

Pedestrian and Bicyclist Safety at Highway-Rail Grade Crossings (Year 1 Report)

Aemal J. Khattak, Ph.D.

Professor and Director
Mid-America Transportation Center
Department of Civil and Environmental
Engineering
University of Nebraska-Lincoln
330E Prem Paul Research Center at Whittier
School
2200 Vine Street, Lincoln 68588 Nebraska
khattak@unl.edu, (402) 472-8126
ORIC ID: 0000-0002-6695-9842

Muhammad Naveed Aman, Ph.D.

Assistant Professor
School of Computing
University of Nebraska-Lincoln
263 Avery Hall
1144 T Street, Lincoln, NE 68588
ORIC ID: 0000-0002-4629-7589

Muhammad Umer Farooq, Ph.D.

Postdoctoral Research Associate
Mid-America Transportation Center,
University of Nebraska-Lincoln
330D Prem Paul Research Center at Whittier
School
2200 Vine Street, Lincoln 68588 Nebraska
mfarooq2@unl.edu, (402) 472-0777
ORIC ID: 0000-0003-3850-0601

Moomal Bukhari

Graduate Research Assistant
School of Computing
University of Nebraska-Lincoln
123 Avery Hall
1144 T Street, Lincoln, NE 68588
mbukhari2@unl.edu
ORIC ID: 0009-0007-5028-6456

A Report on Research Sponsored by

University Transportation Center for Railway Safety (UTCRS)

University of Nebraska Lincoln

September 2024

Technical Report Documentation Page

1. Report No. UTCRS-UNL-O3CY23	2. Government Accession No. 01898076	3. Recipient's Catalog No.	
4. Title and Subtitle Pedestrian and Bicyclist Safety at Highway-Rail Grade Crossings (Year 1 Report)		5. Report Date September 15, 2024	
		6. Performing Organization Code UTCRS-UNL	
7. Author(s) Aemal J. Khattak, M. Naveed Aman, M. Umer Farooq, Moomal Bukhari		8. Performing Organization Report No. UTCRS-UNL-O3CY23	
9. Performing Organization Name and Address University Transportation Center for Railway Safety (UTCRS) University of Nebraska Lincoln (UNL) 1400 R Street Lincoln, NE United States 68588		10. Work Unit No. (TRAIS)	
		11. Contract or Grant No. 69A3552348340	
12. Sponsoring Agency Name and Address U.S. Department of Transportation (USDOT) University Transportation Centers Program 1200 New Jersey Ave. SE Washington, DC, 20590		13. Type of Report and Period Covered Project Report June 1, 2023 – August 31, 2024	
		14. Sponsoring Agency Code USDOT UTC Program	
15. Supplementary Notes			
16. Abstract Published literature is sparse on non-motorist (pedestrians and bicyclists) crashes at Highway-rail Grade Crossings (HRGCs), despite their involvement in rail-related crashes. A critical aspect of crash prediction models for HRGs is crash exposure, which measures activities of interest at specific locations. While data on motor vehicle and train traffic at HRGCs are available from different sources, non-motorist traffic counts are not readily available. Current Federal Railroad Administration (FRA) models focus on vehicular crash exposure, overlooking non-motorized traffic; gathering non-motorized traffic data is crucial for improving crash prediction models. In this study, non-motorist traffic videos were recorded for 1,848 hours from various urban and suburban HRGCs in Nebraska followed by application of the AI-based You Only Look Once Version 8 (YOLOv8) algorithm for automated non-motorist volume detection. Additionally, data on grade crossing characteristics, including population density and land use, were collected to create a comprehensive non-motorist database. Statistical and AI models were developed to analyze non-motorist exposure in terms of daily traffic volumes, utilizing physical, dynamic, and temporal characteristics of HRGCs. The models indicated that sidewalks, improved visibility, and cloudy weather conditions were associated with increased non-motorist traffic volume. Conversely, higher motor vehicle traffic levels, adverse weather conditions (rain and snow), industrial zones, and greater number of traffic lanes were linked with lower non-motorist traffic. This foundational study aims to enhance crash prediction models at HRGCs by incorporating non-motorist traffic factors, potentially improving crash prediction models and non-motorist safety at HRGCs.			
17. Key Words Railroad Grade Crossings, Bicycle Crossings, Crosswalks, Safety, Artificial Intelligence, Object Detection, Data Science		18. Distribution Statement This report is available for download from https://www.utrgv.edu/railwaysafety/research/operations/index.htm	
19. Security Classification (of this report) None	20. Security Classification (of this page) None	21. No. of Pages 106	22. Price

Table of Contents

List of Figures	4
List of Tables	5
List of Abbreviations	5
Disclaimer	6
Acknowledgements	6
Executive Summary	7
1. Introduction	8
1.1. Problem Statement	10
1.2. Background	10
1.3. Objectives	12
1.4. Overall Approach	12
1.5. Research Framework	14
1.6. Organization of the Report	14
2. Literature Review	15
2.1. Pedestrian Exposure	15
2.2. Bicyclist Exposure	31
2.3. Combined Pedestrian and Bicyclist Volume/Exposure Studies	43
3. Data Collection	49
4. Artificial Intelligence for Non-motorist Detection	60
5. Case Studies	70
6. Exposure Prediction Modelling	77
7. Conclusion and Recommendations	91
References	93
Appendix A	103

List of Figures

Figure 1. Research Framework.	14
Figure 2. Research methodology by Griswold et al. (2019).	26
Figure 3. Photographic image of scout hardware (Miovison.com).....	50
Figure 4. Locations of urban and suburban HRGCs with recorded non-motorist activity.	51
Figure 5. Proximity to local amenities within a 0.5-mile radius of HRGCs.	53
Figure 6. Population density buffer zones (1 mile) of HRGCs.	54
Figure 7. Batch II video recording data collection overview: crossing ID 083519T.....	55
Figure 8. Key Studies for Non-motorist Object Detection.	61
Figure 9. Yolov8 overall framework (Sheng et al. 2020).	65
Figure 10. Flowchart of object detection process.	69
Figure 11. Object detection case I.....	70
Figure 12. Object detection case II.	71
Figure 13. Object Detection Case III.	71
Figure 14. Object Detection Case III.	72
Figure 15. Object Detection Case IV.	73
Figure 16. Object Detection Case V.	74
Figure 17. Object Detection Case VI.	74
Figure 18. Object Detection Case VII.....	75
Figure 19 (a) Snapshot of the output showing detection of a person while video is being processed by YOLOv8 model. (b). Snapshot of the output at the end of the video processing by YOLOv8 model.	76
Figure 20. Diagnostic Plots for Final Selected Random Effects Negative Binomial Model (RENB).....	83
Figure 21. Correlation matrix of features.....	87
Figure 22. Coefficient value of the features used in SGD Regressor model.	89
Figure 23. Partial Dependence Plots (PDP) for Each Selected Feature in the SGD Model.....	89

List of Tables

Table 1. Recent pedestrian exposure/volume prediction studies.	16
Table 2. Factors investigated in relation to pedestrian exposure measures.	29
Table 3. Recent bicyclist exposure/volume prediction studies.	32
Table 4. Combined pedestrian and bicyclist volume/exposure studies.....	43
Table 5. Summary of non-motorist exposure measures from the literature review.	48
Table 6. Locations of HRGCs from Batch I of Video Data Collection	56
Table 7. Locations of HRCs from Batch II of Video Data Collection.....	58
Table 8. Locations of HRCs from Batch II of Video Data Collection.....	59
Table 9. Summary statistics of candidate variables for NB and RENB modelling.	79
Table 10. Estimated NB and RENB for Predicting Total Non-Motorist Volume (24-Hr) at HRGCs.....	81
Table 11. Estimated average marginal effects (AMEs) for NB and RENB model.....	84

List of Abbreviations

AREMA	American Railway Engineering and Maintenance-of-Way Association
FHWA	Federal Highway Administration
FRA	Federal Railroad Administration
GIS	Geographic Information System
GPS	Global Positioning System
HRGCs	Highway-Rail Grade Crossings
LiDAR	Light Detection and Ranging
NB	Negative Binomial Model
RENB	Random Effects Negative Binomial Model
RMSE	Root-Mean-Square Error
YOLOV8	You Only Look Once Version 8
USDOT	U.S. Department of Transportation
UTC	University Transportation Centers

Disclaimer

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 U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

Acknowledgements

The authors want to acknowledge the University Transportation Center for Railway Safety (UTCRS) at UTRGV for the financial support provided to perform this study through the USDOT UTC Program under Grant No. 69A3552348340. We are also deeply thankful to the Mid-America Transportation Center for providing access to its facilities and laboratories, which were crucial for conducting this study. Additionally, we extend our appreciation to postdoctoral research associate Dr. MM Shaikul Haque, graduate student Frank Dzawu, and undergraduate student Caitlin Keathley for their invaluable assistance with on-site data collection and manual data verification.

Pedestrian and Bicyclist Safety at Highway-Rail Grade Crossings

Executive Summary

Published literature is relatively sparse on the topic of non-motorist (pedestrians and bicyclists) crashes at rail crossings. Such crashes contribute to the overall fatalities and injuries reported at rail grade crossings and cause disruptions to both the rail and highway networks. These crashes may be associated with several factors, such as risk-prone crossing user behavior, underestimation of approaching train speed, distractions caused by electronic devices, insufficient visibility due to obstructed sightlines, and inadequate infrastructure and maintenance of rail crossings (e.g., malfunctioning warning signals or absence of signage). Estimation of crash prediction models provides insights into the nature of safety and its associated factors. While such models are simplifications of complex phenomena, they are useful for advancing understanding, forecasting future conditions, and for allocation of resources.

A staple of crash predication models is crash exposure, which is a measure of activities at a location of interest (e.g., motor vehicle, train, and non-motorized traffic at rail crossings). While motor vehicle and train traffic information are available for rail crossings, there are no reliable estimates of non-motorist traffic at rail crossings readily available. The currently available Federal Railroad Administration (FRA) crash frequency and severity models for rail crossings cover vehicular exposure, overlooking the dynamic aspects of non-motorized traffic. Similarly, other available models of rail crossing safety are motor vehicle centric. Collection of non-motorized traffic exposure information is needed for developing better crash prediction models. However, doing so for the approximately 212,000 rail crossings across the US is not practical.

This study developed a non-motorized traffic prediction model for HRGCs based on specific rail crossing characteristics. Non-motorist traffic data were collected at urban and suburban rail crossings in Nebraska using video recording devices and AI-based object detection techniques. In addition to non-motorist traffic, data were gathered on rail crossing characteristics and surrounding area features, such as population density, land use, vehicle and train traffic, and HRGC proximity to activity centers. Temporal factors, including temperature, windspeed, visibility, and precipitation, were also recorded to assess their influence on non-motorist activity.

The study aimed to address two key research questions: (a) which crossing characteristics are statistically significant predictors of non-motorist traffic at rail crossings? and (b) what is the predictive accuracy of the developed non-motorist traffic model based on these characteristics? Using this comprehensive dataset, both statistical and AI-driven prediction models were estimated for non-motorist traffic volumes at HRGCs. A comparison of model performance indicators was conducted to select the final model for non-motorist volume prediction at HRGCs.

The selected model indicated that sidewalks, improved visibility, and cloudy weather conditions are associated with increased non-motorist traffic volume. Conversely, higher motorized traffic levels, adverse weather conditions (rain and snow), industrial zones, and more traffic lanes on the crossing road were linked to reduced non-motorist activity. By understanding factors that influence non-motorist traffic at HRGCs, policymakers and planners can develop targeted interventions to enhance safety at HRGCs. The output from these models (i.e., predictions of non-motorist crash exposure) can then be used as input in HRGC crash prediction models focused on non-motorized users.

1. Introduction

Safety stands as the cornerstone of any transportation system, and nowhere is this more evident than at highway-rail grade crossings (HRGCs). Highway-rail grade crossings are crucial intersections where motorists and trains intersect, requiring rigorous caution. These crossings have a higher likelihood of potential hazards and crashes if safety measures are compromised. The consequences of mishaps at these intersections extend far beyond mere inconvenience, often resulting in loss of lives, severe injuries, extensive property damage, and societal unrest. Furthermore, crashes at grade crossings have significant financial consequences that include paying for litigation, fixing infrastructure, and causing delays in trade and commerce. To ensure a safe coexistence between motorists and moving trains, smooth coordination is essential, requiring adherence to safety protocols and the establishment of safety infrastructure. Hence, ensuring safety at highway rail grade crossings is not merely a regulatory obligation, but a cornerstone of sustainable transportation infrastructure. According to preliminary Federal Railroad Administration (FRA) statistics, 2,190 highway-rail grade crossing crashes occurred in

2023. There were 248 crossing fatalities and 752 crossing injuries in 2023 across the U.S. Approximately 84% of all 2023 highway-rail grade crossing collisions occurred in these states (FRA, 2024).

While there has been a lot of emphasis on keeping motorists safe at rail grade crossings, it is crucial not to overlook the safety of non-motorists. The safety of non-motorists, essentially comprising pedestrians and bicyclists, remains a pressing concern. These vulnerable road users navigate these crossings with distinct challenges, facing heightened risks due to their limited visibility and slower speeds (Farooq and Khattak, 2023; Khattak et al. 2023; Zhao et al. 2024). These crossings frequently present significant obstacles to non-motorists despite being equipped with warning devices like lights, gates, and bells; these obstacles could be due to inadequate infrastructure, poor signage, uneven or poorly maintained walking or cycling paths leading to the crossing, confusing road markings or signage, insufficient lighting, particularly during nighttime hours, and insufficient public awareness campaigns. Therefore, the increasing vulnerability of bicycles and pedestrians at highway rail grade crossings emphasizes the need for targeted interventions and all-encompassing safety measures made to satisfy their needs.

Furthermore, crashes involving pedestrians and bicyclists at rail crossings can be further associated with several factors beyond the obstacles described earlier, such as, underestimation of the speed and proximity of approaching trains, distractions caused by electronic devices, insufficient visibility due to obstructed sightlines from vegetation or structures, socio-economic factors such as risky behavior, or attempts to cross tracks while ignoring safety protocols, and inadequate infrastructure and maintenance of rail crossings, including malfunctioning warning signals or absence of clear signage. To state briefly, enhancing non-motorists' safety at highway-rail grade crossings presents a multifaceted challenge due to diverse settings in which these crashes occur, necessitating coordination among various stakeholders. Furthermore, incidents involving pedestrians and bicyclists can occur both at rail crossings and at non-rail crossing locations, each situation exhibiting a distinct crash dynamic.

1.1. Problem Statement

The Federal Railroad Administration (FRA) compiles data on HRGCs, encompassing various physical and dynamic attributes such as location information, warning devices, device configurations, pavement markings, number of lanes on the roadway, type of pavement surface, track configurations, geometric features, as well as train and vehicular traffic volumes. Through the analysis of these datasets, the FRA formulates models to predict the frequency and severity of crashes, facilitating the allocation of resources and the management of crash risks at these crossings. Regrettably, these datasets are devoid of any information pertaining to pedestrian or bicyclist crash exposure. This absence, particularly concerning pedestrian and bicyclist crash exposure data, signifies a notable deficiency in non-motorists' safety at rail crossings.

The lack of crash exposure data significantly hampers our capacity to investigate the risks associated with pedestrian and bicyclist safety at rail crossings. Moreover, the prevailing safety models developed for rail crossings predominantly concentrate on motor vehicle crashes, leaving a significant gap in safety models, developed specifically to predict crashes involving pedestrians and bicyclists. Additionally, the currently established FRA crash frequency and severity models for HRGCs only consider the dynamic elements of train and vehicular traffic, overlooking the dynamic aspects relevant to non-motorized pedestrians and bicyclists.

1.2. Background

In risk analysis “exposure” is a concept describing the opportunity for a random event to occur, that is, the number of trials. Consequently, identifying the appropriate measure of exposure for a particular risk event is extremely important for analyzing the likelihood of its occurrence (Shapiro et al. 1997). For non-motorist safety analysis, this exposure metric should take into consideration the degree to which bicyclists and pedestrians put themselves at risk of being struck by a train or a motor vehicle at a rail crossing. If these criteria are met, the exposure metric can be a reliable explanatory variable for predicting pedestrian and bicyclist crashes at rail crossings.

Exposure is defined as the rate of contact with a potentially harmful agent or event. Non-motorist exposure is therefore defined as the rate of non-motorist contact with potentially harmful situations involving moving vehicles and trains (e.g., highway rail grade crossings, or a highway intersection). In addition, non-motorist risk is defined as the probability that a non-motorist and vehicle/train crash will occur, based on the rate of exposure. Estimating exposure requires precise measurements of pedestrian and bicyclist volumes, yet obtaining these measurements poses considerable challenges. Other units of exposure in non-motorist crash analysis include person-hours, trip counts, and pedestrian or cyclist population counts (Miah et al. 2024; Almasi et al. 2021). The choice of pedestrian exposure measure strongly influences the risk analysis results. Keall (1995) examined pedestrian crash data using the exposure measures “time spent walking” and “number of roads crossed”.

Recent research in the field has highlighted the importance of pedestrian and bicyclist exposure modeling in urban environments. Studies have explored various factors influencing pedestrian behavior, such as land use patterns, built environment characteristics, and transportation infrastructure design. Additionally, advancements in data collection techniques, including GPS tracking, video surveillance, and crowd-sourced data, have enabled researchers to capture detailed information about pedestrian movements and interactions with their surroundings. By synthesizing findings from these studies, we can glean valuable methodologies and insights applicable to our investigation of pedestrian activity at rail crossings (Miah et al. 2024; Almasi et al. 2021; Sze et al. 2019).

Moreover, a growing body of literature has focused on predictive modeling techniques for estimating pedestrian volumes in different contexts. Researchers have employed a variety of approaches, including statistical modeling, machine learning algorithms, and agent-based simulations, to forecast pedestrian flow patterns and identify potential safety hazards (Chen et al. 2020; Miah et al. 2024; Almasi et al. 2021; Sze et al. 2019). While many of these studies have primarily targeted urban highways and public spaces, the principles and methodologies developed can be adapted and extended to study pedestrian and bicyclist behavior and exposure specifically at rail crossings. By examining these predictive models, we can discern best practices and identify opportunities to tailor them to the unique characteristics of rail crossing environments.

The FRA-sponsored “*Grade- Crossing Research Needs Workshop*” (Harrison and DaSilva, 2019) highlights rail crossing stakeholders’ desire to improve data collection on pedestrians and bicyclists at rail crossings. Since no reliable accident exposure measures for pedestrians and bicyclists exist for rail crossings, the research is focused on collecting data on pedestrian and bicyclist traffic at urban and suburban rail crossings in Nebraska to develop a framework for estimating accident exposure at other rail crossings.

1.3. Objectives

This research aims to address the following key objectives:

1. Develop a methodology to collect pedestrian and bicyclist crash exposure data at urban and suburban highway rail grade crossings using video-based data collection devices.
2. Develop AI-based computer vision algorithms to detect pedestrian and bicyclist activities at highway rail grade crossings.
3. Analyze the existing characteristics and dynamics of highway rail grade crossings to identify crash-risk factors specific to pedestrians and bicyclists.
4. Develop statistical and AI-based models to forecast non-motorist volumes at highway rail grade crossings, enabling accurate prediction of future exposure levels.

1.4. Overall Approach

The research project began with an exhaustive analysis of existing literature on pedestrian and bicyclist exposure. Although the literature on non-motorist volume and exposure modeling is extensive, we identified a noticeable gap in research specific to highway-rail grade crossings. Nevertheless, studies centered on other transportation systems and infrastructure can offer valuable insights and establish a foundational understanding for our work. The task also involved producing a tabular summary of the reviewed literature and presenting a systematic review of key research studies along with their findings.

Subsequent steps involved the recording of video data on pedestrian and bicyclist activities at Nebraska HRGCs, utilizing Miovision Scout and City of Lincoln Cameras. To ensure a representative sample and accessibility to the University of Nebraska, we identified and delineated suitable locations for rail crossings, as well as determined optimal times and durations for video data collection. Furthermore, spatial and temporal analyses were conducted to identify patterns and trends in pedestrian and bicyclist activities on rail crossings. Pertinent research questions were investigated, such as: (1) Are there specific times of day, days of the week, or months with higher or lower volumes? (2) Are there particular rail crossings with consistently high or low activity?

Data collection efforts were not limited to pedestrian and bicyclist volumes at HRGCs; they also extended to gathering information about various volume-associated factors. These factors encompassed census related data population density, posted speed limit, crossing facilities, presence of safety measures, lane count on intersecting highways, peak hour travel demand, land-use characteristics, socio-economic factors, and identification of nearby activity centers. Afterwards, we conducted an evaluation of the collected data by selecting suitable metrics to measure pedestrian and bicyclist exposure. While common metrics like person-miles traveled, person-hours exposed, or trip counts are frequently used, our focus was on assessing pedestrian and bicyclist volumes in counts per day and counts per hour. Moreover, exposure analysis was integrated with other pertinent data sources such as census data, roadway characteristics, and land use data. Additionally, to effectively communicate crash exposure-related aspects, we employed data visualization techniques, including the creation of maps, graphs, and visualizations.

Finally, we utilized various statistical and machine learning techniques to estimate models for predicting non-motorist volume at HRGCs. We presented research conclusions, offered guidance on improving pedestrian and bicyclist safety at HRGCs, and identified areas for future investigation.

1.5. Research Framework

The research framework is presented in Figure 1. As depicted in the figure, videos were recorded at urban and suburban rail grade crossings in Nebraska using Miovision and City of Lincoln cameras. Subsequently, AI-based algorithms were utilized to detect non-motorists' volume at rail crossings. Utilizing the count data alongside other rail crossing-associated data, we developed volume prediction models for non-motorists at HRGCs.

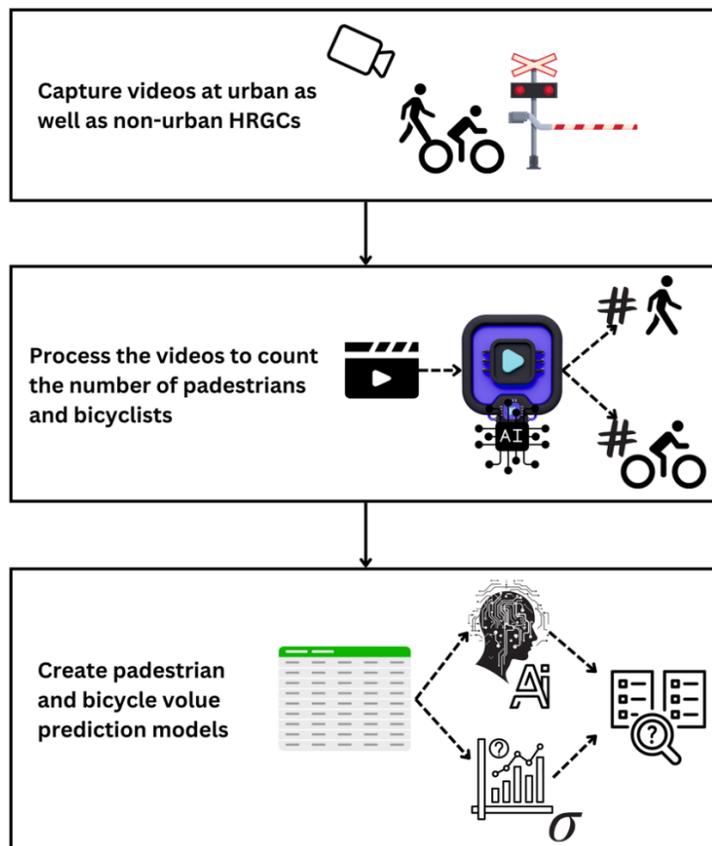


Figure 1. Research Framework.

1.6. Organization of the Report

This report is divided into 6 sections. Section 1 provides an introduction, background information, research aims, and overall approach. Section 2 presents a detailed literature review on non-motorist exposure measures. In Section 3, we elaborate on data collection, including location selection details for video monitoring of highway-rail grade crossings in Nebraska,

accompanied by comprehensive descriptions of the video data. Section 4 outlines the development of AI-based algorithms for non-motorist detection at rail crossings. Section 5 offers insights from case studies to illustrate the study's methodology, along with details on the validation of the algorithms employed for non-motorist detection at rail crossings. Section 6 delves into a detailed discussion on the development of statistical and AI volume prediction models for non-motorists at rail crossings and presents the results. The final section provides conclusions, discusses limitations, and outlines avenues for future research.

2. Literature Review

Non-motorists, that essentially include pedestrian and bicyclist safety and exposure assessment are critical components of urban planning and transportation management. Understanding pedestrian and bicyclist behavior and predicting pedestrian volumes at various locations, including transportation hubs like rail crossings, is essential for designing safe and efficient infrastructure. While the literature on non-motorist volume and exposure modeling is extensive, there is a noticeable gap in research specific to rail crossings. Nonetheless, drawing from studies focused on other transportation systems and infrastructure can provide valuable insights and establish a foundational understanding for our work.

2.1. Pedestrian Exposure

Walking is perhaps the most prevalent and widespread form of physical activity. It necessitates no particular skills or gear and can seamlessly blend into the daily schedules of urban residents across all age groups (Tudor-Locke et al., 2005). Furthermore, walking yields several environmental and social advantages, including alleviating congestion and greenhouse gas emissions, as well as enhancing social cohesion and the overall livability of urban areas. (Yin, 2017).

Many experts and public health officials are interested in establishing walkable communities and promoting walking behavior. Urban land use and transportation planning policies are progressively focusing more on promoting walking and reducing reliance on cars (Babb &

Curtis, 2015). Encouraging walking lies at the core of urban planning theories, which connect higher rates of walking with the overall vibrancy of city life (Southworth, 2005). Jane Jacobs emphasized the significance of walking both at district and street levels for urban vitality, advocating for streets to remain active throughout the day with diverse activities and people (Jacobs, 1961). Her ideas influenced the emergence of new urbanism and smart growth theory. Similarly, promoting walking in cities and public spaces aligns with urban theories concerning the city's visual appeal, and the quality of public spaces (Whyte, 1980). Thus, monitoring walking activities, such as pedestrian volume, can offer valuable insights for researchers and planners to encourage walking behavior (Chen et al., 2020).

However, encouraging safe walking behavior around rail crossings is vital for both urban mobility and public safety. We can highlight the importance of integrating pedestrian-friendly infrastructure with rail systems to ensure safe passage for walkers. This includes designing crossings that are easily accessible, well-lit, and clearly marked to reduce the risk of accidents (Irwin, 2003; Lobb et al., 2003; Alshehri et al. 2023). In this section, we have provided an overview of the latest research on pedestrian exposure studies conducted globally. These tabulated details in Table 1 offer concise insights into the research context, data collection methods, and key findings. Additionally, to enhance comprehension, the studies are further elaborated upon in the subsequent section following Table 1.

Table 1. Recent pedestrian exposure/volume prediction studies.

