

Article

Investigating Unsafe Pedestrian Behavior at Urban Road Midblock Crossings Using Machine Learning: Lessons from Alexandria, Egypt

Ahmed Mahmoud Darwish ^{1,*}, Sherif Shokry ^{2,*}, Maged Zagow ³, Marwa Elbany ⁴, Ali Qabur ⁵,
Talal Obaid Alshammari ⁶, Ahmed Elkafoury ⁷ and Mohamed Shaaban Alfiqui ⁸

¹ Department of Transportation Engineering, Faculty of Engineering, Alexandria University, Alexandria 21533, Egypt

² The Center of Road Traffic Safety, Naif Arab University for Security Sciences, Riyadh 6830, Saudi Arabia

³ Department of Architectural Engineering, Faculty of Engineering, Tanta University, Tanta 31527, Egypt; mzagow@f-eng.tanta.edu.eg

⁴ Transportation and Traffic Engineering, Civil Engineering Department, Faculty of Engineering, Port Said University, Port Said 42526, Egypt; mr_elbany@eng.psu.edu.eg

⁵ Department of Civil and Architectural Engineering, Jazan University, Jazan 45142, Saudi Arabia; aqabur@jazanu.edu.sa

⁶ Department of Civil Engineering, College of Engineering, Jouf University, Sakaka 72388, Saudi Arabia; toalshammari@ju.edu.sa

⁷ Department of Public Works Engineering, Faculty of Engineering, Tanta University, Tanta 3111, Egypt; kaforiengin@f-eng.tanta.edu.eg

⁸ Public Works Department, Faculty of Engineering, Cairo University, Cairo 12613, Egypt; mohammedalfiki@cu.edu.eg

* Correspondence: adarwish@alexu.edu.eg (A.M.D.); sshokry@nauss.edu.sa (S.S.)

Abstract

Examining pedestrian crossing violations at high-risk road midblock crossings has become essential, particularly in high-speed corridors, as a result of accidents at crossings resulting in fatalities. Hence, this article investigates such behavior in Alexandria, Egypt, as a credible case study in a developing country. According to our research methodology, a comprehensive dataset of over 2400 field-observed video recordings was used for real-life data collection. Machine learning (ML) models, such as CatBoost and gradient boosting (GB), were employed to predict crossing decisions. The models showed that risky behavior is strongly influenced by waiting time, crossing time, and the number of crossing attempts. The highest predictive performance was achieved by CatBoost and gradient boosting, indicating strong interpersonal influence within small groups engaging in unsafe road-crossing behavior. In the same context, the Shapley additive explanation (SHAP) values for these variables were 3, 2, and 0.60, respectively. Subsequently, based on SHAP sensitivity analysis, the results show that the total time (s) and age group (40–60 Y) had a significant negative influence on model prediction converging to class 0 (e.g., crossing illegally). The results also showed that shorter exposure times increase the likelihood of crossing illegally. This research work is among the few studies that employ a behavior-based approach to understanding pedestrian behavior at midblock crossings. This study offers actionable insights and valuable information for urban designers and transportation planners when considering the design of midblock crossings.

Keywords: urban mobility; sustainable built environment; active travel; unsafe pedestrian road-crossing behavior; machine learning; urban transport planning



Academic Editors: Roberto Murcio and Honghan Bei

Received: 21 December 2025

Revised: 14 January 2026

Accepted: 18 January 2026

Published: 26 January 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Ensuring the safety and security of pedestrians on roads is a pressing necessity for sustainable urban transport. Devastating consequences result from exponential increases in the number of vehicles worldwide, pushing research to act as a guard of pedestrians' safety [1,2].

Moreover, studying pedestrian movement patterns, pedestrian movement theories, their routes, and their destinations is essential. Pedestrian crossing is one of the main safety issues of pedestrian–vehicle interactions, helping urban planners to proactively implement warrants, reducing risks in midblock sections. Moreover, midblock crossing on major arterial roads should be studied due to complex interactions among higher vehicle speed, higher traffic volumes, and mixed-use environments [3–6].

Developing countries (e.g., low- and middle-income countries (LMICs)) usually suffer from a high fatality rate according to the *Global Status Report on Road Safety 2023*, with about 92% of world road fatalities despite ownership of only 60% of worldwide vehicles [7]. Subsequently, vehicles are the top priority in these countries' considerations in road design and operation [8]. Pedestrians in LMICs suffer from high-risk exposure, especially in a densely populated urban context. As such, those vulnerable groups should receive more attention and care. Challenges include underreported pedestrian accident data because of a lack of communication among authorities in those countries. Moreover, the lack of implementation of protection measures at dangerous locations, especially in and around midblock sections, hinders any positive attitude towards pedestrian safety in these areas [9].

It is impractical to fully split this pedestrian–vehicle interaction along the whole road length in urban areas. With a focus solely on the most hazardous locations, zebra crossings and pedestrian footbridges, subways, crossing buttons, signs, markings, and signals are placed tactically at the most hazardous locations in urban areas [10].

Pedestrian midblock crossings in urban areas of rapidly growing cities in developing countries bring significant safety challenges. Although pedestrian decks, tunnels, and signalized crossings are often present, many pedestrians choose to illegally cross at uncontrolled locations. Over the last decade, the growing rates of pedestrian accident rates in developing cities, especially in Egypt, hit a record [11,12]. Most of these accidents occurred due to illegal or unsafe crossing behaviors, known as crossing violations [13]. Such a phenomenon grows in densely populated urban neighborhoods with arterial roads. In these areas, pedestrians often experience excessive delays waiting for crossing signals. Despite different crossing facilities, such as pedestrian decks or tunnels, many pedestrians do not choose to use them for different reasons [14].

Little work has been performed to assess the effect of modeling illegal pedestrian midblock crossing behavior in Egypt. To the best of the authors' knowledge, there have been no studies investigating illegal crossing behavior at midblock crossings in Alexandria. Furthermore, the machine learning method has not been used for analyzing pedestrian crossing behavior in Egypt. Therefore, gaps remain with regard to methodology, study area, and data source used (e.g., field video dataset analysis).

Only three research studies have been conducted in Egypt investigating pedestrian crossing behavior. Okail (2022) [15] studied pedestrian behaviors in developing countries (Egypt as a case study) using a pedestrian behavior questionnaire and exploratory factor analysis. Moreover, Sayed et al. (2022) [16] investigated driver' behavior, personal characteristics, risk perception, and involvement in crashes using the Driver Behavior Questionnaire.

Bayoumi et al. [17] investigated the Pedestrian Behavior Scale (PBS) in Egypt using confirmatory factor analysis. In addition, they studied pedestrian crossing behavior at five different midblock locations using video recordings, using logistic regression models, not

machine learning models, based on socio-demographic characteristics such as gender, age, using a phone, pedestrian crossing locations, and crossing side.

As a result, the tension between pedestrian needs and safety and operational strategies reveals a major research gap. This gap drives the need to understand the reasons why pedestrians ignore safety, even when crossings exist or no immediate danger is present.

To overcome this, comprehensive research on this relationship should be conducted. Moreover, illegal crossing behavior should also be examined to better understand significant related factors, such as perceived crossing time, exact pedestrian destination, average vehicle speed, distance to the nearest legal designated crossing facilities (e.g., foot over-bridge, pedestrian tunnels, on-road crossing marking/signs, etc.).

Some researchers [18–24] concluded that males adopt IPCB more often than females. Other factors, such as habit and attitude, affect IPCB, as mentioned by refs. [25–27]. Moreover, social impact affects IPCB, as mentioned in refs. [13,21,23]. In addition, traffic characteristics, road and environment, and behavior also affect IPCB, as mentioned in refs. [19,22,24,25,28–33].

The influence of demographic factors on pedestrian safety has been examined in many considerable studies [34–36]. Various valuable attempts have paid much attention to investigating the effectiveness of various pedestrian control strategies [37]. Much less research, however, has addressed crossing violations at midblock locations where many fatal pedestrian crashes occur. Traditional statistical modeling, such as discrete choice models and other classical models, provides a reasonable foundation for the interpretation of behavioral patterns [38].

However, they are often ill-equipped to account for complex, nonlinear interactions among behavior, temporal, and contextual factors. Recently, the ML revolution has expanded the opportunity for researchers to explore these interactions and identify trends that may not be captured through traditional modeling. Nowadays, recent research examines explainable artificial intelligence (XAI) tools to examine illegal pedestrian crossing behavior (IPCB) at midblock locations in depth.

Today, ML, deep learning (DL), the Internet of Things (IOT) integration, and computer vision (CV) are being used to develop predictive pedestrian accident models for exploring factors affecting non-compliant pedestrian crossing behavior (NCPCB) [39–41]. Nitire et al. [42] developed a systematic review of factors for earlier prediction of collisions using artificial intelligence (AI), ML, reinforcement learning (RL), inverse reinforcement learning (IRL), and DL. Furthermore, new data sources such as open sources like Kaggle [43], road safety maps, and computer vision applications using DL facilitate tremendous model development using new AI applications [42].

In addition, several ML methods have been developed and applied to model illegal pedestrian crossing behavior (IPCB) at midblock locations such as AdaBoost, random forest (RF), extremely randomized trees (ERTs), gradient boosting (GB), support vector machine (SVM), semi-supervised learning (SSL), decision trees (DTs), neural network (NN), Bagged CART, Model Trees, extreme learning machine (ELM), multivariate adaptive regression spline (MARS), Bayesian generalized linear model (BGLM), k-nearest neighbor (KNN), naïve Bayes (NB), and artificial neural networks (ANNs) [44].

On the other hand, DL has several methods, such as feedforward neural network (FFNN), recurrent neural network (RNN), radial basis function neural network (RBNN), Kohonen self-organizing neural network (KSONN), modular neural network (MNN), convolution neural network (CNN), Autoencoder, restricted Boltzmann machine (RBM), long short-term memory (LSTM), and gradient descent (GD) [45].

The application of machine learning is widely used in pedestrian crossing behavior analysis, achieving a statistically acceptable level of accuracy. For example,

Jun Cai et al. [46] compared machine learning with traditional analytical models and found that the application of OpenCV image recognition and machine learning methods can analyze the mechanisms of pedestrian crossing behaviors with greater accuracy (about 92%). Moreover, Dungar Singh et al. [47] applied random forest, extreme gradient boosting, and binary logit model, with prediction rates of 81.72%, 77.19%, and 74.95%, respectively, to analyze pedestrian crossing behavior at unsignalized intersections. They concluded that the random forest approach is superior to the logit model.

A research study conducted in Jordan in 2024 [48] included over 2400 individual pedestrians' demographic characteristics collected at nine signalized intersections. The model investigated behavioral measures, timing measures, and context characteristics, primarily using structured field observations and video-based tracking. There have also been studies that have applied ML approaches to model crossing behaviors [49], but these studies mainly focus on how to cross, not on the original decision to cross illegally. Moreover, pedestrian activity patterns were examined and classified using classical ML approaches, including ANN, SVM, DT, and RF, at nine signalized intersections in three cities in Jordan (Amman, Irbid, and Zarqa) [48]. The variables were gender, age, carrying bags, crossing conditions, and mobile phone use.

One of the key measures that have been implemented is the development of a comprehensive midblock crossing pedestrian safety action plan through cooperation between the Texas Department of Transportation (TxDOT) and the Federal Highway Administration (FHWA). In addition, pedestrian safety margin and surrogate safety measures are tested in terms of conflicts and interaction severity [50]. This action plan identifies the riskiest midblock locations across urban corridors based on local crash modification factors (CMFs), incorporating site ranking, treatment prioritization strategies, and enforcement considerations [3]. This analytical study examines the factors directly correlated to the accuracy of the prediction procedure and explores the reasons behind these predictions. This study relied on using SHAP values and permutation importance to examine the factors influencing pedestrian crossing behavior through the analysis process [51]. When a strong prediction is linked to interpretability, the obtained results are more reliable and highly significant.

This research study primarily aims to establish an innate recognition framework for pedestrian safety by identifying the motivations behind illegal crossing behavior in developing urban areas. In addition, it does not merely describe pedestrian patterns, as the findings also provide new ways to reconcile safety requirements with the need to ensure mobility for people, benefiting transportation planners and policymakers.

