	0.65
4	0.62	0.77	0.79	0.86	0.78	0.77	0.64	0.63	0.63	0.66	0.72	0.62	0.78	0.64	0.79	0.59	0.58	0.60	0.69
5	0.45	0.64	0.63	0.73	0.80	0.76	0.68	0.66	0.65	0.68	0.69	0.59	0.75	0.71	0.72	0.68	0.64	0.64	0.67
6	0.40	0.59	0.57	0.66	0.72	0.77	0.70	0.69	0.69	0.70	0.72	0.60	0.71	0.72	0.71	0.64	0.67	0.65	0.66
7	0.27	0.45	0.41	0.51	0.59	0.65	0.66	0.63	0.64	0.65	0.64	0.50	0.60	0.68	0.60	0.59	0.64	0.59	0.56
8	0.28	0.46	0.42	0.51	0.58	0.63	0.63	0.63	0.63	0.64	0.64	0.47	0.58	0.65	0.58	0.57	0.61	0.58	0.56
9	0.28	0.45	0.42	0.51	0.56	0.63	0.62	0.62	0.64	0.64	0.64	0.47	0.57	0.64	0.58	0.56	0.61	0.57	0.55
10	0.30	0.48	0.43	0.53	0.59	0.64	0.63	0.63	0.63	0.68	0.68	0.48	0.59	0.65	0.59	0.58	0.62	0.59	0.58
11	0.38	0.56	0.53	0.61	0.63	0.70	0.65	0.65	0.66	0.70	0.75	0.54	0.64	0.66	0.65	0.57	0.62	0.62	0.65
12	0.35	0.49	0.45	0.51	0.53	0.57	0.51	0.49	0.49	0.50	0.53	0.62	0.56	0.53	0.57	0.48	0.49	0.51	0.52
13	0.45	0.62	0.58	0.67	0.70	0.71	0.64	0.62	0.62	0.64	0.67	0.59	0.76	0.68	0.68	0.63	0.61	0.61	0.62
14	0.25	0.43	0.37	0.48	0.58	0.62	0.64	0.61	0.61	0.63	0.61	0.48	0.59	0.70	0.56	0.62	0.63	0.59	0.53
15	0.48	0.65	0.66	0.72	0.71	0.74	0.67	0.65	0.65	0.66	0.70	0.62	0.72	0.67	0.79	0.59	0.63	0.64	0.68
16	0.21	0.38	0.33	0.43	0.55	0.55	0.55	0.54	0.53	0.55	0.53	0.43	0.55	0.62	0.49	0.64	0.56	0.56	0.51
17	0.20	0.38	0.32	0.42	0.52	0.59	0.61	0.58	0.59	0.60	0.58	0.44	0.53	0.64	0.52	0.56	0.62	0.56	0.51
18	0.24	0.40	0.37	0.45	0.53	0.57	0.55	0.54	0.54	0.56	0.57	0.48	0.54	0.59	0.54	0.56	0.55	0.63	0.61
19	0.37	0.54	0.53	0.60	0.63	0.65	0.58	0.58	0.58	0.61	0.66	0.56	0.63	0.59	0.65	0.56	0.56	0.64	0.73

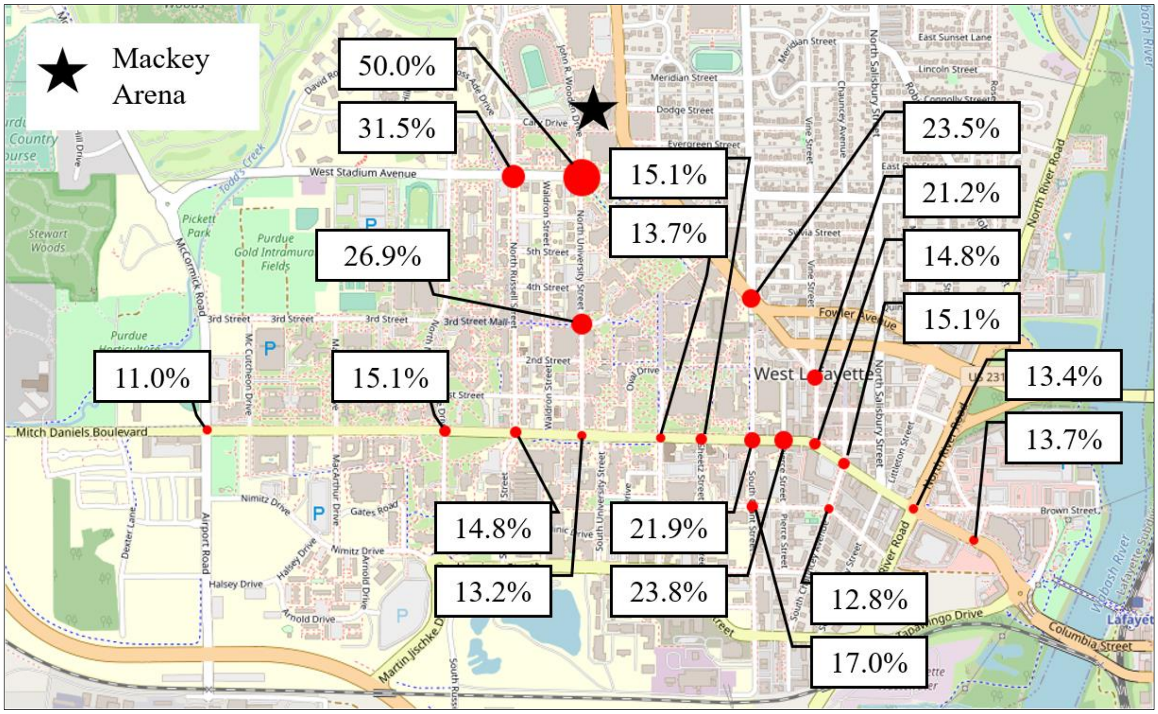
The impact index results for three events show that graduation ceremonies are the most impactful in terms of pedestrian activities across campus, followed by football games and men's basketball games, as seen in Table 12. Intersection-by-intersection impact indices show that intersections impacted the most are those typically within the vicinity of the special event location (i.e., intersections close to Ross-Ade Stadium on football game days, intersections close to Mackey Arena on men's basketball game days, and intersections close to Elliot Hall of Music on graduation ceremony days).

Table 12. Average Campus Impact Index Per Event

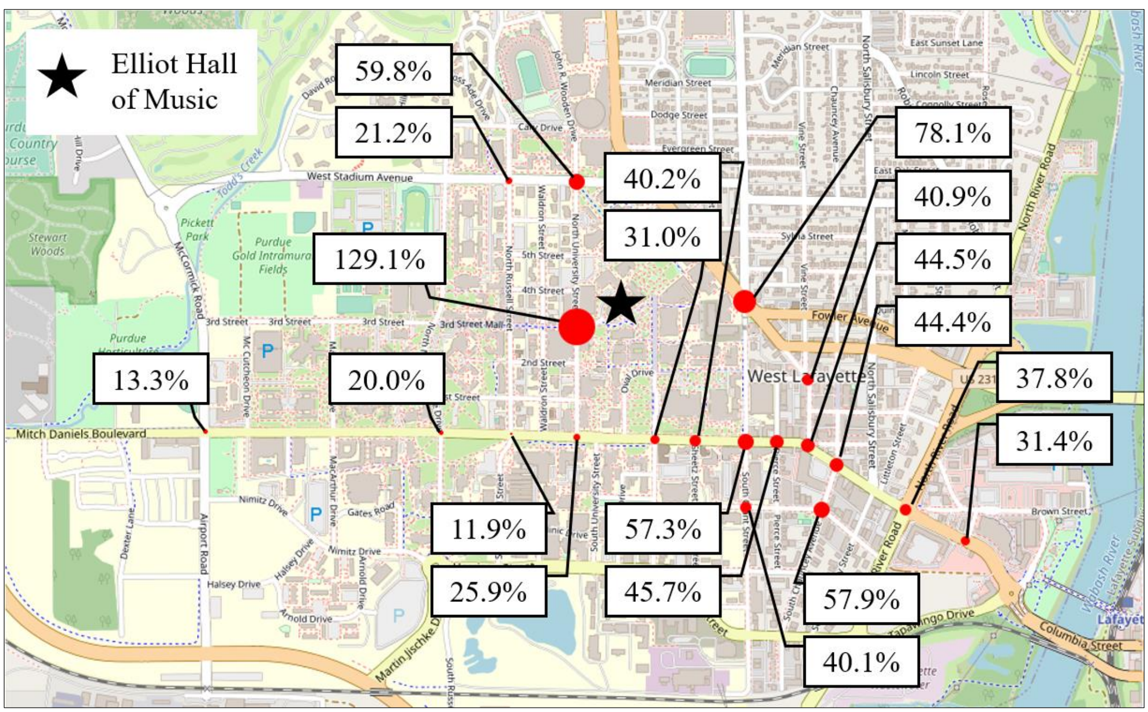
Event	Impact Index
Football Games	50.44%
Men's Basketball Games	20.43%
Graduation Ceremony	53.30%



(a) Average Football Game-day Impact Index



(b) Average Men’s Basketball Game-day Impact Index



(c) Average Graduation Ceremony Impact Index

Figure 16. Average Intersection Impact Index Per Event Type

3.5 Discussion and Conclusions

The aggregate campus pedestrian volume over the analysis period (Figure 11) has several interesting features. Most apparent is the trend in the number of pedestrians, which tends to be higher during academic semesters compared to off-school periods, showing a clear association between pedestrian activities and the university academic calendar (Figure 13). This association can be confirmed in the random plot (Figure 14), where the random value is always negative during off-class periods and positive during in-class periods. The student t-tests further validate differentiating academic semesters and on/off class periods. Moreover, weekdays, weekends, Saturdays, and Sundays are statistically different. Furthermore, an apparent repetition in trend between the four periods is evident. For example, fall 2021 and fall 2022 have a similar tendency, as seen in Figure 17. The similarity indicates that future fall semesters will likely have a similar pattern.

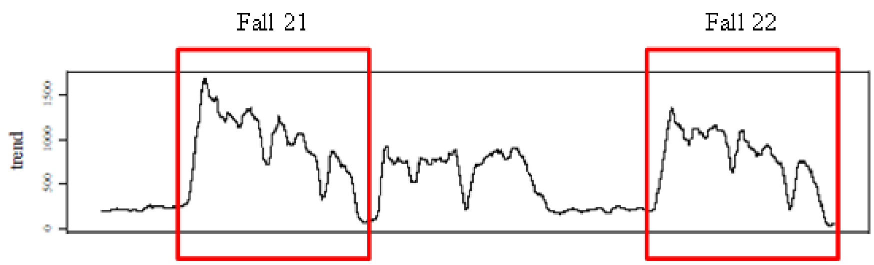


Figure 17. Fall Semester Trend Similarity

In addition, the data shows fall 2021 with higher volumes compared to fall 2022, as seen in Figure 18. That can be related to COVID-19 restrictions being first relaxed during that academic semester resulting in a "bounce back" in activities that were slightly higher than expected volumes, which then returned to normal activities during fall 2022. Similarly, we can also see evidence that pedestrian activities during the same semester and throughout the academic year have a decreasing trend. The decreasing pattern through the same semester period can be explained by students getting closer to finals and student activities therefore decreasing in preparation for exams.

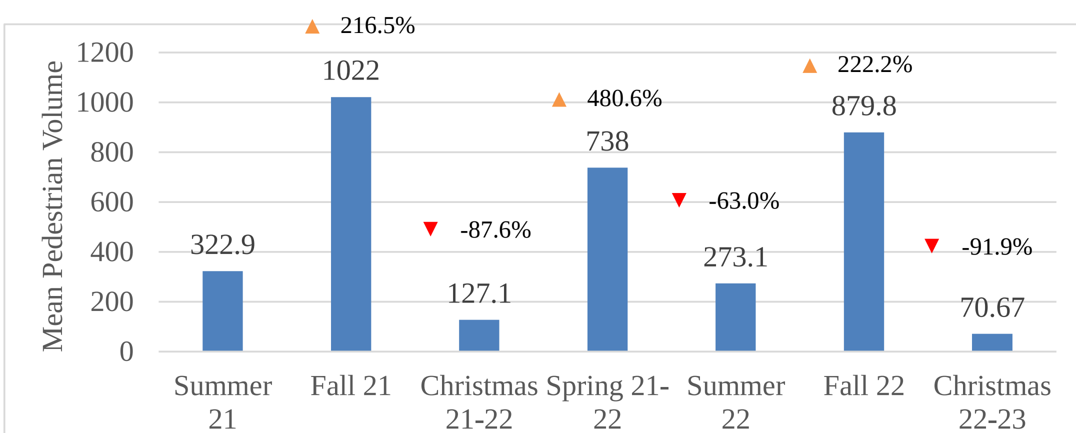


Figure 18. Average Pedestrian Volume Fluctuations

The decreasing pattern from fall to spring semester can be attributed to the weather as spring is typically colder compared to fall. This study was conducted on a university campus. However, non-campus locations also exhibit specific trends and patterns unique to the land-use in question. For example, downtown areas are expected to exhibit unique temporal patterns that recur yearly.