No	Author(s)	Year	Data	Context	Location	Methodology	Results/Highlights
1	Chen et al.	2020	Street View images (SVIs) from Tencent and Baidu Maps, over 700 street segments, A total of 4,507 sampling points along these street segments were analyzed	Estimating pedestrian volume using Street View images: A large-scale validation test	Tianjin, China	Retrieval of Street View images (SVIs), pedestrian detection using the Localized Detection and Classification Framework (LDCF) algorithm, and data aggregation.	Pedestrian volume measured using street view images and machine learning. Automated detection validated against field observation. Street connectivity and volume influence accuracy. Image quality, size, and collection time also impact accuracy.
2	Qin and Ivan	2001	Exposure data on 32	Estimating Pedestrian	Connecticut, USA	General linear modeling and	Pedestrian exposure is associated with

			sites from rural areas in (specific months in 1998 and 1999)	Exposure Prediction Model in Rural Areas		Tukey and Duncan multiple comparison of means methods	sidewalk availability, highway lane count, area type (tourist, campus, downtown, residential), and signal presence.
3	Schneider et al.	2009	Pedestrian counts at 50 intersections along arterial and collector roadways	Pilot Model for Estimating Pedestrian Intersection Crossing Volumes	Alameda County, California, USA	Ordinary least squares (OLS) regression	Total pedestrian intersection crossing per week associated with the proximity to intersection, total employment, number of commercial retail properties, and number of regional transit stations
4	Sze et al.	2019	Travel Characteristic Survey (TCS) data in 2011-2015	Exposure to pedestrian crash based on household survey data: Effect of trip purpose	Hong Kong	Random-parameter negative binomial regression modelling	Total population, walking frequency and walking time were adopted to represent the pedestrian exposure to road crash. Walking frequency and walking time associated with AADT, Zonal population, proportion of population, median household income, road density, number of non-signalized intersection,
5	Jiang et al.	2022	More than 1400 features are constructed from the CBD Melbourne, for pedestrian estimation, covering macro aspects of transportation, socioeconomic, road networks, time, land use and place of interest.	Pedestrian volume prediction with high spatiotemporal granularity in urban areas by the enhanced learning model	Melbourne, Victoria, Australia	Optimal supervised learning model of LightGBM (Light Gradient Boosting Machine), and LightGBM-based enhanced learning model	Pedestrian volume associated with bus stop, railway station, transport terminal, bridge parking type, parking spaces, total employment in block, land use, and time of day

6	Raford and Ragland	2004	1997 economic census employment data, Census 2000 population density data	Innovative Pedestrian Volume Modeling Tool for Pedestrian Safety	Oakland, California USA	Space Syntax Modelling that utilized layout and connectivity of urban street grids to generate “movement potentials”	The highest pedestrian volumes were found in the downtown area, where streets accounted for nearly 5% of total citywide pedestrian volume, but only 1% of total street area.
7	Singleton et al.	2021	pedestrian data from high-resolution traffic signal controller logs, in addition over 10,000 hr of video recorded at 90 signalized intersections	Pedestrian Traffic Signal Data Accurately Estimates Pedestrian Crossing Volumes	Utah USA	Various linear and nonlinear regression models that reflect a factoring approach (Piecewise linear and quadratic models)	Traffic signal data can be successfully used to estimate pedestrian crossing volumes with good accuracy
8	Griswold et al.	2019	Counts taken from 1200 location on the California State Highway System (SHS)	A Pedestrian Exposure Model for the California State Highway System	California, USA	log-linear regression	Associated factors of pedestrian volume are employment density, population density, number of schools, number of street segments, intersections with principal arterial and minor arterial roadways, and four-way intersections
9	Li and Wu	2021	Pedestrian signal and detection events are collected through the MaxView system	Real-time estimation of pedestrian volume at button-activated midblock crosswalks using traffic controller event-based data	Phoenix, Arizona	Maximum Likelihood Estimation (MLE) and Poisson Model and sensitivity analysis	Lower pedestrian volume on weekends than on weekdays; Accurate volume estimation based on the poisson process and MLE
10	Sobriera and Hellinga	2023	Short time counts and pedestrian activity (October 2021 to September 2022)	Estimating Pedestrian Volumes at Each Crosswalk of Intersections: Comparison of Land-Use	The cities of Milton and Toronto in Ontario, Canada, and Pima County,	Aggregated Multinomial Logistic Regression; Optimal land use model	Pedestrian volume associated variables are Population density, commercial land use, presence of schools, presence of transit stops,

				Models and Short-Term Count Methods	Arizona, USA.		crossing width, presence of any type of physical median, presence of a refuge island, road classification, number of paved sidewalks in the corners of the crosswalk, presence of slip lanes
11	Sobriera et al.	2023	Pedestrian count datasets from 1,018 signalized intersections	Comparing Direct Demand Models for Estimating Pedestrian Volumes at Intersections and Their Spatial Transferability to Other Jurisdictions	Milton, Canada; Pima County, U.S.; and Downtown Toronto, Canada	Direct Demand (DD) modelling (Log-linear, Poisson, and Negative Binomial structures)	Key factors associated with pedestrian volume were level of pedestrian activity, land use, and socioeconomics
12	Bosina et al.	2017	Pedestrian activity data collected for morning and evening periods (manual pedestrian tracking) during fall of 2016	Pedestrian flows on railway platforms	Zürich Hardbrücke and Zürich Stadelhofen, Germany	Hypothesis testing on tendencies to minimize walking distances and adapt to platform layouts, allowing for validation and rejection	Entrance/exit usage, waiting position, destination distribution, Platform side change, pedestrian distribution, walking routes, and boarding and alighting behaviors are variables associated with pedestrian volume
13	Chen et al.	2022	Street view images collected from Baidu Map, a popular online mapping service in China (covered a total length of 3,393 kilometers of streets)	Examining the association between the built environment and pedestrian volume using street view images	Shanghai, China	Extraction of pedestrian volume using street view imaging, and machine learning techniques	Key variables: land use (residential, enterprise, commercial, public service, and entertainment), micro-scale-built environment, and street-level characteristics

In a foundational study conducted by Behnam et al. in 1997, uncomplicated quantitative models utilizing land-use data were developed to estimate pedestrian volume on Central Business District (CBD) sidewalks. They employed statistical multiple regression techniques to create two models for CBD circulation. The first model was based on Noon-hour pedestrian volume, while the second model utilized eight land-use variables to estimate average hourly pedestrian volume. A statistical assessment of these models was carried out, indicating that the equations generated could accurately predict pedestrian volume. These models proved to be valuable in the realms of planning, traffic engineering, and the design of pedestrian facilities.

Qin and Ivan (2001) developed pedestrian exposure prediction models in rural areas and investigated the exposure risk of pedestrians to collisions with motor vehicles. They investigated the relationship between the weekly pedestrian exposure in rural areas of Connecticut and factors such as population density, presence of sidewalks, number of lanes, area type, traffic control type, and median household income. General linear modeling and Tukey and Duncan multiple comparison of means methods were used to identify the significant factors. Only the number of lanes, area type, and sidewalk system significantly explained the variation in the resulting pedestrian exposure prediction model. Notably, the study emphasizes the necessity of enhancing pedestrian facilities in areas with high exposure, suggesting future efforts to estimate pedestrian crash models.

Chen et al. (2020) explored the potential of using Street View images (SVIs) and machine learning to automate the assessment of pedestrian volume, a crucial metric for urban walkability. Traditionally, pedestrian volume data collection relied on labor-intensive field observations, which limited coverage and efficiency. The study rigorously validated this new approach against field observations in Tianjin, China, comparing pedestrian volumes extracted from SVIs with manual counts from over 700 street segments. Results suggested that automated pedestrian detection using SVIs held promise, achieving reasonable to good levels of accuracy, although influenced by various factors such as image quality, size, and collection time. The method proved particularly effective in areas with high pedestrian activity and street connectivity. Key findings highlighted the importance of image quality and collection time in the accuracy of automated pedestrian detection. Higher resolution images and closer alignment with field observation periods led to improved results. Additionally, street characteristics, such as

pedestrian volume and connectivity, significantly impacted accuracy. The study underscored the potential of SVIs and machine learning in assessing pedestrian volumes at scale, offering insights for urban planning and research on walking behaviors. However, they suggested that in areas with low pedestrian activity, future studies should consider factors such as weather and image parameters to improve accuracy and usefulness.

Schneider et al. have contributed significantly to pedestrian exposure research. For instance, Schneider et al. (2009) aimed to address the need for better data on pedestrian volumes to enhance pedestrian movement safety, comfort, and convenience. It focused on developing a pilot model for estimating pedestrian intersection crossing volumes in Alameda County, California. The methodology involved gathering weekly pedestrian volumes at 50 intersections with diverse surrounding characteristics. Three alternative model structures were considered, with the final recommended model showing a strong fit (adjusted $R^2 = .897$). Significant factors in the model included population density, job availability, presence of commercial retail properties, and proximity to regional transit stations. The model, developed using ordinary least squares regression, provided a simple yet effective tool for practitioners utilizing geographic information systems (GIS) and basic spreadsheet programs. In addition, validation of the model demonstrated its potential accuracy, with pedestrian volume estimates within 50% of historic manual counts at 30 of 46 comparison intersections. However, the study suggested the need for further refinement and validation, particularly in other communities and with additional variables such as sidewalk coverage, roadway characteristics, and street network density. Despite its limitations, the pilot model offered valuable insights for planning, prioritizing pedestrian projects, and improving safety analyses. They suggested that future research should focus on expanding the model's predictive capability, exploring different statistical approaches, and comparing it with existing pedestrian models to enhance pedestrian transportation planning and infrastructure development.

In a recent study by Sze et al. (2019), pedestrian safety was investigated by analyzing factors contributing to pedestrian exposure and developing a crash prediction model. Using data from the Travel Characteristic Survey (TCS) and crash data from the Transport Information System (TIS) of Hong Kong from 2011 to 2015, the study identified walking frequency as the most effective measure for predicting crash exposure and crash risk. Findings emphasized the

importance of targeted traffic control and management strategies, particularly in densely populated urban areas. Recommendations included improving safety awareness among vulnerable pedestrian groups and promoting walkability through education, enforcement, and traffic control measures. However, the study noted limitations such as the use of aggregated data and suggested exploring broader socio-demographic and environmental factors in future research.

In another recent study by Jiang et al. (2022), which focused on enhancing pedestrian volume prediction in urban areas in Australia, particularly Melbourne. They highlighted that existing model often suffered from limited samples due to costly field sampling. To address this, the study proposed an enhanced learning model for pedestrian volume prediction with high spatiotemporal granularity. Using Melbourne's CBD as a case study, over 1400 features were constructed covering various aspects like transportation, socioeconomics, and land use. The model, based on LightGBM, significantly improved prediction performance compared to traditional supervised learning, reducing root-mean-square error (RMSE) by 41.75% and improving R-squared (R²) by 27.75%. Spatial resolution and combination parameters were found to significantly affect model performance. The study conducted spatiotemporal analysis using GIS maps, offering insights to optimize urban mobility and enhance city management. The research introduced a novel approach to generate unlabeled samples and develop semi-supervised regression models, enhancing pedestrian estimation performance. The proposed model could be extended to other urban mobility-related problems with limited samples. Moreover, it identified important areas for government attention in Melbourne, suggesting sensor installation and safety measures. However, limitations included uneven distribution of sensor stations and the model's applicability limited to Melbourne's CBD.

Raford and Ragland (2004) presented an innovative pedestrian modeling technique called Space Syntax, which was employed to estimate pedestrian volumes and exposure rates for Oakland, California. The aim was to address the lack of detailed pedestrian-exposure data crucial for pedestrian safety planning. By calculating pedestrian volumes and relative risk indices, the study highlighted the importance of accurate data in making informed decisions regarding pedestrian safety measures. Using Space Syntax modeling, the research mapped pedestrian volumes and identified high-risk intersections, revealing that while downtown areas had high

pedestrian volumes, the most dangerous intersections were clustered in less densely populated regions, indicating a complex relationship between pedestrian volume and risk. Despite limitations in data accuracy, the study underscored the utility of Space Syntax in providing reasonably accurate estimations of pedestrian volume, emphasizing its potential for enhancing pedestrian safety planning. For future studies, they offered recommendations suggesting integrating automobile volumes and speeds into risk-modeling approaches to further enhance pedestrian safety assessments. They also proposed refining the Space Syntax model by incorporating more specific land-use characteristics to improve accuracy. Moreover, the study advocated additional analysis exploring the relationship between pedestrian volume and various factors such as criminal activity and retail behavior. Ultimately, the research emphasized the importance of better pedestrian data for informing urban planning decisions and prioritizing pedestrian safety initiatives, highlighting the potential of enhanced data in raising awareness, increasing advocacy, and ultimately creating safer urban environments.

Lam et al. (2014) took advantage of time geography and travel activity data to propose a new pedestrian exposure metric. Making use of the concept of potential path tree (PPT), they developed an individual-based and network constrained pedestrian exposure measure. Using negative binomial regressions to examine crash frequency with exposure, roadway and environmental variables, the proposed metric was compared with other existing pedestrian exposure methods to examine its applicability and potential in road safety analysis. Vanparijs et al. (2015) provided a comprehensive overview of prior research that employed various methods to assess bicyclists' exposure to road crashes. Several noteworthy approaches included evaluating bicyclists' exposure through questionnaires, utilizing data from automated traffic counts, conducting telephone surveys regarding travel behavior, analyzing data from the national travel survey, employing online travel diaries, utilizing video cameras attached to participants' bicycle helmets, and considering travel exposure data (person-trips).

In a recent study by Singleton et al. (2021), pedestrian data sourced from traffic signal controller logs was validated as a reliable method for estimating pedestrian crossing volumes. This validation was achieved through a detailed comparison between pedestrian signal data and observed pedestrian counts gathered from extensive video footage across 90 signalized intersections in Utah. The study involved meticulous data collection efforts, including the

recording of approximately 10,900 hours of video footage and manual counting of pedestrians by trained undergraduate students, resulting in a dataset encompassing roughly 175,000 pedestrian crossings. The collected pedestrian signal data, derived from the Utah Department of Transportation's ATSPM system, underwent thorough processing and merging with pedestrian crossing event data from the videos. Regression models were then developed to gauge pedestrian crossing volumes utilizing pedestrian signal activity metrics. The results unveiled robust correlations between estimated and observed pedestrian volumes, with mean absolute errors reaching as low as 3.0. Consequently, the study concluded that pedestrian signal data stands as a reliable tool for estimating pedestrian crossing volumes, thereby furnishing valuable insights crucial for transportation planning, safety analyses, and health impact assessments. However, the study acknowledged limitations such as potential equipment malfunctions and the necessity of pedestrian push buttons at signalized intersections for data collection.

In addition, three similar studies were performed by Day et al. 2016, Blanc et al. 2015, and Kothuri et al. 2017, that investigated the use of pedestrian data from traffic signal controller logs to estimate walking activities. For example, Day et al. (2016) conducted a study examining the frequency of pedestrian phase activations at a single signalized intersection in Indiana over an 18-month period. They investigated several factors influencing pedestrian signal activity, including time of day, day of the week, weather conditions, seasonal variations, special events, and modifications to the pedestrian phase setup. The authors also highlighted the practicality of continuously recording pedestrian activations over an extended duration with minimal expense. However, they did not conduct a direct comparison between pedestrian activations and observed pedestrian counts.

In addition, Blanc et al. (2015) conducted a 24-hour pilot investigation into pedestrian activity at a signalized intersection in Oregon, equipped with actuated pedestrian crossings on all four sides. Through video analysis, they manually tallied 596 pedestrians, cross-referencing this data with 482 pedestrian phases recorded in traffic signal controller logs. They developed adjustment metrics for individual phases and the intersection as a whole, revealing correlations of 0.83 or higher between pedestrian counts and actuations, indicating the potential of traffic signal data to accurately approximate pedestrian crossing volumes. Moreover, they showcased the applicability of their adjustment metrics to estimate daily and annual average pedestrian

counts. Subsequently, Kothuri et al. (2017) revisited the same intersection two years later to replicate the earlier study. Over nearly three days, they manually counted 818 pedestrians using video footage and recorded 723 pedestrian phases from signal controller logs. Despite minor variations, the adjustment factors remained consistent, and correlations, though slightly lower in some instances, were still substantial (around 0.80, with one crossing at approximately 0.67). These studies indicate that it's possible to use pedestrian signal data to estimate how many pedestrians are crossing, though the research is limited to just a few intersections and a small amount of observation time.

After further examining recent relevant literature, an important study by Griswold et al. (2019) developed a statewide pedestrian exposure model for intersections on the California State Highway System (SHS). For understanding, we have provided a sketch in Figure 2, that presents their methodology. They utilized log-linear regression to estimate annual pedestrian crossing volumes, leveraging a database of over 1,200 count locations, which constituted one of the largest datasets for pedestrian volume modeling. Seventy-five explanatory variables were assessed, with key factors such as land-use variables (employment density, population density, number of schools), roadway network variables, and American Community Survey journey-to-work walk mode share identified. These variables were selected based on their availability and ease of integration into geographic information system analysis. The resulting model estimated pedestrian volumes at over 12,000 intersections, marking it as one of the initial statewide pedestrian volume models, potentially influencing safety studies conducted by Caltrans and local agencies.

Data collection comprised a three-step process, combining short-term crossing counts with long-term count data to develop expansion factors for annual volume estimates. Short-term count data were gathered from 583 locations using video-based methods, supplemented by pedestrian count datasets from local agencies. Long-term counts, collected from automated counters, aided in developing expansion factors, offering insights into pedestrian activity patterns over extended periods. The study's limitations included the use of a convenience sample for short-term count data, which could potentially bias predictions, and inherent errors in applying expansion factors to short-term counts. However, efforts were undertaken to address these limitations through the careful consideration of land-use-based factor groups and the selection of

model structure based on predictive accuracy. Despite these challenges, the developed model represented a significant advancement in pedestrian exposure modeling, providing insights into pedestrian risk on the state highway system and establishing a foundation for future research and safety studies by transportation agencies.

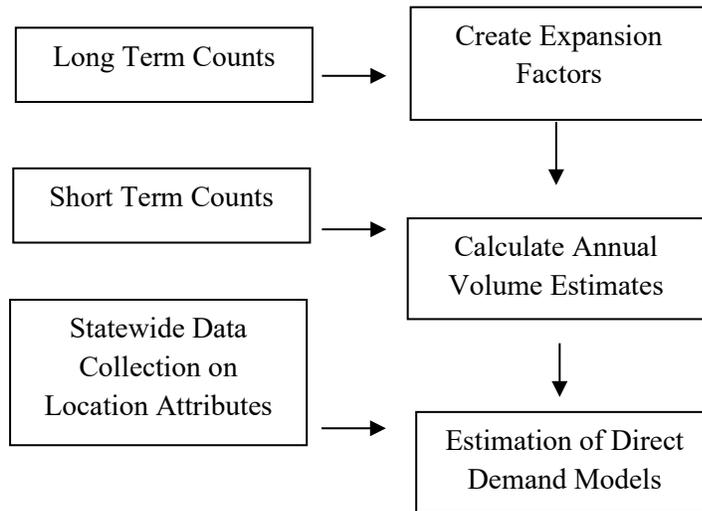


Figure 2. Research methodology by Griswold et al. (2019).

Li and Wu (2021) recently presented a new approach to estimate pedestrian volume at midblock crosswalks, which is vital for enhancing pedestrian signals and safety evaluations. Traditional methods were either time-consuming or costly due to the need for on-site data or expensive sensors. Their method, using button-pushing and signal timing events to model pedestrian arrivals as a Poisson process, provided a cost-effective and scalable solution. They developed two sub-methods tailored for one-stage and two-stage button-activated midblock crosswalks, addressing challenges like missing signal cycles at two-stage button-activated midblock crosswalks (BAMCs). Evaluation using on-site pedestrian volume data from two crosswalks showed promising accuracy, with low mean absolute errors of 2.27 and 1.78 ped/hour for one-stage and two-stage BAMCs, respectively. Sensitivity analysis recommended a one-hour interval for estimation to minimize errors.

The study selected two midblock crosswalks with different crossing strategies for data collection, utilizing event-based data collected through the MaxView system, capturing signal and pedestrian events. Pedestrian signal and button-pushing events were recorded to estimate pedestrian volume. The method addressed challenges such as stochastic button-pushing behavior

and missed pedestrian arrivals during signal phases by modeling pedestrian arrival as a Poisson process and utilizing maximum likelihood estimation (MLE). It offered benefits such as real-time pedestrian volume estimation for signal optimization and safety analyses, leveraging existing infrastructure and reducing the need for additional sensors and ground-truth data collection. However, potential underestimation of pedestrian volume in case of a significant number of pedestrians crossing against signals was noted, suggesting avenues for future research to improve accuracy, including consideration of traffic volume and clustered arrivals, and further validation against other sensors and complex signalized intersections.

Another recent study by Sobreira and Hellinga (2023) highlighted the effectiveness of direct demand (DD) modelling in estimating pedestrian volumes at intersections. The research aimed to qualitatively evaluate existing DD models and investigate their performance across different jurisdictions. Six DD models from various regions were selected and applied to three distinct jurisdictions: Milton, Canada; Pima County, U.S.; and Downtown Toronto, Canada. The models were scrutinized based on their capacity to estimate annual average daily pedestrian traffic (AADPT) at signalized intersections, with observed pedestrian volumes collected for sites in each jurisdiction. The evaluation of DD models unveiled significant disparities in model calibration data size, complexity, and pedestrian activity levels in the original jurisdiction. While some models utilized extensive calibration data, others relied on more limited datasets. Spatial transferability analysis yielded mixed results, with models performing adequately in jurisdictions resembling their original calibration settings, such as Milton. However, accuracy notably declined when applied to jurisdictions like Pima County and Downtown Toronto, characterized by diverse pedestrian activity levels and land use features. Consequently, the study underscores the necessity of aligning model calibration data with the target jurisdiction's characteristics to mitigate substantial estimation errors. The findings suggested that blindly applying DD models to jurisdictions with distinct land use and socioeconomic features can lead to significant inaccuracies.

Furthermore, Sobreira and Hellinga (2024) also conducted a similar study focused on the estimation of pedestrian exposure at intersections. They assessed methods for estimating pedestrian volume at crosswalks when only short-term counts (STCs) or no data were available. To achieve this, they developed a land-use (LU) model and an STC method to estimate

pedestrian volume shares per crosswalk. These methodologies were evaluated using continuous count data from various jurisdictions, with a naive equal-share assumption serving as a benchmark. Their findings indicated that the LU model notably enhanced allocation accuracy by 11.4%, while the STC method, particularly when utilizing multiple days of data, outperformed the naive method by up to 54.9%. Pedestrian counts were gathered through camera-based counting systems with image processing capabilities deployed at intersections. Initially collected as disaggregate raw counts, the data were aggregated into 24-hour counts for analysis. Additionally, 8-hour counts were considered as potential short-term counts for the analysis. Recommendations included utilizing STCs from up to three days for optimal accuracy, while acknowledging the need for further research on spatial model transferability and quantification of STC age impact on accuracy.

As mentioned earlier, in the literature, we did not find any pedestrian exposure-related study specifically for rail crossings. However, the study by Bosina et al. (2017) has some degree of relevance, as they investigated pedestrian volumes for railway platforms. They focused on understanding pedestrian trajectories on railway platforms, analyzing origins, destinations, and routes, aiming to address the challenge of designing platforms to accommodate diverse pedestrian movements. Utilizing two railway stations in Zurich, Zürich Hardbrücke and Zürich Stadelhofen, the research collected data on passenger behaviors and preferences during peak hours on working days. The study examined elements such as train doors, platform accesses, waiting positions, and service facilities to determine their impact on pedestrian trajectories. Data collection involved manual tracking of passenger movements, revealing patterns indicating a tendency to minimize walking distances and adapt behaviors based on station layouts and pedestrian density. Findings suggested that passengers optimize their walking routes on railway platforms, emphasizing the significance of station layout in influencing pedestrian flows.

The characteristics of the built environment, such as infrastructure layout, accessibility, and design elements, strongly influence pedestrian volume. Well-designed environments with clear pathways, efficient connectivity, and amenities tend to attract higher pedestrian traffic. Conversely, poorly designed, or inaccessible areas may deter pedestrians. A notable study by Chen et al. (2022) investigated the link between built environment characteristics and pedestrian volume at a population level, an area of research relatively overlooked compared to individual

walking behaviors. They employed a novel approach utilizing street view images and machine learning to extract citywide pedestrian volume. Focused on Shanghai's Middle Ring Road area, the study analyzed data from 127,921 sampling points along 28,397 street segments, covering 3,393 km. Pedestrian volume estimation, achieved by counting pedestrians in street view images using AI's machine learning interface, demonstrated high accuracy. The findings indicated positive associations between micro-scale environment features such as greenery, open sky, and sidewalk width, and pedestrian volume. Additionally, macro-scale characteristics including density, diversity, design, destination accessibility, and proximity to transit were correlated with pedestrian volume. Statistical analyses, including ordinary least square (OLS) regression and spatial lag model (SLM), underscored significant associations between various built environment factors and pedestrian volume, highlighting the independent influences of both micro- and macro-scale variables in promoting walking behaviors.

In conclusion, our thorough review of pedestrian exposure measures highlights the complexity of this crucial research area for pedestrian safety. Examining diverse methodologies and challenges reveals gaps in understanding, notably regarding pedestrian exposures at highway rail grade crossings. Further investigation and refinement of measurement techniques are warranted. In addition to the review presented above, Table 2 presents key explanatory factors that were studied and found to be associated with pedestrian exposure measures.

Table 2. Factors investigated in relation to pedestrian exposure measures.

Variables Studied	Past Research
Land Use Variable	
Nearby population density	Sobriera and Hellinga, 2023; Singleton and Runa, 2021 Munira et al. 2020; Griswold et al., 2019; Sze et al., 2019; Munira et al., 2017; Wier et al., 2009; Thakuriah, 2010; Schneider et al., 2009; Siddiqui et al., 2012; Lee et al., 2015; Wang et al., 2016; Lam et al., 2014; Ewing and Cervero, 2001; Handy, 2005, Krizek, 2003
Nearby housing unit density	Ewing and Cervero 2001; Handy, 2005
Nearby employment density	Handy, 2005; Shriver, 1997; Ewing and Cervero, 2001
Nearby land use mix; Proximity to mixed-use buildings;	Handy, 2005; Qin and Ivan, 2001 Ewing and Cervero, 2001; Shriver, 1997
Proximity to multistory buildings; Proximity to commercial buildings; Proximity to activity destinations	
Proximity to vacant lots	

Nearby building setback
distances; Proximity to parks

Dynamic Variables	
Average Annual Daily Traffic (AADT1)	Sze et al. 2019; Fagnant and Kockelman, 2016
Average Annual Daily Pedestrian Traffic (AADPT)	Singleton et al. 2021; Hankey et al. 2012

Transportation System Variables	
Four-way Intersection	Singleton and Runa, 2021; Griswold et al. 2019
Sidewalk presence on nearby streets	Handy 2005; Ewing and Cervero 2001 Ewing and Cervero 2001; Shriver, 1997
Nearby sidewalk connectivity	Sobriera and Hellinga, 2023; Handy 2005
Access to multiuse trails	Shriver, 1997
Nearby multiuse trail connectivity	Shriver, 1997 Griswold et al. 2019; Handy, 2005; Shriver, 1997
Access to transit	Krizek, 2003; Ewing and Cervero, 2001
Nearby street network connectivity	Ewing and Cervero, 2001 Ewing and Cervero, 2001
Nearby intersection density	Ewing and Cervero, 2001
Nearby four-way intersections	Handy, 2005; Ewing and Cervero, 2001
Buffer between sidewalk and street on nearby streets	Krizek, 2003; Ewing and Cervero, 2001 Handy, 2005; Ewing and Cervero, 2001
Presence of street trees on nearby streets	Ewing and Cervero, 2001 Ewing and Cervero, 2001
Presence of street lighting on nearby streets	Singleton and Runa, 2021 Singleton and Runa, 2021
Nearby street block length	Singleton and Runa, 2021
Number of arterial roadways nearby	Singleton and Runa, 2021 Singleton and Runa, 2021
Automobile speeds on nearby residential streets	Singleton and Runa, 2021 Singleton and Runa, 2021
Automobile parking spaces in the nearby area	Singleton and Runa, 2021 Singleton and Runa, 2021
Difficulty of crossing nearby streets	Singleton and Runa, 2021 Singleton and Runa, 2021; Ewing and Cervero, 2001
Residential address	Singleton and Runa, 2021
Commercial places	Munira et al. 2021; Handy, 2005; Shriver, 1997
Intersection density	Singleton and Runa, 2021
4-way intersections	Singleton and Runa, 2021
Schools	Singleton and Runa, 2021
Worship places	Singleton and Runa, 2021
Transit stops	Singleton and Runa, 2021
Park acreage	Singleton and Runa, 2021
The intersection is on a major road	Singleton and Runa, 2021 Muniar et al. 2020
Trail length	

**Demographic/Socioeconomic
Factors**

Student status; Larger household of unrelated individuals; Household automobile availability; Household income; Age
Number of employees; Vehicle ownership

2.2. Bicyclist Exposure

According to Statista, in 2022, there were approximately 54.7 million bike riders in the United States, marking a slight uptick from 2021's figure of 51.4 million. The National Bicycle and Pedestrian Documentation Project (NBPD), a collaborative initiative between Alta Planning and Design and the Institute of Transportation Engineers (ITE), strives to establish a standardized approach for bike counts and create a national repository for this data. Additionally, over 22 million bicycles were sold in the European Union in 2021. The global market for bikes and cycling accessories is significant, with the Netherlands boasting over 3,000 bike shops in 2021. Although the industry faced challenges due to the COVID-19 pandemic, sales have largely rebounded, with approximately 3.25 million bikes sold in Great Britain in 2021—a rise of more than 11 percent compared to the previous year. Promoting cyclist safety contributes to creating more inclusive and sustainable communities by encouraging alternative modes of transportation that reduce congestion and emissions.

Like pedestrians, railway crossings also pose significant safety concerns for cyclists due to the potential risks of crashes with trains or other vehicles. Despite the increasing emphasis on cyclist safety in transportation research, there remains a noticeable gap in the literature regarding the specific exposure of cyclists to crashes at rail crossings. The existing literature on cyclist exposure predominantly focuses on overall travel patterns, route preferences, and factors influencing modal choice. While these studies provide valuable insights into general cycling behavior, they often lack the specificity required to address safety concerns at rail crossings.

Understanding cyclists' exposure at railway crossings is crucial for developing targeted safety measures and infrastructure improvements. Although direct studies on this specific aspect

may be scarce, insights from similar transportation infrastructure can offer valuable perspectives. By reviewing existing studies on cyclist exposure in similar contexts, such as intersections, roadways, and shared pathways, we can extrapolate insights applicable to railway crossings. In this section, we have provided an overview of the latest research on cyclist exposure studies conducted globally. These tabulated details in Table 3 offer concise insights into the research context, data collection methods, and key findings. Additionally, to enhance comprehension, the studies are further elaborated upon in the subsequent section following Table 3.

Table 3. Recent bicyclist exposure/volume prediction studies.