This research study, therefore, focuses on crossing behavior at two high-risk midblock sites in Alexandria, Egypt, one of the most congested and populated cities in the Alexandria Governorate. Moreover, it particularly examines the characteristics associated with illegal crossing behavior on Al-Mahmoudiya Road, a newly redeveloped and highly trafficked corridor.

Overall, this study attempts to reshape the understanding of pedestrian behavior at midblock locations in high-speed corridors in a developing country. To achieve the research objectives, therefore, this study is built upon the following sub-objectives:

- Identify the important determinants in relation to legal and illegal pedestrian crossing decisions at midblock locations along major urban roads.
- Develop and compare multiple ML classification models to predict pedestrian crossing behavior with high accuracy.
- Use intelligible tools (e.g., SHAP and permutation importance) to evaluate the comparative impacts of behavioral, demographic, and temporal variables.

- Examine how pedestrians adapt their crossing behavior to compensate for waiting time, crossing speed, grouping, and age in dealing with infrastructural challenges and traffic conditions.
- Assess the generalizability and robustness of ML models in predicting crossing behavior under differing conditions.
- Provide evidence-based support to urban and transportation planners that will assist them in re-designing midblock environments that will reduce violations and improve pedestrian safety.

2. Literature Review

The authors summarize previous related studies focusing on the application of ML and DL on the modeling of IPCB, as mentioned in Table 1. In addition, related studies are categorized based on population and objectives, models, results, and research outcomes. Moreover, the authors divided this review into four main categories: First, modeling pedestrian movements is investigated using AI-aided tools, data collection, simulation models, etc. Second, the authors address factors affecting IPCB, especially near pedestrian crossing facilities. These factors are categorized from different perspectives (demographics, psychological, social influence, traffic characteristics, road and environment, behavioral, policy, and regional and cultural aspects), as mentioned in Table 2. Third, actions taken by developing countries to overcome IPCB are presented. Fourth, models developed for IPCB are discussed, focusing on ML and DL applications.

2.1. Modeling Pedestrian Movements

Understanding pedestrian movements is a crucial point in modeling their illegal crossing behavior. Thus, simulation of pedestrian crossing behavior at midblock locations is especially important. Piyalungka S. et al. [52] developed a model using a DL-enhanced MEIRL prediction trajectory model, which has higher accuracy than traditional DL models. Furthermore, Yongjie Wang et al. [53] also used the MEIRL-DL model and drone videos, addressing the efficiency of autonomous vehicles at midblock locations. The authors in refs. [52,53] concluded that MEIRL-DL has superior accuracy over traditional DL. In addition, the simulation of pedestrian–vehicle conflict was modeled using ML at signalized intersections as a strategic game [8]. Subsequently, a proactive approach to reducing the incidence of crashes and fatalities using AI and surrogate safety assessment was provided [54]. Additionally, a multiagent system was developed to model pedestrian–vehicle interaction [54].

Furthermore, CV-enhanced conflict detection and pedestrian simulation using automated video analysis were investigated [55]. As a result, a novel camera-invariant structure for predicting pedestrian crossing using trajectory data from CCTV footage was developed. It consists of Transformer-based models, graph convolutional networks (GCNs), and a hybrid Transformer +GCN approach [55]. The identification of risky midblock crossing behavior is also vital from an economic point of view. As a result, saving more lives and preventing more crashes from happening needs decision tools to detect the most hazardous locations to implement pedestrian control interventions [10,56]. Prakash S. et al. [57] developed pedestrian path change behavior models using binary models and ML. Song Kim et al. [58] used real-time videos, analyzed using ML, and then validated the output using gradient boosting and logistic regression. Bhagat S. et al. [59] concluded that traditional ML methods are superior to some DL models. They also deduced that the Albert Model achieved the highest accuracy with expert classification.

2.2. Reasons Behind Pedestrians' Illegal Crossing Behavior

Illegal crossing behavior exists in poorly designated pedestrian facilities, especially in developing countries. Developing countries usually neglect pedestrian precautions in roadway network design, giving greater importance to vehicles [8], unlike developed countries, which pay more attention to the most vulnerable road users (pedestrians, cyclists, etc.), who surpass vehicle users in importance [60]. Thus, developing countries suffer from higher fatality rates due to pedestrian crashes based on the Human Development Index (HDI) [61].

Challenges facing developing countries include limited funding, broken authority, insufficient data, and informal urban growth [60]. This endangers pedestrians, the most vulnerable road users, as it fosters illegal actions, which puts them in risky situations where conflicts with vehicles are likely to happen. However, illegal crossing behavior also exists in highly pedestrian-friendly road networks due to other reasons [49]. Pedestrians crossing at undesignated areas exposes them to fatal crash injuries [62,63]. Moreover, pedestrians often focus on time saving and the shortest distance from the boarding/alighting point to the final exact destination. Without properly designated crossings, they tend to engage in illegal behavior at black-spot locations, where there is a high probability of pedestrian accidents [63]. Furthermore, time pressure or being in a hurry encourages pedestrians to engage in NCPCB [1]. Signal timing, pedestrian amenities, and social and psychological factors, including impatience and group behavior, can influence jaywalking [8].

This paper categorizes factors influencing IPCB into eight types, as mentioned in Table 2:

- Demographics (age, gender, income, crossing groups, and walking patterns). The authors in refs. [18–24] concluded that males engage in IPCB more than females. Also, the crossing time was found to be affected by age and gender. In addition, the mean crossing speed for elderly pedestrians varies from 0.82 m/s to 1.37 m/s, and young and middle-aged pedestrians are noted to have crossing speeds ranging from 1.24 m/s to 2.04 m/s and from 1.37 m/s to 2.11 m/s, respectively. The middle-aged pedestrian category has a 60.1% higher likelihood of interrupted crossing than older and younger pedestrians. The male pedestrian category and the middle-aged pedestrian category are more likely to accept the smallest gap between the vehicles, showing the risky nature of their crossing behavior.
- Psychological factors, such as attitude toward risk, habit, subjective norms, perceived behavioral control, wrong perceptions, convenience, and time-saving motivation. The authors in refs. [25–27] concluded that IPCB was mainly affected by habit and attitude.
- Social impact, such as friends'/peers' perceptions, group size, and social acceptance of risky crossing. The authors in refs. [13,21,23] concluded that women's decisions are highly influenced by their own thoughts, while men's risky behavior is inspired by their friends' perceptions.
- Traffic characteristics, such as traffic volume, vehicle speed, number of lanes, presence of vehicles, gap size, headway, and vehicle type. The authors in refs. [13,19,22–24,28–33] concluded that the size of the vehicle has a significant influence on gap acceptance and crossing behavior of pedestrians. They also found that traffic volume, pedestrian red-light time, waiting time, vehicle illegal crossing behavior, and group crossing decreased the probability of violations by pedestrians.
- Road and environment, such as crosswalk type (e.g., zebra, raised, and signalized), crossing length, road width, intersection spacing, lighting, and illegal parking. The authors in refs. [19,24,25,28,30,31] concluded that traditional pedestrian safety measures such as speed cushions or roads narrowing to one lane have a superior impact than the other measures analyzed. Traffic lights or grade-separated solutions (footbridges or tunnels) are good measures for decreasing IPCB.

- Behaviors, such as waiting time, distraction (e.g., phone and headphones), crossing speed, rolling gap acceptance, and crossing in groups. The authors in refs. [18,19,22,24,29,31–33,64] concluded that the crossing time was influenced by gender, age, mobile phone use, clothing type, group crossing, crossing point, crossing path, and the presence of a vehicle. Moreover, the traffic light cycle is an important variable that improves the safety of pedestrian midblock crossings. In addition, parked cars at crosswalks affects the waiting and delay times of pedestrians. Regarding driver behavior, the models indicate that the number of lanes and lane width, crosswalks' width and length, pedestrian crossing time, vehicle speed, time headway, post-encroachment time, and roadside parking are the most significant factors influencing driver-yielding behavior.
- Policies, such as law enforcement, surveillance, road safety campaigns, and engineering measures (e.g., lane narrowing and speed reduction). The authors in refs. [25,30] concluded that these policies should be used to improve observed behavioral control.

Other researchers have identified several factors that affect pedestrians' illegal crossing behavior, such as pedestrian-related social factors like age, gender, employment status, educational level, and group crossing behavior [65–68]. Furthermore, researchers have investigated factors that influence where, how, and when pedestrians illegally cross in midblock sections near pedestrian crossing facilities. Pedestrian speed, waiting time, vehicle time, pedestrian group size, accepted gap size, and crossing nature are the major factors that cause conflict and higher interaction severity [50]. Moreover, mobile distraction is one of the main factors affecting NCPCB [41,69–71].

Moreover, crossing time, pedestrian crossing speed, and distance and time gap perception for safe road crossing within 25 m of overpasses in both directions are the main factors affecting IPCB [72]. Another trend focuses on physical and psychosocial factors, such as vehicle type, pedestrian existence, and traffic volume, which substantially impact IPCB [73]. Pedestrians tend to engage in IPCB when they perceive minimal risk or face a long waiting time [8]. "Self-enforcement feature" is the main factor affecting the use of overpass/underpass foot overbridges (FOBs), while the least cited factor is "Attractiveness" [74]. Moreover, the purpose of the trip, location, time, convenience, and comfort have a strong positive relationship with the use/non-use of underpasses/overpasses [74–76].

Table 1. Previous related studies focused on the application of ML and DL on the modeling of pedestrian crossing behavior.

Study	Objective and Model	Data	Results	Outcomes
Piyalungka S. et al. [52]	Predict pedestrian crossing behaviors at midblock crosswalks and improve safety for autonomous vehicles.	Pedestrians at unsignalized midblock crosswalks, Xi'an, China.	Deep MEIRL enhanced prediction of pedestrian trajectory versus traditional models.	Accuracy of pedestrian crossing behavior prediction.
Cai, Jun et al. [46]	Machine learning models, specifically SVM, can predict pedestrian crossing probabilities and speeds in smart cities.	Pedestrian crossing at signalized intersections in Chinese cities.	SVM predicted pedestrian crossing behavior with the highest accuracy among the tested models.	Pedestrian crossing prediction accuracy and speed.

Table 1. Cont.

Study	Objective and Model	Data	Results	Outcomes
Shaaban K. et al. [77]	Pedestrians illegally cross urban midblock roads by adding a factor of 1.25–1.5 to vehicle speeds before anticipating the gap.	Doha, Qatar.	Pedestrians add a factor of 1.25–1.5 to vehicle speed to judge gaps when crossing illegally.	Pedestrian gap acceptance behavior.
Yongjie Wang et al. [53]	Deep MEIRL and RL.	Drone video and trajectory features (distance, speed, and vehicle type).	Deep MEIRL outperforms MEIRL based on MAE and HD.	Addressing efficient movement at unsignalized midblock crosswalks for autonomous vehicles.
Shengqi Liu et al. [56]	Heatmaps, Association Rules, PCA, and clustering.	STATS19 dataset and spatial/behavioral data, UK.	Identification of high-risk behaviors/locations; infrastructure influences illegal crossing patterns.	Insights to prioritize safety measures that enhance pedestrian safety at unsignalized crossings.
Prakash S. et al. [57]	Binary logit and ML.	Field data, zebra/no-zebra marking, and waiting time.	Path changes are more likely without zebra marking; waiting time reduces path changing.	Waiting time has a negative effect on path-changing behavior, and crossing stages have a positive influence on path-changing behavior.
Dungar Singh1 et al. [47]	k-Nearest neighbors, artificial neural networks, and support vector machines.	A videographic survey, India.	Prediction of pedestrians based on random forest, extreme GB, and binary logit model achieved 81.72%, 77.19%, and 74.95%.	Support for infrastructure-to-vehicle interactions, negotiation of rolling pedestrian behavior, and improvement in pedestrian safety.
Md. Bayezid et al. [78]	CART, RF, XGBoost, and logistic regression.	Survey and crosswalk attributes.	RF best predicts crosswalk use; infrastructure and lighting are key factors.	Support for policymakers to develop more efficient traffic safety measures in Dhaka.
Madhar M. Taamneh et al. [48]	ANN, SVM, DT, and RF.	Video and demographic/spatial data.	RF is the most accurate; local infrastructure/traffic conditions influence compliance.	Measurable solution to improve pedestrian safety dynamically.
YOUNGGUN KIM et al. [55]	Transformer-based models, graph convolutional networks (GCNs), and a hybrid Transformer + GCN method.	Camera-invariant structure for predicting pedestrian crossing directions using trajectory data from CCTV footage.	The Transformer-based model achieved an accuracy of 94.10%, showing its effectiveness in capturing pedestrian intentions across diverse scenarios.	The geometric-invariant model ensures that the system is easily transferable across intersections by collecting fewer data.
Song-Kyoo Kim et al. [58]	Use of real-time video analysis and ML to alert pedestrians.	You Only Look Once algorithm using video footage.	Validation with gradient boosting and logistic regression.	Development of smart city initiatives that prioritize safety through advanced technological solutions.
Manoguid A. et al. [79]	Modified Faster RCNN, Squeeze and Excitation Network, Feature Pyramid Network, and Contrast Limited Adaptive Histogram Equalizer.	Philippines.	An improvement of the unmodified Faster RCNN architecture and Faster RCNN with a ResNet50 backbone.	Detects vehicles, but had difficulty-detecting pedestrians.