Time-series autocorrelation values present a clear relation between values and their lag, where each lag is a 15-minute interval. Moreover, the first lag and the fourth lag have the highest correlation value indicating that hours repeat in pattern. In other words, the first 15-minute interval of an hour have the highest correlation with the following hour first 15-minute interval and then third hour and so on, as seen in Figure 15. This is intuitive, because for a campus town, that period is the time when students are either heading to class or leaving class. Cross-correlation values, seen in Table 11, indicate that intersections within proximity of an intersection have the highest lag correlation. In other words, given volume data of an intersection, one could use that data to predict future pedestrian activities of nearby intersections.

Finally, the impact of sporting events and graduation ceremonies was measured using an impact index. The index shows graduation ceremonies as the most impacting, followed by football and men's basketball games. However, since the impact index is calculated for the same academic session, graduation ceremonies are therefore calculated in comparison to off-class days, resulting in higher index values than the other two events although the magnitude of volumes tend to be lower compared to basketball and football days. Furthermore, the impact for all three events is mostly perceived by intersections within the vicinity of the event (Figure 16).

In this study, ties between seasonal events and pedestrian activities were defined. This effort could help in future work regarding construction of machine learning algorithms to forecast pedestrian activities at both network and intersection levels. This study was conducted on a campus town. However, when sufficient data over prolonged periods are available, the methodology of identifying major attractions and influencing temporal events, like academic calendars and sporting events, in this case, can be performed at any type and size of urban area.

In addition, it is recognized that computer vision and automated counting approaches represent a rapidly evolving field that is expected to continue to enhance the quality and accuracy of counts for all road users. Future work could consider using machine learning approaches to forecast time-series pedestrian activities (i.e., Long-Short-Term-Memory or Autoregressive Integrated Moving Average). Such effort can significantly help in several fields, such as pedestrian signal timings, and appropriate deployment and distribution of law enforcement resources during special events.

3.6 Chapter Summary

This study analyzed time series data of pedestrian volumes on a campus town and identified factors that influence pedestrian movements. The study found that special events and time of day are significant determinants of pedestrian volume. In addition, the study found a significant association between the academic calendar and pedestrian activities. Moreover, the study confirms a repetitive pattern of pedestrian volumes over time, that is, pedestrian volumes tend to have a pattern that recurs. The study suggest that the findings can inform urban planning and designers by highlighting the variation pedestrian activities and gives a prospective to forecasting pedestrian volumes.

CHAPTER 4 MACHINE LEARNING APPROACH FOR FORECASTING PEDESTRIAN WALK-INTERVAL CATEGORIES

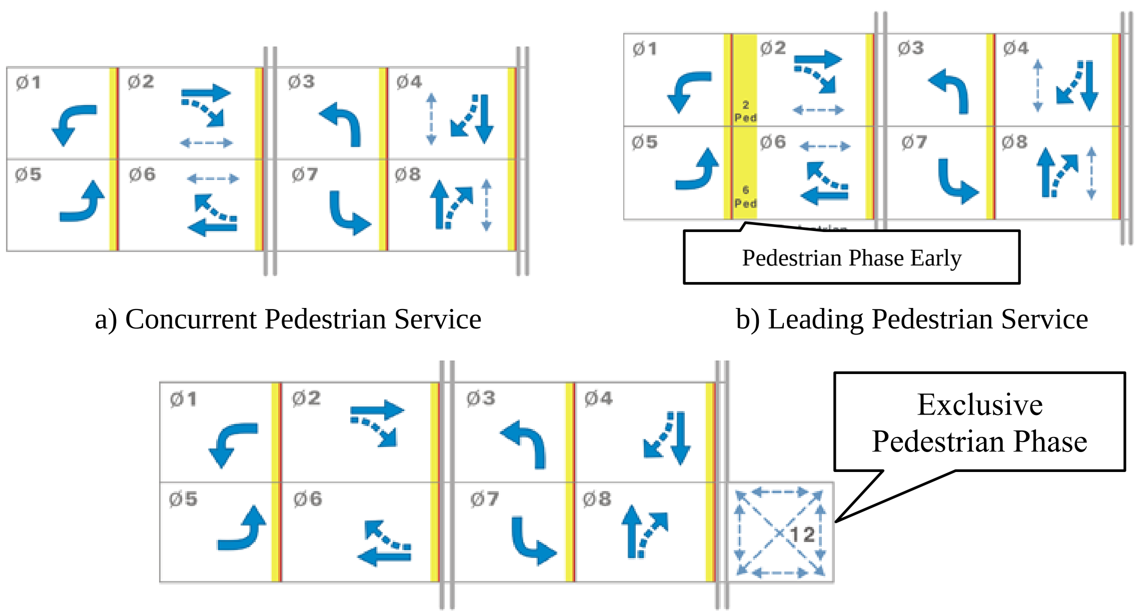
4.1 Introduction

At signalized intersections, the pedestrian phase is typically a fixed duration phase where the number of seconds is always the same regardless of the time of day and corresponding pedestrian volume. The existing practice has three alternatives for serving pedestrians at signalized intersections:

(1) concurrent service in which pedestrians are served concurrently with the adjacent through vehicular movement,

(2) leading pedestrian service in which pedestrians start a few seconds before the adjacent through movement phase, and

(3) exclusive pedestrian service in which pedestrians are served exclusively in all direction while all vehicular movements are halted. Figure 19 presents the phasing sequence of the three pedestrian service types.



a) Concurrent Pedestrian Service

b) Leading Pedestrian Service

c) Exclusive Pedestrian Service

Figure 19. Common Pedestrian Service Phasing Diagrams [1]

The three types of pedestrian service phasing diagrams consist of two intervals for the pedestrian phase: (1) *Walk interval* is the equivalent of the green interval for vehicles and is used to allow pedestrians to move from the curb onto the crosswalk; (2) *Pedestrian Clearance*, also referred to as flashing don't walk (FDW) or change interval: follows the walk interval and informs pedestrians should either complete their crossing if already in the intersection or wait until the next cycle is displayed. Finally, the pedestrian phase ends with the solid *Don't Cross*. The duration of the pedestrian phase, seen in Figure 20 (Walk interval + Clearance interval), is calculated using the following equation:

$$\boldsymbol{G_p = PW + PC}$$

Where:

$\boldsymbol{G_p}$ is the total duration needed for the pedestrian phase.

P.W. is the walk interval duration.

The Manual of Uniform Traffic Control Devices (MUTCD) indicates that the minimum walk duration should be at least 7 seconds but states that a duration as low as 4 seconds may be used if pedestrian volumes are negligible. [2].

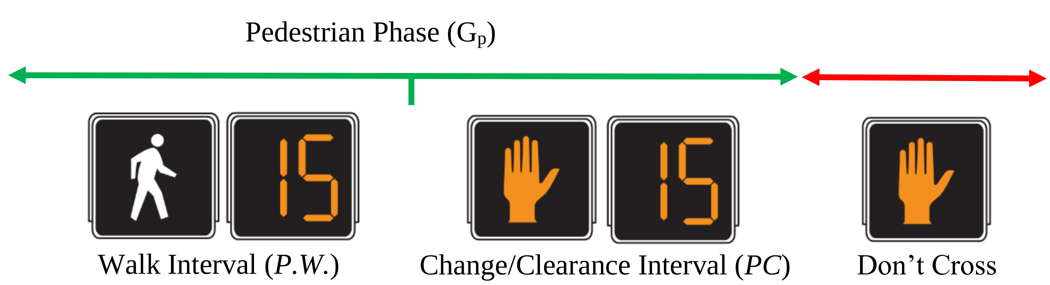


Figure 20. Pedestrian Phase Interval Sequence [2]

PC is the clearance/change interval duration. The duration of this interval is computed as the crossing distance divided by the walking speed. The MUTCD recommends a value of 4.0 feet per second (ft/s) walking speed. The Americans with Disabilities Act (ADA) Accessibility Guidelines for Buildings and Facilities recommended using 3.0 ft/s. Recent work completed by LaPlante and Kaeser has suggested a speed of 3.5 ft/s [2-4].

The walk interval (P.W.) should be designed to accommodate pedestrians' perception-reaction delay and walking time to the crosswalk. There are several factors that could delay a pedestrian in accomplishing this task. The social force model is widely used in defining the factors influencing pedestrian dynamics (i.e., avoiding obstacles and keeping a comfort zone away from other pedestrians). Such characteristics make a pedestrian take some time to exit the curb onto the crosswalk once the walk interval is activated [5,6]. Recent work by Nafakh et al. quantifies the pedestrian volume categories into four timing categories: (1) low pedestrian volume, (2) mid-low volume, (3) mid-high volume, and (4) high pedestrian volume. The suggested four categories are listed in Table 13 below.

Regarding signal timing, the walk interval (P.W.) accommodates the collective behavior of pedestrians waiting to cross. Therefore, the walk interval should provide enough time to allow all waiting pedestrians to move onto the crosswalk from the onset of the walk signal illumination. However, by providing a fixed duration, pedestrians do not always get such needed amount of time. As a result, traffic propagation is often impacted negatively for both pedestrians and vehicles, particularly at intersections with exclusive pedestrian service.

Table 13. Quantifying Pedestrian Walk Interval Classes [7]

Start-up Time (Walk-Interval)	Pedestrian Volume (peds/quad/cycle)			
	Percentile			
	25th	50th	75th	90th
Low Volume: 4 seconds	1	2	4	5
Mid-low Volume: 7 seconds	4	6	8	10
Mid-high Volume: 10 seconds	8	11	15	19.2
High Volume: 15 seconds	11	14	15	24.4

An ideal solution would be to provide a dynamic service in which the pedestrian walk interval is adjusted to accommodate the varying pedestrian volumes throughout a given day. The problem, however, an accurate prediction should be made regarding the expected number of pedestrians at different time intervals. Several studies have considered modeling pedestrian activities at both the macro and intersection levels. However, prediction models were often built based on aggregate counts and are meant for longer durations (i.e., the total number of pedestrians per day). In addition, these studies disregarded directionality (i.e., the total number of pedestrians per intersection instead of the total number of pedestrians per intersection crosswalk or intersection quadrant). Due to the limitations mentioned above, existing prediction models cannot aid in moving from a fixed signal timing to a dynamic timing scheme. A summary of the literature on pedestrian demand prediction models is presented in Table 14.

4.1.1 Study Motivation and Objective

There is a gap in the literature regarding modeling pedestrian demand at the intersection level. Existing studies consider pedestrian demand as an aggregate continuous variable per intersection. In addition, the existing micro-level pedestrian volume prediction models have not considered in quantifying pedestrian demand at finer levels, such as the quadrant level, that is, the total number of pedestrians produced per intersection quadrant. Pedestrian demand, therefore, has never been connected directly to signal timing, and hence the pedestrian walk interval is typically fixed. However, a fixed duration could be problematic, particularly at intersections with exclusive pedestrian service, mainly when pedestrian demand varies from when the walk interval was initially computed, resulting in situations where pedestrians and vehicles often experience needless delay.

Table 14. Summary of Pedestrian Volume Prediction Studies

	Study		
	Kim & Susilo, 2013 [8]	**Miranda-Moreno & Fernandes, 2011 [9]**	**Pulugurtha & Repaka, 2008 [10]**
Spatial Scope	Regional Level of Baltimore, Maryland	Intersection level of Montreal, Canada	Intersection level of Charlotte, North Carolina
Obs. Frequency	24-hour intervals	8 hours of counts on weekdays during the AM peak hour (6-9), the noon period (11-13), and the PM peak hour (15.30-18.30)	per hour between 7 AM and 7 PM
Collection	By interviews from April 2001 to 2002 (NHTS 2001)	Manual counts	Collected by the Charlotte Department of Transportation (CDOT)
Sample Size	3519 households	1018 signalized intersections	176 signalized Intersections
Approach and independent variables	Poisson regression and negative binomial models Variables included: age, driver, education, income, adult driver, residential density, non-residential unit density, degree of urbanism (non-residential/residential)	log-linear and negative-binomial. Data collected using three buffers: (1) 50-m, (2) 150-m, and (3) 400-m buffer. 400m population, 50m commercial space (m2), 150 m open space (m2), 150m subway, 150m bus stations, 400m schools, 400m %major arterials, 400m street segments, four-way intersections, ln(distance to downtown), very warm (>32C), very cold (<-20C), 400m population, 400m employment, 50m commercial space, 150m subway, 150m bus stations, 400m schools, 400m % major arterials, 400m street segments, four way intersections, wind speed, precipitation during period.	Multiple regression analysis through backward elimination Data were collected using three buffers 0.25 mi, 0.50 mi, and 1 mi. Variables included: household units, population, mean income, total employment, vehicular volume, number of lanes, speed limit, presence of median, number of approaches, transit stops and 22 different categories used.

The current study seeks to answer the following question: Which factors influence pedestrian volumes most and how could these variables be used to forecast demand at the quadrant level throughout a given day and then connect it to signal timing accordingly? First, the study suggests classifying pedestrian volumes and treating pedestrian demand thus as classes instead of a continuous variable. Next, the study recommends using machine learning classification models with time series variables to predict pedestrian demand at the intersection quadrant level per 15-minute intervals.