No	Author(s)	Year	Data	Context	Location	Methodology	Results/Highlights
1	Ding et al.	2021	Bicycle trip-related data utilized from London public bicycle rental system, Santander bikes (270 Lower Super Output Areas were selected)	Role of exposure in bicycle safety analysis: Effect of cycle path choice	London, United Kingdom	Bicycle path analysis; simple shortest path model, and weighted shortest path model	Bicycle distance traveled, bicycle trips, bicycle time traveled, path distance, perceived safety level, land use, population characteristics, traffic flow, road infrastructure, environmental conditions, injury severity, location
2	Fournier et al.	2017	Two main data sources: continuous bicycle counters and bike-share systems	A sinusoidal model for seasonal bicycle demand estimation	six cities: Portland, Oregon; Arlington, Virginia; Seattle, Washington; Ottawa, Ontario; Vancouver, British Columbia; and Cambridge, Massachusetts.	Stepwise Linear Regression, and Sinusoidal function modeling	Bicycle demand was associated with seasonal variation, temperature differences, bike share usage, and monthly average daily bicycle count (MADB), and average annual daily bicycle count (AADB)

3	Dadashova and Griffin	2020	Data collected from 350 locations across 12 cities in Texas. Data sources were: Bicycle and Pedestrian Count Exchange Program, Strava, Texas DOT roadway inventory data, American Community Survey, and National Oceanic and Atmospheric Administration	Random parameter models for estimating statewide daily bicycle counts using crowdsourced data	Austin, Dallas, and Houston	Mixed effects models, Random parameter or mixed-effects models with autocorrelated errors, and Fixed effects models	Spatial factors (non-motorized facility type, roadway functional class, openstreet map class), temporal factors, daily strava user counts, weather conditions, socioeconomic factors, and roadway design and characteristics
4	Miah et al.	2023	Data from 2017 to 2019 included 6746 daily bicycle volumes from 178 permanent and short-term count locations	Estimation of daily bicycle traffic using machine and deep learning techniques	Portland, Oregon, USA	Eight modeling techniques ranging from advanced techniques, such as Convolution Neural Network (CNN), Deep Neural Network (DNN), Shallow Neural Network (SNN),	Key variables: anonymous bicycle user activities, built environments, motorized traffic, and sociodemographic characteristics. Strava counts, bike share data, weather data, and spatial variables

						Random Forest (RF), XGBoost, to conventional and simpler approaches, such as Decision Tree (DT), Negative Binomial (NB), and Multiple Linear Regression	
5	Esawey	2014	Bike activity data utilized from 12 permanent count stations in Vancouver during 2009-2011	Estimation of annual average daily bicycle traffic with adjustment factors.	Vancouver, Canada	Regression analysis and error analysis used to estimate annual average daily bicycle (AADB) traffic volumes. Adjustment factors (DFs and MFs) derived from observed bicycle volume data were applied to adjust raw data collected from count stations	Key variables: bicycle volume data, Daily Adjustment Factors (DFs), Monthly Adjustment Factors (MFs), Actual and Estimated AADB Volumes and Error Metrics
6	Esawey	2017	Dataset comprising more than 14,000 daily bicycle volumes collected between 2009 and 2011	Estimation of Daily Bicycle Traffic Volumes Using Spatiotemporal Relationships	Vancouver, Canada	Parametric linear regression, implemented using the ordinary least squares (OLS) method	Key variables: daily cycling volumes, temporal correlation, spatial correlation, and weather conditions

7	Fagnant and Kockelman	2016	Cyclist-count data from 251 locations in the Seattle metropolitan area. conducted by the Puget Sound Regional Council in October 2010, with over 340 three-hour counts collected during that month. Tuesdays through Thursdays, 6 am to 9 am, or 3 pm to 6 pm.	A direct-demand model for bicycle counts: the impacts of level of service and other factors	Seattle metropolitan area	Poisson regression count model, and two negative binomial models	Key variables: cyclist count, population density, employment density, bicycle trail-access, bridge, AADT, temperature, number of lanes, speed limit, presence of parking, and residential area indicator
8	Griswold et al.	2011	Bicycle counts collected at 81 intersections along arterial and collector streets (2008-2009)	Pilot models for estimating bicycle intersection volumes	Alameda County, California	Log-linear regression modeling	Key explanatory variables for modelling were: intersection characteristics, land use, transportation system, socioeconomic characteristics, terrain characteristics, and roadway network characteristics
9	Hochmair et al.	2019	Strava tracking data, which provides GPS information on cycling activities, and sociodemo	Estimating bicycle trip volume from Strava tracking data	Miami-Dade County, Florida	Linear Regression Modelling, and Eigenvector Spatial Filtering (ESF)	Key variables: Strava Bicycle Kilometers Traveled (Strava BKT), length of road Segments, Functional classes, types of bicycle facilities, total length of walkways suitable for Cycling, number of controlled and

			graphic data were obtained from the 5-year (2009–2014) summary of the American Community Survey (ACS)				uncontrolled intersections land use traits, population density, and household income
10	McDaniel et al.	2014	Manual counts conducted by citizen-volunteers standing on assigned street corners for a 2-hour period in the morning and evening (Data split: 90% for calibration and 10% for validation)	Using origin–destination centrality to estimate directional bicycle volumes	Moscow, Idaho	O-D centrality approach, and development of O-D centrality metric; spatial interpolation and visualization	Key variables: distance between origin and destination, origin and destination multipliers, preferred bicycle paths, link impedance, and node impedance
11	Munira et al.	2021	Bicyclist activity data encompassing over 400 variables across three buffer zones (161 m, 804 m, and 1,609 m) within the city of Austin	Estimating Bicycle Demand: Role of a Bikeability Index	Austin, Texas	multiple linear regression model	Key variables were demographics, socioeconomics, infrastructure, transit facilities, major generators, and land use.

Fournier et al. (2017) addressed the challenge of estimating seasonal bicycle demand by devising a sinusoidal model that required minimal calibration data. They utilized data from bike-share systems in multiple cities and 47 permanent bicycle counters to develop this model. By employing sinusoidal functions, they established calibration factors with just two short-term counts, rendering the model applicable even in locales with scant demand data. Successfully estimating both monthly average daily bicycle counts and average annual daily bicycle counts, the model emerged as a valuable asset for transportation planning and infrastructure enhancements.

In a recent study, Dadashova and Griffice (2020) developed a direct-demand model aimed at estimating daily bicycle counts using crowdsourced data alongside socioeconomic and weather indicators, addressing the persistent lack of non-motorized traffic counts that hinder evidence-based decisions in transportation planning and safety. Data were gathered from diverse sources, including the Texas Bicycle and Pedestrian Count Exchange Program, Strava, Texas DOT roadway inventory data, American Community Survey, and National Oceanic and Atmospheric Administration. Approximately 350 locations were utilized across 12 Texas cities, including Austin, Dallas, and Houston; the data collection incorporated both permanent and temporary count stations, reflecting weekly and monthly variations. Fixed and random effects models were developed to identify influential factors, utilizing mixed-effects modeling with autocorrelated errors to predict daily bicycle counts, with cross-validation conducted on 80% of locations for model building and 20% for prediction. Results demonstrated that the mixed-effects model achieved a Mean Absolute Percentage Error (MAPE) of 29%, surpassing simple scaling methods, highlighting the importance of combining information from counts and Strava data rather than relying solely on scaling. Despite recognizing limitations such as the need for further research to enhance model generalizability, reduce estimation error, and simplify model application for practitioners, the study contributes to understanding the utilization of emerging data sources for estimating bicyclist traffic, emphasizing the significance of accounting for external factors in direct-demand modeling.

Miah et al. 2023 addressed the gap in utilizing machine learning (ML) techniques for estimating non-motorized bicycle traffic, an area that has traditionally relied on simple econometric models due to data limitations. Recent advancements in smartphone-based location

data collection offer the potential to apply ML techniques to estimate daily bicycle traffic volumes. Data for the study was gathered from Portland, Oregon, spanning 2017 to 2019. It included 6746 daily bicycle volumes from 178 permanent and short-term count locations. Key variables included anonymous bicycle user activities, built environments, motorized traffic, and sociodemographic characteristics. Various data sources such as Strava counts, bike share data, weather data, and spatial variables were incorporated.

Eight modeling techniques were developed, ranging from advanced ML methods like Convolution Neural Network (CNN), Deep Neural Network (DNN), Shallow Neural Network (SNN), to conventional approaches like Decision Tree (DT), and Multiple Linear Regression. Two variable dimension reduction techniques, Principal Component Analysis (PCA) and random forest variable importance analysis, were employed to prevent over-generalization. K-fold cross-validation was used for model evaluation, and hyperparameter tuning was conducted using grid search and brute force techniques. The study found that SNN and DNN ML techniques yielded higher accuracies in estimating daily bicycle volumes compared to conventional methods. The DNN model without variable reduction outperformed other models with a mean absolute percentage error (MAPE) of 22% and an R-squared (R²) value of 0.86. In addition, strava count, weekends, bike share crossing, average temperature, and several network features were highlighted to be significant predictors of bicyclist volume.

Esawey (2014) delved into the accuracy of estimating annual average daily bicycle (AADB) traffic volumes through adjustment factors, specifically daily and monthly factors (DFs and MFs). He used data gathered from 12 permanent count stations in Vancouver during 2010 and compared the efficacy of using MFs versus seasonal factors, highlighting MFs as the superior choice. The study revealed that combining both DFs and MFs results in an overall error rate of approximately 23%. Additionally, it explored the transferability of MFs across different years, finding that employing factors from the same year yielded higher accuracy. Furthermore, the analysis attributed about 15% of the error to DFs and 11% to MFs, shedding light on the factors influencing estimation accuracy. The findings of this study provided valuable insights into the optimization of estimation methodologies for AADB volumes, particularly in the context of bicycle traffic. Furthermore, the study explored temporal transferability that underscores the importance of considering data from the same year for optimal accuracy in estimating AADB

volumes. Overall, the study contributed to enhancing the reliability of estimation techniques and informed the design of effective data collection programs for bicycle traffic management.

In another study conducted by Esawey (2017), the challenge of missing cycling traffic volume data due to sensor malfunctions was investigated, with a focus on count stations frequently experiencing issues. The study utilized a dataset spanning from 2009 to 2011, containing over 14,000 daily bicycle volumes from 22 count stations in Vancouver, Canada. Correlation analyses revealed significant correlations among most count stations, with cross-correlation analysis confirming strong relationships, particularly when occurring concurrently. Consequently, statistical models were developed to relate daily cycling volumes among neighboring stations, showcasing mean absolute percentage errors generally below 20%, which improved with higher correlation thresholds. The results suggested the efficacy of this approach in estimating missing cycling volumes, potentially aiding monitoring programs and data clearinghouses in addressing sporadic data gaps. Furthermore, Lewin (2011) provided the first comprehensive analysis of five years of detector data for two permanent bicycle count stations (representing four locations) on multi-use paths in Boulder, CO. First, temporal patterns of daily bicycle counts were explored. A strong linear correlation between high temperatures and daily counts was noted with a slight decrease in counts at temperatures greater than 90° F. Counts also decreased on days with rain or snow, although this effect was not linear. In addition, the numbers decreased on weekends at most locations. From this information seasonal factors for bicycle counts were also estimated.

Transportation planning in the US has traditionally prioritized automotive traffic, but there's been a shift towards a multimodal approach to accommodate all users, especially cyclists, who face significant risks on the road. Unfortunately, comprehensive bicycle counts are rare in most municipalities. Fagnant and Kockelman (2016) focused on developing a direct-demand model for estimating peak-period cyclist counts in the Seattle metropolitan area based on various factors, including roadway conditions and trip-generation/attraction factors. The data collection process involved obtaining cyclist-count data from 251 locations in the Seattle area, collected by the Puget Sound Regional Council in 2010. The methodology included developing suitability attributes for major roadway approaches at intersections based on Highway Capacity Manual procedures and employing Poisson and Negative Binomial models for cyclist count estimation.

Results indicated that wider bike lanes and curb lanes, along with lower traffic volumes, created favorable conditions for higher cyclist counts. The negative binomial model showed consistent results with the Highway Capacity Manual's bicycle level of service index, highlighting the impact of roadway features on cyclist counts.

Bicycle volume data are crucial for understanding various aspects such as safety, travel behavior, and development impacts. To address this need, several simple models of bicycle intersection volumes were developed in a study by Grisworld et al. (2011) for Alameda County, California. The county encompasses diverse areas in terms of population density, employment density, and socioeconomic characteristics. The study conducted 2-hour bicycle counts at 81 intersections during the spring of 2008 and 2009, representing arterial and collector roadways. The chosen intersections ensured a wide representation of different built environment characteristics, including proximity to commercial properties, major universities, and variations in terrain and roadway connectivity. Log-linear regression was employed to model bicycle intersection volumes. This approach was chosen for its simplicity and ease of application using geographic information systems and spreadsheet software. The models considered various explanatory variables such as land use, transportation systems, and socioeconomic characteristics. Notably, the modeling process revealed significant associations between bicycle volumes and factors such as proximity to commercial properties, major universities, presence of bicycle facilities, terrain flatness, and roadway network connectivity. Differences in bicycle volumes between weekdays and weekends were also observed, highlighting the importance of temporal considerations in modeling. The results indicated that the developed models offer valuable insights into the factors influencing bicycle volumes at intersections. However, further refinement and testing are necessary to improve the accuracy and applicability of the models, particularly in other communities. While the preliminary models provide a useful starting point for estimating bicycle intersection volumes, they are tailored specifically to Alameda County and may require adaptation for other regions.

Sports and fitness apps on GPS-enabled cell phones and smartwatches have provided rich GPS tracking data for nonmotorized traffic activities like walking, running, and cycling. Leveraging Strava tracking data from Miami-Dade County, Hochmair et al. (2019) examined the factors associated with variations in bicycle ridership across different areas. The study area

included the Urbanized Area of Miami-Dade County, extending into agricultural land and the built environment towards Everglades National Park. The data collection process involved obtaining Strava Metro roll-ups, shapefiles, and sociodemographic data from various sources including the Florida Department of Transportation and the American Community Survey. Linear regression models were estimated to predict non-commute and commute bicycle kilometers traveled per block group, as well as bicycle kilometers traveled on weekends and weekdays. Eigenvector spatial filtering was applied to account for spatial autocorrelation and avoid parameter estimation bias. Explanatory variables included network characteristics, built environment features, and sociodemographic factors.

The study analyzed eight models, combining nonspatial and spatially filtered linear regression, to explore the relationship between bicycle ridership and predictor variables. Results indicated that Strava data, despite potential biases towards male and younger users, offers extensive coverage and detailed insights into cycling behavior. The study found significant associations between network characteristics, built environment features, sociodemographic factors, and bicycle ridership. For instance, bicycle facilities on local and collector roads were associated with increased ridership, while those on arterial roads showed no significant effect. The presence of bay bridges and central road segments had a stronger positive effect on weekday cycling volume compared to weekends. The study provided guidelines for the practical design of bicycle infrastructure improvements, suggesting that innovative approaches may be needed for high-traffic roads to attract more cyclists.

The research conducted by McDaniel et al. (2014), showcased a case study in Moscow, Idaho, a city with high bicycle ridership due to its proximity to the University of Idaho. A new method for estimating directional bicycle volumes across a street network was introduced in their research. This method, termed O-D centrality, was based on a modified version of centrality from graph theory, aiming to quantify the relative importance of links and nodes within a network. Data for calibration and validation were randomly split into two subsamples, with various regression techniques explored for direct demand modeling. The study utilized counts collected manually, and the O-D centrality metric was developed by modifying stress centrality in three keyways: defining preferred bicycle paths, considering a specific subset of origin-destination pairs reachable by bicycle, and incorporating origin and destination multipliers to

represent trip potential. The study compared the performance of O-D centrality against conventional stress centrality in predicting bicycle volumes. O-D centrality exhibited strong explanatory and predictive power, with significantly better model fit compared to stress centrality. The research demonstrated the utility of O-D centrality in estimating and spatially interpolating bicycle volumes across the community, offering advantages over traditional multistep demand models and direct demand models.

In a recent study, Munira et al. (2021) addressed a gap in understanding bicycle traffic demand by introducing a composite measure called the bikeability index (BI). This index aimed to enhance direct demand models for bicycle traffic, particularly in situations where data on bicycle demand was limited. The study explored additional variables, such as the presence of bike-sharing stations, bike signals, and bike-accessible bridges around intersections, to gauge their impact on bicycle volume. Raw datasets from various sources, processed to achieve homogeneous spatial scales, yielded over 400 variables across three buffer zones (161 m, 804 m, and 1,609 m), categorized into seven groups. The study focused on the city of Austin, Texas, developing a bikeability index combining attributes like bicycle route length, comfort, connectivity, destination density, and transit coverage. This index significantly influenced bicycle volume around intersections, alongside sociodemographic variables like African American population, income, and age groups. Regression analysis underscored the importance of built environment features in determining bike traffic behavior. Findings highlighted the significant influence of the bikeability index on bicycle volume, emphasizing built environment features' importance in promoting cycling. Lower-income and older populations showed higher bicycling propensity, and infrastructure variables like bike signals and bicycle-accessible bridges positively impacted bicycle volume.

Ding et al. (2021) aimed to enhance exposure estimation in bicycle safety analysis by utilizing detailed trip data from the London public bicycle rental system. Two modeling approaches, the shortest path method (SPM) and weighted shortest path method (WSPM), were employed to estimate bicycle path choice and distance traveled. Bicycle crash frequency models incorporating these estimates were developed and compared. The Poisson regression method, supplemented by negative binomial (NB) regression to address over-dispersion, was used for bicycle crash analysis, incorporating bicyclist daily traffic (BDT) estimates from SPM and

WSPM. Results showed that WSPM-based models provided a better fit than SPM-based ones, with WSPM exhibiting the best performance. Bicycle crash frequency models incorporating BDT as exposure outperformed those using trip counts or time travel. Factors such as land use, population characteristics, and traffic conditions significantly influenced crash frequency.

2.3. Combined Pedestrian and Bicyclist Volume/Exposure Studies

In the literature review conducted in the preceding sections, we examined individual studies concerning pedestrian and cyclist exposure. However, given that non-motorists encompass both groups, we also looked into research focusing on their exposure across various transportation infrastructures. Table 4 offers a brief overview of combined non-motorized exposure prediction studies. Moreover, in subsequent sections, we discuss these studies in detail. Notably, we observed a gap in the current literature regarding non-motorized exposure predictions specific to rail crossings. Nonetheless, utilizing methodologies from similar studies can be instrumental in advancing our research in this domain.

Table 4. Combined pedestrian and bicyclist volume/exposure studies.

SN.	Author(s)	Year	Data	Context	Location	Methodology	Results/Highlights
1	Ermagun et al.	2018	data collected between January 1, 2014, and February 16, 2016	Bicycle, Pedestrian, and Mixed-Mode Trail Traffic: A Performance Assessment of Demand Models	seven major climatic regions in the continental U.S.	Trail demand modeling by utilizing generalized linear models	Key variables: density, diversity, design, distance to transit, and destination accessibility, average daily bicyclists, average daily pedestrians, average daily mixed-mode traffic, annual average daily bicyclists, annual average daily pedestrians, annual average daily mixed-mode traffic
2	Johnstone et al.	2018	continuous bicycle and pedestrian counts	Annual Average Nonmotorized Traffic	six U.S. cities, namely Arlington,	Grouping Sites and Traffic	Key variables: time periods, days of week, hourly commute, bicycle

			(2002 to 2016)	Estimates from Manual Counts: Quantifying Error	Boulder, Mount Vernon, Portland, San Diego, and Seattle	Distribution Index (AMI)	and pedestrian counts
4	Hankey and Lindsey	2016	peak period counts of pedestrian and bicycle traffic collected (954 separate observations collected at 471 locations)	Facility-Demand Models of Peak Period Pedestrian and Bicycle Traffic: Comparison of Fully Specified and Reduced-Form Models	Minneapolis, Minnesota	Statistically optimal models, reduced-form core models, and reduced-form time-averaged models	Population density, employment density, industrial area, retail area, open space, presence of bicycle facilities, transit stops, proximity to principal arterials, proximity to freeways
5	Hankey et al.	2012	counts of cyclists and pedestrians between 2007 and 2010 at 259 locations	Estimating Use of Non-Motorized Infrastructure: Models of Bicycle and Pedestrian Traffic	Minneapolis, MN	Negative binomial models	weather, neighborhood socio-demographics, built environment characteristics, and road (including presence of bus line) or bicycle facility type
7	Lu et al.	2018	Count data from a non-motorized traffic monitoring campaign at 173 locations	Adding Temporal Information to Direct-Demand Models: Hourly Estimation of Bicycle and Pedestrian Traffic	Blacksburg, VA	Hourly direct-demand models of bicycle and pedestrian traffic (stepwise linear regression approach)	land use features, transportation network characteristics, time of day (as a temporal variable), buffer sizes of independent variables (for stepwise regression), choice of time periods (for sensitivity analysis), choice of day of week (for sensitivity analysis)

In response to the need for enhanced data on pedestrian and cyclist movement, policymakers and scholars collaborated to innovate new methodologies for estimating their traffic volumes. This involved exploring the potential of crowdsourced mobile data, which offered broader spatial and temporal coverage at a lower cost compared to conventional methods. Despite inherent limitations and biases, utilizing such data refined our understanding of pedestrian and cyclist traffic patterns. The recent detailed literature review by Tao et al. (2024) systematically cataloged the utilization of crowdsourced mobile data in estimating pedestrian and cyclist traffic volumes. Based on the earlier studies, it highlighted a predominant reliance on a particular source of commercial fitness app data, notably Strava, over other crowd-sourced sources. Notably, the focus was primarily on estimating cyclist volumes, with relatively fewer studies addressing pedestrian volumes. The prevalent approach involved employing crowdsourced counts as independent variables in direct demand models. These models aimed to predict traffic volumes based on various influencing factors, with crowdsourced data playing a significant role. Studies presented in their review paper demonstrated a strong correlation between variables derived from crowdsourced mobile data and observed counts in statistical models. Moreover, machine learning models revealed the relative importance of crowdsourced data over other factors in predicting traffic volumes. Incorporating crowdsourced mobile data into estimation models consistently enhanced performance, indicating its potential to refine traffic volume estimates.

Ermagun et al. (2018) conducted a study aimed at developing trail demand models across 32 locations in the U.S., representing seven major climatic regions. Their research focused on predicting average daily pedestrians (ADP), average daily bicyclists (ADB), and average daily mixed-mode traffic (ADM) using built environment variables and socio-economic characteristics. Integrated sensors were utilized for data collection, with manual validation counts ensuring data accuracy. Negative binomial regression models were employed for estimating traffic demand, and cross-validation techniques, including leave-one-out cross-validation, were used to assess model performance. The study found differences in correlation between pedestrian and bicycle traffic volumes, moderate accuracy of the models for general planning purposes, and no significant improvement in predicting total demand using separate sensors for bicycles and pedestrians. However, post-validation techniques showed promise in

enhancing prediction accuracy, underscoring the study's contribution to understanding and predicting trail traffic demand for planning and management purposes.

Johnstone et al. (2018) investigated optimal times for conducting manual counts to accurately estimate annual average daily nonmotorized traffic (AADNT) across six U.S. cities. The study aimed to guide manual count programs, commonly used due to limited funding, to enhance data quality and resource utilization. Continuous bicycle and pedestrian counts from these cities were analyzed to estimate AADNT and evaluate estimation errors for various short-duration count scenarios. Results indicated that employing two permanent counters per factor group significantly reduced error, with afternoon counts, particularly from 2:00 to 6:00 p.m., exhibiting the lowest error rates. Additionally, Sunday counts often showed lower error rates compared to Saturdays, contrary to prior findings. The study emphasized the importance of reliable traffic count data for planning bicycle and pedestrian infrastructure, offering recommendations for optimizing manual count programs to enhance data quality and minimize error.

Hankey and Lindsey (2016) undertook a study aimed at developing facility-demand models for pedestrian and bicycle traffic in Minneapolis. Their research, based on peak period counts of pedestrian and bicycle traffic, involved exploring fully specified models as well as reduced-form models. The dataset comprised 954 observations from 471 locations, providing ample spatial density for constructing spatially resolved models. Utilizing stepwise linear regression, they estimated three sets of models: statistically optimal models, core models aligned with theoretical consistency, and time-averaged models. The primary objective was to compare the performance of these models in explaining traffic variations. Independent variables, such as land use and transportation network characteristics, were selected at different spatial scales using a stepwise approach. The models aimed to produce block-level traffic estimates for improved spatial understanding. Results indicated that both reduced-form models and statistically optimal models performed similarly in explaining traffic variations. However, reduced-form models were favored for their simplicity and ease of interpretation, rendering them more practical for application. They suggested designing future sampling campaigns for estimating long-term averages and selecting count locations specifically for spatial modeling to enhance model

performance. On non-motorist volume prediction modelling, other similar studies were also conducted by Hankey et al. in 2011, Lindsey in 2011, and Jones et al. in 2010

Hankey et al. 2012 conducted a study aimed at addressing the notable gap in data and tools for estimating non-motorized traffic on various transportation facilities such as sidewalks, trails, and bike lanes, which according to them, hindered evidence-based decision-making for transportation infrastructure investments. Their comprehensive analysis, focused on Minneapolis, MN, involved gathering cyclist and pedestrian counts from 2007 to 2010 across different locations. They developed scaling factors to extrapolate daily counts from hourly data and constructed regression models considering factors like weather, socio-demographics, and street characteristics to estimate non-motorized traffic. The study revealed that pedestrian traffic exceeded bicycle traffic by 35%, with weather conditions, neighborhood characteristics, and the presence of bicycle facilities being significant correlates. Bicycle traffic displayed an increasing trend over time, particularly on streets with bicycle lanes or off-street facilities, emphasizing the importance of investing in bicycle infrastructure.

The developed models predicted non-motorized traffic for cases where direct counts were unavailable and for assessing the impact of changes in the built environment, such as adding bicycle lanes or altering land use. Additionally, the study identified peak-hour traffic patterns, with the peak occurring between 5:00 pm and 6:00 pm for both cyclists and pedestrians, although mid-day hours contributed more significantly to pedestrian traffic. Furthermore, regression models indicated better fitting for bicycle traffic, suggesting significant influences from neighborhood design features and bicycle facilities. Pedestrian traffic, however, was associated with road classification, proximity to amenities, and neighborhood socio-demographics, showing no evident increase over time. Some aspects similar to this research are also discussed in Lindsey (2011).

Lu et al. (2018) focused on developing direct-demand models to estimate bicycle and pedestrian traffic volumes in Blacksburg, VA; their study was based on hourly traffic data, aiming to overcome the limitations of traditional models that often lack the ability to estimate hazard exposure by time of day. The study introduced Annual Average Hourly Traffic (AAHT) estimates for bicycles and pedestrians, integrating spatial and temporal factors into hour-specific

models and a spatiotemporal model, which demonstrated reasonable goodness-of-fit results. The research elucidated how temporal variability influences spatial traffic patterns, identifying correlations between major land use and transportation variables with non-motorized traffic throughout the day. By emphasizing the significance of considering time of day alongside spatial variables in predicting traffic volumes, the study's spatiotemporal models provided a practical approach for estimating traffic volumes for various time periods.

When modeling pedestrian and bicyclist exposure in transportation research, several units and measures have been used in past research, some of them were also reviewed in the earlier sections. Table 5 outlines some commonly utilized metrics that may be utilized for non-motorist exposure analyses at rail crossings.

Table 5. Summary of non-motorist exposure measures from the literature review.

Exposure Measure	Description
Person-Miles	This measures the exposure of pedestrians or bicyclists in terms of the total distance traveled by each individual, summed across all individuals. It provides a measure of the total exposure to risk.
Trip Counts	This measures exposure in terms of the number of trips made by pedestrians or bicyclists within a certain period. It helps in understanding the frequency of exposure events.
Time Spent	Exposure can also be measured in terms of the time spent by pedestrians or bicyclists at rail crossings/other transportation infrastructures. This includes waiting time at crossings, crossing time, and any other time spent in the vicinity of the rail crossing.
Crossing Events	This measure focuses specifically on the number of times pedestrians or bicyclists cross the intersection. For a rail crossing scenario, it may provide insight into the frequency of interactions between non-motorists and trains.
Number of non-motorists	Simply counting the number of pedestrians and bicyclists passing through or near rail crossings/other transportation infrastructures may provide a basic measure of exposure.
Density Metrics	Metrics such as pedestrian or bicyclist density in the vicinity of rail crossings can also be used to quantify exposure. This considers both the number of non-motorists and the area they occupy.
Risk Index	A risk index may be developed that combines various exposure measures with crash data to assess the relative risk faced by pedestrians and bicyclists at rail crossings.
Non-motorist Volume	Counting the number of pedestrians and bicyclists passing through or near a highway intersection/rail crossings over a specified period may provide a basic measure of pedestrian exposure
Non-motorist Activity Index	This index combines pedestrian and bicyclist volume with other factors such as land use characteristics, infrastructure features, and environmental conditions to quantify pedestrian exposure more comprehensively.