Table 1. Cont.

Study	Objective and Model	Data	Results	Outcomes
Bhagat S. et al. [59]	SVM, XGBoost, BERT Sentence Embeddings, BERT Word Embeddings, and Albert Model.	Iowa, USA.	1. Traditional ML methods exhibited superior overall performance compared with some DL methods. 2. The Albert Model achieved the highest efficiency with expert classifications and original tabular data.	Hybrid approaches combining automated classification with targeted expert review offer a methodology for improving crash data quality.
Eloğlu B. [10]	Decision-support tool comprising two complementary stages for footbridge planning and design with ML methods.	Ankara, Türkiye.	The best-performing model had an accuracy score of 0.92.	The necessity of constructing footbridges in the predicted locations was critically evaluated by the researcher.

Table 2. Factors affecting illegal pedestrian crossing behavior from the literature.

Factor Category	Studies	Detailed Factors
Demographics	[18–24]	Age, gender (males and middle-aged are more likely to commit a violation), group crossing, and walking patterns
Psychological factors	[25–27]	Attitude toward risk, habit, subjective norms, perceived behavioral control, wrong perceptions, convenience, time-saving motivation
Social impact	[13,21,23]	Friends’/peers’ perceptions, group size, and social acceptance of risky crossing
Traffic characteristics	[13,19,22–24,28–33]	Traffic volume, vehicle speed, number of lanes, presence of vehicles, gap size, headway, and vehicle type
Road and environment	[19,24,25,28,30,31]	Crosswalk type (e.g., zebra, raised, and signalized), crossing length, road width, intersection spacing, lighting, and illegal parking
Behavioral	[18,19,22,24,29,31–33,64]	Waiting time, distraction (e.g., phone and headphones), crossing speed, rolling gap acceptance, and crossing in groups
Policy	[25,30]	Law enforcement, surveillance, road safety campaigns, and engineering measures (e.g., lane narrowing and speed reduction)
Regional and cultural	[21,26]	Differences between developed and developing countries, and local norms

2.3. Actions Taken by Developed Countries to Enhance Safety and Decrease IPCB

Smart pedestrian traffic signals, pedestrian-only zones, raised crosswalks, traffic calming, and intelligent transport system applications are the most important countermeasures to ensure pedestrians’ safety [60]. In Sweden, Vision Zero aims at zero fatal or serious injuries, leading to half of the deaths on Swedish roads [80]. Moreover, AI and ML are integrated to predict and mitigate high-risk intersections [81]. Some authors recommend traffic calming measures in midblock sections and awareness campaigns to ensure pedestrian crossing safety maneuvers [73]. Other measures include addressing the consequences of

and common beliefs regarding IPCB and improving riders' awareness and risk perception through education and training [82].

As shown in Table 3, several actions and interventions have been conducted by countries/organizations to enhance traffic safety and decrease IPCB. These actions include midblock pedestrian signals (MPSs), rectangular rapid flashing beacons (RRFBs), pedestrian hybrid beacons (PHBs), active warning systems, road narrowing and speed cushions, raised crosswalks and medians, pavement markings and signage, enforcement and surveillance, and public awareness.

2.4. Models Developed for IPCB

Many researchers use logit models with all subtypes to model illegal pedestrian crossing behavior in midblock sections [50,63,83–87]. Other models, such as the Random Parameter Multinomial Logit Model with Heterogeneity, were developed to study hidden pedestrian behavior in a low-visibility environment and stressed crossing pedestrians [82].

Consequently, there is an urgent need to adopt a safety system approach (SSA) that integrates engineering measures with psychological strategies to help risky demographic groups better comply with legal crossing requirements [82]. On the other hand, Bayesian combined neural networks (BCNNs), permutation-invariant support vector machines (piSVM), and distributional auto-replicative random forest (ditARF) models have been developed to model a dilemma zone passive warning system to model jaywalking [88].

Other researchers have used an ordinal logit model to examine the probability of occurrence based on a vehicle-scaled risk indicator (VSRI) [50]. Other studies have focused on surveys, onsite observations, exploratory factor analysis (EFA), and binary logistic regression (BLR) for modeling IPCB [73]. The Theory of Planned Behavior was applied to understand factors affecting motorist crossing behavior in Vietnam, applying the structural equation modeling (SEM) approach [82]. A multiagent adversarial inverse reinforcement learning (MAAIRL) within a Markov game framework was developed to model pedestrian jaywalking, representing a dynamic decision-making context [8,54]. Moreover, smaller predicted time-to-collision (TTC) and post-encroachment time (PET) values, reduced minimum distances, and faster pedestrian movements in jaywalking compared with non-jaywalking scenarios have been found in urban areas [8].

Table 3. Actions to enhance safety and reduce IPCB.

Actions	Studies	Outcomes
Midblock Pedestrian Signals (MPSs)	[89]	Use of adaptive signal systems at midblock crossings substantially decreases pedestrian–vehicle conflicts and upgrades safety compared with other interventions
Rectangular Rapid Flashing Beacons (RRFBs)	[89–91]	Use of high-visibility flashing beacons at midblock crossings increases driver yielding rates to over 90%
Pedestrian Hybrid Beacons (PHBs)	[89,91]	Signalized crossings with pedestrian-activated phases lower conflicts and improve safety edges
Active Warning Systems	[28,92,93]	LED lights and variable message signs alert drivers to pedestrian presence, increasing yielding rates and reducing conflicts
Road Narrowing and Speed Cushions	[28]	Lane narrowing and installing speed cushions to reduce vehicle speeds at midblock crossings
Raised Crosswalks and Medians	[90]	Raised sidewalks and medians slow down vehicles, providing safe waiting areas for pedestrians

Table 3. *Cont.*

Actions	Studies	Outcomes
Pavement Markings and Signage	[28]	Use of red paint, anti-skid surfaces, and clear markings to increase crosswalk visibility and driver awareness
Enforcement and Surveillance	[94]	Deployment of police, ePolice, and cameras to monitor compliance and deter illegal crossing behavior and non-yielding
Public Awareness	[95]	Campaigns for pedestrians and drivers about safe crossing habits

3. Methodology

3.1. Research Framework

This section narrates the methodology considered by the researchers. This methodology is illustrated in Figure 1, which provides a sequential process to determine the best model. This model will investigate key variables affecting illegal pedestrian crossing behavior (IPCB) in midblock sections on a major roadway. First, a comprehensive review of related studies was conducted, and then the study area was determined based on the researchers' experience. Second, video-recorded data near midblock were collected to be analyzed through ML algorithms. Later, data were refined, read, screened, and filtered. Third, a heatmap of the Pearson correlation coefficient matrix was developed to investigate cross-correlation between explanatory variables affecting IPCB.

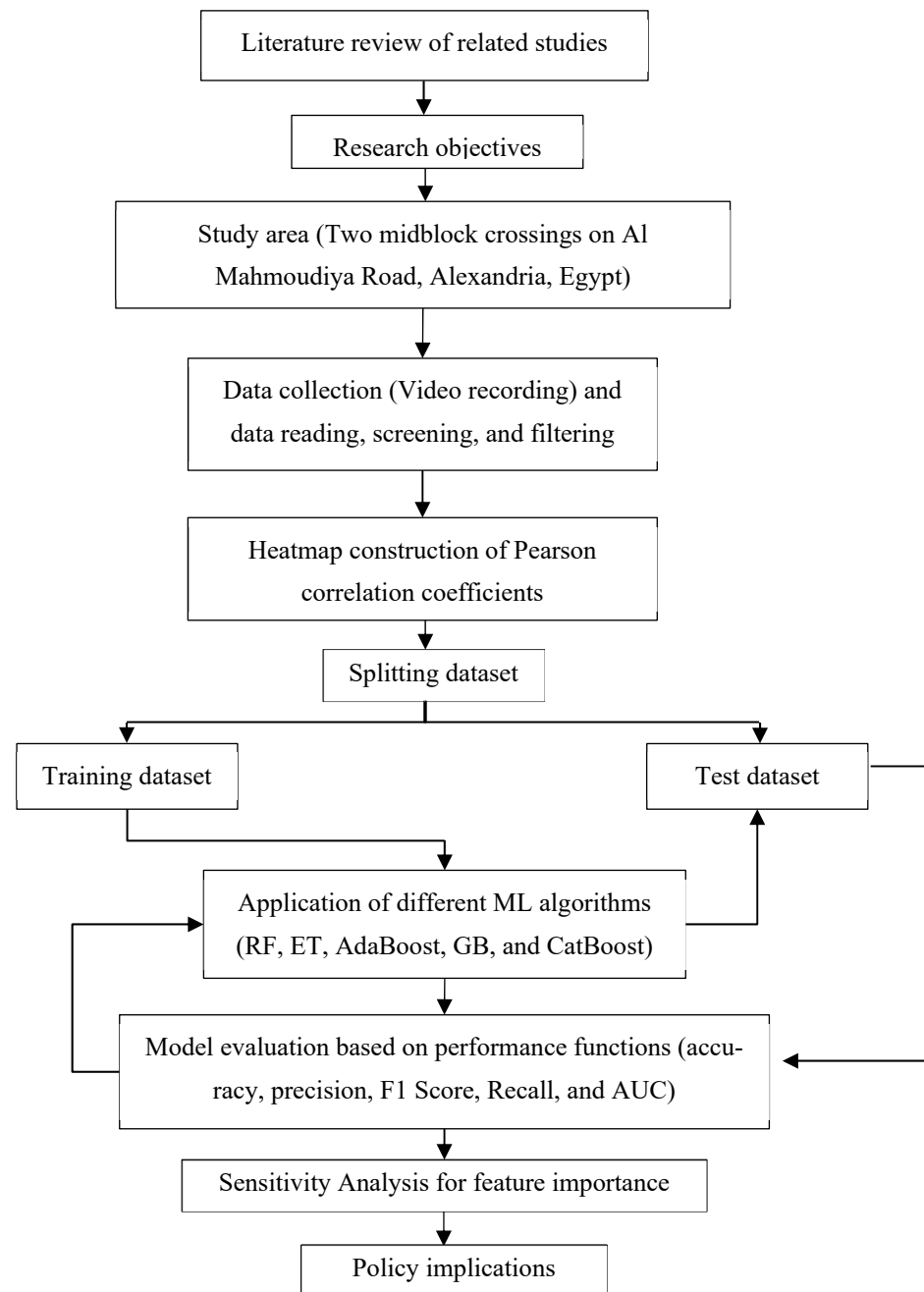


Figure 1. Research methodology.