4.2 Methodology

Pedestrian volume data were collected for each intersection quadrant and classified into the four categories of the walk interval established by Nafakh et al. (2022). Then descriptive data within walking distance of each intersection quadrant were collected and used to build two machine learning classification models. The first model uses a random forest classifier, and the second is a XGBoost classification. Both models use time and lag variables to account for the time series nature of the data. Figure 21 presents the overall framework of this specific study.

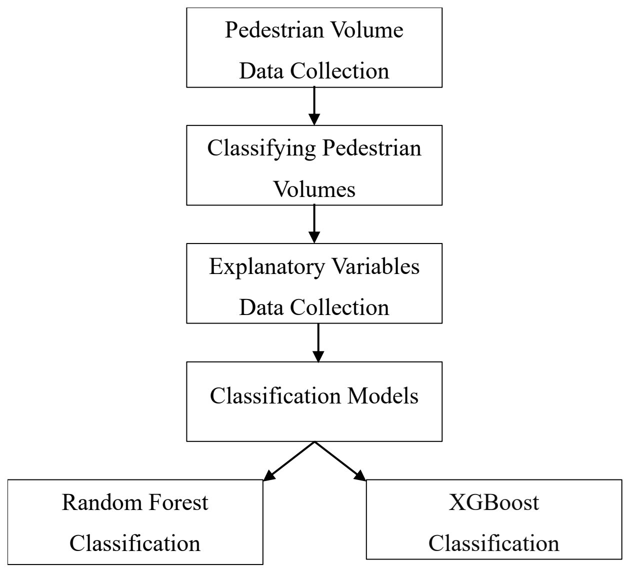
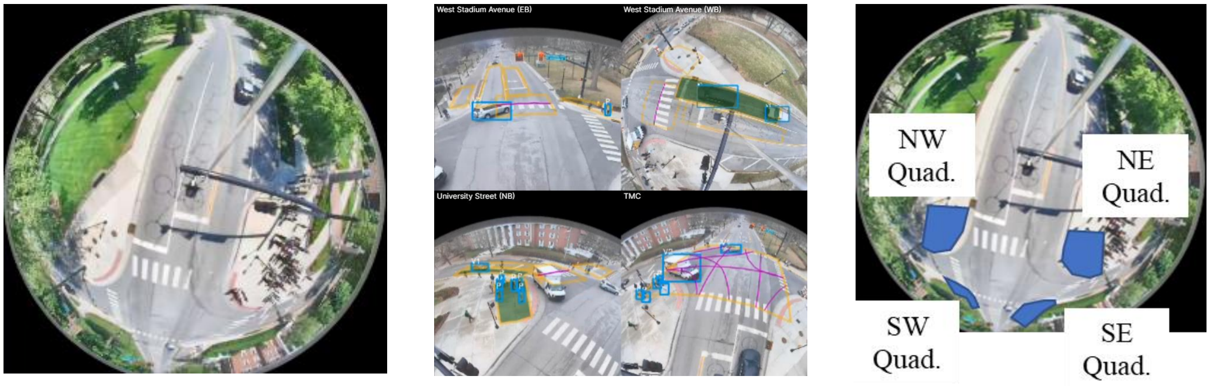


Figure 21. Study Framework

4.2.1 Pedestrian Data Collection and Processing

Pedestrian volume data used in this study were collected per intersection quadrant at 15-minute intervals from 13 cameras at signalized intersections providing exclusive pedestrian service. Data were collected between June 2021 and December 2022 in the City of West Lafayette, Indiana, making the total number of intervals per intersection quadrant per day 96 and per analysis period 55,584. Figure 22 and Table 15 present the camera setup and locations of the 13 intersections used in this study. The cameras were permanently mounted on the traffic light mast arms providing a complete view of all intersection approaches and quadrants. The cameras recorded continuously since the day of installation and allowed for automatic data extraction of volume counts of all roadway users in real-time.

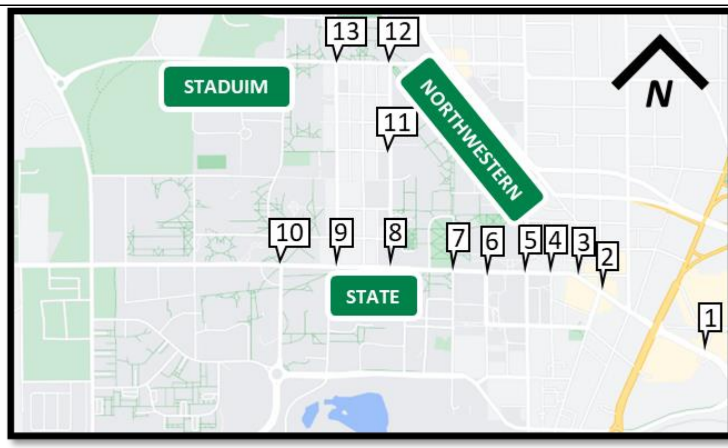
The conversion of video imagery to meaningful data (pedestrian counts per intersection quadrant, in this case) was achieved using Computer Vision algorithms. Computer Vision is a field of artificial intelligence (A.I.) that enables computers and systems to derive meaningful information from digital images, videos, and other visual inputs [11]. In each 15-minute interval, data are aggregated for vehicles' travel direction and per intersection quadrant for pedestrians. Figure 22b shows the detection/tracking view of the cameras used to collect pedestrian volume data. Figure 22c indicates the four quads for which the pedestrian volumes were aggregated.



a) Camera view b) Camera detection/tracking view c) Intersection Quads

Figure 22. Camera Installed at W. Stadium Ave and University St. (Intersection #13)

Table 15. Intersections where Cameras are Mounted at Exclusive Service Intersections.



Intersection		Location	
		Lat	***Long***
1	Roebuck Drive and State Street	40.4212	–86.9019
2	State Street and Chauncey Avenue	40.4233	–86.9069
3	Northwestern Avenue and State Street	40.4240	–86.9082
4	State Street and Andrew Place	40.4240	–86.9092
5	South Grant Street and State Street	40.4239	–86.9103
6	State Street and Sheetz Street	40.4240	–86.9122
7	State Street and Marsteller Street	40.4241	–86.9138
8	State Street and University Street	40.424	–86.9168
9	State Street and Russell Street	40.4242,	–86.9191
10	State Street and S. Martin Jischke Drive	40.4242	–86.9217
11	University Street and 3rd Street	40.4272	–86.9166
12	West Stadium Avenue and University Street	40.4313	–86.9168
13	Russell Street and West Stadium Avenue	40.4242	–86.9191

After collecting data on pedestrian volumes per 15-minute intervals per intersection quadrant for the analysis period, the study classified the volumes using the Nafakh et al. (2022)-recommended values for the walk interval: (1) 4 seconds for low volumes, (2) 7 seconds for mid-low volumes, (3) 10 seconds for mid-high volumes, and (4) 15 seconds for high volumes per cycle. The volumes were aggregated per 15-minute intervals; therefore, the following steps were used to adjust the walk interval values from per cycle to per 15-minute intervals:

Step 1: Cycle length = number of phases*30 seconds (i.e., for an intersection with 4 phases, the cycle length is assumed to be 120):

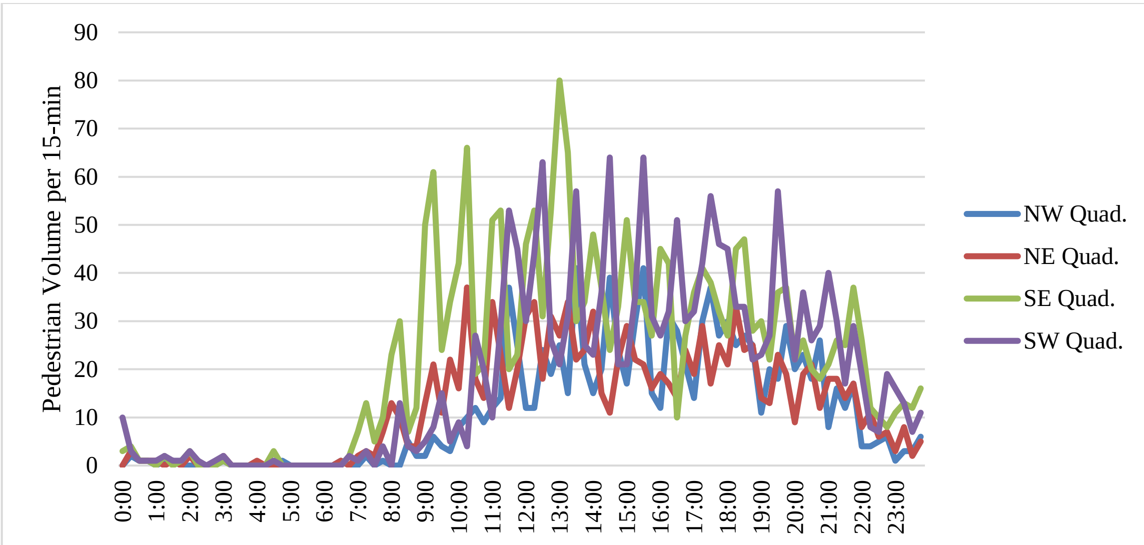
Step 2: Pedestrian volume per 15 minutes based on an assumed number of phases per intersection using Nafakh et al. (2022)'s volumes for the four Walk-Interval categories:

1. Low volume requiring 4 seconds:
   1 pedestrian *(15-min interval / (120 seconds /60 seconds)) = 7.5 pedestrians or below per 15-min interval
2. Mid-low volume requiring 7 seconds:
   4 pedestrians *(15-min interval / (120 seconds /60 seconds)) = between 8 and 30 pedestrians per 15-min interval
3. Mid-high volume requiring 10 seconds:
   8 pedestrians *(15-min interval / (120 seconds /60 seconds)) = between 31 and 60 pedestrians per 15-min interval
4. High volume requiring 15 seconds:
   11 pedestrians *(15-min interval / (120 seconds /60 seconds)) = above 61 pedestrians per 15-min interval

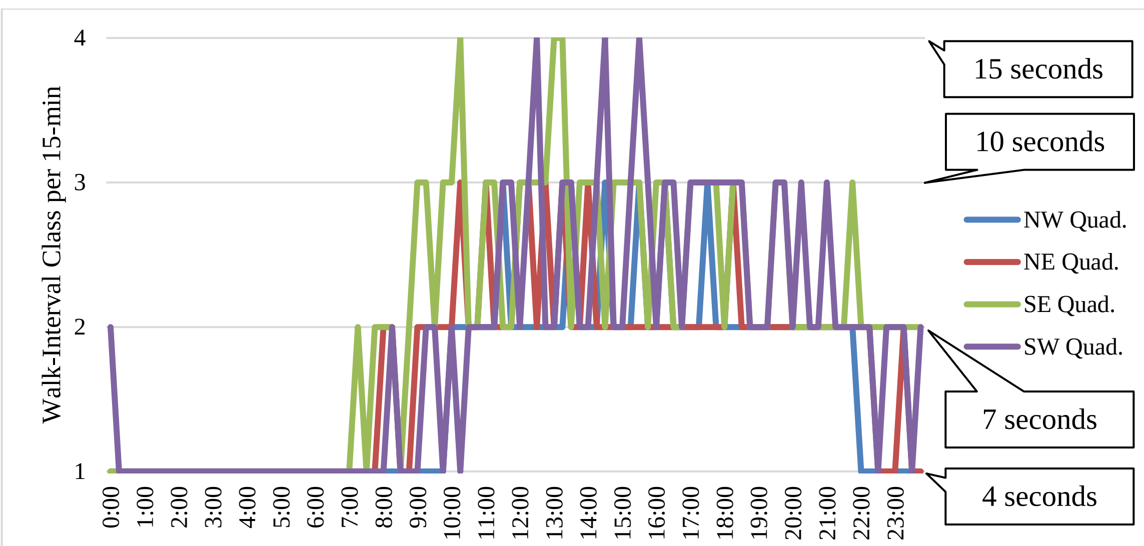
Table 16. Intersection Phasing & Corresponding Walk-Interval Thresholds per 15-min Intervals

#	Intersection	Pedestrian Service Type	Phases	Nafakh et al., 25th %tile per 15-minute values (peds/15-min)	Nafakh et al., 50th %tile per 15-minute values (peds/15-min)
1	Roebuck Dr & State St	Exclusive	3	10, 40, 80	20, 60, 110
2	State St & Chauncey Ave	Exclusive	4	7.5, 30, 60	15, 45, 82.5
3	NW Ave & State St	Exclusive	3	10, 40, 80	20, 60, 110
4	State St & Andrew Pl	Exclusive	4	7.5, 30, 60	15, 45, 82.5
5	S Grant St & State St	Exclusive	4	7.5, 30, 60	15, 45, 82.5
6	State St & Sheetz St	Exclusive	3	10, 40, 80	20, 60, 110
7	State St & Marsteller St	Exclusive	4	7.5, 30, 60	15, 45, 82.5
8	State St & University St	Exclusive	3	10, 40, 80	20, 60, 110
9	State St & Russell St	Exclusive	4	7.5, 30, 60	15, 45, 82.5
10	State St & Martin J Dr	Exclusive	4	7.5, 30, 60	15, 45, 82.5
11	University St & 3rd St	Exclusive	3	10, 40, 80	20, 60, 110
12	Stadium Ave & Univ. St	Exclusive	4	7.5, 30, 60	15, 45, 82.5
13	Russell St & W Stad. Ave	Exclusive	4	7.5, 30, 60	15, 45, 82.5

The walk-interval values were established on a campus town consisting of mainly college-aged pedestrians. Therefore, it is safe to use the conservative 25th percentile volumes to account for the age bias. Therefore, the classification considered in this study uses the 25th percentile values. Figure 23 below presents an example of the conversion from actual volumes to classes for State Street and Chauncey Avenue (#2) on Saturday, September 25, 2021.



a) Pedestrian Volumes per 15-minute Interval per Quadrant



b) Pedestrian volume classes per 15-minute interval per quadrant

Figure 23. Moving from Volume to Classes (Example)

4.2.2 Explanatory Variables -- Data Collection

To forecast pedestrian walk-interval classes, explanatory data were collected. To the best of the authors knowledge, all previous work on pedestrian volume forecasting have used circular buffers within which explanatory variables were collected and then used to explain pedestrian volumes [12-17]. However, in this study, to account for what a pedestrian can realistically reach, walk buffers are used instead of the circular buffers to collect explanatory variables.