Non-motorist Proximity	Measuring the distance between pedestrian and bicyclist pathways and intersections/rail crossings may help assess exposure based on spatial relationships.
------------------------	--

3. Data Collection

Prior to the collection of non-motorist volume data, a planning phase was conducted. Emphasis was placed on video recording proximity to the University of Nebraska Lincoln campus in Lincoln, Nebraska, and surrounding suburban towns including Hickman, and Roca. The selection of rail grade crossings in these locations aimed to ensure accessibility, prioritize the safety of student and staff workers, and optimize available resources. Non-motorist activities at suburban rail crossings were specifically examined, distinct from urban counterparts, which typically exhibit lower traffic volumes and are situated in less densely populated areas, resulting in fewer pedestrian and vehicular crossings. These locations often feature enhanced warning systems, such as extended gate lengths, additional flashing lights, and louder audible signals, attributed to higher train speeds. A total of 21 sites were utilized for data collection, of which 7 were monitored using City of Lincoln cameras, while the remainder utilized Miovision Scout for recording non-motorist volumes over multiple 24-hour periods. Miovision Scout is a portable, camera-based system designed for collecting traffic data, particularly for traffic volume studies. It is equipped with high-definition video capabilities, allowing it to record vehicles, pedestrians, and cyclists as they pass through intersections or road segments. This data are then processed using Miovision's analytics software, which extracts valuable information such as traffic counts, vehicle classifications, and non-motorist volumes. Miovision Scout is commonly used for assessing traffic loads and patterns over specific time periods, which is essential for transportation planning, road safety evaluations, and infrastructure design (Figure 3). Locations of the selected sites can be seen in Figure 4.

A total of 77 24-hour video recordings were captured across different months of the year to ensure that seasonal variations are randomly accounted for in the exposure analysis of rail crossings. Rail grade crossings can be affected by various environmental and weather conditions such as rain, snow, fog, ice, and extreme temperatures. Rain can reduce visibility and make tracks slippery, increasing the stopping distance for trains and the risk of vehicles skidding. Snow and ice can obscure rail signals and signs, and cause difficulties for both trains and vehicles in maintaining traction. Fog significantly reduces visibility, making it harder for non-motorists as well as motorists to see oncoming trains or crossing signals in time. Extreme temperatures can affect the integrity of the rail tracks, causing them to expand or contract, and can also impact the functioning of crossing warning systems. It is crucial to consider these environmental

and weather conditions because they directly affect the safety and behavior of non-motorists at the HRGCs.



Figure 3. Photographic image of scout hardware (Miovision.com).

Poor visibility and slippery surfaces increase the risk of crashes. Non-motorists are particularly vulnerable as they do not have the protection of a vehicle and may be less aware of the dangers posed by crossing tracks in adverse conditions. Ensuring proper signage, warnings, and safety measures can help mitigate these risks and protect all users of rail crossings. Other important environmental and weather-related data were also recorded during the video sessions including, average 24-hour visibility in miles, average 24-hour maximum wind speed in mph, average 24-hour total precipitation in inches, rail-crossing surface conditions, intersecting road surface conditions, and lighting conditions at the crossings during nighttime. Moreover, various factors specific to crossing infrastructure and the surrounding environment were examined. These factors include the presence of sidewalks, the number of highway lanes, and the types of non-motorist activity areas adjacent to the HRGCs, such as residential, commercial, and downtown zones. Additionally, the presence and clarity of signage and signals were considered as well-maintained and prominent signs help alert non-motorists of an approaching train.

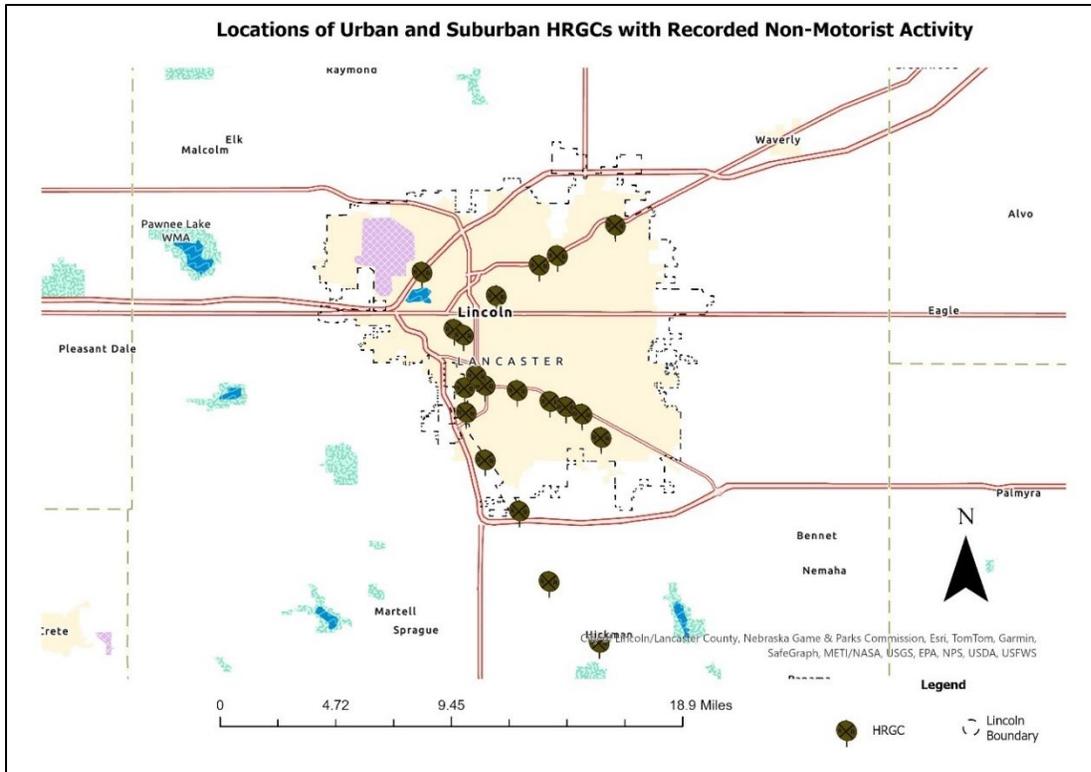


Figure 4. Locations of urban and suburban HRGCs with recorded non-motorist activity.

A critical aspect of the analysis was observing the presence of sidewalks at rail crossings. Sidewalks are essential for providing safe pathways for pedestrians and cyclists, offering a designated space away from vehicular traffic. The number of highway lanes was also evaluated, as crossings with more lanes may present greater risks for non-motorists due to increased traffic volume and complexity in navigating the crossing. The characteristics of zones often influence transportation behaviors; by considering the type of area—whether residential, commercial, or mixed-use—varying levels of pedestrian and cyclist activity that typically occur in different settings were accounted for. The availability of parking infrastructure within a 0.5-mile radius of the crossings was also investigated. The presence of parking lots or structures can influence the flow of pedestrian traffic, as individuals may park their vehicles and then walk to nearby destinations, increasing the likelihood of non-motorist activity at the crossings. Additionally, the type of intersection near the crossings was an important factor, as complex intersections with multiple roads converging can create additional hazards for non-motorists, which can affect non-motorist crash exposure at HRGCs. The proximity of schools within a 0.5-mile radius of the rail crossings was another significant element in the data collection. Schools generate substantial non-motorist traffic, including children walking or biking to and from school. Industrial areas generally experience higher vehicular traffic, with

limited pedestrian infrastructure and fewer amenities such as parks or retail centers that typically attract non-motorists. Consequently, non-motorist activity, including walking or cycling, tends to be lower in these zones. Therefore, industrial areas within a 0.5-mile radius of HRGC locations were also considered in this study to assess their impact on non-motorist traffic. Figure 5 identifies the proximity to local amenities within a 0.5-mile radius of HRGCs.

While analyzing the recorded videos, train traffic was also monitored during the day and nighttime to understand the activity of the trains. Initially, the recorded videos were identified and incorporated in the database based on the names of streets/highways they were intersecting, and later, their actual location was matched with U.S. DOT crossing IDs to get the correct Latitude and Longitudes of each crossing. For volume counts, data was both manually counted and then Algorithms were used for non-motorist counts of pedestrian and bicyclist. For identification purposes, each video was given a batch number. A total of 3 Batches of videos were recorded and distributed among team members. Batch I videos were captured via city of Lincoln cameras, while Batch II and Batch III videos were recorded using Miovision Scout. Each data point in the database, representing a 24-hour recording period, includes the start and end dates of the video capture.

Additionally, the start time and end time of each video were recorded in 24-hour format. For each recorded period, counts of pedestrians and cyclists were documented separately, along with a combined count of non-motorists. The team ensured the accurate inclusion of key physical attributes of rail crossings in the data, cross-referencing with the FRA's inventory dataset for HRGCs and verifying on-site observations. Photos were taken for recordkeeping purposes to document these aspects comprehensively. These important physical attributes verified on the site include the number of crossbuck assemblies, stop and yield signs, posted speed limits on highways and streets intersecting rail crossings, number of main tracks, siding tracks, yard tracks, transit tracks, track signalization, advance warning signs, low ground clearance signs, pavement markings, channelization devices, exempt signs, ENS signs, count of crossing gate arms, gate configuration, type and number of flashing lights, number of warning bells, highway traffic signal interconnection, highway traffic signal preemption, presence of a paved highway, crossing surface description, smallest crossing angles, and road at crossing type.

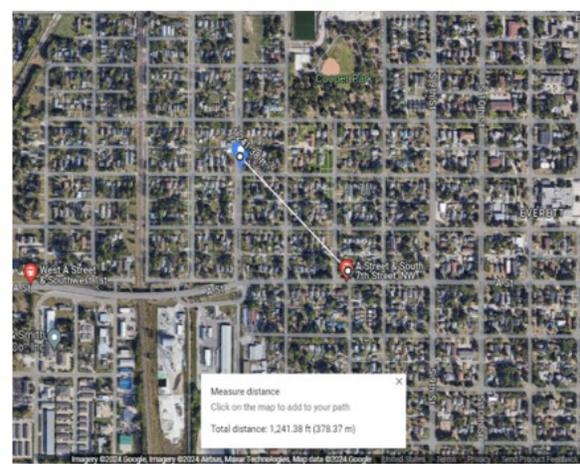
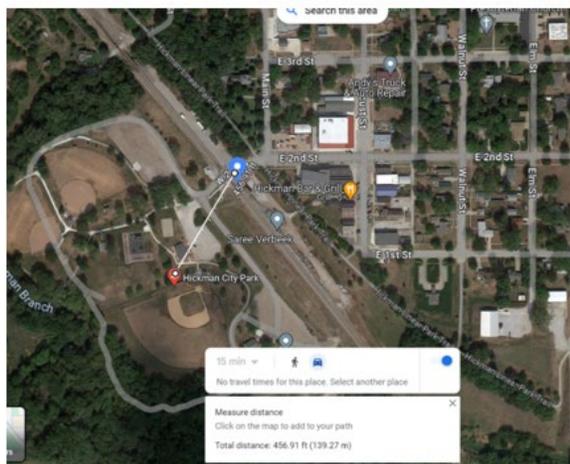
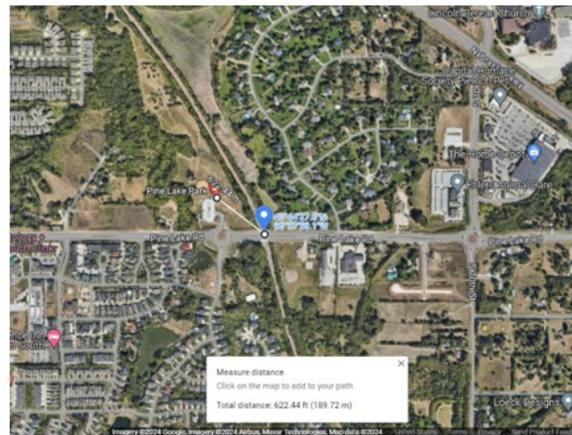
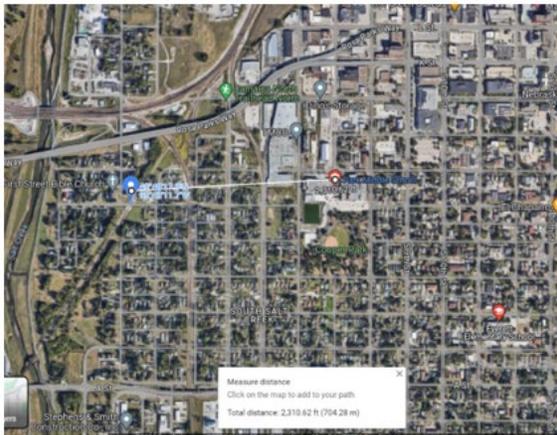


Figure 5. Proximity to local amenities within a 0.5-mile radius of HRGCs.

The practical constraints involved in data collection, including resource limitations and logistical challenges, guided the selection of 21 crossings. These crossings were chosen to provide a representative sample that captures the variability in environmental and situational factors affecting non-motorist volumes. Despite the limited number of locations, each of the 21 selected grade crossings underwent video monitoring for extended hours, with a minimum of 100 hours of video footage per location. This approach included recording various environmental and situational factors known to influence non-motorist volumes. These factors included day and night variations, diverse weather scenarios, seasonal changes, as well as daily and weekly activity patterns. Moreover, the study accounted for temporal fluctuations associated with holidays, school schedules, and commuting patterns. Additionally, the

analysis considered event-driven impacts such as festivals, sporting events, and community gatherings, along with disruptions caused by construction or maintenance activities affecting pedestrian and cyclist routes. Furthermore, economic factors, such as local shopping hours and market days, were also evaluated for their influence on commuting and recreational activities.

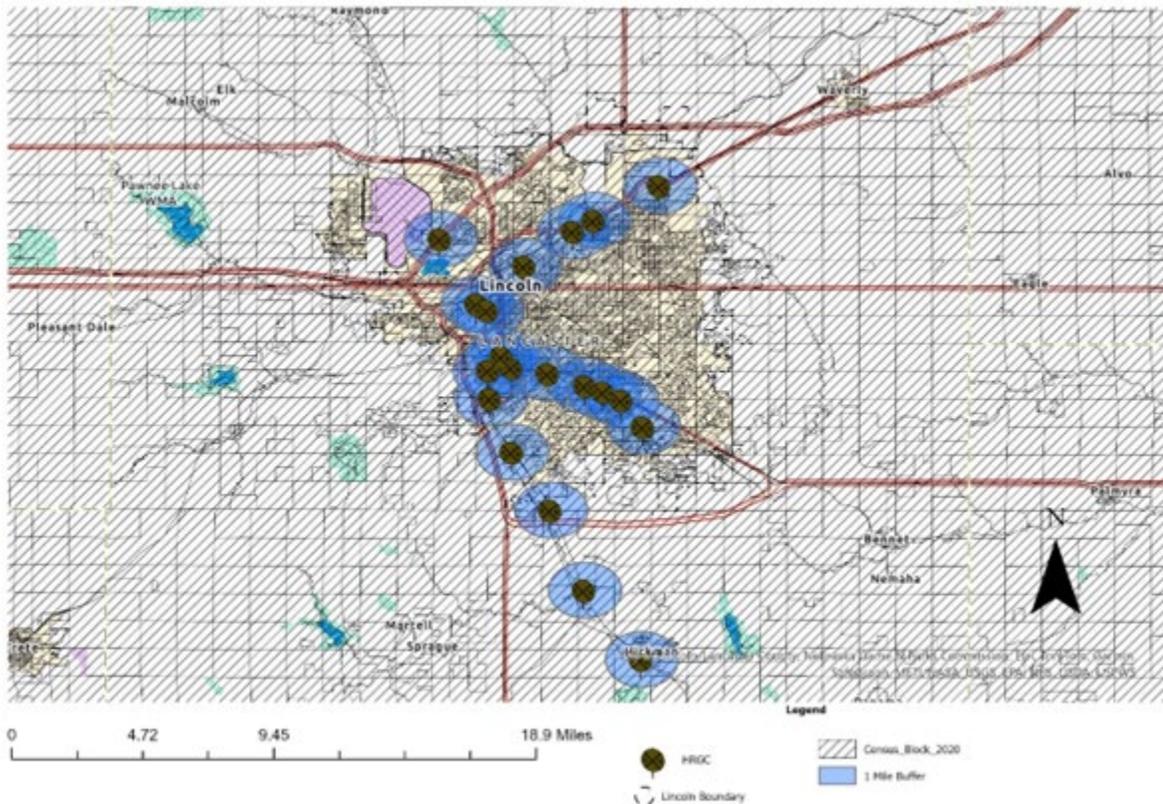
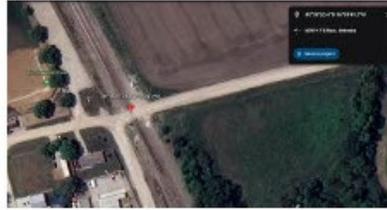


Figure 6. Population density buffer zones (1 mile) of HRGCs.

Figure 6 indicates buffer zone of 0.5-mile radius for population density estimation around selected HRGCs. First, the locations of the HRGCs were plotted using geographic coordinates in ArcGIS Pro. Next, the buffer tool was applied to generate a 0.5-mile radius around each HRGC, creating circular zones. The most recent available census data, in shapefile format, were then imported and overlaid onto the map. Using the spatial join or intersect tool, the population data within each buffer zone was extracted, allowing for the calculation of total population within these areas.



(a) HRGC's location identification via satellite imagery



(b) Road type/condition assessment



(c) Miovision scout installation



(d) Assessment of HRGC's warning devices and types



(e) Footage excerpt from scout video recording

Figure 7. Batch II video recording data collection overview: crossing ID 083519T.

Figure 7 provides detailed insights into the on-site data collection process for crossing ID 083519T. In part (a), the image shows how the location was chosen based on various factors such as ease of access and suitability for data collection. Part (b) highlights the evaluation of the road surface condition at the crossing. In part (c), an image captures the installation of the Miovision Scout, which is securely strapped to a pole facing the HRGCs to ensure optimal data capture. Part (d) showcases the assessment of the HRGC warning devices, including gates, lights, and other safety features. Lastly, part (e) presents a screenshot from the video footage recorded by the Scout, offering a visual excerpt from the on-site data collection.

Table 6. Locations of HRGCs from Batch I of Video Data Collection

Intersection/Proximity/Street Location Name (Miovision Scout)	Crossing ID	Details	Location Details
14th and NE Pkwy (Near NDOT Office)	083884M	Primary Operations: BNSF Railway Company (BNSF) County or City: LANCASTER In or Near: In City City: LINCOLN Street or Intersection: US HIGHWAY RR Mile post: 59.539 Nearest RR: HILL ST JCT Crossing Type: Public Crossing P: Highway	Latitude: 40.77153 Longitude: -96.701536
17th and Antelope Valley (Near UNL Passport Office on Y Street)	598814X	Primary Operations: Omaha, Lincoln & Beatrice Railway Company (OLB) County or City: LANCASTER In or Near: In City City: LINCOLN Street or Intersection: "Y" STREET RR Mile post: 0 Nearest RR: 553900 Crossing Type: Public Crossing P: Highway	Latitude: 40.824485 Longitude: -96.695099
27th and NE Pkwy	083886B	Crossing Number: 083886B Primary Operations: BNSF Railway Company (BNSF) County or City: LANCASTER In or Near: In City City: LINCOLN Street or Intersection: S 27TH STREET RR Mile post: 58.55 Nearest RR: LANCASTER Crossing Type: Public Crossing P: Highway	Latitude: 40.768124 Longitude: -96.682724
35th and Cornhusker Hwy	064129E	Crossing Number: 064129E Primary Operations: BNSF Railway Company (BNSF) County or City: LANCASTER In or Near: In City	Latitude: 40.84264 Longitude: -96.669739

		City: LINCOLN Street or Intersection: ADAMS ST RR Mile post: 56.61 Nearest RR: HAVELOCK Crossing Type: Public Crossing P: Highway	
40th and NE Pkwy	083890R	Crossing Number: 083890R Primary Operations: BNSF Railway Company (BNSF) County or City: LANCASTER In or Near: In City City: LINCOLN Street or Intersection: S 40TH STREET RR Mile post: 57.379 Nearest RR: HILL ST JCT Crossing Type: Public Crossing P: Highway	Latitude: 40.762074 Longitude: -96.663078
48th and NE Pkwy	083891X	Crossing Number: 083891X Primary Operations: BNSF Railway Company (BNSF) County or City: LANCASTER In or Near: In City City: LINCOLN Street or Intersection: S 48TH STREET RR Mile post: 56.81 Nearest RR: HILL ST JCT Crossing Type: Public Crossing P: Highway	Latitude: 40.758927 Longitude: -96.65361
56th and Old Chenny	083895A	Crossing Number: 083895A Primary Operations: Union Pacific Railroad Company (UP) County or City: LANCASTER In or Near: In City City: LINCOLN Street or Intersection: SOUTH 56TH STREET RR Mile post: 56.23 Nearest RR: LANCASTER Crossing Type: Public Crossing P: Highway	Latitude: 40.754636 Longitude: -96.644355
NW 12th and Cornhusker	815572E	Crossing Number: 815572E Primary Operations: Union Pacific Railroad Company (UP) County or City: LANCASTER In or Near: In City City: LINCOLN Street or Intersection: NW12TH STREET RR Mile post: 54.97 Nearest RR: Crossing Type: Public Crossing P: Highway	Latitude: 40.838539 Longitude: -96.738899

Table 6-8 provides details on the locations and proximity of the HRGCs selected for the study, including their crossing IDs, primary operations, city and county information, street addresses, milepost details, crossing access type, and latitude and longitude coordinates for each crossing.

Table 7. Locations of HRCs from Batch II of Video Data Collection

Intersection/Proximity/Street Location Name (Miovision Scout)	Crossing ID	Details	Location Details
Grant Street and Roy Street	083519T	Primary Operations: BNSF Railway Company (BNSF) County or City: LANCASTER In or Near: Near City City: GRANT Street or Intersection: ROY ST RR Mile post: 196.589 Nearest RR: HICKMAN Crossing Type: Public Crossing	Latitude: 40.77153 Longitude: -96.701536
Hill Street and Old Cheney Road	074406N	Primary Operations: BNSF Railway Company (BNSF) County or City: LANCASTER In or Near: Near City City: LINCOLN Street or Intersection: OLD CHENEY RD RR Mile post: 203.99 Nearest RR: HILL ST JCT Crossing Type: Public Crossing P: Highway Crossing	Latitude: 40.7556 Longitude: -96.71278
Pinelake Road and S 66 th Street	083897N	Primary Operations: Union Pacific Railroad Company (UP) County or City: LANCASTER In or Near: Near City City: LINCOLN Street or Intersection: PINE LAKE ROAD RR Mile post: 55.12 Nearest RR: COLLEGEVIEW Crossing Type: Public Crossing P: Highway Crossing	Latitude: 40.74082 Longitude: -96.633079
West Pioneer Blvd and Jamaica North Trail	924642S	Primary Operations: BNSF Railway Company (BNSF) County or City: LANCASTER In or Near: In City City: LINCOLN Street or Intersection: W PIONEER BLVD RR Mile post: 205.019 Nearest RR: LINCOLN TERMINAL Crossing Type: Public Crossing P: Highway	Latitude: 40.77017 Longitude: -96.7136
Custer Street and N 70th Street	074929T	Primary Operations: BNSF Railway Company (BNSF) County or City: LANCASTER In or Near: In City City: LINCOLN Street or Intersection: N 70TH ST RR Mile post: 53.72 Nearest RR: HAVELOCK Crossing Type: Public Crossing P: Highway	Latitude: 40.866286 Longitude: -96.624685

44th ST and Cornhusker Hwy	074860A	Primary Operations: BNSF Railway Company (BNSF) County or City: LANCASTER In or Near: In City City: LINCOLN Street or Intersection: 44TH ST RR Mile post: 55.919 Nearest RR: HAVELOCK Crossing Type: Public Crossing P: Highway	Latitude: 40.848341 Longitude: -96.658883
----------------------------	---------	---	--

Table 8. Locations of HRCs from Batch II of Video Data Collection

Intersection/Proximity/Street Location Name (Miovision Scout)	Crossing ID	Details	Location Details
Saltillo Road and South 27 th Street	083516X	Primary Operations: BNSF Railway Company (BNSF) County or City: LANCASTER In or Near: Near City City: LINCOLN Street or Intersection: SALTILLO RD RR Mile post: 199.6 Nearest RR: HILL ST JCT Crossing Type: Public Crossing P: Highway	Latitude: 40.697374 Longitude: -96.681406
Yankee Hill Road and South 14 th Street	083512V	Primary Operations: BNSF Railway Company (BNSF) County or City: LANCASTER In or Near: Near City City: LINCOLN Street or Intersection: S 14TH STREET RR Mile post: 201.929 Nearest RR: HILL ST JCT Crossing Type: Public Crossing P: Highway	Latitude: 40.727573 Longitude: -96.701518
W 2 nd Street	083524P	Primary Operations: BNSF Railway Company (BNSF) County or City: LANCASTER In or Near: In City City: HICKMAN Street or Intersection: 2ND STREET RR Mile post: 193.63 Nearest RR: HICKMAN Crossing Type: Public Crossing P: Highway	Latitude: 40.619626 Longitude: -96.634196
Calvert and 10 th	064365J	Primary Operations: BNSF Railway Company (BNSF) County or City: LANCASTER In or Near: In City City: LINCOLN Street or Intersection: CALVERT & 10TH RR Mile post: 60.612	Latitude: 40.777375 Longitude: -96.706812

		Nearest RR: HILL ST JCT Crossing Type: Public Crossing P: Highway	
F street and South 1 st Street	064344R	Primary Operations: BNSF Railway Company (BNSF) County or City: LANCASTER In or Near: In City City: LINCOLN Street or Intersection: 1ST F ST RR Mile post: 60.51 Nearest RR: LINCOLN TERMINAL Crossing Type: Public Crossing P: Highway	Latitude: 40.804798 Longitude: -96.719906
C street and South 5 th Street	064355D	Primary Operations: BNSF Railway Company (BNSF) County or City: LANCASTER In or Near: In City City: LINCOLN Street or Intersection: 5TH & C ST RR Mile post: 60.661 Nearest RR: LINCOLN TERMINAL Crossing Type: Public Crossing P: Highway	Latitude: 40.80143 Longitude: -96.7144

4. Artificial Intelligence for Non-motorist Detection

In the early stages of non-motorist detection, traditional image processing techniques formed the backbone of the methodologies used. These methods, being foundational studies, primarily relied on manually extracted features from images and utilized classical machine learning algorithms for the detection of pedestrians and cyclists. Edge detection was one of the pioneering techniques in image processing, used to identify the boundaries within images. Early works, such as those by Marr and Hildreth (1980), introduced the concept of edge detection by detecting zero-crossings in the second derivative of the image intensity, laying the groundwork for subsequent methods.

For Motion analysis, another critical approach involved identifying moving objects based on changes in pixel intensity over time. Pioneering studies by researchers like Aggarwal and Cai (1997) provided comprehensive surveys on human motion analysis, highlighting the importance of understanding motion patterns for detecting and tracking non-motorists. These methods typically used frame differencing and optical flow to track the motion of objects, but they often struggled with occlusions and varying lighting conditions. Traditional non-motorist detection

techniques often relied on manually extracted features, such as shape, texture, and color (Wang and Adelson,1993). Wren et al. (1997) proposed the Pfunder system, which used statistical models of color and shape to detect and track people in real-time. Although effective to some degree, these methods were constrained by their reliance on predefined features and often struggled to accommodate the variability in pedestrian appearances and the complexity of backgrounds.

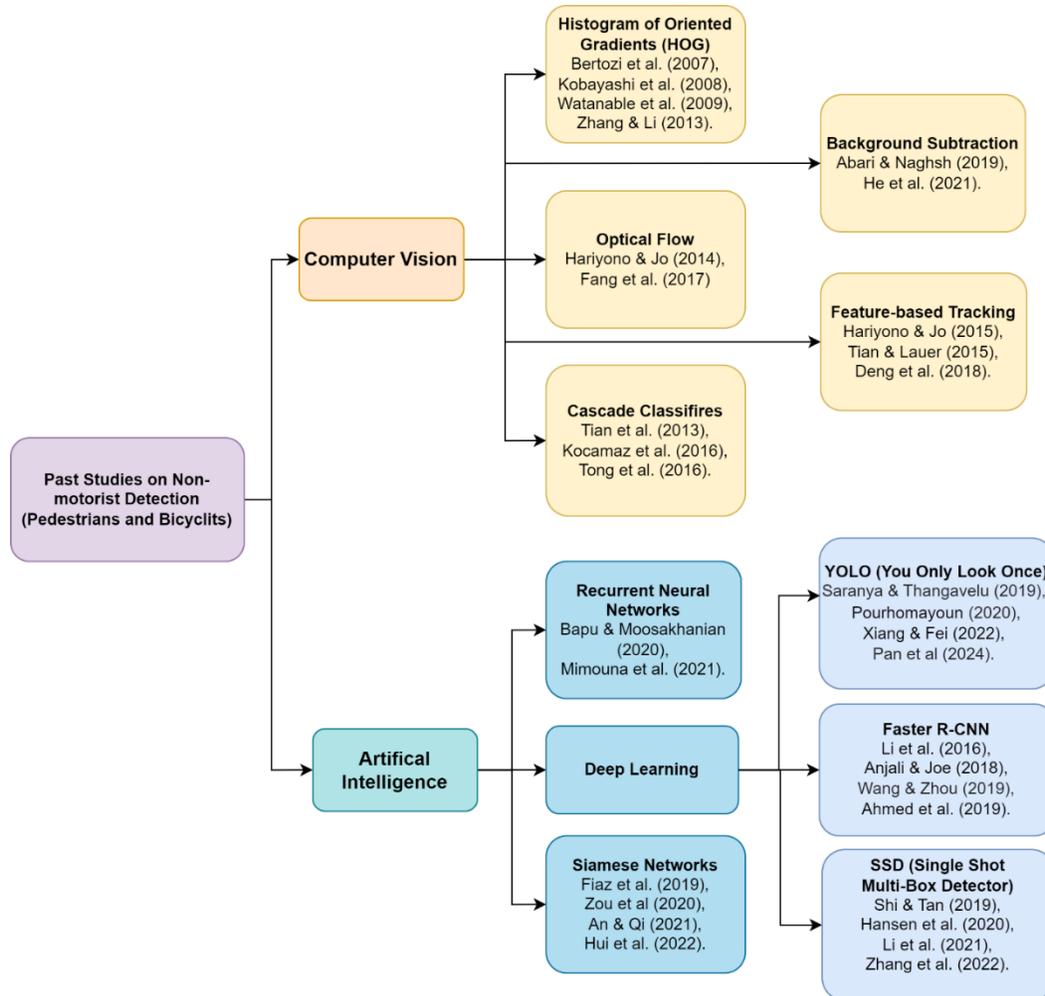


Figure 8. Key Studies for Non-motorist Object Detection.