This research study differentiates between two crossing behaviors (legal and illegal), where legal crossing behavior involves FOB usage and illegal crossing behavior denotes road crossing in midblock sections. Fourth, data were split into two main groups: the training dataset and the test dataset. Fifth, five ML algorithms were applied for model development (RF, ET, CatBoost, AdaBoost, and GB). These models were evaluated based on several performance functions such as accuracy, precision, AUC, F1 Score, and Recall. Finally, sensitivity analysis was conducted for feature importance investigations using SHAP. In other words, we quantified the contribution of each feature to model predictions with positive or negative impact, as well as the impact value. ML models have superior model accuracy over discrete choice models, providing deeper insights into IPCB investigation. This framework also contributes to model sensitivity, focusing on the most effective factors in model prediction. This sheds light on enhancing pedestrian safety in a sustainable built environment in a developing country.

3.2. Machine Learning Models

ML is a field of data science and AI that enables systems to learn and improve from provided data without surplus software intervention. Moreover, it is classified into these categories: supervised learning, unsupervised learning, semi-supervised learning, self-supervised learning, and reinforcement learning [96,97].

- **Supervised ML:** The model is trained on a labeled dataset [98], classified into the following algorithms:
 - Regression algorithms predict product values by discovering linear relationships (e.g., linear regression, RF, and GB).
 - Classification algorithms predict categorical outcome variables by labeling portions of input data (e.g., logistic regression, KNN, and SVM).
 - Naïve Bayes classifiers enable categorization for huge datasets (e.g., DT).
 - Neural networks replicate how the human brain works with neural networks, with a substantial number of linked nodes that can assist in tasks such as natural language translation, image recognition, speech identification, and image construction.
 - RF algorithms predict a value or category by combining the results from a few decision trees.
- **Unsupervised ML:** Algorithms such as Apriori, Gaussian mixture models (GMMs), and principal component analysis (PCA), which make inferences from unlabeled datasets, enabling pattern identification and predictive modeling. Here are the most used methods [99,100]:
 - K-means clustering assigns data points close to a given centroid into K groups, and K represents the cluster-based size and level of granularity.
 - Hierarchical clustering, including agglomerative and divisive clustering.
 - Probabilistic clustering solves density estimation by grouping data points based on their likelihood of belonging to a particular distribution.
- **Self-supervised ML:** Self-supervised learning (SSL) [101] enables models to train themselves on unlabeled data, instead of requiring massive annotated and/or labeled datasets.
- **Semi-supervised learning (SeSL):** It offers a combination of supervised and unsupervised learning and is trained on a tiny, labeled dataset and a large unlabeled dataset. The SeSL model [100] can use unsupervised learning to recognize data clusters, so it uses supervised learning to label the clusters. The following subsections narrate the techniques used in this paper for ML, such as RF, ET, AdaBoost, GB, DT, and cat boosting.

3.2.1. Random Forest (RF)

It is a supervised learning procedure combining the outcomes of multiple DTs to reach a single output, avoiding overfitting. It applies the bagging technique to create multiple replaced, randomized datasets [101]. Arunabha Banerjee et al. [102] used RF to model fourteen locations in Indian cities to study the use or non-use of foot overbridges (FOBs). They divided the data into 80% for training and 20% for testing. About 100 to 500 trees were tried in RF; the maximum tree depth was fixed at 40. Furthermore, four distinct scenarios were developed using RF. Training data model performance metrics included model accuracy, which ranged from 0.7 to 0.96; area under the curve (AUC), which ranged from 0.7 to 0.98; and mean square error (MSE), which ranged from 0.03 to 0.18. Moreover, for the test dataset, accuracy ranged from 0.74 to 0.96 and AUC from

0.73 to 0.97. Subasish Das et al. [103] applied RF to classify pedestrian crash types. Model accuracy reached 0.6 for the training and test datasets.

3.2.2. Extra Trees (ETs)

The conventional batch-mode supervised learning algorithm emphasizes learning scenarios defined by many numerical input variables and a singular target variable, which may be either categorical or numerical. Furthermore, it converts tree bagging into the ET algorithm in case of an infinite number of trees. As shown in Equation (1), an infinite ensemble of ETs presents an estimate in the form of [104]

$$\hat{y}(x) = \sum_{i_1}^N \dots \sum_{i_n=0}^N I_{(i_1, \dots, i_n)}(x) \sum_{x \in \{x_1, \dots, x_n\}} \lambda_{(i_1, \dots, i_n)}^x \prod_{x_j \in x} x_j \quad (1)$$

where the real-valued parameters $\lambda_{(i_1, \dots, i_n)}^x$ depends on input x_i and output y_i .

For a one-dimensional input space, the model is transformed into a piecewise linear model, as shown in Equation (2).

$$\hat{y}(x) = \sum_{i_1=0}^N I_{(i_1)}(x) \sum_{x \in \{x_1\}} \lambda_{(i_1)}^x \prod_{x_j \in x} x_j \quad (2)$$

3.2.3. AdaBoost

AdaBoost combines weak learning algorithms frequently in a sequence slightly more reliable than random guessing that can be “boosted” into a randomly accurate “strong” learning algorithm. It can also be a technique acting as functional gradient descent, as a logistic regression estimate, and as a repetitive game-playing algorithm. It has the potential for success since it has a “good” base learner with a moral sequence of base classifiers [105].

3.2.4. Gradient Boosting (GB)

It constructs the new base learners to be highly correlated with the negative gradient of the loss function, linked to the entire ensemble [106]. The authors in ref. [107] narrate the first attempt at a greedy function estimate, later enhanced and updated by others. Furthermore, this algorithm depends on the choice of the base learner model and the loss function, as well as negative gradient computation, followed by fitting the new learner and then finding the best gradient descent step. Finally, this algorithm involves updating the function estimate [108].

Banerjee et al. [102] used the GB algorithm to model the use or non-use of pedestrian FOBs in Indian cities. They proved that GB is a better algorithm than RF and the generalized linear model in terms of accuracy of training and test data. Subasish Das et al. [103] developed the Pedestrian and Bicycle Crash Analysis Tool (PBCAT). Furthermore, they used an extreme GB algorithm to catalogue crash types from the formless textual data of pedestrian crash locations in Texas, USA. They proved that the XGBoost classifier has higher accuracy than RF and SVMs, with 77% for trained data and 72% for test data.

Researchers developed a categorization model with respectable accuracy using the Shapley (SHAP) additive explanation explainable ML approach. Moreover, a binary classification task is formulated for legal and illegal crossing behavior using/not using existing FOBs. Basically, this algorithm is a summation of a group of weak learners. Furthermore, the data are split into 25% for testing and 75% for training. The following equation shows the total loss minimization function [109]:

$$\text{Min } l^{(t)} = \min(\sum_{i=1}^n l(y_i, \hat{y}_i^{(t)}) + \sum_{i=1}^T (\gamma T + 0.5\lambda \sum_{j=1}^T w_j^2)) \quad (3)$$

where $l^{(t)}$ is the chosen loss function, γ is the parameter of the penalty for complexity, γ is the degree of the regularization of function f , T stands for the number of leaves of the tree model, and w_j is the weight of the j th leaf of the tree model.

3.2.5. Decision Tree (DT)

It is used for building classification models to predict the possible value of a specific known attribute based on the values of input attributes. It is represented by a direct graph with nodes (input attributes with several branches) and leaves (target attribute value). Entropy is calculated in Equation (4) [48].

$$E(S) = - \sum_{i=1}^c p_i * \log_2(p_i) \quad (4)$$

where S is the decision tree building dataset, c is the target attribute's number of classes, and P_i is the instance ratio of class I to the total dataset instances. The IG for feature A is calculated as shown in Equation (5):

$$IG(S, A) = E(S) - \sum_{v=1}^n \frac{|S_v|}{|S|} * E(S_v) \quad (5)$$

where n is the value's feature, S_v denotes the instances, $|S_v|$ the total number of instances in $|S_v|$, and $|S|$ is the total number of instances in S .

3.2.6. CatBoost

This algorithm is an application of binary DT as a base forecaster. The following equation clarifies the learning task's goal to train a function H [110]:

$$l(H) := E L(y, H(x)) \quad (6)$$

where L is a smooth loss function and (x, y) is a test data point from the dataset.

3.3. Model Validation

To create two separate datasets (training—75%; independent test—25%), the dataset was first randomly partitioned into two subsets. The training dataset was used exclusively for model fitting and hyperparameter tuning. A 10-fold cross-validation technique was used to both optimize the model parameters and determine the model's in-sample predictive performance.

Upon selecting the final model, it was assessed on the previously held-out independent test dataset, which had not been used at any stage of model training, tuning, or selection. By using this two-stage evaluation process, cross-validation supports model selection, while the independent test dataset provides an unbiased estimate of the model's out-of-sample predictive performance.

3.4. Case Study and Data Collection

In Egypt, frequent non-compliant pedestrian crossing behavior is noticed and noted as a crossing violation. Thus, this behavior contributes to a widespread increase in pedestrian injuries and fatalities. Dense traffic volumes and higher crossing flows exacerbate the problem, especially in high-speed major arterials in urban areas [111]. These issues have created an environment where pedestrians are at even greater risk. Most pedestrians in Egypt experience real danger, as many feel compelled to cross illegally, avoiding the use of legal means. This underscores the importance of implementing measures that have already proven effective elsewhere in reducing such risky behavior. When pedestrian crossings are

safer and easier to use, people tend to rely on them more, and the rate of illegal crossing behavior declines accordingly.

In Egypt, in 2024, based on the annual bulletin results of cars and train accidents, about 2125 pedestrians (crossing the road) died, which indicates that they are the most vulnerable road users [112], while vehicle passengers or drivers (the passenger) came in the last place with a number of 1174 deaths in 2024 [112].

Also, 20752 pedestrians were injured out of 76362 total injuries in Egypt [112]. Alexandria had a considerable share about 3.9% of total Egyptian injuries, out of which about 37.7% were pedestrians. Moreover, 7.3% of total Egyptian road accident deaths were from Alexandria. In addition, about 66% of accident deaths in Alexandria were pedestrians. Contrarily, Alexandria suffers from governmental ignorance, as it only has 1.1 ambulance centers per 100,000 inhabitants and 0.6 ambulances per 25,000 inhabitants, but it is the second largest city in Egypt [112]. Thus, this situation underscores the urgent need to study the causes of pedestrian accidents and illegal crossing behavior [104–106].

Alexandria has ongoing traffic congestion challenges, and some mitigation efforts may be challenged by unsafe pedestrian behavior [113]. Thus, the city introduced several safety measures, such as pedestrian bridges, tunnels, and signalized crossings, to reduce fatal and serious crashes. Despite the notable efforts, illegal crossing behavior is still widely common at the observed locations, even along high-speed corridors such as Al-Mahmoudiya Road. Two midblock crossings along Al-Mahmoudiya Road were selected as realistic case studies. These locations experience severely risky pedestrian crossing behavior, which underscores the urgent need for action.

Site 1 is in Hadara, and Site 2 is in the Hagar Nawatiya area along Al-Mahmoudiya Road, as illustrated in Figure 2. Item (a) in Figure 2 shows the location of Egypt in the middle east, while item (b) indicates the location of Alexandria city in Egypt, and item (c) indicates Alexandria boundary, while item (d) illustrates the location of the study area location in Alexandria city. The two sites captured in the video recordings at Site 1 and Site 2 are illustrated in Figures 3 and 4, respectively. The concrete narrow road barrier fails to hinder illegal pedestrian crossing behavior. Moreover, both study sites experience high traffic rates, with a higher speed limit, enhancing IPCB-related accidents.