Walk buffers are irregular shapes replicating the area a pedestrian can walk within using existing sidewalk infrastructure. Explanatory variables are collected within three walk buffers of each intersection quadrant to explain and aid in classifying pedestrian volumes. The three buffers are a 5-minute walk buffer, a 10-minute walk buffer, and a 15-minute walk buffer. The walk buffers are constructed using the Esri ArcGIS Pro Network Analysis Service Area tools built based on the roadway network shape file obtained through OpenStreetMaps of the State of Indiana [18,19]. In addition, each buffer was constructed assuming a pedestrian speed per The Americans with Disabilities Act (ADA) Accessibility Guidelines for Buildings and Facilities recommended speed of 3.0 ft/s. Figure 24 presents the three buffers for each intersection.

Within each walking buffer, two types of explanatory variables were collected to explain the variability in pedestrian volume classes: (1) time-dependent and (2) fixed variables. Data for each variable were collected per walk buffer per intersection quadrant. Figure 25 presents a closer look at the intersection quadrant level of detail, and Table 17 presents the variables considered in this study to explain and predict pedestrian volumes.

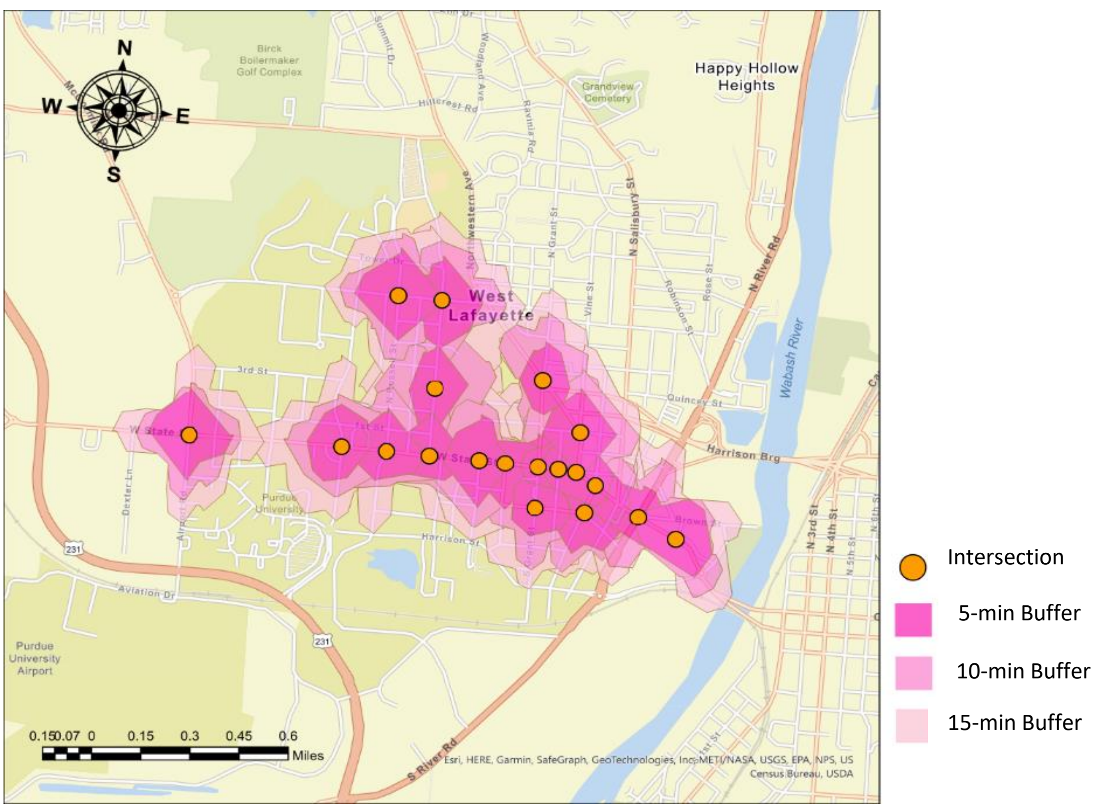


Figure 24. Walk Buffers Around Campus Intersections

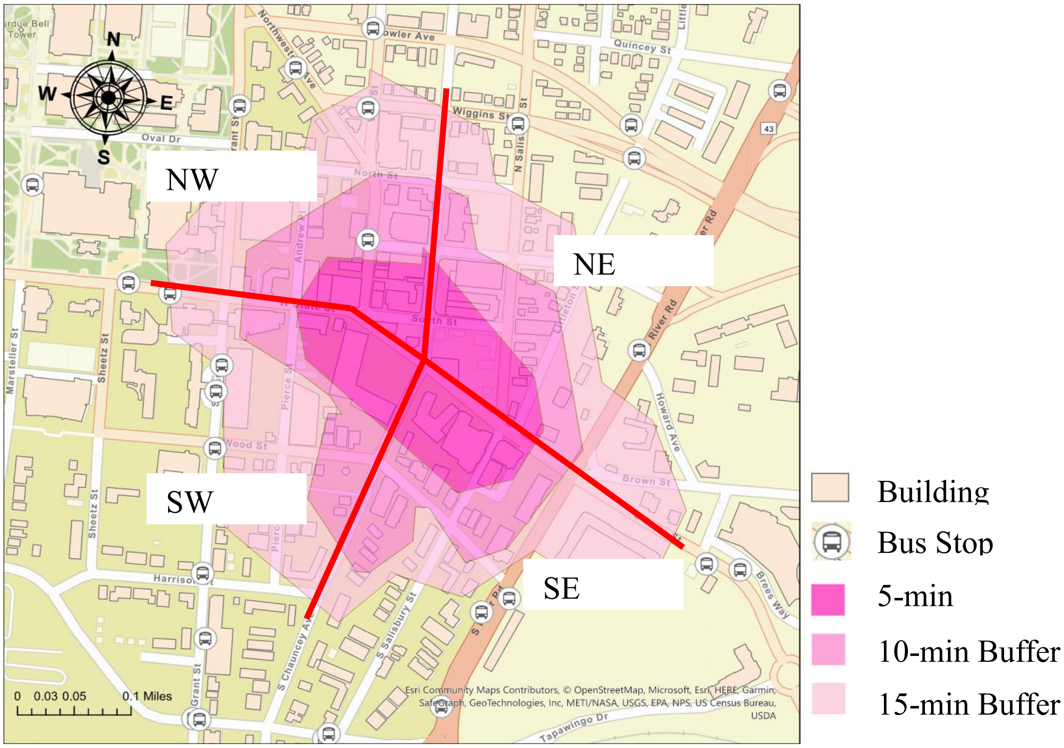


Figure 25. State Street and Chauncey Avenue (#2) Quadrant Walk Buffers

The academic calendar variables were obtained through the Purdue University Office of the Registrar website. Bus stop locations obtained through CityBus shapefiles website. Weather data variables were obtained through the National Oceanic and Atmospheric Administration (NOAA) website. The land use and land use hours were obtained through OpenStreetMaps shapefiles, Google Maps hours, and then verified through site visits [20-22].

Table 17. Study Explanatory Variables

Variable	Type	Observation Frequency	Unit
Day of month	Time-dependent	1-31	Per day
Time index	Time-dependent	96 times per day (15-minute intervals)	1-96
Day of week	Time-dependent	Once every 24 hours	1-7
Month of the year	Time-dependent	12 times per year	1-12
Semester index	Time-dependent	Four times per year	1. Summer 2. Fall 3. Christmas 4. Spring
Special events index	Time-dependent	On game and graduation days	1. Basketball 2. Football 3. Graduation Day
Vacation index	Time-dependent	On academic vacation days	0 or 1
Bus Stops	Fixed	-	Number of bus stops within walking buffer
Precipitation	Time-dependent	15-minute intervals	Inches
Temperature Min	Time-dependent	15-minute intervals	Fahrenheit
Temperature Max	Time-dependent	15-minute intervals	Fahrenheit
Traffic	Time-dependent	15-min intervals	Total vehicles
Dorms	Time-dependent	Per Semester	SQFT
Dining courts	Time-dependent	15-minute intervals	SQFT
Gym	Time-dependent	15-minute intervals	SQFT
Libraries	Time-dependent	15-minute intervals	SQFT
University buildings	Fixed	-	SQFT
Restaurants	Time-dependent	15-minute intervals	SQFT
Bars	Time-dependent	15-minute intervals	SQFT
Other Commercial	Time-dependent	15-minute intervals	SQFT
Single Unit Residential	Fixed	-	SQFT
Residential buildings	Fixed	-	SQFT
Other	Fixed	-	SQFT
Religious Facilities	Fixed	-	SQFT
Hotels	Fixed	-	SQFT
Lag Variable	Time-dependent	15-minute intervals	SQFT

4.2.3 Machine Learning Classification Models

Given that the response variable is a categorical variable with four possible outcomes: (1) low pedestrian volume, (2) mid-low volume, (3) mid-high volume, and (4) high volume, classification models are most appropriate for prediction [23]. It is possible to use multiple classification techniques, or classifiers, to predict a categorical response. In this research, two approaches were used to forecast pedestrian demand (1) machine learning random forest classification and (2) machine learning XGBoost classification.

The two classification models are used to predict pedestrian volume class using the collected explanatory variables, which use 80% of the data set for training and 20% for testing. The training data set consists of 2,134,426 observations, and the testing data set of 533,606 observations. The classifiers were used following the steps seen in Figure 26.

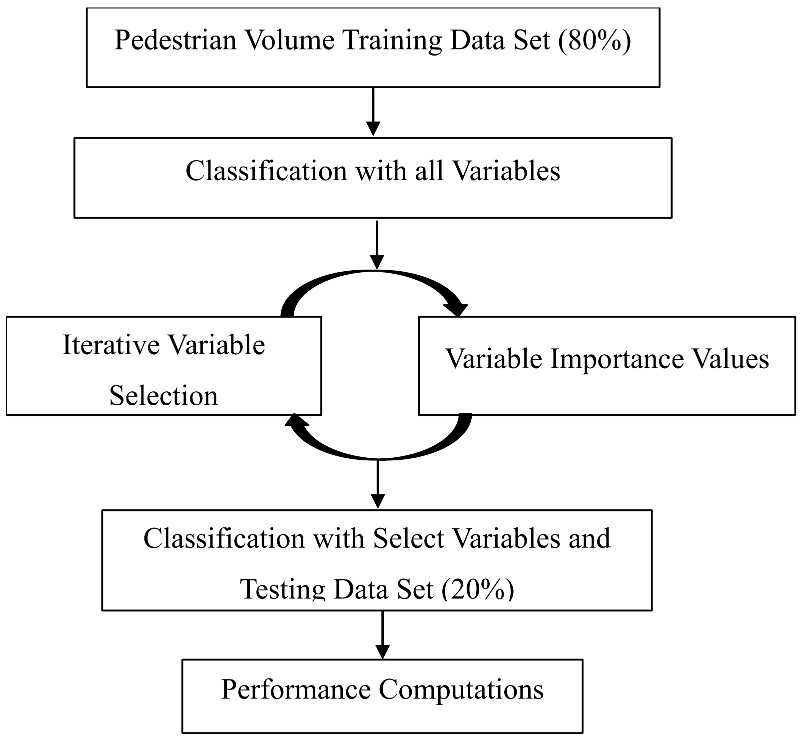


Figure 26. Classification Model Steps

Classifiers were built for each walk buffer: (1) one for the 5-minute walk buffer data, (2) one for the 10-minute walk buffer data, and (3) one for the 15-minute walk buffer. The three models for each approach were then compared using the confusion matrix, normalized confusion matrix, sensitivity, specificity, precision, negative predictive value, accuracy, and variable normalized importance (equations shown as follows): .