Classical machine learning algorithms, including Support Vector Machines (SVMs) and decision trees, were employed to classify detected features as non-motorists or background elements. Early applications of SVMs in pedestrian detection by Papageorgiou and Poggio

(2000) demonstrated the potential of machine learning for this purpose. These approaches, however, were constrained by their reliance on handcrafted features and the computational complexity involved in training the models. Despite their innovative nature, early non-motorist detection methods faced significant limitations. Complex backgrounds and varying lighting conditions posed substantial challenges, as these methods lacked robustness and adaptability. As noted by Gavrila (1999) in a comprehensive review of pedestrian detection methodologies, these traditional approaches struggled to achieve high detection accuracy and real-time performance in dynamic environments.

In recent years, several methods have been developed and refined to enhance the precision and efficacy of pedestrian detection systems. Traditional feature-based algorithms, like Histogram of Oriented Gradients (HOG), are one popular technique. Using gradient orientation histograms extracted from photos, pedestrian shapes and silhouettes are recognized using HOG. This technique can be difficult to use in situations with complex backdrops and occlusions, but it works well in structured contexts. The Region-based Convolutional Neural Network (R-CNN) and its variations, including Fast R-CNN and Faster R-CNN, represent another cutting-edge method (Kim et al. (2018), (Wang & Zhou (2019))). In order to detect pedestrians, these two-stage object detection algorithms first create region recommendations, which are subsequently classified. Specifically, faster R-CNN improves speed and accuracy by directly integrating the region proposal network (RPN) into the detection network, leading to a notable improvement in pedestrian detection jobs performance. Figure 8 shows some recent studies on object detection.

The Single Shot MultiBox Detector (SSD) is an additional notable method (Chang et al. (2019), Kumar et al. (2020)). SSD is a single-stage detector that, in contrast to R-CNN variations, can predict bounding boxes and class scores during a single network forward pass. Since this method drastically cuts down on detection time, it can be used in real-time applications. In order to recognize things at many scales and reliably identify pedestrians of different sizes, SSD employs a sequence of convolutional layers.

Since its conception, the YOLO (You Only Look Once) family of models has experienced substantial advances, with each iteration improving its capabilities and performance for object detection tasks (Vijayakumar & Vairavasundaram (2024)). By redefining object recognition as a single regression issue and enabling the network to predict bounding boxes and

class probabilities straight from entire photos in a single pass, the original YOLO model, which debuted in 2016, revolutionized object detection and made real-time detection possible.

Subsequent versions of YOLO, such as YOLOv2 (YOLO9000) and YOLOv3, introduced further enhancements. YOLOv2 brought improvements like batch normalization, high-resolution classifiers, and the use of anchor boxes, significantly boosting performance and detection accuracy. YOLOv2, sometimes called YOLO9000, brought in many improvements, such as batch normalization, high-resolution classifiers, and the application of anchor boxes to boost performance and detection accuracy. Over 9000 object categories could be detected by YOLOv2, a considerable improvement over the initial model. Darknet-53, a more powerful and in-depth feature extractor, was introduced by YOLOv3 to further improve the architecture. Additionally, multi-scale predictions were used, which greatly enhanced the ability to detect small objects. Furthermore, logistic regression was added to YOLOv3 for class prediction, which improved its capacity to manage overlapping objects and class imbalance problems.

Building on these developments, YOLOv4 and YOLOv5 continued to advance the state-of-the-art in object detection. By using cutting-edge methods like the Path Aggregation Network (PANet), Mish activation function, and Cross-Stage Partial connections (CSPDarknet53), YOLOv4 significantly increased speed and accuracy. To further improve performance, it also used cutting-edge data augmentation techniques like Self-Adversarial Training (SAT) and Mosaic.

This pattern of improvement is maintained in YOLOv5, the subsequent iteration. The streamlined and lightweight architecture of YOLOv5 is well renowned for improving accuracy and speed (Zhao et al. 2023; Chen et al. 2023; Li et al. 2023). Advanced features like integrated hyperparameter evolution, auto-learning bounding box anchors, and the use of new modules like the Focus layer for improved feature extraction are also included. Thanks to these improvements, YOLOv5 is now extremely effective for real-time applications, especially in settings where quick and precise object identification is required.

Following the developments of YOLOv5, the YOLO family continues to evolve with the introduction of YOLOv6 and YOLOv7, both of which brought more performance and efficiency improvements (Vijayakumar & Vairavasundaram (2024)). YOLOv6 was created, particularly for

industrial applications, to maximize the trade-off between detection accuracy and inference speed (Li et al. (2022)). It incorporates cutting-edge parts like EfficientRep, a RepVGG block variation that streamlines the architecture without sacrificing performance. Global Attention Mechanism (GAM) approaches are also used by YOLOv6 to enhance feature extraction and reaction to changing object sizes and occlusions. It improves the overall efficiency and accuracy of the model by utilizing the CSPNet backbone for improved gradient flow and feature reuse. YOLOv6 is especially well-suited for real-time object detection jobs in intricate situations due to these enhancements.

With emphasis on higher detection accuracy and speed, YOLOv7 is a significant advancement in YOLO models' development. A more effective feature pyramid network (FPN) and the addition of a Path Aggregation Network (PAN) for improved multi-scale feature fusion are two of the novel architectural enhancements brought forth by YOLOv7. Also, to handle objects of various scales and enhance localization accuracy, YOLOv7 integrates sophisticated techniques such as Spatial Pyramid Pooling (SPP). In order to further improve bounding box predictions, it also makes use of a novel dynamic anchor box allocation technique. Thanks to its improvements, YOLOv7 is now among the fastest and most accurate models out there, making it ideal for a variety of uses, including surveillance and autonomous driving.

The most recent model in Ultralytics' "You Only Look Once" (YOLO) series of object identification models is called YOLOv8 (Vijayakumar & Vairavasundaram (2024)). Figure 9 represents the overall framework of YOLOv8. Its goal is to preserve efficiency and convenience of use while offering high-performance object identification and tracking capabilities. Building on the innovations of its predecessors, YOLOv8 integrates cutting-edge technologies to improve speed and accuracy in real-time applications.

The capacity of YOLOv8 to strike a balance between performance and resource efficiency is one of the key factors making it a strong option for object recognition and tracking. The model is ideal for a range of applications, from autonomous driving to security monitoring, because it is tailored for both high accuracy and low latency. YOLOv8's enhanced detection skills over earlier iterations and other models in the field are a result of its utilization of cutting-edge approaches including mosaic data augmentation and an anchor-free detection head.

and 1,176 hours of recorded video from 11 locations across six states, collecting data on traffic, rail signal activations, train events, and trespass incidents. The AI detected over 29,000 trespass events, with detailed information on each event, such as date, time, type, weather, path, and video clips. Manual validation ensures data accuracy. The study also included two in-depth, year-long case studies at a New Jersey grade crossing and a North Carolina right-of-way, providing temporal and spatial analyses and discussing AI-informed mitigation strategies (Zaman et al. 2024).

Comprehensive Workflow for Detecting Non-Motorists in Miovision Videos Using YOLOv8

The YOLOv8 model requires a multi-step technical method to detect non-motorists in videos recorded with Miovision, such as bicyclists and pedestrians. Installing the prerequisites tools and libraries required to execute the YOLOv8 model and handle video data is the first step in setting up the required software environment. This basic configuration guarantees that the system is prepared to handle video inputs and effectively identify objects.

Preparing the YOLOv8 model is the next step after the environment is configured. Either a pre-trained model that has previously been tuned for object identification can be loaded, or a new model can be trained using bespoke datasets that contain pedestrians and bicyclists particularly. The decision is based on the training data's accessibility and the particular needs of the detection task. We used a pre-trained model of YOLOv8 for non-motorists' detection and counting.

As soon as the model is prepared, the video input is handled. Miovision videos are processed frame by frame so that the YOLOv8 model may examine every frame separately. Effective capture and management of the video streams in this step necessitate the use of video processing libraries. The YOLOv8 model identifies and categorizes the objects included within. Making the inference based on the YOLOv8 model is an important step in the procedure and used to locate and identify non-motorists in every video frame. The types and positions of the objects spotted within each frame are among the detection findings that the model outputs. The detected pedestrians and bicyclists are then highlighted in the video frames using these results.

The findings of the detection are finally displayed. By doing this, the process of detecting bicyclists and pedestrians is simplified by projecting the detection outputs into the video frames. These frames with annotations can be saved for further examination or shown in real time. Understanding can be greatly improved by including figures and images at important stages of this process. Screenshots of the video frames containing items that the model has detected demonstrate the model's performance, for example, displaying the command line interface during installation aids in visualizing the setup.

The entire process of using YOLOv8 for object detection in Miovision videos is covered in this thorough description, which also emphasizes the significance of each stage and its technological complexities in order to achieve accurate and efficient non-motorist detection. A step-by-step process is given below.

A. Setting Up the Environment

The first step is to install all the dependencies

- Make sure Python is installed.
- Install any essential packages, Ultralytics YOLO, OpenCV, and other libraries.

```
pip install ultralytics opencv-python
```

B. Getting the Model Ready

Second step is to download the pre-trained model YOLOv8 model file “yolov8n.pt” and save it in your project directory where your main.py file is saved. Then we load the model using the following command.

```
from ultralytics import YOLO

# Load pre-trained YOLOv8 model

model = YOLO('yolov8n.pt')
```

C. Processing Video Input

To load and open a video as input, OpenCV is used which captures frames from the Miovision videos.

```
import cv2

video_path = 'path_to_video.mp4'

cap = cv2.VideoCapture(video_path)

while cap.isOpened():

    ret, frame = cap.read()

    if not ret:

        break

    # Process each frame with YOLOv8

    results = model(frame)
```

D. Executing Inference

Run each frame through the YOLOv8 model to identify bicycles and pedestrians.

```
results = model(frame)
```

E. Displaying Outcomes

In this step we render the detection results on each frame, with the option to store the video output.

```
for result in results:

    # result.render() modifies the frame in-place

    frame = result.render()

    # Show the frame with detection boxes

    cv2.imshow('YOLOv8 Detection', frame)
```

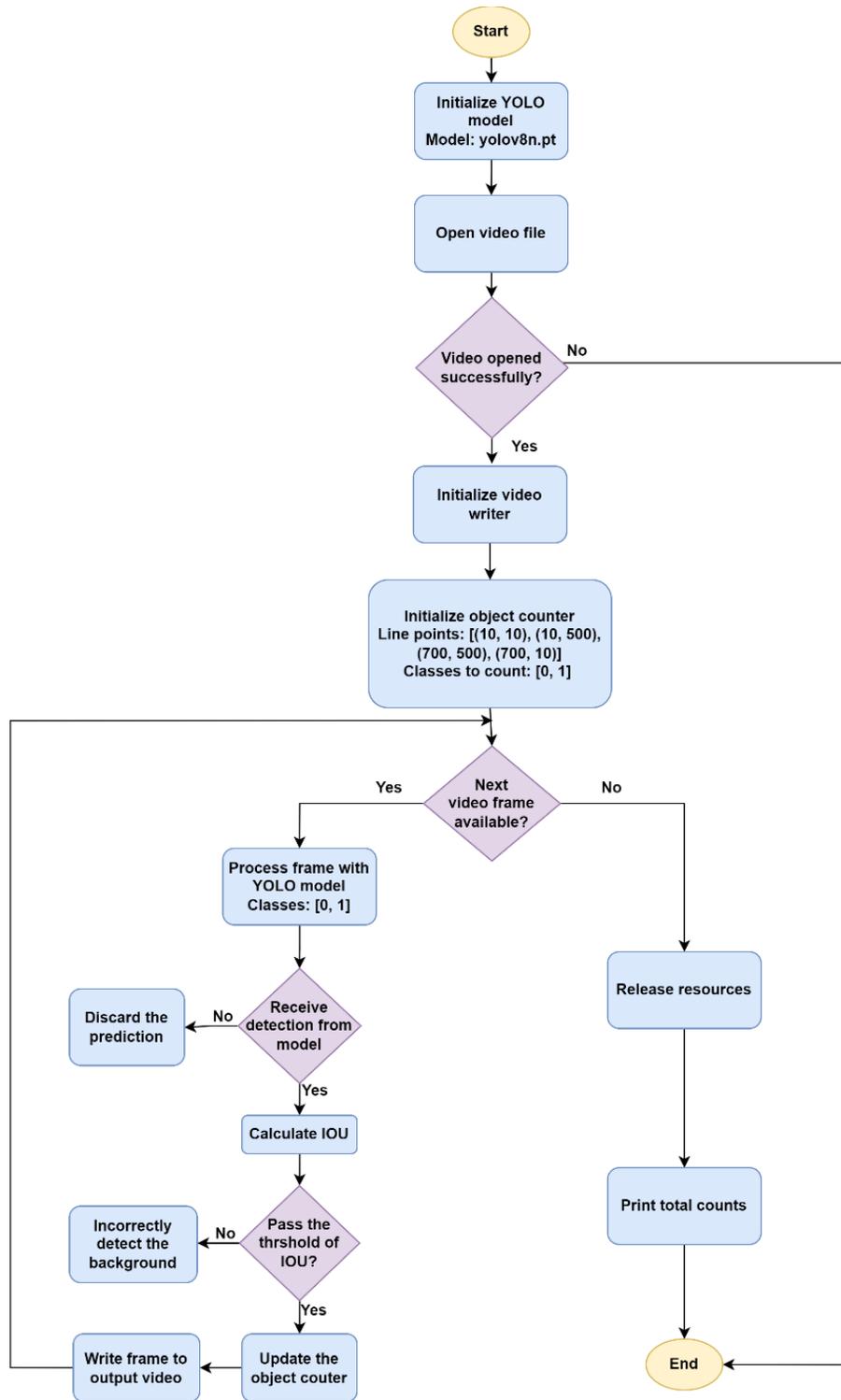


Figure 10. Flowchart of object detection process.

Figure 10 represents the flowchart for the object detection process.

5. Case Studies

The videos recorded through Miovision Scout were analyzed using YOLOv8 for object detection. While YOLOv8 generally performed well in identifying non-motorists, some instances revealed limitations in its accuracy. Factors such as lower video quality, adverse weather conditions (e.g., rain, higher wind speeds), and suboptimal camera angles occasionally led to detection errors. YOLOv8's performance can be affected by several factors. For example, lower video resolution can reduce the model's ability to discern finer details, while poor weather conditions like rain can obscure visibility, complicating object identification. Additionally, camera angles and motion blur caused by wind or camera shake can contribute to inaccuracies in detection. To enhance the prediction model's accuracy, adjustments were made to account for these detection errors. In cases where YOLOv8's object detection was not 100% accurate due to uncontrollable factors, such as rain or blurred video, manual corrections were applied. This involved using manual counts to adjust the total non-motorist counts and removing instances where poles, or channelization devices were incorrectly labeled as people due to lighting and shadows, as illustrated in Figure 11 and 12. These corrections ensured that the final count for analysis was more accurate and reflective of actual non-motorist volumes.



Figure 11. Object detection case I.



Figure 12. Object detection case II.

It is also noted as the person moves further away from the camera and then come back, the tracking of the person(s) is lost for a short time but significant enough for the algorithm that it assigns the person(s) a new ID and counts it as a new person. This scenario is captured in the following two screenshots in Figure 13 and Figure 14, case III.



Figure 13. Object Detection Case III.



Figure 14. Object Detection Case III.

Two people were seen crossing the intersection at 40th Street and NE Parkway in the scenario shown in Figure 15, case IV (Crossing ID 083890R). This image, which was obtained from a video that was captured on May 14, 2024, at around 4:43 PM, shows how well the identification algorithm performed in real-time in detecting pedestrians who were not vehicles at the crossing.

At first, the system misclassified one of the people as a cyclist. This happened because the person was not riding the bicycle; instead, they were walking along it. The person was automatically classified as a cyclist by the algorithm when it saw that they were riding a bicycle. Nevertheless, a change was made to the algorithm's recognition, correctly recognizing the person as a pedestrian rather than a cyclist, following additional examination of the movement patterns.

This situation emphasizes how crucial it is to continuously improve detection algorithms in order to distinguish between minute differences in human behavior, such as strolling alongside a bicycle and riding one. The system's accuracy depends on its capacity to rectify these misclassifications, guaranteeing that non-motorists are accurately recognized and categorized in real-time. This degree of specificity is essential for boosting the overall efficacy of monitoring systems intended to track and safeguard non-motorist activities, as well as for strengthening safety protocols at rail crossings and intersections.



Figure 16. Object Detection Case V.

Another scenario is shown in the following screenshot in Figure 17 where the bicyclist is detected and labeled as a person due to the angle of the bicycle. The algorithm cannot detect the bicycle from the front or rear view as the side view is not clear for it to detect. The algorithm cannot detect the bicycle due to low resolution of the video.



Figure 17. Object Detection Case VI.

The scenario in Figure 18 illustrates a case of low visibility brought on by the camera's focus settings. This problem occurs when there is a hazy or unclear image due to either misalignment or obstruction of the camera's focus. This leads to the loss of crucial visual information, which adversely

affects the algorithm's capacity to recognize crucial components, including non-motorists or items at the crossing.

This particular incident happened at Crossing ID 064355D, where the lower definition of the video made it difficult for the algorithm to identify individuals or objects in the area. These kinds of low visibility situations can make the algorithm overlook important detections that it would have otherwise picked up with standard camera settings. This constraint emphasizes how critical it is to set up and focus the cameras optimally in order to ensure the accuracy of the detection system, especially in situations where having excellent visuals is essential for spotting possible risks or keeping an eye on non-motorist activities at crossings.



Figure 18. Object Detection Case VII.

Figure 16 (a) and (b) demonstrate the output of the detection algorithm in operation, emphasizing in particular the recognition of non-motorists like cyclists and pedestrians. These pictures demonstrate how the algorithm interprets visual information from crossings and successfully determines whether people are present in the scene. The algorithm's capacity to follow movement and improve safety monitoring at crossings is demonstrated when it uses computer vision techniques to identify and classify non-motorists in the clip.

A crucial part of the detection system are the screenshots, which provide a visual representation of the algorithm's performance under real-time conditions. Bounding boxes or labels are usually used to indicate the precise locations of the non-motorists within the frame for each detection. In order to evaluate the detection system's accuracy and dependability and make sure that non-motorist behavior is reliably observed and recorded, especially at high-risk or high-traffic crossings—visual feedback is essential.

These pictures give a clear picture of the algorithm's functionality and show how, under different circumstances, it reliably detects and tracks non-motorists, improving safety.

```
0: 448x640 1 person, 16.3ms
Speed: 3.9ms preprocess, 16.3ms inference, 3.3ms postprocess per image at shape (1, 3, 448, 640)

0: 448x640 1 person, 22.0ms
Speed: 6.0ms preprocess, 22.0ms inference, 3.0ms postprocess per image at shape (1, 3, 448, 640)

0: 448x640 1 person, 17.1ms
Speed: 3.0ms preprocess, 17.1ms inference, 3.0ms postprocess per image at shape (1, 3, 448, 640)

0: 448x640 2 persons, 18.0ms
Speed: 4.0ms preprocess, 18.0ms inference, 4.0ms postprocess per image at shape (1, 3, 448, 640)

0: 448x640 1 person, 18.0ms
Speed: 4.0ms preprocess, 18.0ms inference, 3.5ms postprocess per image at shape (1, 3, 448, 640)

0: 448x640 1 person, 17.1ms
Speed: 6.0ms preprocess, 17.1ms inference, 6.0ms postprocess per image at shape (1, 3, 448, 640)

0: 448x640 (no detections), 17.2ms
Speed: 4.0ms preprocess, 17.2ms inference, 3.1ms postprocess per image at shape (1, 3, 448, 640)
```

(a)

```
0: 448x640 (no detections), 14.1ms
Speed: 4.9ms preprocess, 14.1ms inference, 5.0ms postprocess per image at shape (1, 3, 448, 640)

0: 448x640 (no detections), 14.3ms
Speed: 3.0ms preprocess, 14.3ms inference, 2.0ms postprocess per image at shape (1, 3, 448, 640)

0: 448x640 (no detections), 13.6ms
Speed: 4.0ms preprocess, 13.6ms inference, 2.3ms postprocess per image at shape (1, 3, 448, 640)

0: 448x640 (no detections), 14.1ms
Speed: 4.0ms preprocess, 14.1ms inference, 3.8ms postprocess per image at shape (1, 3, 448, 640)
Video frame is empty or video processing has been successfully completed.
Current counts: 2
```

(b)

Figure 19 (a) Snapshot of the output showing detection of a person while video is being processed by YOLOv8 model. (b). Snapshot of the output at the end of the video processing by YOLOv8 model.

6. Exposure Prediction Modelling

6.1. Statistical Modelling

Non-motorist traffic volumes are count data, making the Poisson family of models suitable for representing the relationship between total daily non-motorist traffic count at HRGCs per 24-hour period and the explanatory variables. The variance of recorded daily non-motorist traffic counts across all HRGCs was significantly larger than the respective means, necessitating the use of a Negative Binomial (NB) Model in place of a Poisson model (Greene, 2008; Mukherjee and Mitra, 2020; Farooq, 2023). The probability mass function of the NB model is typically defined as follows:

$$P(Y = y) = \frac{\Gamma(y+r)}{\Gamma(y+1)\Gamma(r)} \left(\frac{\mu}{\mu+\theta}\right)^y \left(\frac{\theta}{\mu+\theta}\right)^r \quad \text{Eq (i)}$$

Where, Y is the count of non-motorists, μ is the mean count, θ is the dispersion parameter, and $r = \frac{\mu^2}{\theta}$ is a parameter related to the variance. In addition, the mean of the distribution is given by μ , and the variance is given by $\mu + \frac{\mu^2}{\theta}$, which accounts for over-dispersion. In practice, the model is often expressed in a log-linear form (Greene, 2008), where the log of the expected count is modeled as a linear function of predictors:

$$\log(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} \quad \text{Eq (ii)}$$

Where μ_i is the expected count for the i -th observation, $X_{i1}, X_{i2}, \dots, X_{ik}$ are the predictor variables, and $\beta_0, \beta_1, \beta_k$ are the respective coefficients. The NB model's flexibility in handling varying dispersion levels makes it suitable for diverse applications by modeling the mean count as a function of predictors while allowing its variance to differ. This approach enhances predictive accuracy by capturing the underlying data distribution. Random parameter models, which extend traditional NB models, account for unobserved heterogeneity by allowing certain parameters to vary across observations. However, difficulties in obtaining significant random parameters during model estimation in this study suggested that the variability might be insufficient to justify their inclusion, indicating that the fixed parameter NB model may adequately capture the variability in non-motorist volume. Average Marginal Effects (AMEs) are valuable in count models as they provide an intuitive measure of how predictor variables affect the expected count of the response variable. Unlike raw coefficients, AMEs show the average change in the predicted count for a one-unit change in each predictor, making them easier to interpret. This is crucial for applications like traffic safety

and urban planning, where understanding the impact of factors such as traffic volume and visibility can enhance decision-making (Farooq and Khattak, 2023). The 'margins' package in R calculates AMEs for various regression models, including Negative Binomial models, by applying the 'margins' function to a fitted model from 'glm.nb'. (Leeper, 2017; Walker et al. 2024; Zhao et al. 2024). The output includes estimates, standard errors, z-values, and p-values, for each predictor, facilitating interpretation.

In addition to the Negative Binomial (NB) model, we considered the use of Random Effects Negative Binomial (RENB) models to address the panel nature of our data. Panel data, characterized by multiple observations over time for the same entities, often contain repeated measures where certain attributes remain constant. In our dataset, many location-based attributes, such as the presence of sidewalks, proximity to parks, educational institutions, transit stops, parking facilities, residential areas, and industrial areas within 0.5 miles of crossings, were unchanged across observations for the same crossing locations.

Given these repeated measures, a random effects model is particularly useful as it accounts for the within-group correlation by allowing for random variations across the entities (in this case, the crossing locations). This approach assumes that unobserved heterogeneity is uncorrelated with the observed variables and captures the influence of these stable, location-based characteristics on non-motorist volumes. By incorporating random effects, we can more accurately model the variability across different crossings and account for the unobserved factors that might influence non-motorist counts. Whether using a Poisson model or a negative binomial (NB) model, both assume that non-motorist counts at rail crossings for any given year are independent. However, due to the presence of location-specific effects and possible serial correlation in the data, it is more appropriate to handle it as a time-series cross-sectional panel with M locations over T periods. Failing to account for this structure may lead to underestimation of the standard errors, as each observation provides less information than initially assumed. This, in turn, inflates the t-statistics and leads to flawed statistical inferences.

If spatial effects exist in the data, the RENB model can be adopted by introducing a random location-specific effects term into the relationship between the expected numbers of accidents ($\tilde{\mu}_{it}$) and the covariates, \mathbf{X}_{it} , of an observation unit i in a given time period t , i.e.

$$\tilde{\mu}_{it} = \mu_{it} \delta_i \tag{Eq (iii)}$$

where δ_i is a random location-specific effect. To ensure a positive value, the term $\tilde{\mu}_{it}$ can be rewritten as

$$\tilde{\mu}_{it} = \mu_{it} \delta_i = \exp(\mathbf{X}_{it} \boldsymbol{\beta} + \mu_i) \tag{Eq (iv)}$$

where β is the coefficient vector to be estimated, μ_i the random effects across location and $\exp(\mu_i)$ is gamma distributed with mean 1 and variance k , where k is also the overdispersion parameter in the NB model. The number of accidents at an intersection i for a given year t , i.e. n_{it} is independently and identically NB distributed with parameters $\delta_i \mu_{it}$ and ϕ_i , where $\mu_{it} = \exp(\mathbf{X}_{it}\beta)$. Hence n_{it} has mean $\delta_i \mu_{it} / \phi_i$ and the variance $(\delta_i \mu_{it} / \phi_i) / z$, where $z = 1 / (1 + \delta_i / \phi_i)$. Additionally, in order to account for the variation of location over time, z is assumed to be a beta-distributed random variable with distributional parameters (a, b) . Using the results from the derivation of Hausman et al. (1984), the probability density function of the RENB model for the i th intersection will be

$$P(n_{i1}, \dots, n_{iT} | \mathbf{X}_{i1}, \dots, \mathbf{X}_{iT}) = \frac{\sqrt{a+b} \sqrt{a+\sum_T \mu_{it}} \sqrt{b+\sum_T n_{it}}}{\sqrt{a} \sqrt{b} \sqrt{a+b+\sum_T \mu_{it} + \sum_T n_{it}}} \prod_T \frac{\sqrt{\mu_{it} + n_{it}}}{\sqrt{\mu_{it}} \sqrt{n_{it} + 1}} \quad \text{Eq (v)}$$

Table 9. Summary statistics of candidate variables for NB and RENB modelling.