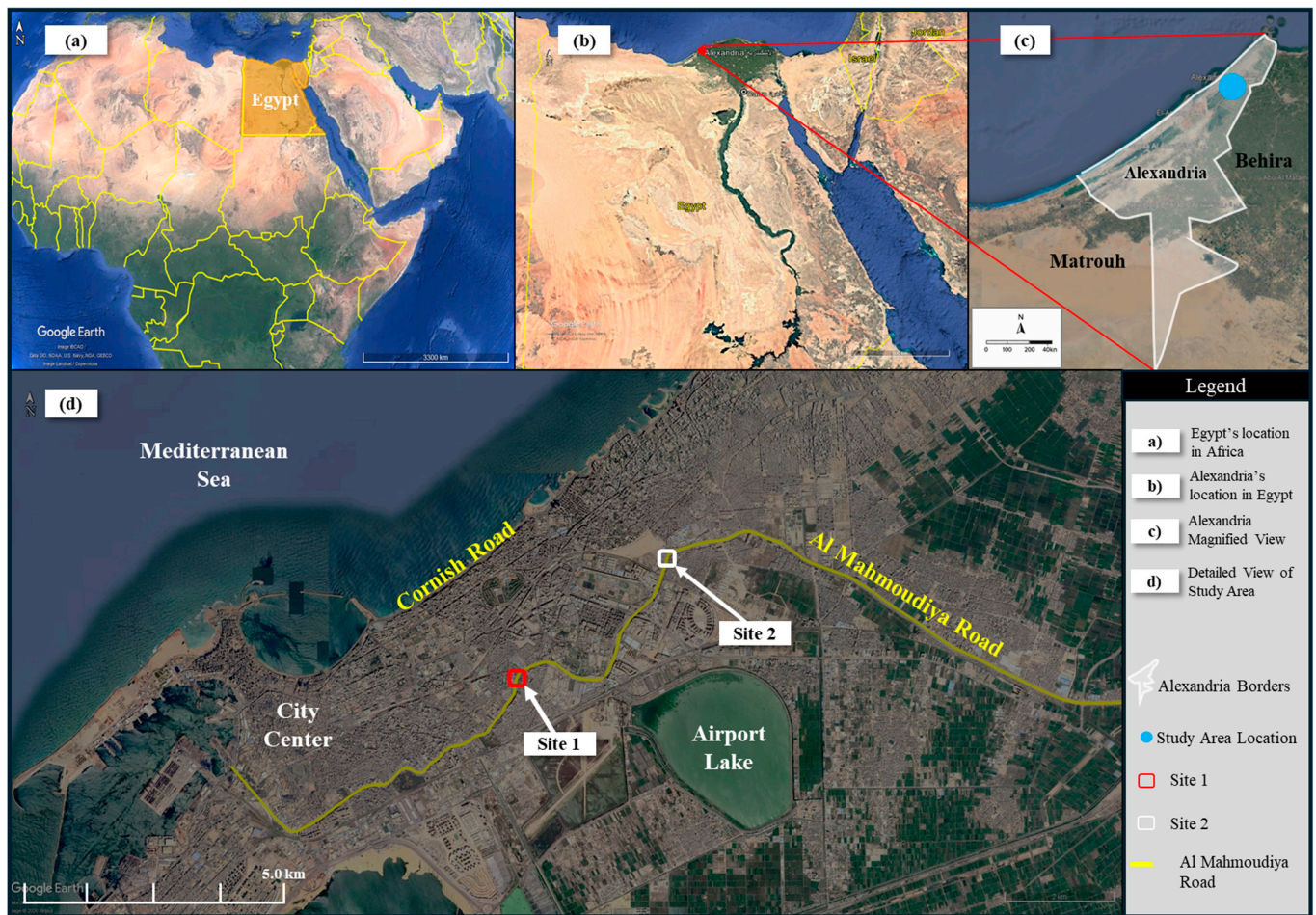


Figure 2. Locations of the selected sites along Al-Mahmoudiya Road in Alexandria, Egypt.

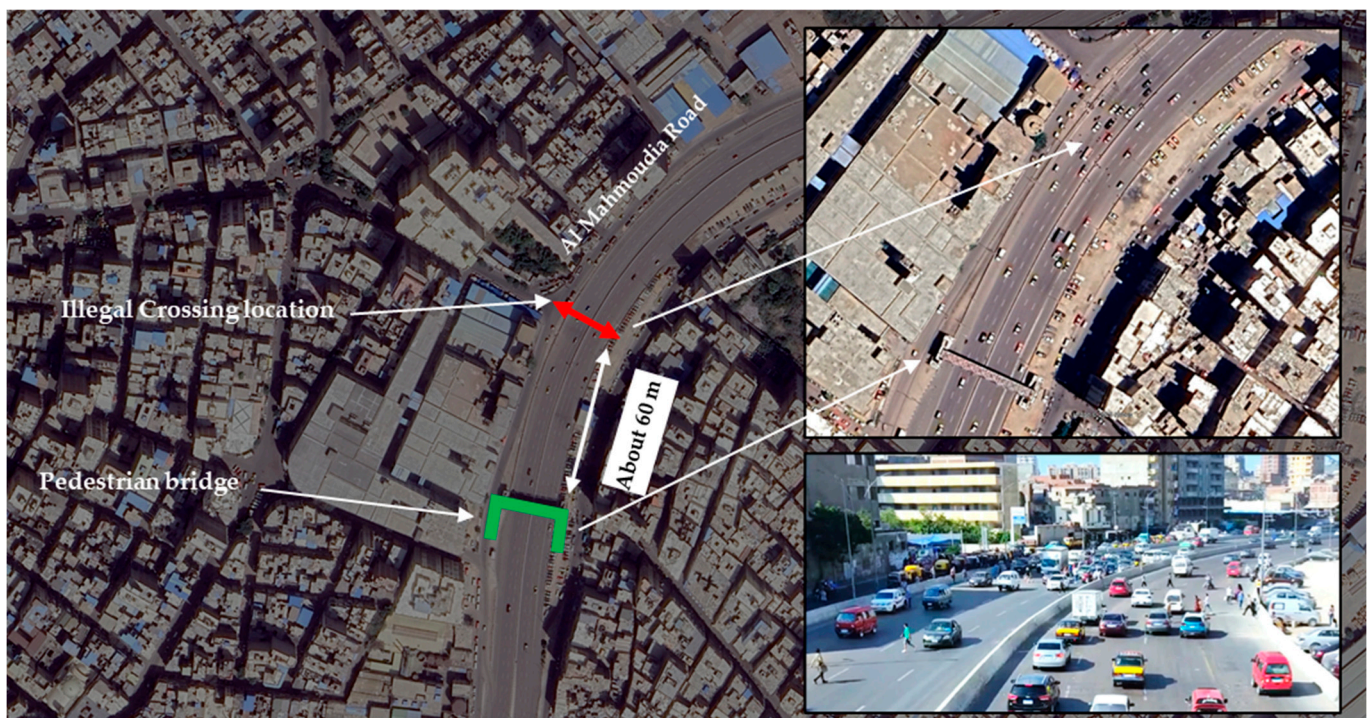


Figure 3. The characteristics and a screenshot from the recorded videos at Site 1.

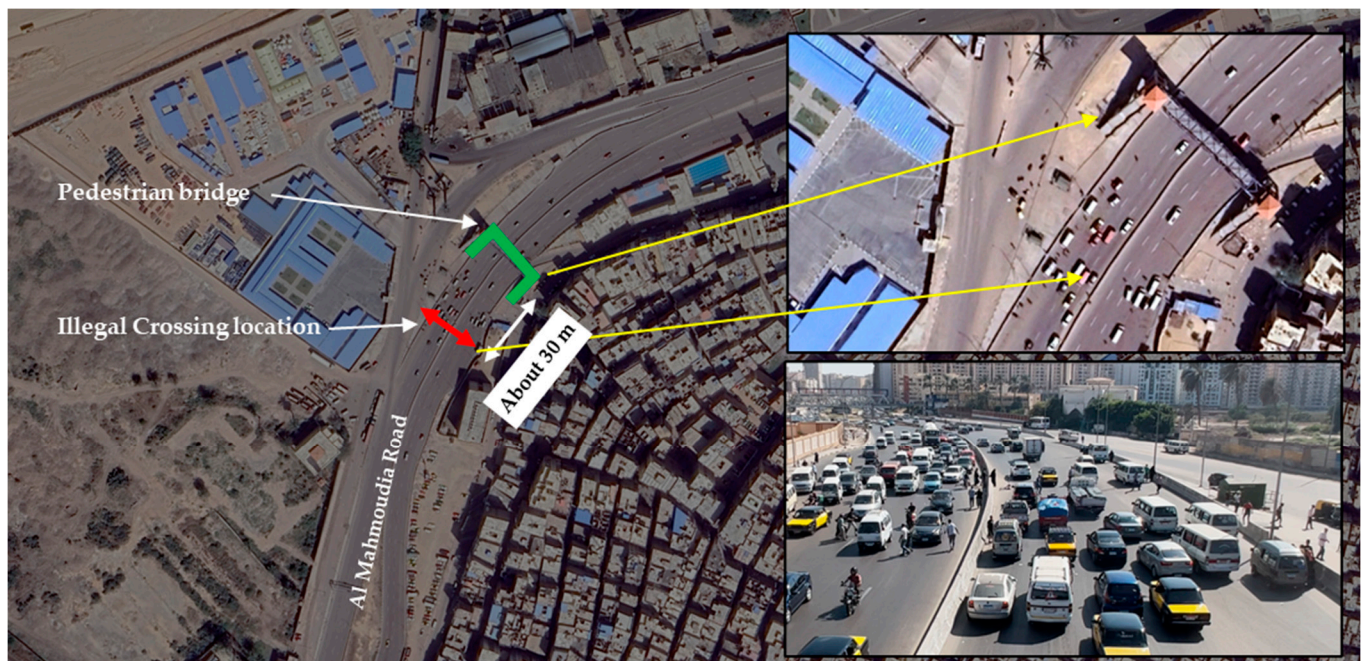


Figure 4. The characteristics and a screenshot from the recorded videos at Site 2.

About 2400 individual pedestrian data were extracted from two sites using video tracking in October 2023, with a detailed evaluation of pedestrian crossing behavior. The recordings reveal the exact waiting time and interactions with vehicular traffic in the background of the video. A structured observational survey was incorporated to document pedestrian crossing behavior and was conducted through manual documentation by trained field coders and subsequent video observations. The observations were made during peak hour to account for the pedestrian interactions under heightened stress. The recorded data were segmented into demographic characteristics: gender and age group; crossing behavior: legal vs. illegal; situational factors: use of cell phones while crossing and group size; risk signs: number of crossing attempts; and temporal features: waiting time, time to cross, and total time crossing.

Each pedestrian was assigned a distinct Pedestrian ID and coded consistently across all observations in the dataset. Moreover, key performance measures were calculated, including average pedestrian waiting time, crossing time, and vehicle–pedestrian interaction.

The model includes nine independent variables: gender, age, cell phone usage during crossing, pedestrian group size, number of crossing maneuvers attempted, the presence of children accompanying the pedestrian, conflicts with vehicles, pedestrian waiting time (in seconds), and pedestrian crossing time (in seconds). Key findings from descriptive statistics are shown in Figure 5a–i. Figure 5a indicates the gender distribution. Figure 5b illustrates the location distribution. Age distribution is highlighted in Figure 5c, while the split of the data according to children accompanied with the pedestrian during crossing is presented in Figure 5d. The usage of cell phone analysis during crossing is described in Figure 5e. The analysis of platoon crossing is demonstrated in Figure 5f. Figure 5g illustrates the number of crossing trials conducted by the pedestrian before successfully completing the road crossing. Figure 5h,i presents the analysis of total crossing time (including waiting and crossing time) and the crossing time, respectively. The data analysis are summarized as follows:

- A total of 80% of the pedestrians showed illegal crossing behavior.

- A total of 77% of pedestrians at the Hadra midblock crossing crossed illegally, while 83% of pedestrians at Hagar Nawatiya crossed illegally, showing the worse FOB accessibility of the second crossing.
- A total of 58% of the pedestrians were male, while 42% were female.
- A total of 80% of males crossed illegally, and so did females.
- The largest age group was 20 to 40 years old, representing 63% of the sample, 80% of whom engaged in illegal crossing behavior, showing the risky choice tendency of this age group. The same manners were found in the age group 40–60 Y, with 95% of them crossing illegally.
- A total of 6% of pedestrians were using a cell phone while crossing. Moreover, 84% of them crossed illegally. Furthermore, about 80% of pedestrian crossing without using a cell phone crossed illegally, indicating that cell phones are not the main reason for risky behavior.
- A total of 61% of pedestrians crossed the road individually, while 25% crossed in pairs, and 14% crossed in groups of three or more. About 81%, 78%, and 80% of individual pedestrian groups consisting of one person, two persons, and three or more persons crossed illegally, respectively. This pattern shows that the pedestrian has the intention to cross illegally for various reasons.
- A total of 11% of pedestrians had a child with them while crossing. Almost 80% of them engaged in illegal crossing behavior.
- The average waiting time before crossing illegally was approximately 4.88 s.
- The average crossing time was 34.5 s for crossing illegally, compared with 44.6 s for crossing legally via pedestrian bridges.