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{Fulse Negative}}$$

$$\text{Specificity} = \frac{\text{True Negative}}{\text{True Negative} + \text{Fulse Postivie}}$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{Fulse Negative}}$$

$$\text{Negative Predictive Value} = \frac{\text{True Negative}}{\text{True Negative} + \text{Fulse Negative}}$$

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Total Observations}}$$

4.3 Summary Statistics of the Data

Between June 1st, 2021, and December 31st, 2022, a total of 35,620,413 pedestrian movements were recorded, with an average campus 15-minute pedestrian volume of 640.84, a minimum volume of 0, and a maximum volume of 7509. The standard deviation was 852.77, with 80% of observed intervals below 500 pedestrian movements per 15 minutes. Figure 27 presents an overview of the aggregate 13 intersections activities through the analysis period. Of the 13 intersections, University St. and 3rd St. (#11), located in the center of campus and close to the Mackey Basketball Arena and Ross-Ade football stadium, had the highest average pedestrian activities at 85.61 pedestrians per 15-minutes, the interval with the maximum number of pedestrians throughout the analysis period at 1780, and the highest volume standard deviation at 170.54. Table 18 presents the descriptive statistics of the 13 intersections. After all volumes had been classified into one of the four classes: (1) low volume, (2) mid-low volume, (3) mid-high volume, (4) high volume, the frequency of observations is listed in Table 19.

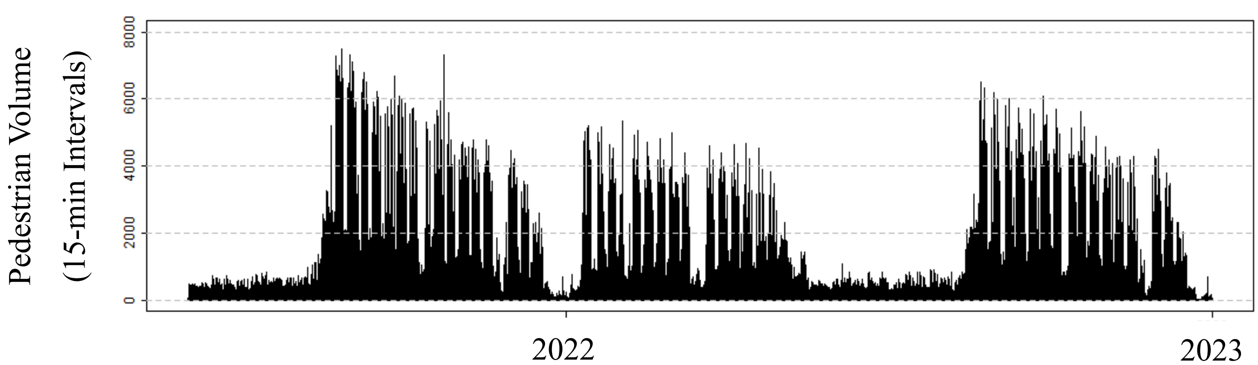


Figure 27. Analysis Period Campus Pedestrian Activities View

Table 18. Descriptive Statistics of Pedestrian Volumes at Exclusive Service Intersections

Int.	Min	1st Q	Median	Mean	3rd Q	Max	Std. Dev.
1	0.00	1.00	5.00	6.78	10.00	148.00	6.88
2	0.00	7.00	36.00	53.44	84.00	517.00	55.84
3	0.00	8.00	47.00	67.59	110.00	638.00	69.66
4	0.00	5.00	30.00	61.10	91.00	668.00	76.73
5	0.00	5.00	28.00	60.62	84.00	714.00	82.97
6	0.00	2.00	12.00	31.56	39.00	544.00	50.92
7	0.00	3.00	15.00	39.65	47.00	900.00	69.95
8	0.00	2.00	13.00	32.94	39.00	758.00	56.94
9	0.00	2.00	10.00	24.79	31.00	653.00	39.74
10	0.00	2.00	8.00	17.43	23.00	341.00	24.99
11	0.00	2.00	16.00	85.61	81.00	1780.00	170.54
12	0.00	1.00	6.00	26.96	25.00	1330.00	54.59
13	0.00	1.00	8.00	30.60	42.00	1193.00	51.17

Table 19. Pedestrian Volume Class Frequency

Class	Observations	Percentage
1 (low volume)	1,766,197	66.20%
2 (mid-low volume)	576,623	21.61%
3 (mid-high volume)	205,208	7.69%
4 (high volume)	120,004	4.50%

4.4 Results

4.4.1 Time Series Analysis

The time series decomposition plot in Figure 28 presents the observed aggregate campus values and the trend using a moving average. Aggregate campus autocorrelation function values (ACF) measuring the correlation of values with their former time step values (lags) are above 0.600 for the first 13 lags, where each lag is the aggregate pedestrian volume for a 15-minute interval. On average, the maximum correlation occurs across the full analysis period at the fourth lag at a correlation of 0.881. In addition, a clear descending pattern is evident across the 50 lags, moving from a positive to a negative correlation around the 30th lag. Figure 29 presents a plot of the first 50 lag intervals

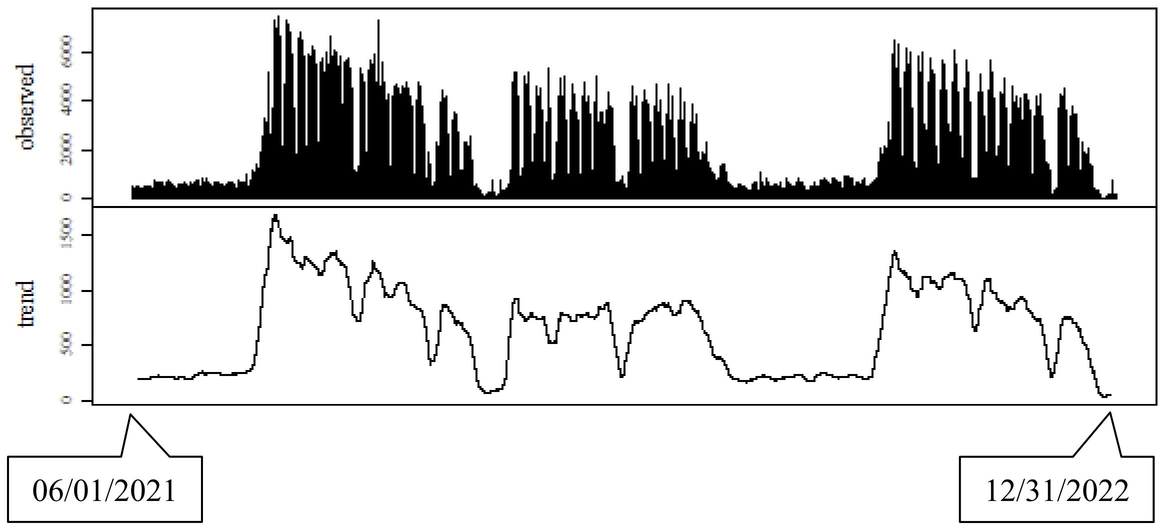


Figure 28. Decomposition of Campus Wide Additive Time Series

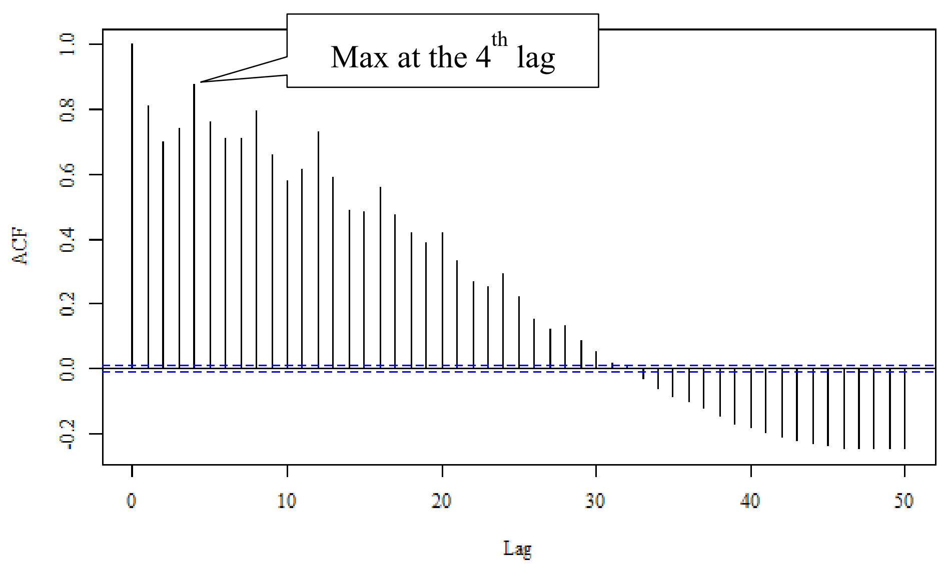


Figure 29. Aggregate Campus Pedestrian Volume Autocorrelation

4.4.2 Classification Models

Variable selection plays a vital role in classification models. The variables selected in this study were based on the variable importance values of an initial model, including all collected variables for the three walk buffer models: (1) 5-minute walk buffer classifier, (2) 10-minute walk buffer classifier, and (3) 15-minute walk buffer classifier. The final variables selection consists of (1) month of year, (2) Time index, (3) lag (the highest correlated lag, that is the 4th lag value of the response), (4) vehicular traffic volume, (5) restaurants, (6) residential buildings, (7) university buildings, (8) number of bus stops, (9) temperature minimum, (10) temperature maximum, (11) precipitation, (12) vacation index, (13) semester index, (14) day of week, (15) day of month. Figure 30 below presents the normalized variable importance values of the final model variables for all three buffers. The three models' confusion matrix and weighted confusion matrix, showing true positive, true negative, false positive, and false negative for each class of the four walk-interval classes, are provided in Table 20 below. In addition, the three models are compared using model precision, negative predictive value, sensitivity, specificity, and accuracy. Figure 31 presents the values of each test for the three models.

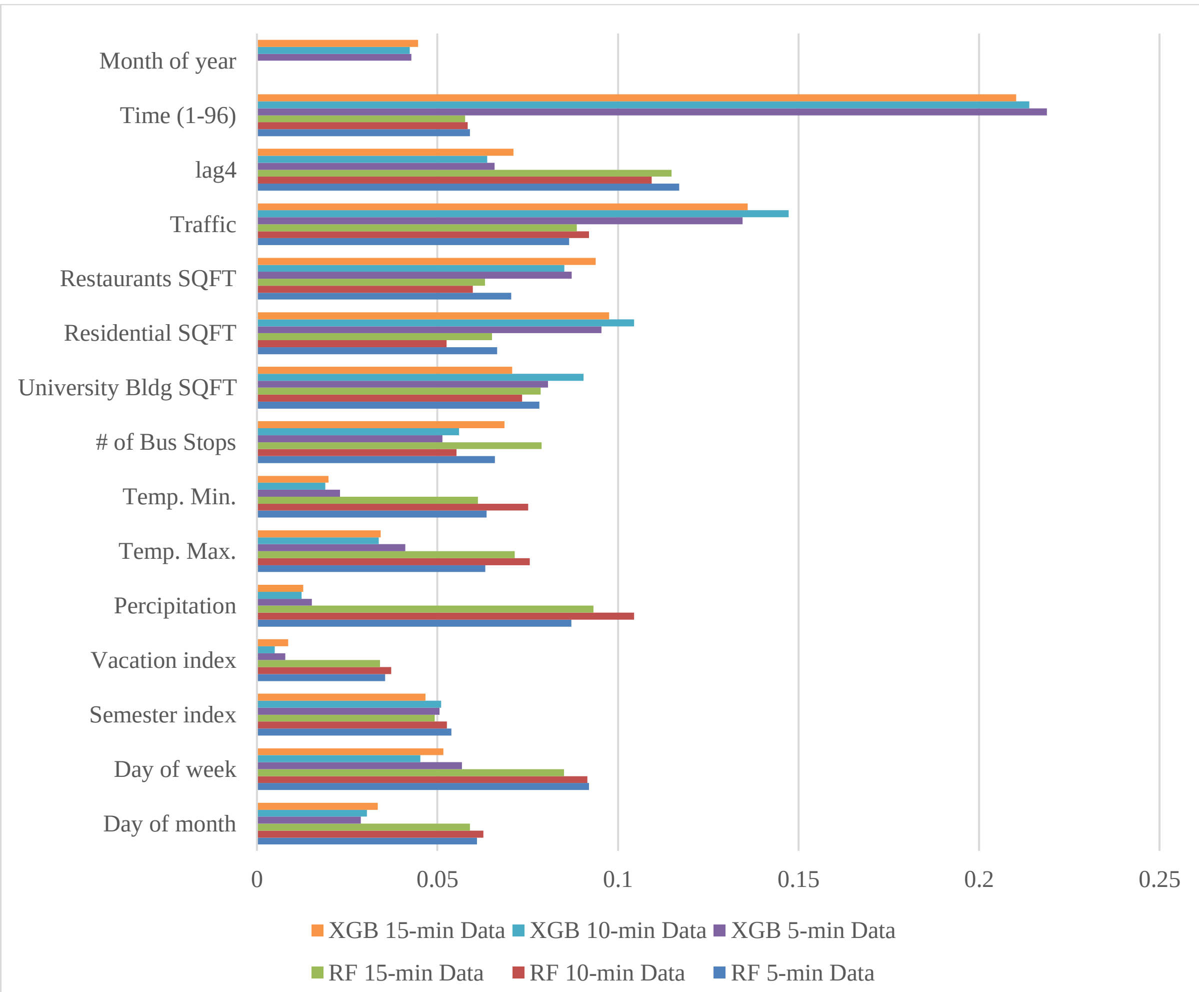


Figure 30. Selected Variables Normalized Importance per Classification Model

Table 20. Confusion and Weighted Confusion Matrices

		Actual			
		1	2	3	4
Predicted	1	327650	32050	1012	301
	2	23645	73972	18210	2352
	3	320	8722	17405	6390
	4	73	920	5627	14957