Variable Symbol	Variable Description (Coding)	Mean	S.D.	Min.	Max.
Csng_Lgtng	Crossing lightning (1 if crossing is illuminated at night, 0 otherwise)	0.8831	0.3233	0	1
Weekday_S	Weekday indicator (1 if the recorded day was a weekday, 0 otherwise)	0.7012	0.4606	0	1
Avg_Temp	Average daily temperature (°F)	67.019	11.6465	24.63	77.94
Prcptn	Average daily precipitation (inches)	0.1237	0.3749	0	1.5
Visibility	Average visibility (miles)	9.8312	8.3354	4	10
Max_W_Spd	Maximum daily wind speed (mph)	18.467	6.8394	7	33
Clear_Cndtn1	clear conditions indicator (1 if the weather is clear, 0 otherwise)	0.6363	0.4842	0	1
Rain_II	Rain indicator (1 if rain is recorded, 0 otherwise)	0.1428	0.3522	0	1
Cloudy_I	cloudy indicator (1 if cloudy weather is recorded, 0 otherwise)	0.1168	0.3233	0	1
Snow_III	Snow indicator (if snowy weather is recorded, 0 otherwise)	0.0129	0.11322	0	1
Rd_Dry	Dry road condition indicator (1 if road surface is dry, 0 otherwise)	0.7272	0.4453	0	1
Rd_Wet	Dry road condition indicator (1 if road surface is dry, 0 otherwise)	0.2337	0.42323	0	1
Rd_Wet_II	Wet road with dry conditions before/after indicator (1 if the road surface is wet although it is dry during the 24 hours period, 0 otherwise)	0.7273	0.4454	0	1
TtlPed_Cnt_24	Total pedestrian count in 24-hr period	19.3377	25.4674	0	135
TtlByc_Cnt_24	Total bicyclist count in 24-hr period	15.5455	22.4173	0	144
Total_NM	Total non-motorist count in 24-hr period	34.8831	41.5983	0	242
Ttl_Train_Trf	Total train traffic	4.98701	8.69317	0	43
Day_Trn	Total daily trains	3.48640	5.46157	0	26
Ngt_Trn	Total night-time trains	1.58667	3.9702	0	21
Pop_Dnsty	Population density (within 1-mile radius of the HRGCs location)	525.052	282.32	112	1174

Pop_L_150	Population density is less than 150 indicator (1 if yes, 0 otherwise)	0.12988	0.33617	0	1
Som_Actv	Some activity around HRGCs indicator (1 if there is some activity observed withing 0.5-mile of HRGCs such as a concert, event, celebration)	0.38961	0.48766	0	1
Sidewalk	Sidewalk indicator (1 if there is a sidewalk at HRGCs, 0 otherwise)	0.58442	0.4928	0	1
Ped_XingSgnl	Pedestrian signal indicator (1 if there is a pedestrian signal, 0 otherwise)	0.57143	0.49487	0	1
Well_Mntn_Crs Pth	Well-maintained crossing path indicator (1 if there is wwell-maintained crossing infrastructure at crossings, 0 otherwise)	0.58442	0.49282	0	1
Clearly Marked Crossings	Clearly marked crossings indicator (1 if yes, 0 otherwise)	0.58442	0.49282	0	1
Trnst_Stop	Transit stop within 0.5-mile of HRGCs indicator (1 if yes, 0 otherwise)	0.51948	0.49962	0	1
School	School within 0.5-mile of HRGCs indicator (1 if yes, 0 otherwise)	0.36364	0.48105	0	1
College	College within 0.5-mile of HRGCs indicator (1 if yes, 0 otherwise)	0.05195	0.22192	0	1
University	University within 0.5-mile of HRGCs indicator (1 if yes, 0 otherwise)	0.02597	0.15906	0	1
SO_GvnmtOfc	Other government office within 0.5-mile of HRGCs indicator (1 if yes, 0 otherwise)	0.6363	0.48105	0	1
Bsns_Cntr	Business center within 0.5-mile of HRGCs indicator (1 if yes, 0 otherwise)	0.24765	0.43112	0	1
Tourst_Atcn	Tourist attraction within 0.5-mile of HRGCs indicator (1 if yes, 0 otherwise)	0.12987	0.33616	0	1
Prkg_Spc	Public/commercial parking spaces within 0.5-mile of HRGC indicator (1 if yes, 0 otherwise)	0.33766	0.47291	0	1
Resd_Area	Residential area within 0.5-mile of HRGCs indicator (1 if yes, 0 otherwise)	0.94804	0.22192	0	1
Bycl_Trail	Bike trail within 0.5-mile of HRGCs indicator (1 if yes, 0 otherwise)	0.32468	0.46821	0	1
Parks	Parks within 0.5-mile of HRGCs indicator (1 if yes, 0 otherwise)	0.48052	0.49962	0	1
IndustrialA	Non-residential/industrial area within 0.5 mile of HRGCs indicator (1 if yes, 0 otherwise)	0.0389	0.1935	0	1
Urban_SubU	Urban crossing indicator (1 if the crossing is in urban area, 0 if the crossing is in sub-urban area)	0.87013	0.33616	0	1
AT5K	AADT between 1 and 5,000 indicator (1 if yes, 0 otherwise)	0.2987	0.45769	0	1
A5KT10K	AADT between 5,000 and 10,000 indicator (1 if yes, 0 otherwise)	0.038	0.1935	0	1
A10KT26K	AADT between 10,000 and 26,000 indicators (1 if yes, 0 otherwise)	0.3766	0.48485	0	1
Mx_T_Spd	Maximum timetable speed	25.793	22.2269	10	79
Crbk_Asb1	one crossbuck assembly indicator (if crossing has one Crossbuck assembly, 0 if otherwise)	0.0389	0.1935	0	1
Crbk_Asb2	two crossbuck assembly indicator (if crossing has two Crossbuck assembly, 0 if otherwise)	0.3897	0.48766	0	1
Traffic_Lanes	Number of traffic lanes	2.7922	1.1435	2	5
HighwayPaved	Highway paved indicator (1 if yes, 0 otherwise)	0.9611	0.1935	0	1
Nbr_Main_Trk	Total number of main tracks	0.83117	0.67232	0	2
GateArms_Prs	gate arms indicator (1 if there are gate arms on crossing, 0 otherwise)	0.87013	0.33616	0	1

Table 9 presents descriptive statistics on key candidate variables for NB modelling. Table 10 presents the results of three negative binomial models on daily pedestrian traffic volume, bicyclist traffic volume, and combined pedestrian and bicyclist traffic volume (total non-motorist traffic volume). The rationale behind estimating separate models for pedestrians and bicyclists was to understand if the factors affecting daily pedestrian volume differ from those influencing bicyclist traffic at HRGCs. For NB modeling, variables significant at the 90% confidence level were considered. Model performance metrics, such as Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Root Mean Square Error (RMSE), null and residual deviances, and the degree of overdispersion (θ), are also presented. The combined non-motorist model indicates that daily non-motorist traffic volume increases with greater visibility, the presence of sidewalks at HRGCs, and during cloudy conditions.

Table 10. Estimated NB and RENB for Predicting Total Non-Motorist Volume (24-Hr) at HRGCs

Variables Symbol	Total Non-Motorist Volume (24-Hr) Negative Binomial Model			Total Non-Motorist Volume (24-Hr) Random Effects Negative Binomial Model		
	Estimate	Z value	P value	Estimate	Z value	P value
Intercept	-4.03501	—	—	-1.7543	—	—
Road_Wet_II	-0.56371	-2.652	0.007999	-0.1735	-2.066	0.03885
Sidewalk	0.82362	2.880	0.003979	0.9037	2.332	0.0197
IndustrialA	-1.8025	-3.353	0.000799	-1.9033	-2.392	0.0168
Traffic_Lanes_5	-1.2576	-4.095	0.000001	-1.3884	-1.982	0.0475
Cloudy_I	0.8065	2.880	0.003971	0.2315	2.009	0.04454
Visibility	0.7543	3.394	0.000689	0.4976	2.327	0.0200
A5KT10K	-1.8461	-3.708	0.000209	-1.6330	-2.572	0.0101
A10KT26K	-0.1346	—	—	-0.2130	—	—
Model Performance Metrics						
Null Deviance	177.322			176.524		
Residual Deviance	86.967			84.214		
Degrees of Freedom	76			77		
Theta	1.888			1.9124		
2* log-likelihood	-640.455			-632.624		
AIC	660.45			654.60		
BIC	683.8926			680.41		
RMSE	36.14914			35.874		
Area Under the curve (AUC)	0.9221			0.9337		
Percentage of Correct Prediction	94.805%			92.22%		

Note: See Table 5 for the variable's coding details.

Conversely, higher average annual daily traffic and wet road conditions are associated with reduced non-motorist traffic volume. A relationship was observed between maximum train timetable speed and non-motorist volume. Average marginal effects (AMEs) for these results are provided in Table 11.

In evaluating the two models (Table 10), it is evident that both models perform well in identifying significant predictors. However, the RENB model provides a precise representation of the data, making it a better choice for the final statistical model in this analysis. The estimates from both models highlight similar trends regarding the impact of various factors on non-motorist volume. For instance, road wetness (Road_Wet_II) negatively affects non-motorist volumes in both models, though the impact is slightly weaker in the RENB model, where the estimate is -0.1735 compared to -0.56371 in the NB model. This difference suggests that while wet roads reduce non-motorist traffic on HRGCs, accounting for location-specific variability (through random effects) diminishes the strength of this association, reflecting more realistic outcomes.

Sidewalk availability plays a crucial role in increasing non-motorist volumes at HRGCs, with both models showing a significant positive effect. The RENB model's estimate of 0.9037 slightly exceeds the NB model's estimate of 0.82362, indicating that the positive influence of sidewalks is better captured when random effects are considered. This suggests that, across different crossings, the presence of sidewalks consistently promotes higher non-motorist activity, and the RENB model better captures this generalized trend by controlling for unobserved heterogeneity. Other variables, such as industrial areas (IndustrialA) and traffic lanes (Traffic_Lanes_5), show negative impacts on HRGCs' non-motorist volumes, with both models providing closed estimates. However, the RENB model adjusts these effects slightly, which aligns with the understanding that certain site-specific characteristics (e.g., industrial activities and road configurations) may vary across locations, affecting non-motorist behavior differently. The inclusion of random effects allows for these variations to be better accounted for, resulting in more precise estimates of their impact.

The variable Cloudy_I, representing cloudy weather conditions, also shows a positive relationship with non-motorist volumes. While the effect is stronger in the NB model (estimate = 0.8065), it is reduced to 0.2315 in the RENB model, suggesting that the initial NB model may have overestimated the influence of cloudy weather. The random effects adjustment in the RENB model provides a more moderated and likely more precise estimate, reflecting the fact that weather conditions may influence crossings differently depending on other local factors. From a statistical standpoint, the overall performance of the RENB model surpasses that of the NB model. Despite having a slightly lower percentage of correct predictions (92.22% for RENB vs. 94.805% for NB), the RENB model compensates for this with improvements in other areas. For instance, the RENB model demonstrates a lower residual deviance (84.214 vs. 86.967 for the NB model) and improved log-likelihood (-632.624 vs. -640.455), which indicates a better fit to the data. Additionally, the lower AIC (654.60 for RENB vs. 660.45 for NB)

and BIC (680.41 for RENB vs. 683.8926 for NB) further confirm that the RENB model is more efficient, balancing model complexity and goodness-of-fit better than the NB model.

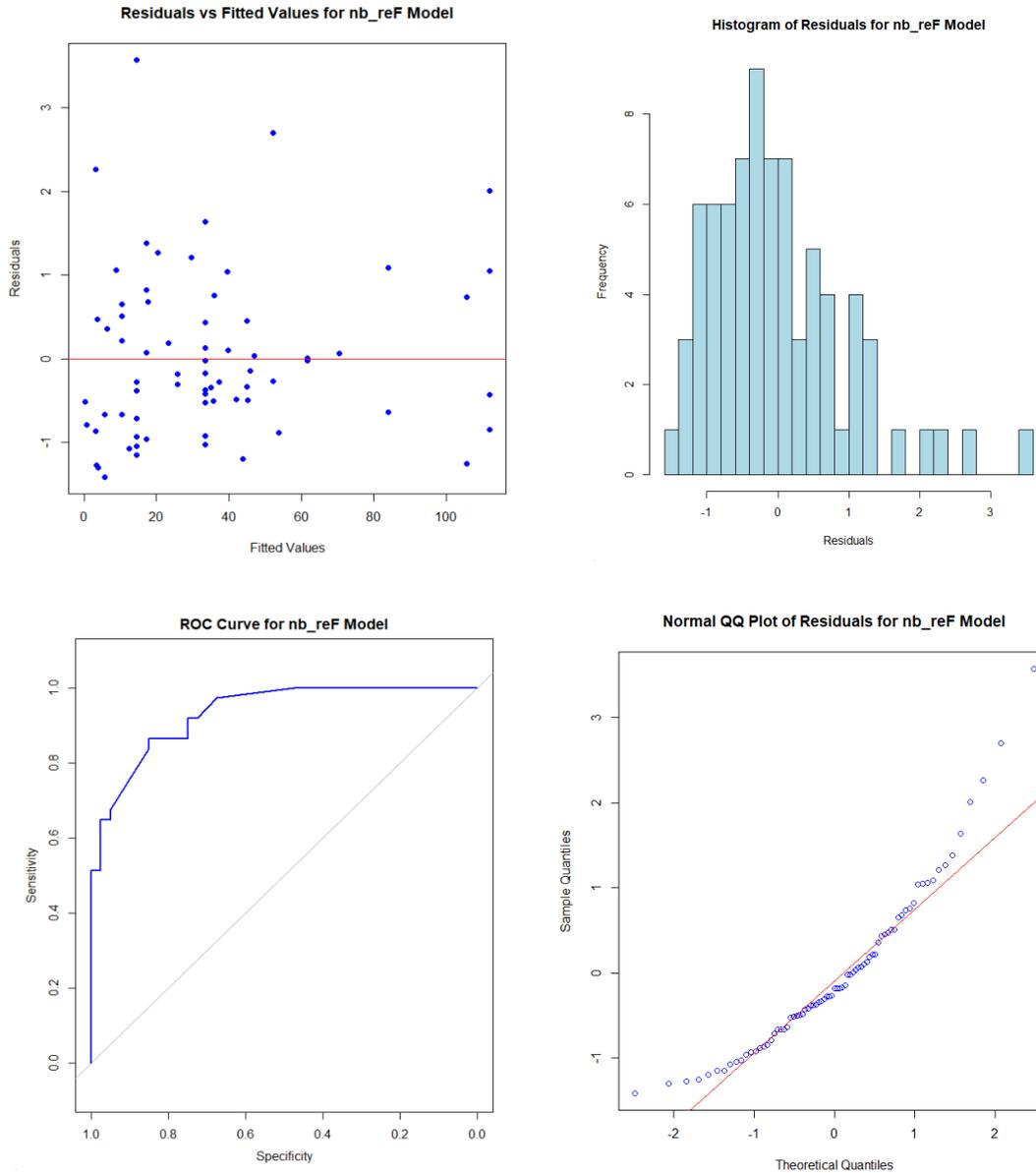


Figure 20. Diagnostic Plots for Final Selected Random Effects Negative Binomial Model (RENB).

Importantly, the slightly lower RMSE in the RENB model (35.874 vs. 36.14914 for the NB model) reflects its superior predictive accuracy. The consideration of random effects in the RENB model is key to its stronger performance. The diagnostic plots in Figure 20 provide critical insights into the

performance of the RENB model for HRGCs’ non-motorist volume prediction. The ROC curve, which measures the model's ability to distinguish between different outcomes, shows an area under the curve (AUC) close to 1 (ROC curve close to top-left corner), indicating a high level of predictive accuracy for the RENB model. This supports the conclusion that the RENB model effectively captures the relationship between predictors and non-motorist volume at HRGCs. Furthermore, the normal QQ plot assesses the normality of the residuals, with points falling along the diagonal line suggesting that the residuals are approximately normally distributed (Figure 20). The RENB model shows a reasonable adherence to normality, which indicates that the assumption of normally distributed random effects is not violated. This further strengthens the model's reliability for prediction. The histogram of residuals provides a visual check of residual distribution and looks into any deviation from normality; a bell-shaped histogram for RENB model suggests that the residuals are normally distributed. Furthermore, the residuals are centered around zero and show a symmetric spread, indicating that the model is well-calibrated, with no significant over- or underestimation across the range of predicted values. The residuals vs. fitted values plot reveals how well the RENB model fits the data. The absence of any clear pattern or trend in this plot suggests that the residuals are randomly distributed, implying that the model adequately explains the variability in non-motorist volume. This indicates that no significant systematic biases remain in the fitted values, highlighting the robustness of the RENB model.

Table 11. Estimated average marginal effects (AMEs) for NB and RENB model

Variables Symbol	Total Non-Motorist Volume (24-Hr) Negative Binomial Model			Total Non-Motorist Volume (24-Hr) Random Effects Negative Binomial Model		
	Estimate	Z value	P value	Estimate	Z value	P value
Road_Wet_II	-20.1572	-2.426	0.0153	-5.7953	-2.8148	0.0078
Sidewalk	29.4524	2.5871	0.0097	30.1655	1.9789	0.0479
IndustrialA	-64.4544	-3.0923	0.0020	-63.779	-2.0784	0.0377
Traffic_Lanes_5	-44.9713	-3.547	0.0004	-46.602	-1.9214	0.0454
Cloudy_I	28.8394	2.5341	0.0113	7.7628	1.9945	0.0401
Visibility	26.9716	3.1534	0.0016	16.6618	2.0653	0.0389
A5KT10K	-66.0145	-3.2421	0.0012	-54.7043	-2.1248	0.0336
A10KT26K	-4.8155	—	—	-7.0101	—	—

Note: See Table 9 for the variable’s coding details.

The Average Marginal Effects (AMEs) in Table 11 provide valuable insights into the influence of various factors on the total non-motorist volume at HRGCs. Both the Negative Binomial (NB) and Random Effects Negative Binomial (RENB) models indicate similar trends, but the RENB model offers a more detailed view by accounting for location-specific variability. Wet road conditions (Road_Wet_II) significantly reduce non-motorist volumes in both models, with a decrease of 20 in the NB model and about 6 in the RENB model. The smaller effect in the RENB model suggests that accounting for random

effects produces a more moderated, realistic estimate. Sidewalks have a strong positive impact on non-motorist traffic, with AMEs of 29 in the NB model and 30 in the RENB model, highlighting the importance of sidewalks in promoting non-motorist activity at HRGCs. Industrial areas are associated with a significant decrease in non-motorist volume, with both models showing a reduction of around 64 non-motorists, implying that these areas are less favorable for pedestrian and cyclist traffic. Traffic lanes (Traffic_Lanes_5) also reduce non-motorist activity, with AMEs of around 45 in both models, indicating that wider roads act as barriers. Cloudy weather increases non-motorist counts, with a larger effect in the NB model (29) than the RENB model (8), suggesting that the NB model may overestimate this impact. Higher visibility consistently leads to more non-motorist traffic, with AMEs of 27 in the NB model and 17 in the RENB model, reflecting that better visibility encourages non-motorist activity. Lastly, higher motorized traffic (A5KT10K) significantly reduces non-motorist volumes, with a reduction of 66 in the NB model and 55 in the RENB model. This consistent negative effect shows that higher traffic volumes create less favorable conditions for non-motorists at HRGCs.

6.2. Non-Motorist AI Modelling

Overview of Stochastic Gradient Descent (SGD) Model

To determine the most important aspects, AI modeling was used in this research to forecast the non-motorist count based on a variety of features. Numerous models were put to the test, such as neural networks, random forests, and elastic nets. SGD regression eventually produced the most dependable and understandable findings, despite the fact that each of these models—Elastic Net balancing regularization, Random Forest handling non-linear relationships, and Neural Networks collecting complicated patterns—has advantages of its own. The SGD model was the best option for our dataset because of its great predictive performance, simplicity, and efficiency.

The Stochastic Gradient Descent algorithm is used by the Stochastic Gradient Descent regression model, a linear regression model, to minimize the cost function, which is often the mean squared error for regression jobs. The primary principle underlying SGD is to iteratively alter the model parameters by taking small steps proportionate to the negative gradient of the cost function with respect to the parameters, which are then evaluated on a randomly selected subset (or just one) of the training data.

The linear regression model in SGD takes the following mathematical form:

$$\hat{y} = w^T x + b \quad \text{Eq (vii)}$$

Where:

\hat{y} is the predicted value.

- w is the vector of model weights.
- x is the input feature vector.

- b is the bias term (intercept).
- The goal of SGD is to minimize the cost function, which in the case of linear regression is usually the Mean Squared Error (MSE):

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad \text{Eq (viii)}$$

Where:

- n is the number of training examples.
- \hat{y}_i is the predicted value for the i_{th} training example.
- y_i is the true value for the i_{th} training example.

Gradient Descent Update Rule

The SGD algorithm updates the model parameters w and b using the following rules:

$$w := w - \eta \cdot \nabla_w J(w, b) \quad \text{Eq (ix)}$$

$$b := b - \eta \nabla_b J(w, b) \quad \text{Eq (x)}$$

where:

- η is the learning rate, a hyperparameter that controls the size of the steps taken during the update.
- $J(w, b)$ is the cost function (MSE in this case).
- $\nabla_w J(w, b)$ is the gradient of the cost function with respect to w .
- $\nabla_b J(w, b)$ is the gradient of the cost function with respect to b .

For a single training example (x_i, y_i) the gradients are computed as follows:

$$\nabla_w J(w, b) = -2(x_i)(y_i - \hat{y}_i) \quad \text{Eq (xi)}$$

$$\nabla_b J(w, b) = -2(y_i - \hat{y}_i) \quad \text{Eq (xii)}$$

Feature Selection

The initial selected characteristics for the model are Pop_Dnsty, Pop_L_150, Csng_Lgtng, Weekday_S, Avg_Temp, Precptn, Visibility, Max_W_Spd, Clear_Cndtn1, Rain_II, Snow_III, Cloudy_I, Rd_Dry_I, Rd_Wet, Rd_Wet_II, Well_Mntn_CrsPth, Bycl_Trail, Sidewlks, A5KT10K, Total_NM, MT20K, and Mx_T_Spd. Figure 21 shows the correlation matrix and matching heat map between these traits and the dependent variable, Total Nonmotorist.

These features—Rd_Wet_II, Rd_Wet, Well_Mntn_CrsPth, Pop_L_150, Rd_Dry_I, A5KT10K, and Total_NM—were chosen based on this matrix to be eliminated because their correlation with the dependent variable, Total Nonmotorist, was weak or negative, indicating that they had little effect on the predictive capacity of the model. The final features used in the SGD model are: Population Density (Pop_Dnsty), Crossing Lighting (Csng_Lgtng), Weekday Indicator (Weekday_S), Average Temperature (Avg_Temp), Precipitation (Precptn), Visibility (Visibility), Maximum Wind Speed (Max_W_Spd), Clear Conditions (Clear_Cndtn1), Rain Indicator (Rain_II), Snow Indicator (Snow_III), Cloudy Weather

Indicator (Cloudy_I), Presence of Bicycle Trail (Bycl_Trail), Sidewalk Indicator (Sidewlks), Maximum Timetable Speed of Trains (MT20K), and Maximum Train Speed (Mx_T_Spd).

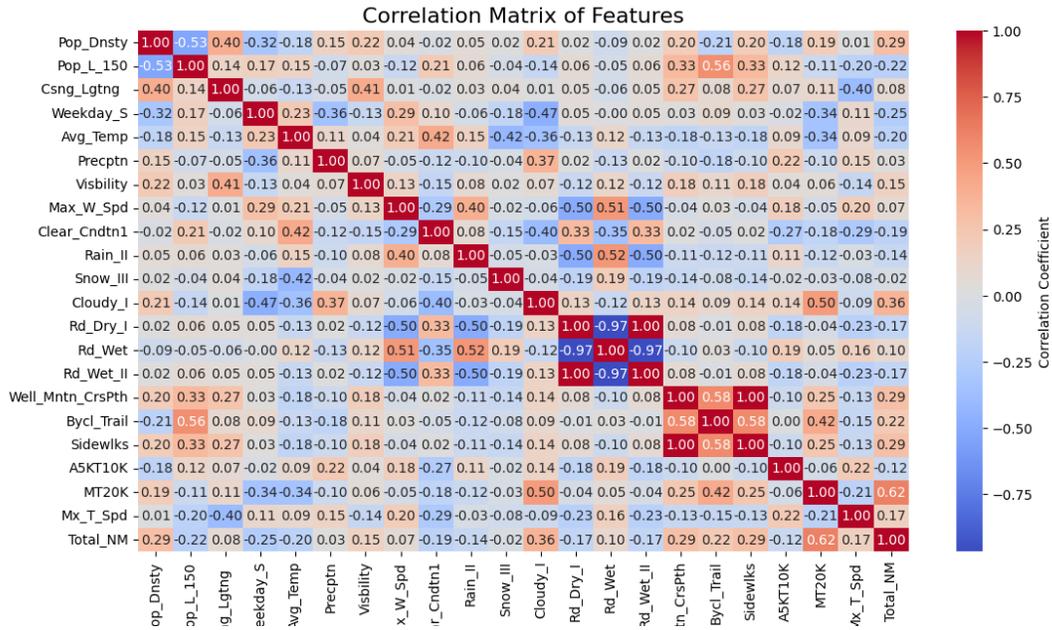


Figure 21. Correlation matrix of features.

Key Model Parameters and Representation

Parameters required to represent the model.

- Coefficients for the input characteristics are represented by w (Weights Vector).
Weights (coefficients): [2.117927 2.0422475 -2.4856343 1.6880354 -1.2662231 4.788772
4.230943 4.7556744 -5.0440063 2.6243181 4.422939 -3.6651185
7.3166375 27.027609 15.793775]
- Bias Term, b : The model's intercept.
Bias (intercept): [37.18875431]
- The learning rate, or η_t , regulates the step size for each weight update.

$$\eta_t = \frac{\eta_0}{(1+t)^{pt}} \quad \text{Eq (xiii)}$$

Where:

- η_0 is the initial learning rate and is set to 0.01.
- t is the iteration number.
- pt is set to 0.25, which controls how quickly the learning rate decreases.

- Function of Cost, $J(w, b)$:
For regression tasks, this is frequently the mean squared error (MSE). We use MSE.

5. Number of Iterations (Epochs): The total number of times the training dataset is run through the algorithm. We use 1000 epochs.
6. Mini-batch Size: The quantity of training examples utilized in a single SGD update iteration (optional, since SGD typically uses one sample each update). In our model we used mini-batch size=1, (Stochastic gradient descent).

Model Performance Metrics

Accuracy:

To assess the accuracy of our model, we set a 9-unit threshold. This cutoff was established using our dataset's nonmotorist count standard deviation, which came out to be 41.6. We rounded the threshold value to nine units, which is equal to one-fifth of the standard deviation. We were able to establish a useful and empirically based criterion for evaluating the model's performance thanks to this method.

Procedure for Computing SGD Regressor's AIC and BIC:

Do the Residual Sum of Squares (RSS) calculation: The total of the squared discrepancies between the observed and predicted values is used to achieve this.

Count the number of parameters. This is the number of weights (coefficients) plus the intercept for an SGDRegressor.

Utilizing the following formulas, calculate AIC and BIC:

$$\begin{aligned} \text{AIC} &= n \cdot \ln\left(\frac{\text{RSS}}{n}\right) + 2k && \text{Eq (xiv)} \\ \text{BIC} &= n \cdot \ln\left(\frac{\text{RSS}}{n}\right) + k \cdot \ln(n) && \text{Eq (xv)} \end{aligned}$$

where:

- n is the number of data points (samples).
- k is the number of parameters (coefficients + intercept).
- RSS is the residual sum of squares.

The trained SGD model's accuracy was assessed using several performance indicators. The average squared difference between the expected and actual values is 28.3687, which is the Mean Squared Error (MSE) for the test set. With a R-squared (R^2) value of 0.8438, the model appears to be a strong fit, explaining 84.38% of the variation in the dependent variable (Total Nonmotorist). Furthermore, the model's custom accuracy, which measures the percentage of predictions that fall within ten units of the actual values, was 91.67% within a 10-unit threshold. AIC and BIC, which penalize model complexity and aid in model selection by balancing fit and simplicity, were used to further evaluate the model's goodness-of-fit. The model's respective AIC and BIC values were 72.14 and 79.90. The model works well, capturing a considerable portion of the variability in the data while keeping an error rate that is comparatively low, according to these measures taken together.

Interpretation of Feature Coefficients

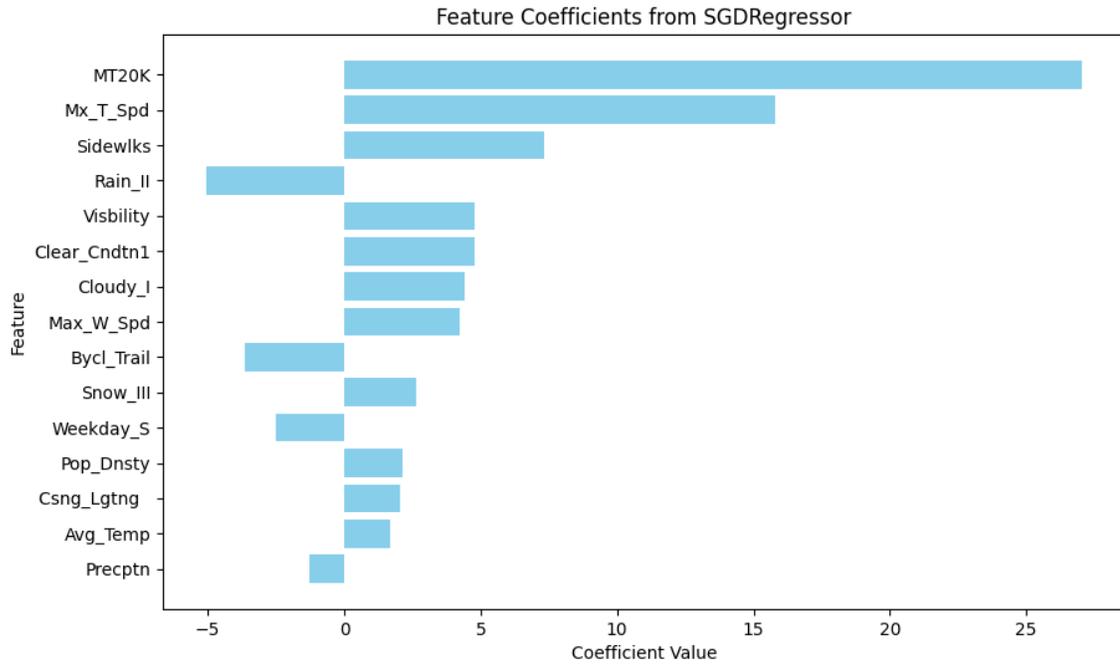


Figure 22. Coefficient value of the features used in SGD Regressor model.