Participants' age group and gender were collected through direct self-report rather than observer estimation or inference. Respondents selected their age category and gender from a predefined list of options, thereby eliminating subjective classification by researchers and reducing the risk of systematic misclassification. Records with missing or inconsistent demographic information were excluded from the analysis. Accordingly, the dataset reflects age and gender based on participants' self-identification rather than external assumptions, which minimizes classification bias and enhances the reliability of age and gender as explanatory variables.

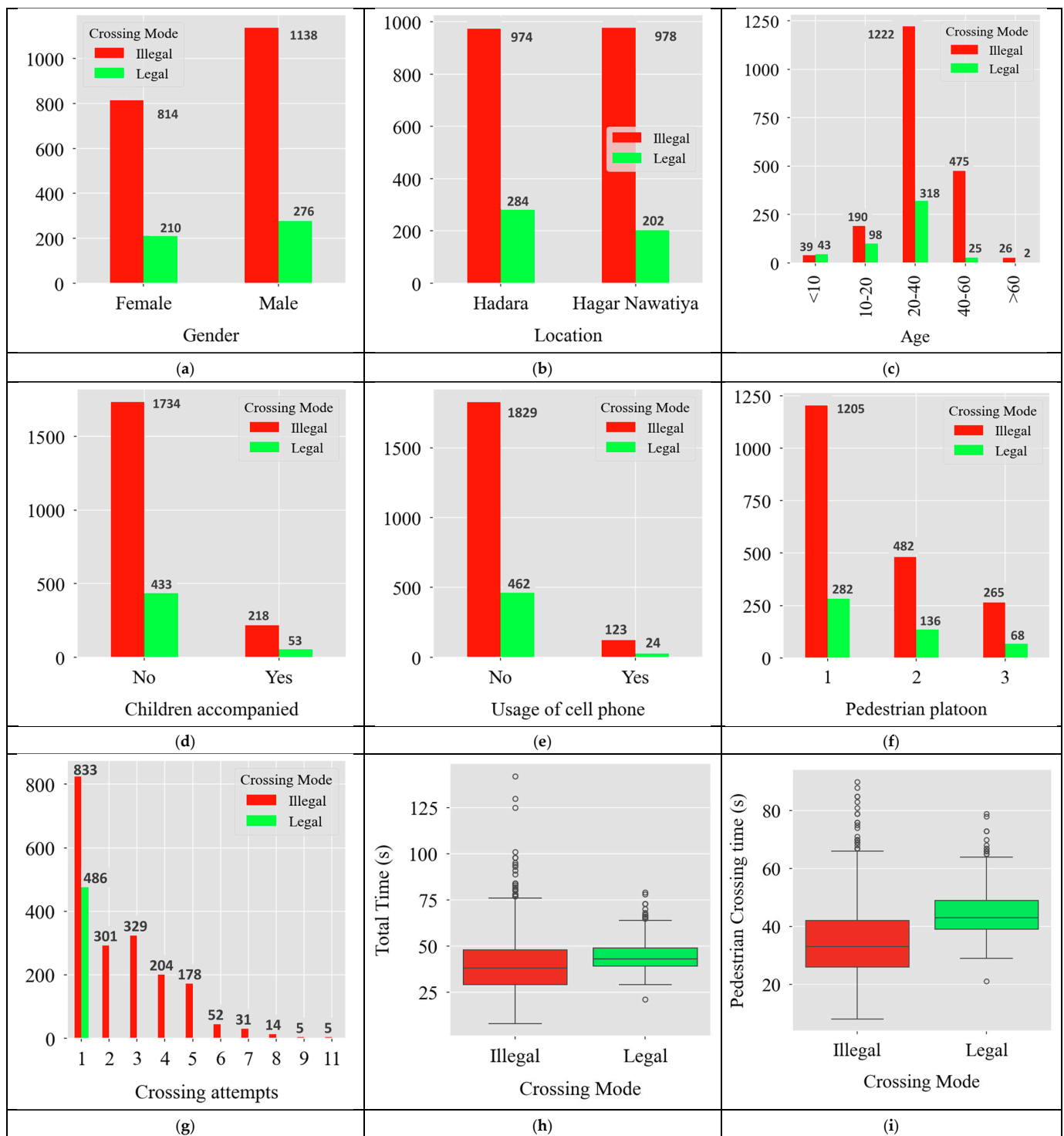


Figure 5. Analysis of variables versus legal or illegal crossing behavior.

All behavioral variables were defined before coding to ensure consistency and replicability in both data collection and analysis. The term “crossing attempt” was defined as any instance in which a pedestrian moved from the sidewalk into the roadway with the intention of crossing, including unsuccessful attempts. A “successful crossing” was recorded when a pedestrian reached the opposite curb or a designated refuge area without external help. The term “conflict” referred to an interaction between a pedestrian and a vehicle in which the two entities approached one another closely in time and space, such that one or both engaged in observable evasive behavior. This included driver responses,

such as braking, abrupt deceleration, swerving, or stopping, as well as pedestrian responses, including hesitation, retreat, acceleration, or evasive movement to avoid collision.

The dependent variable was defined as the crossing mode (legal or illegal). The number of crossing attempts refers to the observed attempts by a pedestrian trying to cross between conflicting vehicles before finally deciding to cross the road after finding an acceptable gap from the pedestrian's perspective. This variable was included in the model as a continuous variable. This episode-based classification ensures that the model captures the decision process associated with each crossing event rather than aggregating behavior across persons, which would mask short-term situational effects such as gap availability, traffic flow, and time pressure.

3.5. Representativeness and Generalizability of the Case Study

This research study focused on two midblock locations along the Al-Mahmoudiya Road corridor in Alexandria. These sites were selected because they showed considerable pedestrian risky crossing behavior, but they share key characteristics with many midblock locations in large urban areas across low- and middle-income countries. These shared characteristics include (a) high vehicular traffic volumes, (b) elevated vehicle speeds, (c) limited formal at-grade pedestrian infrastructure, (d) strong pedestrian demand generated by surrounding mixed land uses (e.g., shopping, banking, dining, and daily errands), and (e) pressure to minimize waiting time and walking effort.

However, several site-specific attributes may differ across locations and could explain variations in observed crossing behavior, such as the placement of pedestrian bridges, roadway width, and the presence or continuity of median barriers. These physical features influence perceived convenience and accessibility and may shape pedestrians' crossing decisions differently across different contexts.

The research results should be interpreted as behaviorally representative of midblock crossing in high-density, high-speed urban arterials characterized by high pedestrian demand and significant perceived considerations associated with the physical effort required to use grade-separation crossing facilities.

Therefore, this study offers insights into the key behavioral mechanisms that influence pedestrians' decisions to cross outside designated facilities under constrained urban conditions, particularly those related to time pressure, exposure duration, and physical effort needed for grade-separation crossings.

4. Modeling Results

First, the researchers studied the intercorrelation between the variables measured to determine which among them have the highest importance to be included in the models. Thus, a Pearson correlation matrix was developed, as shown in the heatmap in Figure 6. This heatmap shows that pedestrian crossing time and total crossing time are the most important variables (Pearson correlation coefficient of 0.95), followed by the number of crossing attempts and pedestrian waiting time to cross, to be included in any developed model.

Data were divided into two parts, training data (representing 75% of the data), and test data (representing 25% of the data), for model training using several classifiers. The prediction models were assessed using a confusion matrix. It involved four class occasions (true positive (TP), true negative (TN), false positive (FP), and false negative (FN)). Furthermore, the accuracy rate is one of the most important measures to evaluate a classifier's performance. It represents the ratio of correctly grouped occurrences with respect to all occurrences, according to Equation (7). Moreover, another important measure is Recall, which represents the ratio of occurrences that have been correctly categorized as positive accord-

ing to all positive occurrences and can be calculated as mentioned in Equation (8), while precision and F1 Score can be estimated with Equations (9) and (10), respectively [48,109].

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

$$F1 \text{ score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (10)$$

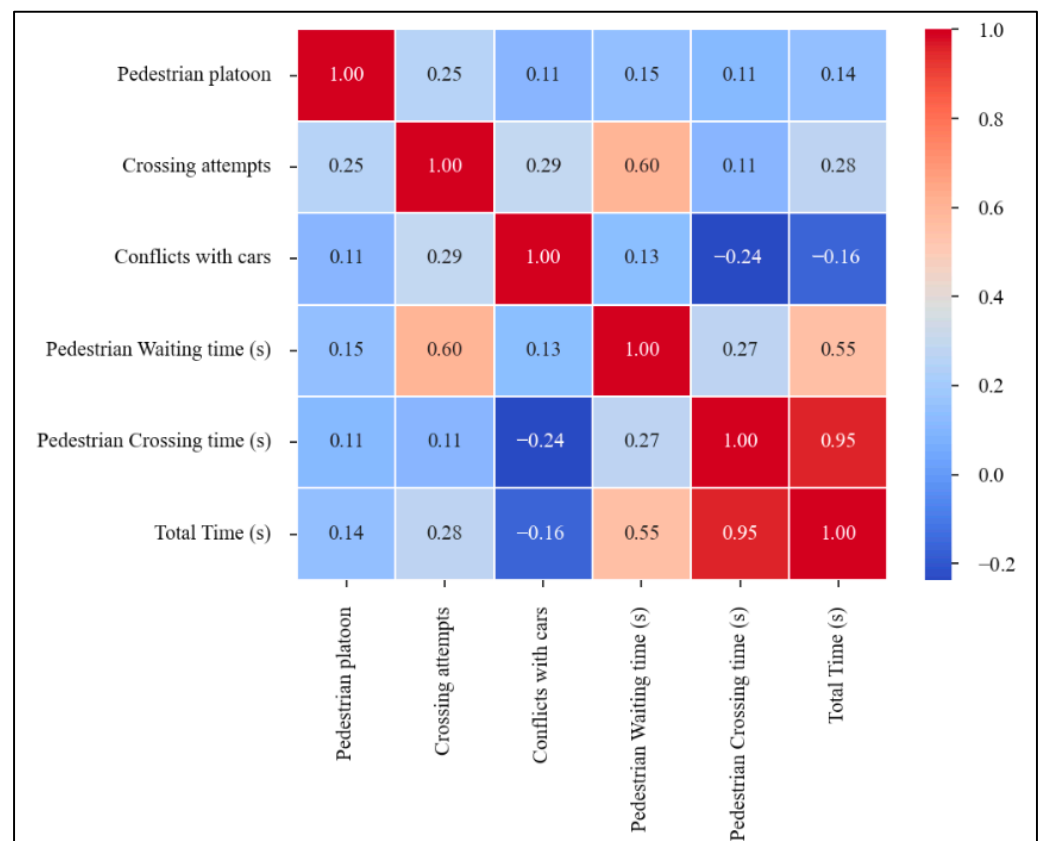


Figure 6. Heatmap of Pearson correlation coefficients.

Model accuracy is one of the most important measures for model assessment. Table 4 shows the accuracy and F1 Score of different ML classifiers. It is worth noting that LGBM, Bagging, LinearSVC, calibrated classifier, and DT have the highest model accuracy and F1 Score, with almost 0.97.

Table 4. Machine learning models' assessment.