a) 5-min data RF Matrix

		Actual			
		1	2	3	4
Predicted	1	0.91	0.09	0.00	0.00
	2	0.20	0.63	0.15	0.02
	3	0.01	0.27	0.53	0.19
	4	0.00	0.04	0.26	0.69

b) 5-min data RF Weighted Matrix

		Actual			
		1	2	3	4
Predicted	1	328654	30622	1020	260
	2	23096	75199	18228	2833
	3	382	10331	15032	6689
	4	109	974	5466	14711

c) 10-min data RF Matrix

		Actual			
		1	2	3	4
Predicted	1	0.91	0.08	0.00	0.00
	2	0.19	0.63	0.15	0.02
	3	0.01	0.32	0.46	0.21
	4	0.01	0.05	0.26	0.69

d) 10-min data RF Weighted Matrix

		Actual			
		1	2	3	4
Predicted	1	328229	29849	1068	278
	2	23869	75159	18026	2689
	3	359	9947	15111	7068
	4	98	898	5760	15198

e) 15-min data RF Matrix

		Actual			
		1	2	3	4
Predicted	1	0.91	0.08	0.00	0.00
	2	0.20	0.63	0.15	0.02
	3	0.01	0.31	0.47	0.22
	4	0.00	0.04	0.26	0.69

f) 15-min data RF Weighted Matrix

		Actual			
		1	2	3	4
Predicted	1	329978	29819	975	241
	2	22586	76102	17298	2193
	3	277	8401	17835	6324
	4	66	844	5311	15356

g) 5-min data XGBoost Matrix

		Actual			
		1	2	3	4
Predicted	1	0.91	0.08	0.00	0.00
	2	0.19	0.64	0.15	0.02
	3	0.01	0.26	0.54	0.19
	4	0.00	0.04	0.25	0.71

h) 5-min data XGBoost Weighted Matrix

		Actual			
		1	2	3	4
Predicted	1	329809	29553	971	223
	2	22748	76686	17590	2332
	3	297	8033	17571	6533
	4	53	894	5287	15026

i) 10-min data XGBoost Matrix

		Actual			
		1	2	3	4
Predicted	1	0.91	0.08	0.00	0.00
	2	0.19	0.64	0.15	0.02
	3	0.01	0.25	0.54	0.20
	4	0.00	0.04	0.25	0.71

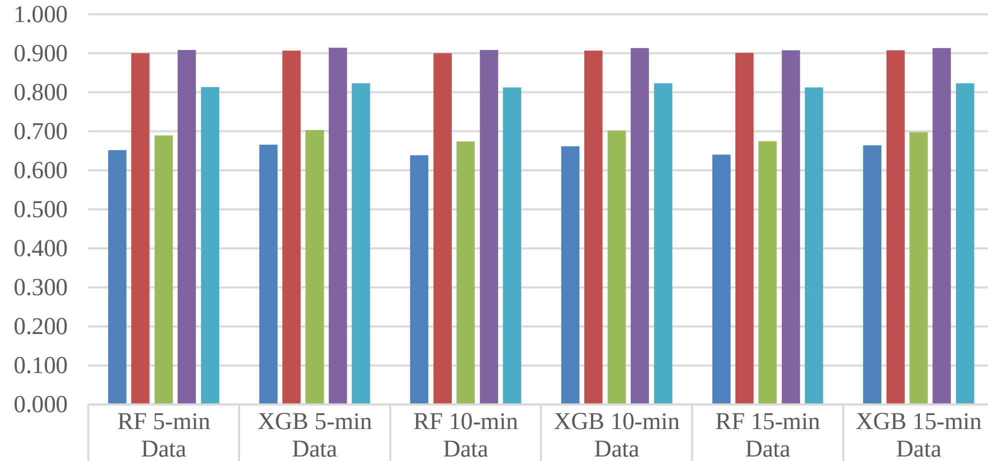
j) 10-min data XGBoost Weighted Matrix

		Actual			
		1	2	3	4
Predicted	1	329420	28852	930	222
	2	23087	76927	17405	2324
	3	319	8440	17425	6301
	4	81	947	5659	15267

k) 15-min data XGBoost Matrix

		Actual			
		1	2	3	4
Predicted	1	0.92	0.08	0.00	0.00
	2	0.19	0.64	0.15	0.02
	3	0.01	0.26	0.54	0.19
	4	0.00	0.04	0.26	0.70

l) 15-min data XGBoost Weighted Matrix



	RF 5-min Data	XGB 5-min Data	RF 10-min Data	XGB 10-min Data	RF 15-min Data	XGB 15-min Data
Precision	0.652	0.666	0.638	0.662	0.640	0.664
Negative Predicitive Value	0.900	0.907	0.900	0.907	0.901	0.907
Sensetivity	0.689	0.703	0.674	0.701	0.675	0.698
Specificity	0.909	0.914	0.908	0.914	0.908	0.913
Accuracy	0.813	0.823	0.813	0.823	0.813	0.823

Figure 31. Random Forest Classification Performance Measures

4.5 Discussion and Conclusion

It is apparent that the trend in the number of pedestrians, which tends to be higher during academic semesters compared to off-school periods, shows a clear association with the university academic calendar, as seen in Figure 32. This association also indicates the time series nature of pedestrian volumes, reinforced by the autocorrelation function plot, as seen in Figure 29. Therefore, time variables were of clear association and included as variables in the models trained. Categorizing pedestrian volumes per the walk-interval categories changed the volume prediction problem to a classification problem. Hence, a machine learning random forest and XGBoost models were trained to predict pedestrian volumes and proved to do so accurately. The early data processing and time series analysis proved to be beneficial in pointing out explanatory data that eventually aided in building an accurate prediction model. Although three geospatial scales of explanatory variables were used, all six developed models proved to accurately predict the volume category with a minimum accuracy of 81.3%. Accuracy is a good measure of the overall fit of the developed models; however, sensitivity should also be considered in scenarios where safety is the main concern, like the pedestrian crossing context. Sensitivity in this case would reflect a measure of how many instances the model was able to accurately predict demand. In other words, the model's ability to predict positive cases. The minimum sensitivity of the six models is at 67% indicating a good prediction and forecasting ability, as seen in Figure 33. The final set of explanatory variables used can be summarized into four main groups: (1) time variables, (2) weather variables, (3) transit variables, and (4) land use variables. Furthermore, with the three different geospatial data models providing similar accuracy, the transferability and applications of this methodology can be easily implemented at other places with only one defined geospatial scale for data collection (i.e., 5-minute walk buffer scale) hence easing and speeding up the process.

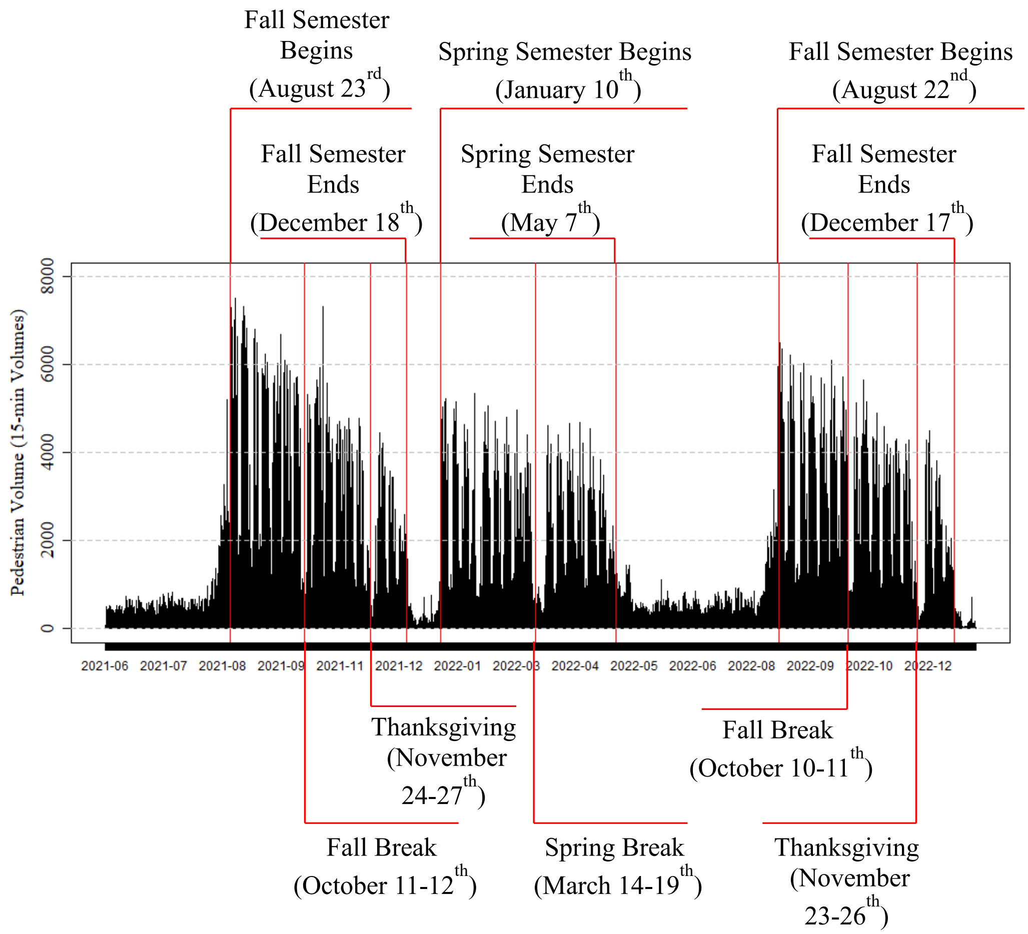
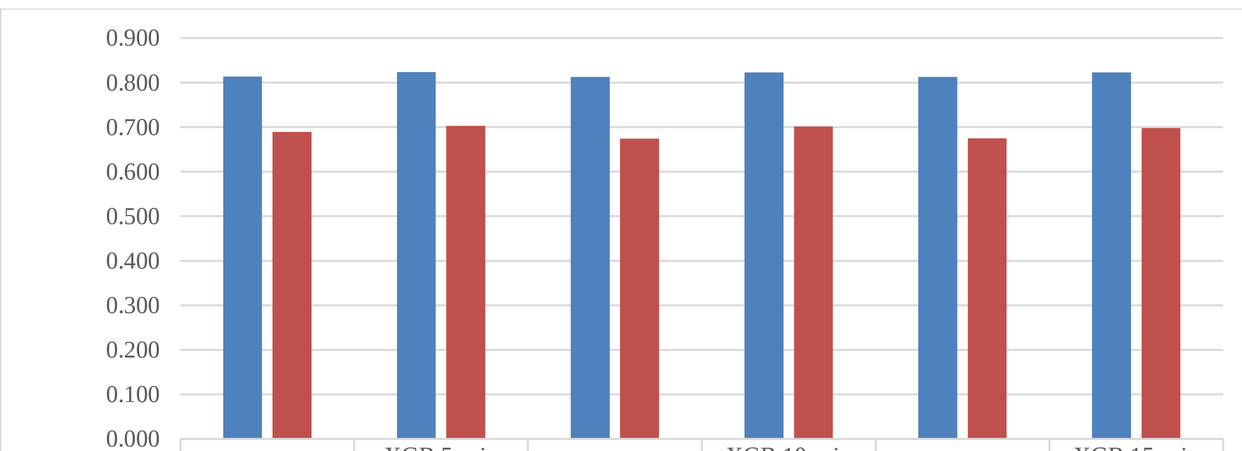


Figure 32. Aggregate Pedestrian Activities and University Calendar



	RF 5-min Data	XGB 5-min Data	RF 10-min Data	XGB 10-min Data	RF 15-min Data	XGB 15-min Data
Accuracy	0.813	0.823	0.813	0.823	0.813	0.823
Sensitivity	0.689	0.703	0.674	0.701	0.675	0.698

Figure 33. Classification Accuracy per Model

In conclusion, this study demonstrates the effectiveness of machine learning classification models in predicting pedestrian volumes using time series, weather, transit, and land use variables. By analyzing a large dataset of pedestrian volumes and corresponding explanatory variables, a model has been developed to reliably predict pedestrian volumes for a given location per 15-minute intervals with an accuracy of 82.3%. The findings suggest that time series variables, such as weather conditions and time of day, play a critical role in predicting pedestrian volumes, while land use variables, such as proximity to public transport and commercial areas, also contribute significantly.

The significance of this study lies in its potential to improve pedestrian safety and mobility by providing accurate pedestrian volume predictions, which can inform signal timing design at intersections. The developed model can help traffic engineers and city planners optimize signal timings thereby reducing pedestrian wait times and ensuring pedestrian safety by minimizing pedestrian-vehicle exposure. The ability to predict pedestrian volumes accurately can also aid in the planning of future developments and public transportation infrastructure by identifying high-traffic areas where pedestrian facilities and amenities may be needed.