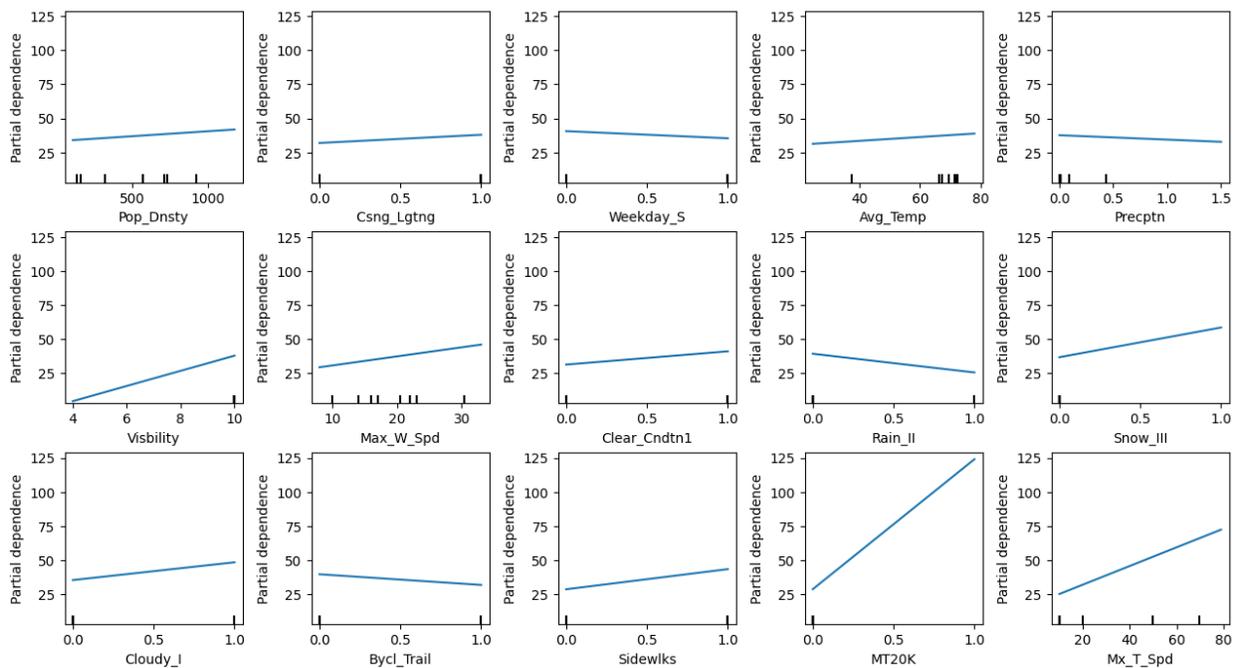


Figure 23. Partial Dependence Plots (PDP) for Each Selected Feature in the SGD Model.

Figure 22 shows the coefficient value of the features used in SGD Regressor model. The Partial Dependence Plots (PDP) for every feature chosen in the Stochastic Gradient Descent (SGD) regression model are displayed in Figure 23. Keeping other parameters constant, each subplot shows the relationship

between a particular feature and the expected outcome (Total Nonmotorist). Plots aid in the visualization of the effects of individual feature modifications on the model's predictions, offering insights into the behavior of the model and the influence of features.

The model's chosen characteristics offer insightful information about the variables influencing non-motorist counts. Greater population density (Pop_Dnsty) implies more traffic from pedestrians and cyclists, which increases non-motorist activity. Comparably, non-motorist counts are positively impacted by Crossing Lighting (Csng_Lgtng). Well-lit crossings make walking and bicycling safer, especially at night, which encourages more people to use non-motorized transportation.

In contrast, Weekday Indicator (Weekday_S) has a negative effect, meaning non-motorist counts are lower during weekdays compared to weekends. This could be because on weekdays, people are more likely to commute by motorized transportation for work or school, while weekends may see more recreational pedestrian and cyclist activity. Average Temperature (Avg_Temp) shows a positive correlation with non-motorist counts, as warmer temperatures generally encourage outdoor activities like walking or cycling.

Although it seems contradictory, there could be a positive association between the maximum wind speed and the number of nonmotorists due to particular behavioral, infrastructural, and geographical factors. This unexpectedly finding can be explained by a number of factors, including urban wind tunnels, weather coincidences, the presence of bike aficionados, and protective infrastructure in high-traffic regions.

The Clear Conditions Indicator (Clear_Cndtn1), which measures clear weather, has a favorable impact on non-motorist activity since it creates the perfect environment for cycling and strolling outside. Since cloudy weather frequently brings cooler temperatures and nice weather for outdoor activities without the hassle of rain or snow, it has a favorable impact on the number of non-motorists and does not considerably decrease their numbers. Non-motorist counts are negatively impacted by snow (Snow_III). Walking and cycling become hazardous in snowy circumstances, which decreases non-motorist activity. On the other hand, having sidewalks encourage more people to walk and increase the number of non-motorists by providing safe, designated areas for them.

The model demonstrates that, in contrast to predictions, the existence of bicycle trails is linked to a drop in the total population of nonmotorists, which includes both pedestrians and cyclists. This negative coefficient might be the result of bike lanes that are specifically designed to discourage foot traffic. If pedestrians perceive that bike trails are solely for bicycles, they may choose to use other routes or forms of transportation instead. Furthermore, places with bike routes might have better public transportation, which would encourage non-motorists to convert to more effective modes of transportation and lessen the need for walking or riding. Bicycle pathways may occasionally be constructed for recreational purposes, drawing riders who don't substantially increase daily nonmotorist counts, which would account for part of the negative correlation.

Lastly, the Maximum Timetable Speed of Trains (MT20K) exhibits a positive correlation with the number of non-motorists. This is probably due to the fact that regions with greater train speeds frequently have better developed infrastructure, which increases the volume of pedestrian traffic near stations and

crossings. Higher train speeds may be used in these well-infrastructure, high-activity locations to reduce disruptions and make sure that frequent train crossings don't impede non-motorist activities.

Comparison and Final Selection of Nonmotorist Exposure Model

In selecting the most suitable model for predicting non-motorist traffic at HRGCS, both AI and statistical models were evaluated in this study. While AI models, such as the Stochastic Gradient Descent (SGD) regression, offers powerful predictive capabilities, their application in this study revealed some limitations. The SGD model, for instance, delivered impressive performance, with an R-square value of 0.8438 and a custom accuracy of 91.67%. However, its complexity and the relatively small dataset size suggested that a more interpretable model might yield better insights into the relationships between various factors affecting non-motorist volume. Moreover, AI models generally focus on optimizing predictive accuracy but can lack transparency in interpreting the underlying effects of individual variables, which is crucial for transportation research. On the other hand, statistical models such as the Negative Binomial (NB) model and the Random Effects Negative Binomial (RENB) model offer both strong predictive performance and greater interpretability. The NB model provided a slightly higher accuracy, with a 94.805% correct prediction rate and an Area Under the Curve (AUC) of 0.9221. Despite this, the RENB model proved to be a better fit for explaining the variability in non-motorist traffic at different HRGC locations. The RENB model accounted for unobserved heterogeneity through random effects, which adjusted for site-specific characteristics that might not have been captured in the simpler NB model. As a result, the RENB model achieved a residual deviance of 84.214 (lower than the NB model's 86.967), a better log-likelihood (-632.624 compared to -640.455), and improved model efficiency metrics such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). While the NB model demonstrated higher overall accuracy, the RENB model provided a better overall understanding of the data. For instance, variables like road wetness (Road_Wet_II) and cloudy weather (Cloudy_I) exhibited weaker associations with non-motorist volume in the RENB model than in the NB model. This reflects the importance of adjusting for location-specific variability, as the RENB model captures more realistic effects by accounting for unobserved factors inherent to each HRGC. The positive effect of sidewalks (Sidewlks) on non-motorist volume was slightly stronger in the RENB model, indicating its ability to generalize across locations more effectively than the NB model. Moreover, factors such as industrial areas (IndustrialA) and traffic lanes (Traffic_Lanes_5) were better represented in the RENB model, where random effects captured how site-specific characteristics might affect non-motorist behavior differently across crossings. Ultimately, the RENB model was selected as the final model for this study. Despite the NB model's higher predictive accuracy, the RENB model offered a more comprehensive and reliable framework for explaining the impact of various factors on non-motorist volume across different HRGCs. By capturing the variability inherent in location-based fixed variables through random effects, the RENB model presents a more precise representation of the underlying data patterns, making it the preferred choice for non-motorist traffic prediction at HRGCs.

7. Conclusion and Recommendations

The study underscores the critical need for comprehensive non-motorist data collection and analysis at HRGCs to improve safety measures and crash prediction models. Video-based data collection

and AI-based computer vision algorithms were utilized to capture and analyze non-motorist activity at HRGCs. The use of AI-based object detection algorithms, specifically the YOLOv8 model, demonstrated effective identification and measurement of pedestrian and bicyclist volumes. While the model performed well overall, some discrepancies were noted, particularly in distinguishing between pedestrians and cyclists under certain conditions. These errors highlight the need for further refinement and validation of the model to ensure accurate data collection.

The study's predictive models using NB regression and AI models provide valuable insights for planning and safety of HRGC operations. The findings reveal that the presence of sidewalks and better visibility conditions are associated with higher volumes of both pedestrians and bicyclists. Conversely, higher traffic levels, and higher traffic lanes are linked to reduced non-motorist activity, suggesting that non-motorists are deterred by these factors due to perceived or actual safety risks. Similarly, the positive association between observed activities near rail crossings and non-motorist volume indicates that events or amenities can attract more bicyclists, likely due to increased interest or necessity to travel in those areas. By understanding the factors that influence non-motorist volumes, policymakers and planners can develop targeted interventions to enhance safety at HRGCs. For example, improving infrastructure such as sidewalks, and considering the impact of traffic and weather conditions to reduce the risk of crashes involving non-motorists. It is important to acknowledge that this study utilized a sample size of 77 data points, each representing a 24-hour recording period (totaling 1,848 hours of recorded data). The limited data points were primarily due to the time-intensive nature of Miovision installation, data processing, algorithm execution, and manual verification of counts. Consequently, fixed parameter Negative Binomial (NB) models were estimated. While larger sample sizes generally enhance the reliability of random parameter estimates and increase the ability to detect variability, our study still provides valuable insights. Despite the modest sample size, the collected data represent diverse conditions and scenarios encountered at HRGCs. The use of the fixed parameter NB model is a strategic choice that ensures robust parameter estimation without the instability that random parameters might introduce in smaller datasets. This modeling approach effectively balances model complexity and data constraints.

In conclusion, this research offers a foundational framework for integrating non-motorist data into HRGC safety analyses and crash prediction models. By addressing the previously overlooked non-motorist exposure, this study contributes to a detailed understanding of HRGC safety and highlights the need for continued efforts to protect vulnerable road users. Future research should focus on refining data collection methods and extending case studies, improving AI-based detection algorithms, and exploring additional factors that may influence non-motorist behavior and safety at HRGCs.

References

- Abari, M. E., & Naghsh-Nilchi, A. (2019). Toward a pedestrian detection method by various feature combinations. *International Journal of Knowledge-Based and Intelligent Engineering Systems*, 23(3), 191-201.
- Aggarwal, J. K., & Cai, Q. (1997). Human motion analysis: A review. *Computer Vision and Image Understanding*, 73(3), 428-440.
- Ahmed, S., Huda, M. N., Rajbhandari, S., Saha, C., Elshaw, M., & Kanarachos, S. (2019). Pedestrian and cyclist detection and intent estimation for autonomous vehicles: A survey. *Applied Sciences*, 9(11), 2335.
- Almasi, S. A., Behnood, H. R., & Arvin, R. (2021). Pedestrian crash exposure analysis using alternative geographically weighted regression models. *Journal of Advanced Transportation*, 2021, 1-13.
- Alshehri, A. H., Farooq, M. U., Farhan, A., Yosri, A. M., Alanazi, F., Deifallah, A. F., & Okail, M. A. (2023). A surrogate safety analysis at sharp gore areas of diverging freeway ramps using micro simulation under congested traffic conditions. *Scientific Reports*, 13(1), 22465.
- An, N., & Qi Yan, W. (2021). Multitarget tracking using Siamese neural networks. *ACM Transactions on Multimedia Computing Communications and Applications*, 17(2s), 1-16.
- Anjali, S. N. J., & Joe, N. (2018). Faster RCNN for concurrent pedestrian and cyclist detection. *SSRG International Journal of Electronics and Communication Engineering (SSRG IJECE)*, 5(5), 1-6
- Babb, C., & Curtis, C. (2015). Institutional practices and planning for walking: A focus on built environment audits. *Planning Theory & Practice*, 16(4), 517-534.
- Bapu Sridhar, S., & Moosakhanian, A. (2021). Pedestrian intent prediction using deep machine learning (Master's thesis, Chalmers University of Technology, Gothenburg, Sweden). Department of Electrical Engineering, Chalmers University of Technology.
- Behnam, J., & Patel, B. G. (1977). Pedestrian volume estimation by land-use variables. *Transportation Engineering Journal of ASCE*, 103(4), 507-520.
- Benham J., Patel B. G. A method for estimating pedestrian volume in a central business district. *Transportation Research Record: Journal of the Transportation Research Board*, 1977. 629: 22–26.
- Bertozzi, M., Broggi, A., Del Rose, M., Felisa, M., Rakotomamonjy, A., & Suard, F. (2007, September). A pedestrian detector using histograms of oriented gradients and a support vector machine classifier. In *IEEE Intelligent Transportation Systems Conference* (pp. 143-148). IEEE.
- Blanc B., Johnson P., Figliozzi M., Monsere C., & Nordback K. Leveraging signal infrastructure for nonmotorized counts in a statewide program: Pilot study. *Transportation Research Record: Journal of the Transportation Research Board*, 2015. 2527: 69-79.
- Bosina, E., Meeder, M., & Weidmann, U. (2017, May). Pedestrian flows on railway platforms. In *Proceedings of the 17th Swiss Transport Research Conference*.

- Chen, L., Lu, Y., Sheng, Q., Ye, Y., Wang, R., & Liu, Y. (2020). Estimating pedestrian volume using Street View images: A large-scale validation test. *Computers, Environment and Urban Systems*, 81, 101481.
- Chen, L., Lu, Y., Ye, Y., Xiao, Y., & Yang, L. (2022). Examining the association between the built environment and pedestrian volume using street view images. *Cities*, 127, 103734.
- Chen, Y., Wang, C., & Zhang, C. (2023). A robust lightweight network for pedestrian detection based on YOLOv5-x. *Applied Sciences*, 13(18), 10225.
- Cheng, Y., Chen, C., & Gan, Z. (2019, October). Enhanced single shot multibox detector for pedestrian detection. In *Proceedings of the 3rd International Conference on Computer Science and Application Engineering* (pp. 1-7).
- Cottrill, C. D., & Thakuriah, P. V. (2010). Evaluating pedestrian crashes in areas with high low-income or minority populations. *Accident Analysis & Prevention*, 42(6), 1718-1728.
- Dadashova, B., & Griffin, G. P. (2020). Random parameter models for estimating statewide daily bicycle counts using crowdsourced data. *Transportation Research Part D: Transport and Environment*, 84, 102368.
- Dadashova, B., Griffin, G. P., Das, S., Turner, S., & Sherman, B. (2020). Estimation of average annual daily bicycle counts using crowdsourced strava data. *Transportation research record*, 2674(11), 390-402.
- Day, C. M., Premachandra, H., & Bullock, D. M. (2016). Rate of pedestrian signal phase actuation as a proxy measurement of pedestrian demand. *Transportation Research Record*.
- Deng, Q., Tian, R., Chen, Y., & Li, K. (2018, June). Skeleton model based behavior recognition for pedestrians and cyclists from vehicle scene camera. In *2018 IEEE Intelligent Vehicles Symposium (IV)* (pp. 1293-1298). IEEE.
- Ding, H., Sze, N. N., Guo, Y., & Li, H. (2021). Role of exposure in bicycle safety analysis: Effect of cycle path choice. *Accident Analysis & Prevention*, 153, 106014.
- El Esawey, M. (2017). Estimation of daily bicycle traffic volumes using spatiotemporal relationships. *Journal of Transportation Engineering, Part A: Systems*, 143(11), 04017056.
- Ermagun, A., G. Lindsey, and T. Hadden Loh. 2018. Bicycle, pedestrian, and mixed-mode trail traffic: A performance assessment of demand models. *Landscape and Urban Planning* 177 (May): 92–102. <https://doi.org/10.1016/j.landurbplan.2018.05.006>.
- Esawey, M. E. (2014). Estimation of annual average daily bicycle traffic with adjustment factors. *Transportation Research Record*, 2443(1), 106-114.
- Ewing, R., and R. Cervero. Travel and the built environment: A synthesis. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1780, TRB, National Research Council, Washington, D.C., 2001, pp. 87–113.
- Fagnant, D. J., & Kockelman, K. (2016). A direct-demand model for bicycle counts: The impacts of level of service and other factors. *Environment and Planning B: Planning and Design*, 43(1), 93-107.
- Farooq, M. U., and A. J. Khattak. Investigating highway–rail grade crossing inventory data quality’s role in crash model estimation and crash prediction. *Applied Sciences*, 2023. Vol. 13, No. 20, 11537.

Farooq, M. U. (2023). The effects of inaccurate and missing highway-rail grade crossing inventory data on crash and severity model estimation and prediction (Doctoral dissertation, The University of Nebraska-Lincoln).

Farooq, M. U., Madson, C., & Khattak, A. (2023). Analyzing Major Elements of Crash Injury Severity Involving Priority-I Detriments of Vision-Zero Plan. Presented at the International Road Federation Global R2T Conference, Phoenix, AZ, November 2023.

Farooq, M. U., and A. J. Khattak. 2023. A Heterogeneity-Based Temporal Stability Assessment of Pedestrian Crash Injury Severity Using an Aggregated Crash and Hospital Data Set. Presented at the 102nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2023.

Farooq, M. U., Ahmed, A., & Saeed, T. U. (2021). A statistical analysis of the correlates of compliance and defiance of seatbelt use. *Transportation research part F: Traffic Psychology and Behaviour*, 77, 117-128.

Farooq, M. U. (2017). An Analysis of Legislation and Level of Compliance of Key Crash Risk Factors (Masters dissertation, National University of Sciences and Technology, Pakistan).

Federal Railroad Administration (FRA). Safety data and reporting. Available online: <https://railroads.dot.gov/safety-data> (accessed 3 May 2024).

Fiaz, M., Mahmood, A., & Jung, S. K. (2019). Deep siamese networks toward robust visual tracking. *Visual Object Tracking with Deep Neural Networks*.

Fournier, N., Christofa, E., & Knodler Jr, M. A. (2017). A sinusoidal model for seasonal bicycle demand estimation. *Transportation research part D: transport and environment*, 50, 154-169.

Gavrila, D. M. (1999). The visual analysis of human movement: A survey. *Computer Vision and Image Understanding*, 73(1), 82-98.

Greene, W. Functional forms for the Negative Binomial model for count data. *Economics Letters*, 2008. Vol. 99, No. 3, pp. 585-590.

Grembek O., Bosman C., Bigham J. M., Fine S., Griswold J. B., Medury A., Sanders R. L., Schneider R. J., Yavari A., Zhang Y., Ragland D.R. Pedestrian Safety Improvement Program: Final Technical Report. Prepared by the UC Berkeley Safe Transportation Research and Education Center for the California Department of Transportation, Sacramento, CA, 2014.

Griswold, J. B., Medury, A., & Schneider, R. J. (2011). Pilot models for estimating bicycle intersection volumes. *Transportation research record*, 2247(1), 1-7.

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.

Handy, S. Critical Assessment of the Literature on the relationships among transportation, land use, and physical activity. Transportation Research Board of the National Academies, Washington, D.C., 2005. http://trb.org/downloads/sr282_papers/sr282Handy.pdf.

Hankey S., Lindsey G. Facility-demand models of peak-period pedestrian and bicycle traffic: A comparison of fully-specified and reduced-form models. *Transportation Research Record: Journal of the Transportation Research Board*, 2016. 2586: 48–58.

Hankey S., Lindsey G., Wang X., Borah J., Hoff K., Utecht B., Xu Z. Estimating Use of non-motorized infrastructure: Models of bicycle and pedestrian traffic in Minneapolis, MN. *Landscape and Urban Planning*, Vol. 107, 2012, pp. 307–316.

Hankey S., Lu T., Mondschein A., Buehler R. Spatial models of active travel in small communities: Merging the goals of traffic monitoring and direct-demand modeling. *Journal of Transport and Health*, Vol. 7, 2017, pp. 149–159.

Hansen, J., Takeda, K., Schmidt, G., & Abut, H. (Eds.). (2020). *Vehicles, Drivers, and Safety (Vol. 2)*. Walter de Gruyter GmbH & Co KG.

Hariyono, J., & Jo, K. H. (2015, September). Detection of pedestrian crossing road. In *2015 IEEE International Conference on Image Processing (ICIP)* (pp. 4585-4588). IEEE.

Hariyono, J., Hoang, V. D., & Jo, K. H. (2014). Moving object localization using optical flow for pedestrian detection from a moving vehicle. *The Scientific World Journal*, 2014(1), 196415.

Harrison, J., & daSilva, M. (2019). 2017 Grade crossing research needs Workshop (No. DOT-VNTSC-FRA-18-01). United States Department of Transportation. Federal Railroad Administration. Office of Research, Development, and Technology.

Hausman, J. A., Hall, B. H., & Griliches, Z. (1984). Econometric models for count data with an application to the patents-R&D relationship.

Haynes M., Andrzejewski S. GIS based bicycle & pedestrian demand forecasting techniques. Presentation for US Department of Transportation, Travel Model Improvement Program. Fehr & Peers Transportation Consultants, Los Angeles, CA, 2010. <https://tmip.org/content/gis-based-bicycle-and-pedestrian-demand-forecasting-and-traffic-count-programs>.

He, F., Karami Olia, P., Jamili Oskouei, R., Hosseini, M., Peng, Z., & BaniRostam, T. (2021). Applications of deep learning techniques for pedestrian detection in smart environments: A comprehensive study. *Journal of advanced transportation*, 2021(1), 5549111.

Hochmair, H. H., Bardin, E., & Ahmouda, A. (2019). Estimating bicycle trip volume for Miami-Dade County from Strava tracking data. *Journal of Transport Geography*, 75, 58-69.

Hui, L., Wang, L., Tang, L., Lan, K., Xie, J., & Yang, J. (2022, October). 3D siamese transformer network for single object tracking on point clouds. In *European Conference on Computer Vision* (pp. 293-310). Cham: Springer Nature Switzerland.

Irwin, D. (2003, November). Safety criteria for light rail pedestrian crossings. In *9th National Light Rail Transit Conference* (Vol. 2003, pp. 266-288). Portland, Oregon: Transportation Research Board.

Jacobs, J. (2010). *Dark age ahead: Author of the death and life of great American cities*. Vintage Canada.

Jain, S. (2020). *Advances in computational intelligence techniques*. M. Sood, & S. Paul (Eds.). Springer Singapore.

Jiang, F., Ma, J., & Li, Z. (2022). Pedestrian volume prediction with high spatiotemporal granularity in urban areas by the enhanced learning model. *Sustainable Cities and Society*, 79, 103653.

Johnstone D., Nordback K., & Kothuri S. Annual average nonmotorized traffic estimates from manual counts: Quantifying error. *Transportation Research Record: Journal of the Transportation Research Board*, 2018. 2672(43): 134-144.

Jones M. G., Ryan S., Donlan J., Ledbetter L., Arnold L., Ragland D. Seamless Travel: Measuring bicycle and pedestrian activity in San Diego County and its relationship to land use, transportation, safety, and facility type. Task Order 6117. Prepared by Alta Planning & Design and UC Berkeley Safe Transportation Research & Education Center, California Department of Transportation, Sacramento, CA, 2010.

Jones, M. G., Ryan, S., Donlon, J., Ledbetter, L., Ragland, D. R., & Arnold, L. S. (2010). Seamless travel: measuring bicycle and pedestrian activity in San Diego County and its relationship to land use, transportation, safety, and facility type (No. UCB-ITS-PRR-2010-12).

Keall, M. D. (1995). Pedestrian exposure to risk of road accident in New Zealand. *Accident Analysis & Prevention*, 27(5), 729-740.

Khattak, A. J., Farooq, M. U., & Farhan, A. (2024). Motor vehicle drivers' knowledge of safely traversing highway-rail grade crossings. *Transportation research record*, 2678(7), 604-621.

Kim, J. H., Batchuluun, G., & Park, K. R. (2018). Pedestrian detection based on faster R-CNN in nighttime by fusing deep convolutional features of successive images. *Expert Systems with Applications*, 114, 15-33.

Kobayashi, T., Hidaka, A., & Kurita, T. (2008). Selection of histograms of oriented gradients features for pedestrian detection. In *Neural Information Processing: 14th International Conference, ICONIP 2007, Kitakyushu, Japan, November 13-16, 2007, Revised Selected Papers, Part II 14* (pp. 598-607). Springer Berlin Heidelberg.

Kocamaz, M. K., Gong, J., & Pires, B. R. (2016, March). Vision-based counting of pedestrians and cyclists. In *2016 IEEE winter conference on applications of computer vision (WACV)* (pp. 1-8). IEEE.

Kothuri, S., Nordback, K., Schrope, A., Phillips, T., & Figliozzi, M. (2017). Bicycle and pedestrian counts at signalized intersections using existing infrastructure: Opportunities and challenges. *Transportation research record*, 2644(1), 11-18.

Krizek, K. Operationalizing neighborhood accessibility for land use travel behavior research and regional modeling. *Journal of Planning Education and Research*, Vol. 22, 2003, pp. 270-287.

Kumar, A., Zhang, Z. J., & Lyu, H. (2020). Object detection in real time based on improved single shot multi-box detector algorithm. *EURASIP Journal on Wireless Communications and Networking*, 2020(1), 204.

Kuzmyak J. R., Walters J., Bradley M., Kockelman K. M. Estimating bicycling and walking for planning and project development: A guidebook. National Cooperative Highway Research Program Report 770, Transportation Research Board, Washington, D. C., 2014.

Lam, W. W., Yao, S., & Loo, B. P. (2014). Pedestrian exposure measures: A time-space framework. *Travel Behaviour and Society*, 1(1), 22-30.

- Lee, J., Abdel-Aty, M., & Jiang, X. (2015). Multivariate crash modeling for motor vehicle and non-motorized modes at the macroscopic level. *Accident Analysis & Prevention*, 78, 146-154.
- Leeper, T. J. Interpreting regression results using average marginal effects with R's margins. Available at the Comprehensive R Archive Network (CRAN), 2017. Vol. 32, pp. 1-32.
- Lewin, A. (2011). Temporal and weather impacts on bicycle volumes (No. 11-2536).
- Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., ... & Wei, X. (2022). YOLOv6: A single-stage object detection framework for industrial applications. arXiv preprint arXiv:2209.02976.
- Li, M. L., Sun, G. B., & Yu, J. X. (2023). A pedestrian detection network model based on improved YOLOv5. *Entropy*, 25(2), 381.
- Li, X., & Wu, Y. J. (2021). Real-time estimation of pedestrian volume at button-activated midblock crosswalks using traffic controller event-based data. *Transportation Research Part C: Emerging Technologies*, 122, 102876.
- Li, X., Flohr, F., Yang, Y., Xiong, H., Braun, M., Pan, S., & Gavrilu, D. M. (2016, June). A new benchmark for vision-based cyclist detection. In 2016 IEEE Intelligent Vehicles Symposium (IV) (pp. 1028-1033). IEEE.
- Li, Z., Dong, Y., Wen, Y., Xu, H., & Wu, J. (2021). A deep pedestrian tracking SSD-based model in the sudden emergency or violent environment. *Journal of Advanced Transportation*, 2021(1), 2085876.
- Lindsey, G. H. 2011. Forecasting use of nonmotorized infrastructure: Models of bicycle and pedestrian traffic in Minneapolis, Minnesota." *Transportation Research Board 90th Annual Meeting* 11–2749. <https://trid.trb.org/view/1092646>.
- Lobb, B., Harre, N., & Terry, N. (2003). An evaluation of four types of railway pedestrian crossing safety intervention. *Accident Analysis & Prevention*, 35(4), 487-494.
- Lu, T., A. Mondschein, R. Buehler, and S. Hankey. 2018. Adding temporal information to direct-demand models: Hourly estimation of bicycle and pedestrian traffic in Blacksburg, VA. *Transportation Research, Part D: Transport & Environment* 63:244–260. <https://doi.org/10.1016/j.trd.2018.05.011>.
- Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 207(1167), 187-217.
- McDaniel S., Lowry M. B., & Dixon M. Using origin–destination centrality to estimate directional bicycle volumes. *Transportation Research Record: Journal of the Transportation Research Board*, 2014. 2430: 12-19.
- Miah, M. M., Hyun, K. K., & Mattingly, S. P. (2024). A review of bike volume prediction studies. *Transportation Letters*, 1-28.
- Miah, M. M., Hyun, K. K., Mattingly, S. P., & Khan, H. (2023). Estimation of daily bicycle traffic using machine and deep learning techniques. *Transportation*, 50(5), 1631-1684.
- Mimouna, A., Khalifa, A. B., Alouani, I., Taleb-Ahmed, A., Rivenq, A., & Amara, N. B. (2021, February). LSTM-based system for multiple obstacle detection using ultra-wide band radar. In 13th

International Conference on Agents and Artificial Intelligence, ICAART 2021 (pp. 418-425). SCITEPRESS-Science and Technology Publications.