Model	Accuracy	F1 Score
LGBM Classifier	0.97	0.97
SGD Classifier	0.96	0.96
Bagging Classifier	0.97	0.97
Linear SVC	0.97	0.97
Calibrated Classifier CV	0.97	0.97
Decision Tree Classifier	0.97	0.97

Table 4. *Cont.*

Model	Accuracy	F1 Score
Quadratic Discriminant Analysis	0.92	0.93
SVC	0.95	0.95
Logistic Regression	0.96	0.96
Perceptron	0.95	0.95
Random Forest Classifier	0.93	0.93
Linear Discriminant Analysis	0.93	0.93
Extra Tree Classifier	0.91	0.91
K Neighbors Classifier	0.9	0.9
Label Propagation	0.89	0.89
Label Spreading	0.89	0.88
Ridge Classifier	0.9	0.9
Ridge Classifier CV	0.9	0.9
Nearest Centroid	0.69	0.72

Furthermore, Table 5 shows the model results for the best estimators. It is noted that RF, ETs, and DT have the highest training accuracy, with 0.99, but GBoost and CatBoost have the highest accuracy for testing models to estimate the likelihood of pedestrian crossing behavior in midblock sections. In addition, precision, as mentioned above, refers to model accuracy and effectiveness. Therefore, DT and ETs have the highest training precision, while CatBoost has the highest test precision.

Table 5. Machine learning models' results.

	Classifier	Training Acc	Test Acc	Training F1 Score	Test F1 Score	Training Precision	Test Precision	Training Recall	Test Recall	Training Acc	Test Acc
1	RF	0.99	0.94	0.98	0.91	0.97	0.92	0.99	0.9	0.99	0.9
2	ETs	0.99	0.92	0.98	0.88	0.98	0.9	0.99	0.86	0.99	0.86
3	AdaBoost	0.86	0.83	0.81	0.78	0.78	0.76	0.87	0.84	0.87	0.84
4	GBoost	0.97	0.97	0.96	0.95	0.94	0.93	0.98	0.97	0.98	0.97
5	DT	0.99	0.96	0.98	0.94	0.98	0.93	0.99	0.94	0.99	0.94
6	CatBoost	0.98	0.97	0.98	0.96	0.96	0.94	0.99	0.97	0.99	0.97

The ROC curves were created by translating the model-predicted probabilities into class labels using different thresholds. Moreover, the AUC (area under the curve) is a metric to quantify a model's performance. Each fold gives an AUC value, which is also illustrated on the label of each ROC curve. Finally, Figures 7 and 8 show different ML models calculated for the training and test sets. For the training set, RF, DT, and CatBoost are the most suitable models that have the highest AUC value (0.99), while AdaBoost has the lowest AUC value (0.87). Moreover, for the test set, CatBoost and GBoost have the highest AUC value (0.97). All developed models have a consistent split ratio (25% for test data) and the same number of folds. More folds will determine the standard deviation of the average AUC value for each model type. The AUC model variation for the same type of model under different seeds demonstrates the initial conditions' impact and model randomness in the training and validation process.

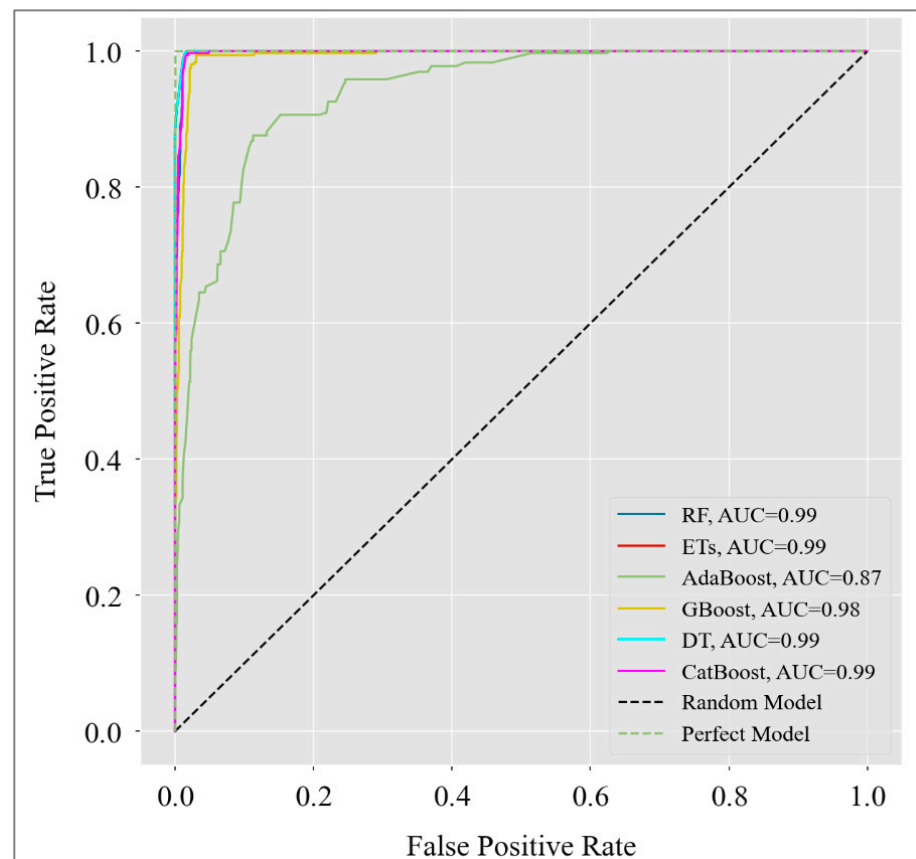


Figure 7. ROC curve for training data.

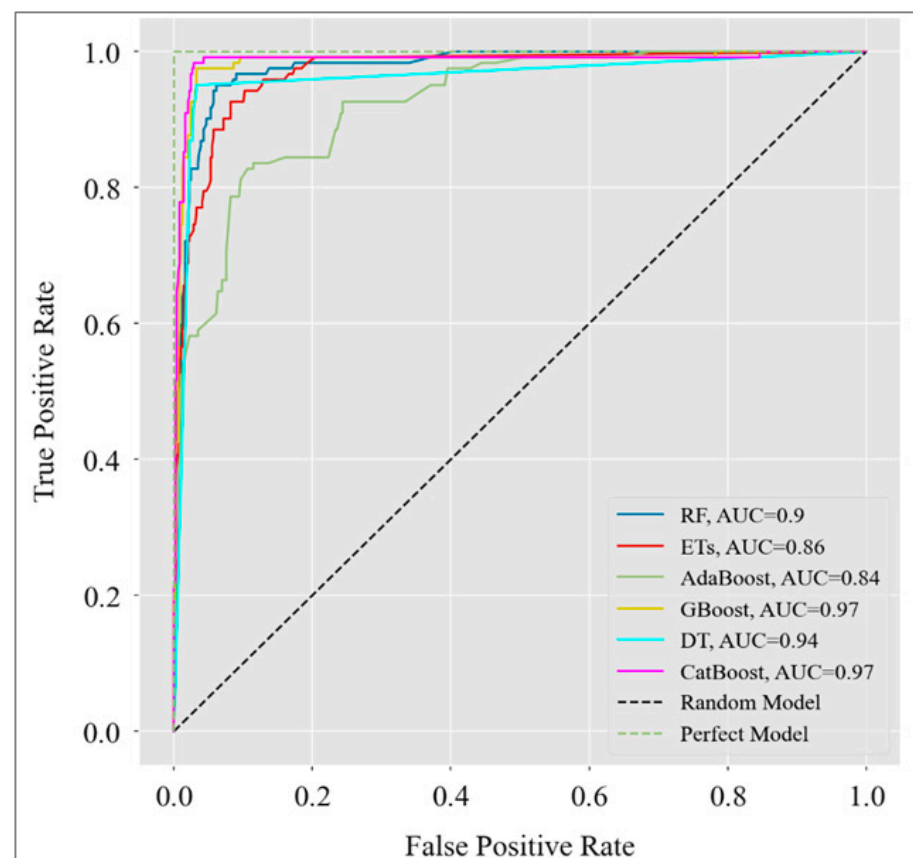


Figure 8. ROC curve for test data.

Sensitivity Analysis and Feature Importance

Permutation importance is a technique for evaluating the influence of a variable on the efficacy of predictive models, fundamentally relying on the concept of randomizing each feature's value and analyzing the resultant impact on model performance. Other feature importance techniques, such as permutation importance, are model-agnostic, indicating their applicability across any model. The model is first trained on the original dataset, followed by an evaluation of its performance.

Subsequently, for each variable, the values inside that variable column are randomly permuted while maintaining the integrity of the other variables, and the performance is re-evaluated to document the variation in model efficacy. Furthermore, the permutation relevance of a variable is often defined as the disparity between the original and after-shuffling performance.

To mitigate the influence of randomness, it is necessary to repeat this operation numerous times, and the mean variation in performance is regarded as the permutation relevance of that variable. After model evaluation, SHAP is used to quantify the contribution of each feature to model predictions.

Feature importance was ranked within the CatBoost model, as it was the best-performing model with the highest AUC, ROC, accuracy, and precision.

As shown in Figure 9, pedestrian crossing time (s) has the highest SHAP value (almost 3), followed by total time (s), (with a SHAP value of almost 2) and age (SHAP of almost 0.6), indicating that pedestrian crossing time (s) has the most substantial impact on model predictive outcomes. As shown in Figure 10, pedestrian crossing time (s) has the highest impact on pushing the model prediction towards a higher class (e.g., legal crossing behavior). Contrarily, total time and age category (40 to 60 years) have the highest negative impact on the model, which pushes the model towards the lower class (e.g., illegal crossing behavior).

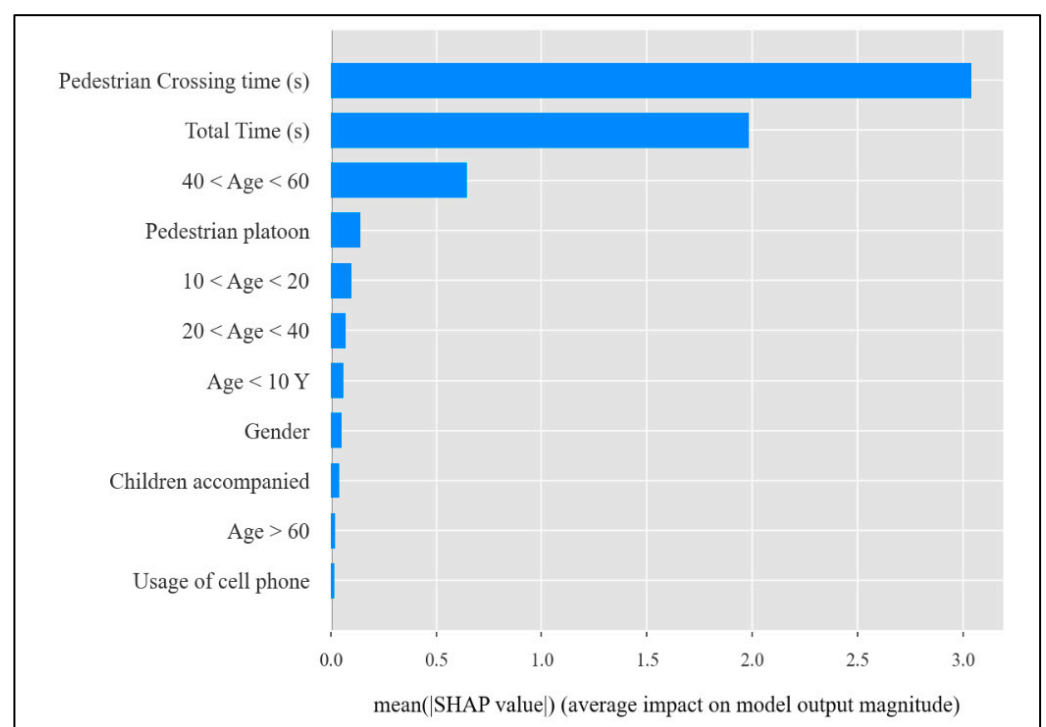


Figure 9. SHAP for different the variables considered (CatBoost model).