Overall, the results of this study demonstrate the power of machine learning in improving transportation planning and management. By incorporating time series and land use variables into a random forest classification model, we have shown that it is possible to predict pedestrian volumes accurately, providing valuable insights into pedestrian behavior that can inform signal timings and infrastructure planning. This research highlights the importance of using data-driven approaches to tackle transportation-related challenges, ultimately leading to safer, more sustainable, and more efficient urban environments in the current era of HDVs and in the prospective era of CAVs or mixed stream.

4.6 Chapter Summary

This chapter discussed the effectiveness of machine learning classification models in predicting pedestrian volumes using time series, weather, transit, and land use variables. By using a large dataset of pedestrian volumes and corresponding explanatory variables, the study develops a model that accurately predicts pedestrian volumes for a given location per 15-minute intervals per intersection quadrant with high accuracy.

CHAPTER 5 CONCLUDING REMARKS

The work presented in this report explores the critical role that emerging data on pedestrian demand and behavior can play in optimizing signal timings and improving efficiency at signalized intersections. The data used in this framework is obtained through permanently mounted cameras at 19 signalized intersections, automatically counting pedestrians and traffic in real time. By quantifying pedestrian demand and identifying key factors that influence pedestrian volumes, traffic engineers can use this framework to understand the needs and behaviors of pedestrians better and improve intersection performance accordingly.

This study used historical data on pedestrian volumes surrounding environment, and signal timings to develop accurate machine-learning models that can forecast pedestrian demand and optimize signal timings to prioritize pedestrian needs. Summaries of the three key aspects of this report, are provided below.

5.1 Objectively Quantifying the Pedestrian Walk Interval

This part of the study quantitatively analyzes the pedestrian walk interval duration based on varying pedestrian volumes at 12 signalized intersections through ten months. In addition, data on the storage area and offset from the pedestrian push button to the crosswalk were used to explain the variability in pedestrian start-up time. As a result, the built statistical model can help designers in identifying proper walk interval timing on an intersection-by-intersection basis. In addition, designers now have quantitative data for new construction to support prioritizing close-to-crosswalk push-button locations to help minimize pedestrian start-up time.

5.2 Time Series Analysis of Pedestrian Activities

This part of the study presents an aggregate overview of network-level pedestrian activities on a campus town over a prolonged period. The aggregate campus pedestrian volume over the analysis period has several useful features. Most apparent is the trend in the number of pedestrians, which tends to be higher during academic semesters than off-school periods showing a clear association between pedestrian activities and the university academic calendar. This study also points out time-series significant correlation and cross-correlation values indicating how former timesteps can be used to forecast present or future demand. In addition, intersections with cameras can be used to predict demand at nearby intersections without cameras. Finally, this study concluded with an approach to measure the impact of special events, such as graduation ceremonies and sporting events, on pedestrian activities across the network. Detailed information on pedestrian demand at the network level provides city planners and traffic engineers with accurate performance measures at both the network and intersection levels allowing for proper assessment and budget allocation for pedestrian infrastructure based on need.

5.3 Machine Learning Algorithms Forecasting Needed Pedestrian Walk-Interval

This part of the study demonstrates the effectiveness of machine learning classification models in predicting pedestrian volumes using time series, weather, transit, and land use variables. By analyzing a large dataset of pedestrian volumes consisting of approximately 23 million observations and corresponding explanatory variables, the developed model accurately forecasts pedestrian volumes for a given location per 15-minute intervals with an accuracy of 82.3%. The findings suggest that time series variables, such as weather conditions and time of day, and land use variables, such as proximity to public transport and commercial areas, play a critical role in predicting pedestrian volumes. This framework can help traffic engineers and city planners optimize signal timings, reducing pedestrian wait times and ensuring pedestrian safety by minimizing the risk of pedestrian-vehicle exposure. Reliable prediction of pedestrian volumes can also aid in planning future developments and public transportation infrastructure by identifying high-traffic areas where pedestrian facilities and amenities may be needed.

5.4 Overall Summary, Limitations, and Future Research Directions

The results of this report indicate that large-scale data on pedestrian behavior and volumes is now a reality that should be utilized to ensure equitable service. The findings of studies included in this report indicate that several key factors can significantly impact pedestrian behaviors and volumes at signalized intersections, including built environment features such as the location of the push-button, the time of day, day of the week, weather conditions, and proximity to public transit and other pedestrian-oriented destinations. By incorporating these factors into machine learning classification models, the developed framework could reliably forecast pedestrian demand, allowing for optimal signal timings to improve intersection performance and pedestrian safety.

The findings suggest that machine learning algorithms can provide a powerful tool for transportation planners and designers to predict pedestrian demand and optimize signal timings to improve safety and efficiency at intersections more accurately. By incorporating data-driven approaches into transportation planning and design, planners can ensure that the transportation systems are designed to be both sustainable and equitable, prioritizing the needs of pedestrians and other vulnerable road users.

The developed methodology in this report uses video analytics to convert videos into pedestrian counts and thereafter uses these counts to forecast the needed walk interval duration. A shortcoming of this approach is that it assumes all pedestrians have an equal need of time and therefore, pedestrians with disabilities may be neglected. Future work should investigate ways to incorporate pedestrians with disability into to the framework of detecting and forecasting demand. Further, pedestrian demand forecasting and signal timing can consider areas beyond only signalized intersections by accounting for pedestrian behavior and demand in other areas of the transportation network. Future studies should aim to incorporate data on pedestrian behavior and demand across the entire transportation network, including non-signalized intersections, pedestrian crossings, and other areas where pedestrians may be present.

In addition, future work can tie the pedestrian forecasting approaches used in this report to autonomous vehicle (AV) applications in the prospective era of CAVs. Forecasting pedestrian demand can be a valuable tool for AVs to enhance the road's overall efficiency and safety. Such information can help in adjusting AVs speed, trajectory, and timing to ensure safe and efficient travel. For example, if a high volume of pedestrians in a particular area is forecasted at specific time of the day, the AV can slow down and give pedestrians more time to cross the street, or even take an alternative route to avoid vehicle-pedestrian conflict areas.

Despite these limitations, this study provides valuable insights into using emerging large-scale data on pedestrians and machine learning algorithms for forecasting pedestrian demand and optimizing signal timings at intersections. By using data-driven approaches to transportation planning and design, the safety and efficiency of the transportation network can be improved while also prioritizing the needs and behaviors of pedestrians and other vulnerable road users. In conclusion, using machine learning algorithms to forecast pedestrian demand and optimize signal timings is a promising approach for improving intersection performance and prioritizing pedestrian safety. While further research is needed to refine and keep up-to-date the models and account for changing conditions in real time, the potential benefits of data-driven transportation planning and design are clear and warrant further investigation and investment.

CHAPTER 6 SYNOPSIS OF PERFORMANCE INDICATORS

6.1 Part I of USDOT Performance Indicators

Over the study period for this project, three (2) transportation-related courses were offered that were taught by the PIs. Two graduate students and a post-doctoral researcher participated in the research project during the study period. During the study period, one (1) transportation-related advanced degree (doctoral) program and one (1) transportation-related M.S. program utilized the CCAT grant funds from this research project to support the graduate students. One graduate student graduated in May 2023 and the other is expected to graduate in December 2023. The post-doctoral researcher was appointed a faculty member at the University of Wisconsin Madison.

6.2 Part II of USDOT Performance Indicators

Research Performance Indicators:
One (1) journal publication and one (1) conference presentations were produced from this project. The research from this applied research project was disseminated to 31 people in attendance (from industry, government, and academia) through the conference presentation at the 2023 Transportation Research Board Annual Meeting in Washington, DC.

Leadership Development Performance Indicators:
This research project generated 1 academic engagements and 5 industry engagements (meetings with the officials of West Lafayette). The PIs held positions in 2 national organizations that address issues related to this research project. One of the CCAT students who worked on this project holds a membership position in a related ASCE committee related to the subject of this research. The post-doctoral researcher holds a position in a TRB committee related to the subject of this research. In addition, the graduate students were elected as the Purdue president of the Institute of Transportation Engineers, received the School of Civil Engineering Graduate Student Engagement Award, received the Outstanding Student of the Year award by the United States Department of Transportation, and received the Outstanding Graduate Student Service Award.

Education and Workforce Development Performance Indicators:
The methods, data and/or results from this study were incorporated (or, are being incorporated) in the syllabi for the Fall 2022, Spring 2023, and Fall 2023 versions of the following courses at Purdue University:
(a) CE 561: Transportation Systems Evaluation, a mandatory graduate level course at Purdue’s transportation engineering graduate programs (average 10 students at each course offering),
(b) CE 299: Smart Mobility, an optional undergraduate level course at Purdue’ civil engineering B.S. program, (average 12 students),
(c) CE 398: Introduction to Civil Engineering Systems, a mandatory undergraduate level course at Purdue University’s civil engineering program, (average 85 students at each course offering).

These students will soon be entering the workforce. Thereby, the research helped enlarge the pool of people trained to develop knowledge and utilize the at least a part of the technologies developed in this research, and to put them to use when they enter the workforce. Based partly on a recognition of his contributions to this study, the post-doctoral researcher on this project earned a faculty position at the University of Wisconsin Madison.

Collaboration Performance Indicators:
There was collaboration with other agencies and institutions provided matching funds. The CCAT PI collaborated with various professors at Purdue and outside Purdue on An Indiana DOT-funded project related to this study, titled "An Assessment of a Displaced Pedestrian Crossing for Multilane Arterials, under SPR 4301. The collaboration included Professors Darcy Bullock and Sarah Hubbard on a related INDOT-funded project. The outcome of the collaboration was a research report, 1 journal paper, and 3 conference posters.
Collaboration report: Nafakh, A. J., Zhang, Y., Hubbard, S., & Fricker, J. D. (2021). Assessment of a displaced pedestrian crossing for multilane arterials (Joint Transportation Research Program Publication No. FHWA/IN/JTRP-2021/16). West Lafayette, IN: Purdue University. https://doi.org/10.5703/1288284317318
The study report for this collaboration project was selected as the recipient of the AASHTO High-Value Research Project for 2022.

The outputs, outcomes, and impacts are described in Chapter 7.

CHAPTER 7 STUDY OUTCOMES AND OUTPUTS

7.1 Outputs

7.1.1 Publications, conference papers, or presentations (from major conference or similar event)

(a) Publications
Title of publication: A Quantitative Approach for Timing the Pedestrian Walk Interval.
Full Citation: Nafakh, A. J., Bullock, D. M., & Fricker, J. D. (2022). A Quantitative Approach for Timing the Pedestrian Walk Interval. Journal of Transportation Technologies, 12(4), 732-743.
Link where published: https://www.scirp.org/pdf/jtts_2022093013481727.pdf

(b) Presentations
Title of presentation: A Quantitative Approach for Timing the Pedestrian Walk Interval.
Full Citation: Nafakh, A. J., Bullock, D.M., & Fricker, J. D. (2023). A Quantitative Approach for Timing the Pedestrian Walk Interval, Transportation Research Board Annual Meeting, Washington, DC.

7.2 Outcomes

This project produced outcomes that could influence road agencies' transportation system design or operational policies. These are:

- This study presented a quantitative analysis of the pedestrian walk interval duration given pedestrian volume conducted on 12 signalized intersections across the City of West Lafayette, Indiana, for ten months. In addition, data on the storage area and offset from the pedestrian push button to the crosswalk was used to explain the variability in pedestrian start-up time.
- This study analyzed time series data of pedestrian volumes on a campus town and identified factors that influence pedestrian movements.
- This study provides valuable insights into using emerging large-scale data on pedestrians and machine learning algorithms for forecasting pedestrian demand and optimizing signal timings at intersections.
- The study demonstrates that by using data-driven approaches to transportation planning and design, the safety and efficiency of the transportation network can be improved while also prioritizing the needs and behaviors of pedestrians and other vulnerable road users.

7.3 List of impacts

The results of this report indicate that large-scale data on pedestrian behavior and volumes is now a reality that should be utilized to ensure equitable service. The findings of studies included in this report indicate that several key factors can significantly impact pedestrian behaviors and volumes at signalized intersections, including built environment features such as the location of the push-button, the time of day, day of the week, weather conditions, and proximity to public transit and other pedestrian-oriented destinations. By incorporating these factors into machine learning classification models, the developed framework could reliably forecast pedestrian demand, allowing for optimal signal timings to improve intersection performance and pedestrian safety.

The findings suggest that machine learning algorithms can provide a powerful tool for transportation planners and designers to predict pedestrian demand and optimize signal timings to improve safety and efficiency at intersections more accurately. By incorporating data-driven approaches into transportation planning and design, planners can ensure that the transportation systems are designed to be both sustainable and equitable, prioritizing the needs of pedestrians and other vulnerable road users. A list of specific impacts from this research project, are as follows:

- The built statistical model can aid designers in identifying proper walk interval timing on an intersection-by-intersection basis. In addition, designers are herein provided quantitative data aiding in the selection of proper timings and to support prioritizing close-to-crosswalk push-button locations that could help minimize pedestrian start-up time.
- The study found that special events and time of day are significant determinants of pedestrian volume. In addition, the study found a significant association between the academic calendar and pedestrian activities. Moreover, the study confirms a repetitive pattern of pedestrian volumes over time, that is, pedestrian volumes tend to have a pattern that recurs.
- This study demonstrates the effectiveness of machine learning classification models in predicting pedestrian volumes using time series, weather, transit, and land use variables.