Miranda-Moreno L. F., Fernandes D. Modeling of pedestrian activity at signalized intersections: Land use, urban form, weather, and spatiotemporal patterns. *Transportation Research Record: Journal of the Transportation Research Board*, 2011. 2264: 74–82.

Mukherjee, D., and S. Mitra. Identification of pedestrian risk factors using negative binomial model. *Transportation in Developing Economies*, 2020. Vol. 6, No. 1, 4.

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.

Munira, S., Sener, I. N., & Zhang, Y. (2021). Estimating bicycle demand in the Austin, Texas area: Role of a bikeability index. *Journal of Urban Planning and Development*, 147(3), 04021036.

Munira, Sirajum, and Ipek N. Sener. Use of direct-demand modeling in estimating nonmotorized activity: A meta-analysis. *Safety through Disruption (Safe-D) National University Transportation Center (UTC) Program* (2017).

Pan, L., Diao, J., Wang, Z., Peng, S., & Zhao, C. (2024). HF-YOLO: Advanced pedestrian detection model with feature fusion and imbalance resolution. *Neural Processing Letters*, 56(2), 90.

Papageorgiou, C., & Poggio, T. (2000). A trainable system for object detection. *International Journal of Computer Vision*, 38(1), 15-33.

Pourhomayoun, M. (2020). Automatic traffic monitoring and management for pedestrian and cyclist safety using deep learning and artificial intelligence. *IEEE Transactions on Intelligent Transportation Systems*, 21(5), 1848-1857.

Pulugurtha S. S., Repaka S. R. Assessment of models to measure pedestrian activity at signalized intersections. *Transportation Research Record: Journal of the Transportation Research Board*, 2008. 2073: 39–48.

Pushkarev, B., & Zupan, J. M. (1971). Pedestrian travel demand. Presented at the 50th Annual Meeting of the Highway Research Board, Washington, D.C., United States, January 18-22, 1971.

Qin, X., & Ivan, J. N. Estimating pedestrian exposure prediction model in rural areas. *Transportation Research Record: Journal of the Transportation Research Board*, 2001. 1773: 89-96.

Raford, N., & Ragland, D. Space syntax: Innovative pedestrian volume modeling tool for pedestrian safety. *Transportation Research Record: Journal of the Transportation Research Board*, 2004. 1878: 66-74.

Saranya, K. C., & Thangavelu, A. (2019). Vulnerable road user detection using YOLO v3. *International Journal of Advanced Computer Science and Applications*, 10(12).

Schneider R. J., Arnold L. S., Ragland D. R. A pilot model for estimating pedestrian intersection crossing volumes. *Transportation Research Record: Journal of the Transportation Research Board*, 2009. 2140: 13–26.

- Schneider R. J., Henry T., Mitman M. F., Stonehill L., Koehler J. Development and application of the San Francisco pedestrian intersection volume model. *Transportation Research Record: Journal of the Transportation Research Board*, 2012. 2299: 65–78.
- Shapiro, R. J., J. F. Holmes, N. M. Gordon, J. F. Long, and D. H. Weinberg. 1997 Population profile of the United States. U.S. Department of Commerce, Economic and Statistics Administration and Bureau of the Census. Sept. 1998. <http://www.census.gov/prod/3/98pubs/p23-194.pdf>. Accessed May 2023.
- Sheng, W., Shen, J., Huang, Q., Liu, Z., & Ding, Z. (2024). Multi-objective pedestrian tracking method based on YOLOv8 and improved DeepSORT. *Mathematical Biosciences and Engineering: MBE*, 21(2), 1791-1805.
- Shi, W., Bao, S., & Tan, D. (2019). FFESSD: An accurate and efficient single-shot detector for target detection. *Applied Sciences*, 9(20), 4276.
- Shriver, K. Influence of environmental design on pedestrian travel behavior in Four Austin neighborhoods. In *Transportation Research Record 1578*, TRB, National Research Council, Washington, D.C., 1997, pp. 64–75.
- Siddiqui, C., Abdel-Aty, M., & Choi, K. (2012). Macroscopic spatial analysis of pedestrian and bicycle crashes. *Accident Analysis & Prevention*, 45, 382-391.
- Singleton, P. A., & Runa, F. (2021). Pedestrian traffic signal data accurately estimates pedestrian crossing volumes. *Transportation research record*, 2675(6), 429-440.
- Sobreira, L. T. P., & Hellinga, B. (2023). Comparing direct demand models for estimating pedestrian volumes at intersections and their spatial transferability to other jurisdictions. *Transportation research record*, 2677(10), 260-271.
- Sobreira, L. T. P., & Hellinga, B. (2024). Estimating pedestrian volumes at each crosswalk of intersections: Comparison of land-use models and short-term count methods. *Journal of Transportation Engineering, Part A: Systems*, 150(2), 04023136.
- Southworth, M. (2005). Designing the walkable city. *Journal of Urban Planning and Development*, 131(4), 246-257.
- Statista (2024). Cycling-Statistics & Facts. <https://www.statista.com/topics/1686/cycling/#topicOverview>. Accessed June 2024.
- Sze, N. N., Su, J., & Bai, L. (2019). Exposure to pedestrian crash based on household survey data: Effect of trip purpose. *Accident Analysis & Prevention*, 128, 17-24.
- Tao, T., Lindsey, G., Stern, R., & Levin, M. (2024). The use of crowdsourced mobile data in estimating pedestrian and bicycle traffic: A systematic review. *Journal of Transport and Land Use*, 17(1), 41-65.
- Tian, H., Duan, Z., Abraham, A., & Liu, H. (2013). A novel multiplex cascade classifier for pedestrian detection. *Pattern Recognition Letters*, 34(14), 1687-1693.
- Tian, W., & Lauer, M. (2015, September). Fast cyclist detection by cascaded detector and geometric constraint. In *2015 IEEE 18th International Conference on Intelligent Transportation Systems* (pp. 1286-1291). IEEE.

- Tong, B., Fan, B., & Wu, F. (2016). Convolutional neural networks with neural cascade classifier for pedestrian detection. In *Pattern Recognition: 7th Chinese Conference, CCPR 2016, Chengdu, China, November 5-7, 2016, Proceedings, Part I 7* (pp. 243-257). Springer Singapore.
- Tudor-Locke, C., Bittman, M., Merom, D., & Bauman, A. (2005). Patterns of walking for transport and exercise: A novel application of time use data. *International Journal of Behavioral Nutrition and Physical Activity*, 2, 1-10.
- Vanparijs, J., Panis, L. I., Meeusen, R., & De Geus, B. (2015). Exposure measurement in bicycle safety analysis: A review of the literature. *Accident Analysis & Prevention*, 84, 9-19
- Vijayakumar, A., & Vairavasundaram, S. (2024). Yolo-based object detection models: A review and its applications. *Multimedia Tools and Applications*, 1-40.
- Walker, C. L., Khattak, A. J., Farooq, M. U., Cecava, J., & Anderson, M. R. (2024). Investigation of winter weather crash injury severity using winter storm classification techniques. *Transportation Research Interdisciplinary Perspectives*, 24, 101073.
- Wang, J. Y., & Adelson, E. H. (1993, June). Layered representation for motion analysis. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (pp. 361-366). IEEE.
- Wang, K., & Zhou, W. (2019). Pedestrian and cyclist detection based on deep neural network fast R-CNN. *International Journal of Advanced Robotic Systems*, 16(2), 1729881419829651.
- Wang, X., Yang, J., Lee, C., Ji, Z., & You, S. (2016). Macro-level safety analysis of pedestrian crashes in Shanghai, China. *Accident Analysis & Prevention*, 96, 12-21.
- Watanabe, T., Ito, S., & Yokoi, K. (2009). Co-occurrence histograms of oriented gradients for pedestrian detection. In *Advances in Image and Video Technology: Third Pacific Rim Symposium, PSIVT 2009, Tokyo, Japan, January 13-16, 2009. Proceedings 3* (pp. 37-47). Springer Berlin Heidelberg.
- Whyte, W. H. (1980). *The social life of small urban spaces*. Washington, D.C.: The Conservation Foundation.
- Wier, M., Weintraub, J., Humphreys, E. H., Seto, E., & Bhatia, R. (2009). An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. *Accident Analysis & Prevention*, 41(1), 137-145.
- Wren, C. R., Azarbayejani, A., Darrell, T., & Pentland, A. P. (1997). Pfnder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), 780-785.
- Xiang, Y., & Fei, Q. (2022). Car, Cyclist and Pedestrian Object Detection Based on YOLOv5. *International Journal of New Developments in Engineering and Society*, 6(3).
- Yin, L. (2017). Street level urban design qualities for walkability: Combining 2D and 3D GIS measures. *Computers, Environment and Urban Systems*, 64, 288-296.
- Zaman, A., Huang, Z., Li, W., Qin, H., Kang, D., & Liu, X. (2024). Development of railroad trespassing database using artificial intelligence (No. DOT/FRA/ORD-24/09). United States Department of Transportation Federal Railroad Administration.

Zhang, L. H., & Li, L. (2013). Improved pedestrian detection based on extended histogram of oriented gradients. *Applied Mechanics and Materials*, 347, 3815-3820.

Zhang, Q., Hu, X., Yue, Y., Gu, Y., & Sun, Y. (2022). Multi-object detection at night for traffic investigations based on improved SSD framework. *Heliyon*, 8(11), e11570.

Zhao, L., Farooq, M. U., & Khattak, A. J. (2024). Data accuracy matters: Improving highway-rail grade crossings crash predictions through inventory verification. *Transportation Research Record*, 03611981241270179.

Zhao, S., Tian, Y., Hao, N., Zhou, J., & Zhang, X. (2023, May). Improved YOLOv5 algorithm for intensive pedestrian detection. In *International Conference on Computational & Experimental Engineering and Sciences* (pp. 587-594). Cham: Springer Nature Switzerland.

Zou, H., Cui, J., Kong, X., Zhang, C., Liu, Y., Wen, F., & Li, W. (2020, October). F-siamese tracker: A frustum-based double siamese network for 3D single object tracking. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 8133-8139). IEEE.

Appendix A

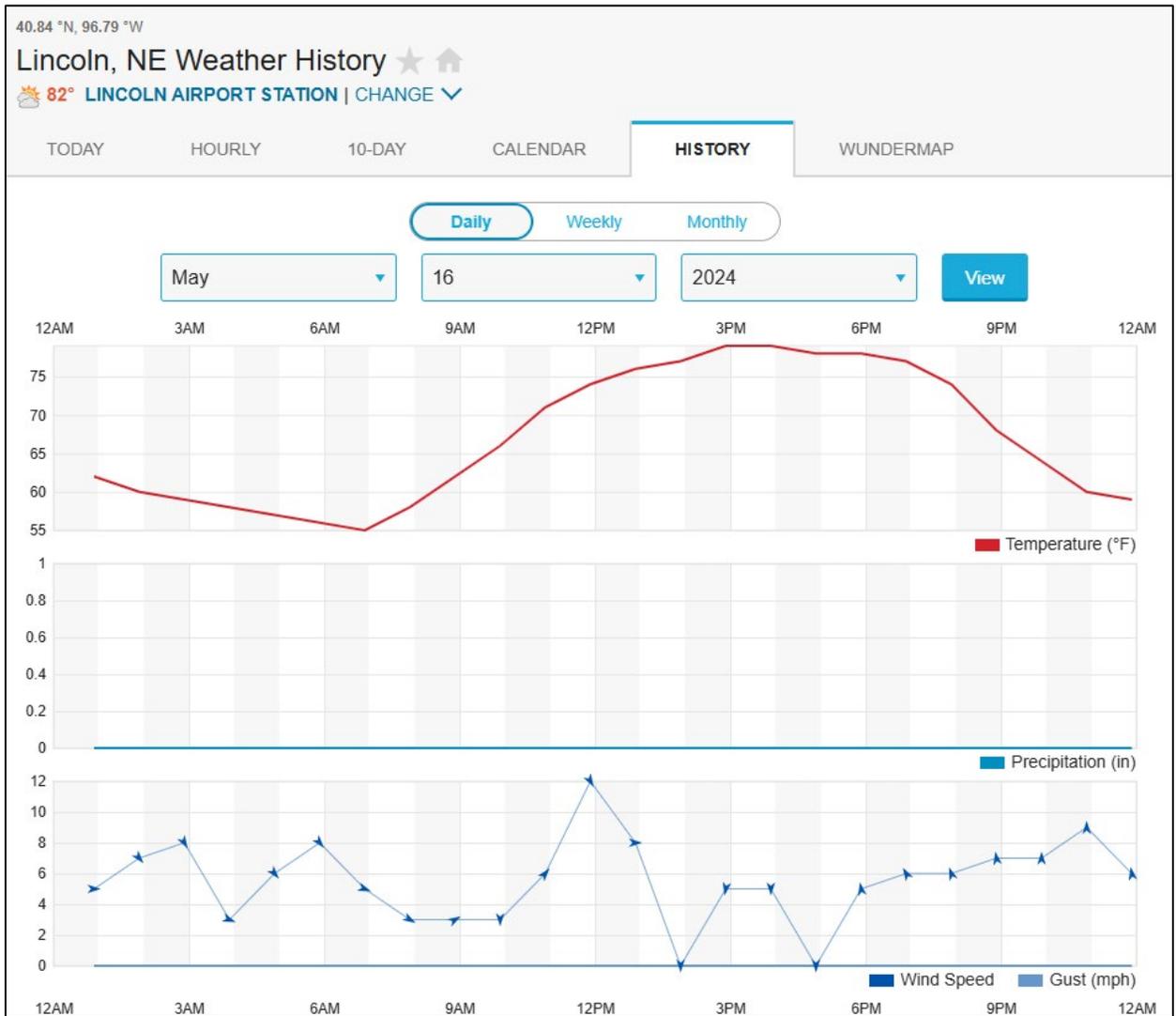
U. S. DOT Crossing Inventory Form 71

U. S. DOT CROSSING INVENTORY FORM					
DEPARTMENT OF TRANSPORTATION FEDERAL RAILROAD ADMINISTRATION				OMB No. 2130-0017	
<p><small>Instructions for the initial reporting of the following types of new or previously unreported crossings: For public highway-rail grade crossings, complete the entire inventory Form. For private highway-rail grade crossings, complete the Header, Parts I and II, and the Submission Information section. For public pathway grade crossings (including pedestrian station grade crossings), complete the Header, Parts I and II, and the Submission Information section. For Private pathway grade crossings, complete the Header, Parts I and II, and the Submission Information section. For grade-separated highway-rail or pathway crossings (including pedestrian station crossings), complete the Header, Part I, and the Submission Information section. For changes to existing data, complete the Header, Part I Items 1-3, and the Submission Information section, in addition to the updated data fields. Note: For private crossings only, Part I Item 20 and Part III Item 2.K. are required unless otherwise noted. An asterisk * denotes an optional field.</small></p>					
A. Revision Date (MM/DD/YYYY) ____/____/____	B. Reporting Agency <input type="checkbox"/> Railroad <input type="checkbox"/> Transit <input type="checkbox"/> State <input type="checkbox"/> Other		C. Reason for Update (Select only one) <input type="checkbox"/> Change in Data <input type="checkbox"/> Re-Open <input type="checkbox"/> New Crossing <input type="checkbox"/> Date Change Only <input type="checkbox"/> Closed <input type="checkbox"/> Change in Primary Operating RR		D. DOT Crossing Inventory Number
Part I: Location and Classification Information					
1. Primary Operating Railroad		2. State		3. County	
4. City / Municipality <input type="checkbox"/> In <input type="checkbox"/> Near		5. Street/Road Name & Block Number (Street/Road Name) * (Block Number)		6. Highway Type & No.	
7. Do Other Railroads Operate a Separate Track at Crossing? <input type="checkbox"/> Yes <input type="checkbox"/> No If Yes, Specify RR			8. Do Other Railroads Operate Over Your Track at Crossing? <input type="checkbox"/> Yes <input type="checkbox"/> No If Yes, Specify RR		
9. Railroad Division or Region <input type="checkbox"/> None		10. Railroad Subdivision or District <input type="checkbox"/> None		11. Branch or Line Name <input type="checkbox"/> None	
12. RR Milepost ____ ____ ____ (prefix) (nnnn.nnn) (suffix)		13. Line Segment *		14. Nearest RR Timetable Station *	
15. Parent RR (if applicable) <input type="checkbox"/> N/A		16. Crossing Owner (if applicable) <input type="checkbox"/> N/A		17. Crossing Type <input type="checkbox"/> Public <input type="checkbox"/> Private	
18. Crossing Purpose <input type="checkbox"/> Highway <input type="checkbox"/> Pathway, Ped. <input type="checkbox"/> Station, Ped.		19. Crossing Position <input type="checkbox"/> At Grade <input type="checkbox"/> RR Under <input type="checkbox"/> RR Over		20. Public Access (if Private Crossing) <input type="checkbox"/> Yes <input type="checkbox"/> No	
21. Type of Train <input type="checkbox"/> Freight <input type="checkbox"/> Intercity Passenger <input type="checkbox"/> Commuter		22. Average Passenger Train Count Per Day <input type="checkbox"/> Less Than One Per Day <input type="checkbox"/> Number Per Day _____		23. Type of Land Use <input type="checkbox"/> Open Space <input type="checkbox"/> Farm <input type="checkbox"/> Residential <input type="checkbox"/> Commercial <input type="checkbox"/> Industrial <input type="checkbox"/> Institutional <input type="checkbox"/> Recreational <input type="checkbox"/> RR Yard	
24. Is there an Adjacent Crossing with a Separate Number? <input type="checkbox"/> Yes <input type="checkbox"/> No If Yes, Provide Crossing Number			25. Quiet Zone (FRA provided) <input type="checkbox"/> No <input type="checkbox"/> 24 Hr <input type="checkbox"/> Partial <input type="checkbox"/> Chicago Excused Date Established _____		
26. HSR Corridor ID <input type="checkbox"/> N/A		27. Latitude in decimal degrees (WGS84 std: nn.nnnnnnn)		28. Longitude in decimal degrees (WGS84 std: -nnn.nnnnnnn)	
29. Lat/Long Source <input type="checkbox"/> Actual <input type="checkbox"/> Estimated		30.A. Railroad Use *		31.A. State Use *	
30.B. Railroad Use *		31.B. State Use *		30.C. Railroad Use *	
30.D. Railroad Use *		31.C. State Use *		30.D. Railroad Use *	
31.D. State Use *		32.A. Narrative (Railroad Use) *		32.B. Narrative (State Use) *	
33. Emergency Notification Telephone No. (posted)		34. Railroad Contact (Telephone No.)		35. State Contact (Telephone No.)	
Part II: Railroad Information					
1. Estimated Number of Daily Train Movements					
1.A. Total Day Thru Trains (6 AM to 6 PM)		1.B. Total Night Thru Trains (6 PM to 6 AM)		1.C. Total Switching Trains	
1.D. Total Transit Trains		1.E. Check if Less Than One Movement Per Day <input type="checkbox"/> How many trains per week?			
2. Year of Train Count Data (YYYY)		3. Speed of Train at Crossing 3.A. Maximum Timetable Speed (mph) _____ 3.B. Typical Speed Range Over Crossing (mph) From _____ to _____			
4. Type and Count of Tracks Main _____ Sidling _____ Yard _____ Transit _____ Industry _____					
5. Train Detection (Main Track only) <input type="checkbox"/> Constant Warning Time <input type="checkbox"/> Motion Detection <input type="checkbox"/> AFO <input type="checkbox"/> PTC <input type="checkbox"/> DC <input type="checkbox"/> Other <input type="checkbox"/> None					
6. Is Track Signaled? <input type="checkbox"/> Yes <input type="checkbox"/> No		7.A. Event Recorder <input type="checkbox"/> Yes <input type="checkbox"/> No		7.B. Remote Health Monitoring <input type="checkbox"/> Yes <input type="checkbox"/> No	
FORM FRA F 6180.71 (Rev. 08/03/2016)		OMB approval expires 11/30/2022		Page 1 OF 2	

U. S. DOT CROSSING INVENTORY FORM

A. Revision Date (MM/DD/YYYY)		PAGE 2		D. Crossing Inventory Number (7 char.)	
Part III: Highway or Pathway Traffic Control Device Information					
1. Are there Signs or Signals? <input type="checkbox"/> Yes <input type="checkbox"/> No					
2.A. Crossbuck Assemblies (count)		2.B. STOP Signs (R1-1) (count)	2.C. YIELD Signs (R1-2) (count)	2.D. Advance Warning Signs (Check all that apply; include count) <input type="checkbox"/> None	
				<input type="checkbox"/> W10-1 <input type="checkbox"/> W10-3 <input type="checkbox"/> W10-11 <input type="checkbox"/> W10-2 <input type="checkbox"/> W10-4 <input type="checkbox"/> W10-12	
2.E. Low Ground Clearance Sign (W10-5)		2.F. Pavement Markings		2.G. Channelization Devices/Medians	2.H. EXEMPT Sign (R15-3)
<input type="checkbox"/> Yes (count _____) <input type="checkbox"/> No		<input type="checkbox"/> Stop Lines <input type="checkbox"/> Dynamic Envelope <input type="checkbox"/> RR Xing Symbols <input type="checkbox"/> None		<input type="checkbox"/> All Approaches <input type="checkbox"/> Median <input type="checkbox"/> One Approach <input type="checkbox"/> None	<input type="checkbox"/> Yes <input type="checkbox"/> No
2.I. Other MUTCD Signs		<input type="checkbox"/> Yes <input type="checkbox"/> No Specify Type _____ Count _____ Specify Type _____ Count _____ Specify Type _____ Count _____	2.K. Private Crossing Signs (if private)	2.L. LED Enhanced Signs (List types)	
			<input type="checkbox"/> Yes <input type="checkbox"/> No		
3. Types of Train Activated Warning Devices at the Grade Crossing (specify count of each device for all that apply)					
3.A. Gate Arms (count)		3.B. Gate Configuration	3.C. Cantilevered (or Bridged) Flashing Light Structures (count)	3.D. Mast Mounted Flashing Lights (count of masts)	3.E. Total Count of Flashing Light Pairs
Roadway _____ Pedestrian _____		<input type="checkbox"/> 2 Quad <input type="checkbox"/> Full (Barrier) <input type="checkbox"/> 3 Quad Resistance <input type="checkbox"/> 4 Quad <input type="checkbox"/> Median Gates	Over Traffic Lane _____ <input type="checkbox"/> Incandescent Not Over Traffic Lane _____ <input type="checkbox"/> LED	<input type="checkbox"/> Incandescent <input type="checkbox"/> LED <input type="checkbox"/> Back Lights Included <input type="checkbox"/> Side Lights Included	
3.F. Installation Date of Current Active Warning Devices: (MM/YYYY)		3.G. Wayside Horn		3.H. Highway Traffic Signals Controlling Crossing	3.I. Bells (count)
____/____/____ <input type="checkbox"/> Not Required		<input type="checkbox"/> Yes Installed on (MM/YYYY) ____/____/____ <input type="checkbox"/> No		<input type="checkbox"/> Yes <input type="checkbox"/> No	
3.J. Non-Train Active Warning				3.K. Other Flashing Lights or Warning Devices	
<input type="checkbox"/> Flagging/Flagman <input type="checkbox"/> Manually Operated Signals <input type="checkbox"/> Watchman <input type="checkbox"/> Floodlighting <input type="checkbox"/> None				Count _____ Specify type _____	
4.A. Does nearby Hwy Intersection have Traffic Signals?		4.B. Hwy Traffic Signal Interconnection	4.C. Hwy Traffic Signal Preemption	5. Highway Traffic Pre-Signals	6. Highway Monitoring Devices (Check all that apply)
<input type="checkbox"/> Yes <input type="checkbox"/> No		<input type="checkbox"/> Not Interconnected <input type="checkbox"/> For Traffic Signals <input type="checkbox"/> For Warning Signs	<input type="checkbox"/> Simultaneous <input type="checkbox"/> Advance	<input type="checkbox"/> Yes <input type="checkbox"/> No Storage Distance * _____ Stop Line Distance * _____	<input type="checkbox"/> Yes - Photo/Video Recording <input type="checkbox"/> Yes - Vehicle Presence Detection <input type="checkbox"/> None
Part IV: Physical Characteristics					
1. Traffic Lanes Crossing Railroad		2. Is Roadway/Pathway Paved?	3. Does Track Run Down a Street?	4. Is Crossing Illuminated? (Street lights within approx. 50 feet from nearest rail)	
<input type="checkbox"/> One-way Traffic <input type="checkbox"/> Two-way Traffic <input type="checkbox"/> Divided Traffic		<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Yes <input type="checkbox"/> No	
5. Crossing Surface (on Main Track, multiple types allowed) Installation Date * (MM/YYYY) ____/____/____ Width * _____ Length * _____					
<input type="checkbox"/> 1 Timber <input type="checkbox"/> 2 Asphalt <input type="checkbox"/> 3 Concrete <input type="checkbox"/> 4 Concrete and Rubber <input type="checkbox"/> 5 Rubber <input type="checkbox"/> 6 Metal <input type="checkbox"/> 8 Unconsolidated <input type="checkbox"/> 9 Composite <input type="checkbox"/> 10 Other (specify) _____					
6. Intersecting Roadway within 500 feet?			7. Smallest Crossing Angle	8. Is Commercial Power Available? *	
<input type="checkbox"/> Yes <input type="checkbox"/> No If Yes, Approximate Distance (feet) _____			<input type="checkbox"/> 0° - 29° <input type="checkbox"/> 30° - 59° <input type="checkbox"/> 60° - 90°	<input type="checkbox"/> Yes <input type="checkbox"/> No	
Part V: Public Highway Information					
1. Highway System		2. Functional Classification of Road at Crossing		3. Is Crossing on State Highway System?	4. Highway Speed Limit System? MPH
<input type="checkbox"/> (D1) Interstate Highway System <input type="checkbox"/> (D2) Other Nat Hwy System (NHS) <input type="checkbox"/> (D3) Federal AID, Not NHS <input type="checkbox"/> (D8) Non-Federal Aid		<input type="checkbox"/> (0) Rural <input type="checkbox"/> (1) Urban <input type="checkbox"/> (1) Interstate <input type="checkbox"/> (5) Major Collector <input type="checkbox"/> (2) Other Freeways and Expressways <input type="checkbox"/> (3) Other Principal Arterial <input type="checkbox"/> (6) Minor Collector <input type="checkbox"/> (4) Minor Arterial <input type="checkbox"/> (7) Local		<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Posted <input type="checkbox"/> Statutory
7. Annual Average Daily Traffic (AADT)		8. Estimated Percent Trucks	9. Regularly Used by School Buses?	6. LRS Milepost *	
Year ____ AADT ____		_____ %	<input type="checkbox"/> Yes <input type="checkbox"/> No Average Number per Day _____	_____	
				10. Emergency Services Route	
				<input type="checkbox"/> Yes <input type="checkbox"/> No	
Submission Information - This information is used for administrative purposes and is not available on the public website.					
Submitted by _____ Organization _____ Phone _____ Date _____					
Public reporting burden for this information collection is estimated to average 30 minutes per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed and completing and reviewing the collection of information. According to the Paperwork Reduction Act of 1995, a federal agency may not conduct or sponsor, and a person is not required to, nor shall a person be subject to a penalty for failure to comply with, a collection of information unless it displays a currently valid OMB control number. The valid OMB control number for information collection is 2130-0017. Send comments regarding this burden estimate or any other aspect of this collection, including for reducing this burden to: Information Collection Officer, Federal Railroad Administration, 1200 New Jersey Ave. SE, MS-25 Washington, DC 20590.					

Historical Weather Report by Weather Underground



Summary

Temperature (°F)	Actual	Historic Avg.	Record	▲
High Temp	79	75	93	
Low Temp	55	51.2	30	
Day Average Temp	66.96	63.1	-	
Precipitation (in)	Actual	Historic Avg.	Record	▲
Precipitation (past 24 hours from 11:54:00)	0.08	4.60	-	
Dew Point (°F)	Actual	Historic Avg.	Record	▲
Dew Point	50.54	-	-	
High	58	-	-	
Low	45	-	-	
Average	50.54	-	-	
Wind (mph)	Actual	Historic Avg.	Record	▲
Max Wind Speed	12	-	-	
Visibility	10	-	-	
Sea Level Pressure (in)	Actual	Historic Avg.	Record	▲
Sea Level Pressure	28.63	-	-	
Astronomy	Day Length	Rise	Set	▲
Actual Time	14h 30m	6:09 AM	8:40 PM	
Civil Twilight		5:38 AM	9:11 PM	
Nautical Twilight		4:59 AM	9:50 PM	
Astronomical Twilight		4:15 AM	10:33 PM	
Moon: waxing gibbous		2:05 PM	3:06 AM	