REFERENCES

Chapter 2 References

[1] Fruin, J. J. (1971). Pedestrian planning and design. Metropolitan Association of Urban Designers and Environmental Planners, Inc., Northwestern University, Evanston, IL.

[2] Nafakh, A. J., Bullock, D. M., & Fricker, J. D. (2022). A Quantitative Approach for Timing the Pedestrian Walk Interval. Journal of Transportation Technologies, 12(4), 732-743.

[3] Manual on Uniform Traffic Control Devices (MUTCD) for streets and highways (2009). Including revision 1. U.S. Dept of Transportation, Washington, D.C.

[4] Traffic Control Devices Handbook. (2001). J.L. Pline editor. Institute of Transportation Engineers, Washington, D.C., 2001.

[5] LaPlante, J., & Kaeser, T. P. (2007). A history of pedestrian signal walking speed assumptions. In 3rd Urban Street Symposium: Uptown, Downtown, or Small Town: Designing Urban Streets That Work Transportation Research Board Institute of Transportation Engineers (ITE) U.S. Access Board.

[6] Bennett, S., Felton, A., & Akçelik, R. (2001, December). Pedestrian movement characteristics at signalized intersections. In 23rd Conference of Australian Institutes of Transport Research (CAITR 2001).

[7] Onelcin, P., and Alver. Y. (2017)The crossing speed and safety margin of pedestrians at signalized intersections, Transportation Research Procedia 22: 3-12.

[8] Furth, P. G., Wang, Y. D., & Santos, M. A. (2019). Multi-stage pedestrian crossings and two-stage bicycle turns: Delay estimation and signal timing techniques for limiting pedestrian and bicycle delay. Journal of Transportation Technologies, 9(04), 489.

[9] Knoblauch, R. L., Pietrucha, M. T., & Nitzburg, M. (1996). Field studies of pedestrian walking speed and start-up time. Transportation Research Record 1538(1), 27-38.

[10] Choi, J., Tay, R., Kim, S., & Jeong, S. (2019). Behaviors of older pedestrians at crosswalks in South Korea. Accident Analysis & Prevention, 127, 231-235.

[11] Langlois, J. A., Keyl, P. M., Guralnik, J. M., Foley, D. J., Marottoli, R. A., & Wallace, R. B. (1997). Characteristics of older pedestrians who have difficulty crossing the street. American Journal of Public Health, 87(3), 393-397.

[12] Fugger, T. F., Randles Jr, B. C., Stein, A. C., Whiting, W. C., & Gallagher, B. (2000). Analysis of pedestrian gait and perception-reaction at signal-controlled crosswalk intersections. Transportation Research Record 1705(1), 20-25.

[13] Kivi, S. B., Norouzizadeh, Z., Rohani, A., Moosavy, S. M., & Nafise, Z. (2021). Investigation of the Association between Type of Walking and Sense of Community (CaseStudy:ALAM AL-HUDA Pedestrian Zone).Open Access Library Journal,8(8),1-21.

[14] Easa, S. M., & Cheng, J. (2013). Reliability analysis of minimum pedestrian green interval for traffic signals. Journal of Transportation Engineering 139(7), 651-659.

[15] Helbing, D., & Molnar, P. (1995). Social force model for pedestrian dynamics. Physical Review E 51(5), 4282.

[16] Johansson, F., Peterson, A., & Tapani, A. (2015). Waiting pedestrians in the social force model. Physica A: Statistical Mechanics and its Applications, 419, 95-107.

[17] Apple, Inc. (n.d.) [West Lafayette, Indiana, USA]. Retrieved June 5th, 2022 from Apple Maps.

Chapter 3 References

[1] NHTSA (2018). Fatality Analysis Reporting System (FARS). National Highway Traffic Safety Administration, Washington, D.C. https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars

[2] Singleton, P. A., & Runa, F. (2021). Pedestrian traffic signal data accurately estimates pedestrian crossing volumes. Transportation Research Record 2675(6), 429–440. https://doi.org/10.1177/0361198121994126

[3] Ryus, P., Ferguson, E., Laustsen, K. M., Schneider, R. J., Proulx, F. R., Hull, T., & Miranda-Moreno, L. (2014). Guidebook on pedestrian and bicycle volume data collection (No. qt11q5p33w). Institute of Transportation Studies, UC Berkeley.

[4] Nordback, K., et al. (2013). Estimating annual average daily bicyclists: Error and accuracy. Transportation Research Record 2339(1), 90–97.

[5] Niska, A., et al. (2012). VTI Rapport 743: (VTI Report 743: Evaluating pedestrian and cycle traffic. Development of a harmonized method for monitoring the municipal proportion of pedestrian and cycle traffic through travel surveys and cycle counts). Swedish National Road and Transport Research Institute, Linköping, Sweden.

[6] Nafakh, A. J., Bullock, D. M., & Fricker, J. D. (2022). A quantitative approach for timing the pedestrian walk interval. Journal of Transportation Technologies, 12(4), 732-743.

[7] Day, C. M., Premachandra, H., & Bullock, D. M. (2011). Rate of pedestrian signal phase actuation as a proxy measurement of pedestrian demand. Transportation Research Board Annual Meeting.

[9] Kim, N. S., & Susilo, Y. O. (2013). Comparison of pedestrian trip generation models. Journal of Advanced Transportation, 47(4), 399-412.

[10] Ercolano, J. M., Olson, J. S., & Spring, D. M. (1997). Sketch-plan method for estimating pedestrian traffic for Central Business Districts and suburban growth corridors. Transportation Research Record 1578(1), 38-47.

[11] What is Computer Vision? IBM. (n.d.). Retrieved January 16th, 2023, from https://www.ibm.com/topics/computer-vision

[12] Kalpić, D., Hlupić, N., Lovrić, M. (2011). Student’s t-Tests. In: Lovric, M. (eds) International Encyclopedia of Statistical Science. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04898-2_641

Other Resources

Congress.gov, Library of Congress. Infrastructure Investment and Jobs Act. (2021, November 15th). Retrieved January 13th, 2023, from https://www.congress.gov/117/plaws/publ58/PLAW-117publ58.pdf

National Center for Statistics and Analysis. (2020, March). Pedestrians: 2018 data (Traffic Safety Facts. Report No. DOT HS 812 850). National Highway Traffic Safety Administration.

National Center for Statistics and Analysis. (2020, July, revised). Bicyclists and other cyclists: 2018 data. (Traffic Safety Facts. Report No. DOT HS 812 884). National Highway Traffic Safety Administration.

Danish Road Directorate (Vejdirektoratet) (2004). Rapport nr. 289: Trafiktællinger: Planlægning, udførelse og efterbehandling (Report 289: Traffic counts: Planning, implementation, and

data reduction). Copenhagen, Denmark.
Nafakh, A. J., Zhang, Y., Hubbard, S., & Fricker, J. D. (2021). Assessment of a displaced pedestrian crossing for multilane arterials (Joint Transportation Research Program Publication No. FHWA/IN/JTRP-2021/16). West Lafayette, IN: Purdue University. https://doi.org/10.5703/1288284317318

Chapter 4 References

[1] National Academies of Sciences, Engineering, and Medicine. 2015. Signal timing manual – Second Edition. Washington, DC: The National Academies Press. https://doi.org/10.17226/22097.
[2] Manual on Uniform Traffic Control Devices (MUTCD) for streets and highways (2009). Including revision 1. U.S. Dept of Transportation.
[3] Traffic Control Devices Handbook. J.L. Pline editor. Institute of Transportation Engineers, Washington, D.C., 2001.
[4] LaPlante, J., & Kaeser, T. P. (2007). A history of pedestrian signal walking speed assumptions. In 3rd Urban Street Symposium: Uptown, Downtown, or Small Town: Designing Urban Streets That Work Transportation Research Board Institute of Transportation Engineers (ITE) U.S. Access Board.
[5] Johansson, F., Peterson, A., & Tapani, A. (2015). Waiting pedestrians in the social force model. Physica A: Statistical Mechanics and its Applications, 419, 95-107.
[6] Helbing, D., & Molnar, P. (1995). Social force model for pedestrian dynamics. Physical Review E 51(5), 4282.
[7] Nafakh, A. J., Bullock, D. M., & Fricker, J. D. (2022). A quantitative approach for timing the pedestrian walk interval. Journal of Transportation Technologies, 12(4), 732-743.
[8] Kim, N. S., & Susilo, Y. O. (2013). Comparison of pedestrian trip generation models. Journal of Advanced Transportation 47(4), 399-412.
[9] Miranda-Moreno, L. F., & Fernandes, D. (2011). Modeling of pedestrian activity at signalized intersections: land use, urban form, weather, and spatiotemporal patterns. Transportation Research Record, 2264(1), 74-82.
[10] Pulugurtha, S. S., & Repaka, S. R. (2008). Assessment of models to measure pedestrian activity at signalized intersections. Transportation Research Record, 2073(1), 39-48.
[11] What is Computer Vision? IBM. (n.d.). Retrieved January 16th, 2023, from https://www.ibm.com/topics/computer-vision
[12] Hankey, S., Lu, T., Mondschein, A., & Buehler, R. (2017). Spatial models of active travel in small communities: Merging the goals of traffic monitoring and direct-demand modeling. Journal of Transport & Health 7, 149-159.
[13] Sanders, R. L., Frackelton, A., Gardner, S., Schneider, R., & Hintze, M. (2017). Ballpark method for estimating pedestrian and bicyclist exposure in Seattle, Washington: Potential option for resource-constrained cities in an age of big data. Transportation Research Record 2605(1), 32-44.
[14] Griswold, J. B., Medury, A., Schneider, R. J., Amos, D., Li, A., & Grembek, O. (2019). A pedestrian exposure model for the California state highway system. Transportation Research Record 2673(4), 941-950.
[15] Munira, S., Sener, I. N., & Dai, B. (2020). A Bayesian spatial Poisson-lognormal model to

examine pedestrian crash severity at signalized intersections. Accident Analysis & Prevention 144, 105679.
[16] Singleton, P. A., Park, K., & Lee, D. H. (2021). Varying influences of the built environment on daily and hourly pedestrian crossing volumes at signalized intersections estimated from traffic signal controller event data. Journal of Transport Geography 93, 103067.
[17] Schneider, R. J., Schmitz, A., & Qin, X. (2021). Development and validation of a seven-county regional pedestrian volume model. Transportation Research Record 2675(6), 352-368.
[18] Esri. (2021). ArcGIS Pro. Redlands, CA: Environmental Systems Research Institute
[19] OpenStreetMap contributors. (2022). Indiana Roadway Layer. Retrieved from https://www.openstreetmap.org/#map=13/40.4264/-86.8995
[20] Registration calendars. Registration Calendars - Office of the Registrar - Purdue University. (n.d.). Retrieved February 26, 2023, from https://www.purdue.edu/registrar/calendars
[21] Past weather by ZIP code - data table. NOAA Climate.gov. (n.d.). Retrieved February 26, 2023, from https://www.climate.gov/maps-data/dataset/past-weather-zip-code-data-table
[22] GTFS/GIS data download. CityBus of Greater Lafayette Indiana. (2021, November 23). Retrieved February 27, 2023, from https://gocitybus.com/gtfs-gis-data-download
[23] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: Springer.

APPENDIX 1

CCAT Project Title: Translation of Driver-Pedestrian Behavioral Models at Semi-Controlled Crosswalks into a Quantitative Framework for Practical Self-Driving Vehicle Applications – Part B (Pedestrian Data Analytics)

Published Related Work

Nafakh, A. J., Bullock, D. M., & Fricker, J. D. (2022). A Quantitative Approach for Timing the Pedestrian Walk Interval. Journal of Transportation Technologies, 12(4), 732-743. doi: 10.4236/jtts.2022.124042.

Abstract

At a typical signalized intersection, the pedestrian phase consists of a walk interval and a change/clearance interval, during which pedestrians are given the right of way. The walk interval is intended to allow pedestrians to exit the curb ramp and enter the crosswalk. The clearance interval will enable them to cross entirely to the other side of the road. Unfortunately, the literature is quite vague on how long the walk interval should be and provides values ranging from 4 to 15 seconds based on qualitative pedestrian demand ranging from Negligible to High. To provide some quantitative guidance for walk interval selection, this paper reports on a study that collected 1,500 pedestrian movement data from 12 signalized intersections with varying pedestrian demand, pedestrian storage areas, and pedestrian push-button locations. The data were used to propose a quantitative model for designers to select the appropriate walk interval. Specifically, this paper seeks to add values to the Traffic Operations Handbook walk-interval guidelines as to how many pedestrians are considered “negligible volume” and can be accommodated by the 4 second minimum time, how many pedestrians are considered “typical volume” and require 7 to 10 seconds, and how many pedestrians are considered “high volume” and require 10 to 15 seconds, or perhaps longer. In addition to examining pedestrian demand, this paper looks at the impact of storage areas and pedestrian push-button location on pedestrian start-up time and, consequently, an appropriate walk interval.