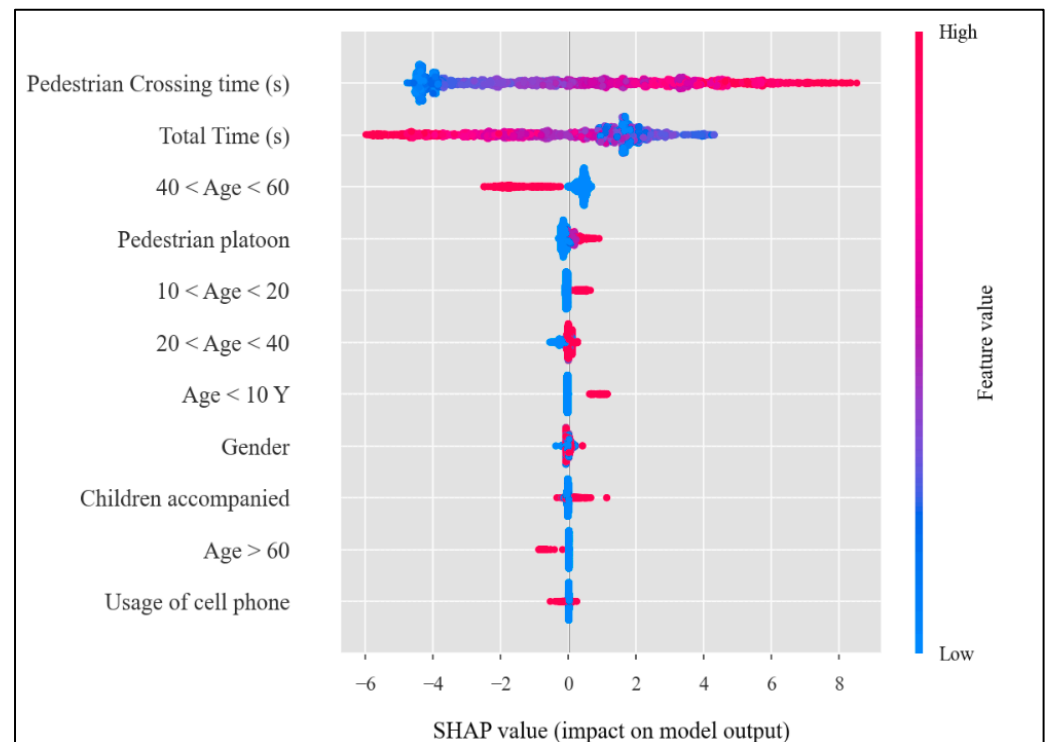


Figure 10. Sensitivity analysis of variables using SHAP (CatBoost model).

5. Discussion and Policy Implications

Based on the results, a deeper understanding and actionable insights into pedestrians' crossing behavior can obviously be gained. Different observations, e.g., high traffic speed and few crossing opportunities, were examined. Moreover, the case when high urban density creates an obstacle to crossing was also considered. The CatBoost and gradient boosting results show that crossing outside pedestrian crossings is not simply an impulsive decision. Compared with other similar studies, as shown in Table 6, the results for XGBoost reach 77%, and those for RF reach about 81%, indicating results similar to this paper.

Table 6. Comparison of results between different models.

STUDY	Model Used	Results
Albert S. [114]	Stepwise multiple linear regression (MLR) vs. artificial neural network (ANN) for gap-acceptance model	ANN, $R^2 = 0.79$ (better than MLR's $R^2 = 0.52$)
Cai C. et al. [115]	Reinforcement learning with post-encroachment time	Success rate > 0.8
Zhang C et al. [116]	Neural networks, random forests, and other ML + unsupervised clustering	Accuracy of over 90% was reported for many tasks
Refs. [41,46,48,117,118]	Random forest, XGBoost, binary logit, SVM, k-NN, Convolutional Neural Networks (CNN), the light GBM, and ANN were also tested	RF, ~81.7%; XGBoost, ~77.2%; logit, ~75.0%, CNN, 94.93%, the light GBM, 80%,
Jin C. et al. [119]	XGB and Light Gradient Boosting (LGB)	Success rate > 0.8
Younggun Kim et al. [55]	GCN and Transformer	Accuracy ranged from 81% to 87%
Song-Kyoo Kim et al. [58]	Advanced ML-based Pedestrian Crossing Alert System	Accuracy (GB = 0.9, LogR = 94%, RFR = 89%, and SVM = 89%)

Table 6. *Cont.*

STUDY	Model Used	Results
Sudesh Bhagat et al. [59]	Different ML models	Accuracy (XGBoost = 87%, sVM = 86%, BERT Word Embeddings = 82%, BERT Sentence Embeddings, and Albert Model = 88%)
Qinyu Sun et al. [120]	Different ML models	Accuracy (RF and SVM = 87% and LSTM = 91%)

The considered crossings are crossings in the normal sense of the term which have a normal shape depending on numerous observable factors but which are not necessarily considered in pedestrian safety. The crossing time was a factor strongly correlated with the probability of crossing outside crossings; this highlights how important the notion of time is for pedestrians to move efficiently in the urban environment. They have to continuously balance roads with high traffic density, the advantage of time saving given by a direct crossing bringing them to a point where it is possible to cross the roadway, and the increased travel time they need in order to safely access ad hoc facilities and crossroads with heavy traffic several times. It should never be seen as risky behavior on their part; the findings of this study reveal that in terms of pedestrian movement, this is rational on the part of pedestrians in their constrained environment. The machine learning models emphasize that crossing illegally is not a random or impulsive violation but an action shaped by underlying factors. Such type of crossing represents adaptive behavior shaped by the shortcomings of the current system. The outputs not only produce accurate predictions, but they also provide actionable insights that function as a practical behavioral decision map.

The total crossing time reinforces this finding, indicating that pedestrians explicitly evaluate the likelihood of crossing before beginning the action. Those who believe they may be exposed for long periods are more likely to use legal facilities, but those who think they can cross quickly—either based on their own experience or because of traffic behavior—may exhibit risky behavior initially. The 40–60 bracket is an important observation: this demographic could be more risk-averse but also more limited in their physical effort to access pedestrian bridges. Conventional safety interventions often overlook subtle concepts related to physical discomfort, a concept that emerged clearly in the model as a consideration of importance.

The propensity to overfit of the random forest and decision tree models indicates that some learning methods are extremely sensitive to aspects of site-specific characteristics; however, the stability of findings for CatBoost, gradient boosting, and SHAP reinforces the stability of the identified behavioral drivers. The SHAP results verify clear directional effects: shorter waiting time and more crossing attempts. Furthermore, reduced overall exposure time consistently shifts the model toward illegal behavior.

The proposed models show clear evidence of the social amplification effect. Pedestrians in small groups tended to take more risks, drawing confidence from the presence of others, even without consciously relying on their support. Although vehicles were included as features in the dataset, the models identified them as being weak indicators of risk in pedestrian crossing.

In dense urban corridors, this is to be expected, as close encounters are part of the norm. When danger is an everyday occurrence, near misses become trivial components of routine street navigation. The heatmap supports these findings and shows close correlations among waiting times, attempts to cross, and crossing speeds. All of these indicate how pedestrians adapt their navigation.

Based on the results, we can conclude that pedestrians' behavior at midblock crossings in evolving urban centers cannot be explained purely by the concept of environmental

friction, by the mere presence of infrastructure, or by logical adaptation in isolation. It is a combination of these forces working in union.

Considering the data collected up to October 2023, the ML models demonstrate the ability to capture real-time decisions made at specific moments. They not only produce accurate information but also classify data based on predicted behavior to support the design of future models.

The key insights are as follows:

- Time efficiency supersedes pedestrian decision making: Crossing time was consistently demonstrated to be the strongest predictor; pedestrians prioritize exposure time to crossing rules over their formally defined behaviors.
- Crossing illegally is a planned act, not reactive: Pedestrian decision making factored in the speed of traffic, gaps, and distance, representing a consideration of tradeoffs that reflect a decision-making process that responds to a context.
- Short waiting times indicate higher risk aversion: Individuals who wait a brief time before beginning to cross are more likely to indicate a willingness to take risks when crossing fast-moving traffic streams.
- Multiple crossing attempts reinforce risk-taking behavior, rather than just one attempt: All subsequent crossing attempts indicate pedestrian behavior is learning-adaptable, becoming more confident with less sensitivity to danger.
- Larger groups indicate a higher likelihood of taking risks: Individuals crossing in pairs or small groups increase collective confidence and reduce hesitation while engaging in illegal middle-of-the-street crossing behavior.
- Age of pedestrians gives complexity to decision making: Individuals aged 40–60 have mixed behavior, being more cautious but more deterred by the physical demand to use pedestrian bridges, increasing the likelihood of violation.
- Vehicle conflicts are desensitized in dense environments: The frequency of near conflicts is small and has limited relevance as a predictor, indicating pedestrians are desensitized to near conflict hazards.
- Illegal crossing behavior results from friction within the built environment, not ignorance about crossing laws: Pedestrians' awareness, for example, did not dictate pedestrian crossing to be legal; instead, pedestrians' understanding led them to cross towards medians according to distance and placement.
- Predictable flow of traffic properties leads to violations: Pedestrians are more likely to ignore the law and traverse through gaps in a predictable flow of traffic movement.
- Time on the roadway provides greater influence than the volume of traffic: It was not the number of cars in the proportional distance that was observed to have an influence but the assigned time in the roadway, whether the crossing behavior was legal or illegal, which was strongly associated with crossing influence.
- Risky crossing develops in response to dysfunction of infrastructure: Illegal midblock crossing behaviors develop through systemic shortcomings rather than deviant disposition.
- Behaviors share patterns across similar midblock conditions: The relative stability of variable importance across models indicates that these findings are transferable across similar road conditions within other developing countries.

Policy Implications

Based on the main findings of this research study, applicable related policies can be developed for transportation and urban policymakers in developing nations as follows.

First, the findings from this investigation hold major implications for transportation and urban policy in developing cities around the world. First, planning decisions must be

based on actual pedestrian behavioral patterns instead of design intentions. Pedestrian bridges should be built at locations where pedestrians walk across the street, not in places where planners expect them to walk. This means that you need to systematically observe pedestrian behavior, base your intervention decisions on data, and be willing to modify the original corridor layouts. Planners themselves should pay much attention to minimizing vertical circulation costs (ramps, elevators, slope, or linked access) that significantly increase the time cost of crossing legally. In many places, reducing this cost would be sufficient to make pedestrian cross legally.

Second, we should consider reducing perceived inconvenience to be a guiding principle of midblock safety interventions. Providing a shaded walkway, a pathway that feels like a shortcut, good lighting, or an escalator will motivate pedestrians to cross a bridge by decreasing the travel effort.

Policies to decrease vehicle speed when vehicles travel down higher-risk segments of the road, adjust the width of road lanes, implement speed enforcement cameras, and deploy raised medians to create other potential vehicle speed changes (while you adjust the road design) will influence pedestrian risk and crossing opportunity.

Third, transportation agencies need to focus on behavioral responses. Campaigns targeting an audience, especially middle-aged women who have well-known mixed observance of caution and constraints, can increase awareness of the illegal consequences of crossing a roadway.

The fourth aspect is that bridge design guidelines should use universal design guidelines to ensure that entire user groups, such as older pedestrians, people with mobility limitations, and parents with children, can use the bridge. Providing rest areas, manageable slopes, or visually continuous routes can make legally designated paths more visually appealing.

Finally, planning for a sustained future should involve simulation-based evaluation tools. Policies can benefit from the incorporation of pedestrian behavior models, like those developed in this study or models that may have been developed previously, into a platform to evaluate design alternatives and predict behavioral responses before implementation. Implementing evidence-based approaches can promote better decision making for designs, focused enforcement efforts, and integrated safety procedures that directly observe the behavioral mechanisms verified through the built models.

This research study offers valuable policy implications, which should be applicable in similar contexts in which the data were gathered. The analysis was conducted using only a few sites located in one large metropolitan area, and the sample size reflects the behavior of a specific group of individuals within a limited geographic, operational, and cultural context.

Therefore, while developing cities may have similar characteristics to the study sites, the results should be verified. The results can serve as an indication of the mechanisms impacting behavior in areas with high urban population density and high pedestrian demand but limited formal crossing opportunities.

When applying these results to other locations, it would not, therefore, be wise to consider them universally applicable recommendations, and local adaptation and empirical validation should take place prior to any policy implementation elsewhere.

6. Conclusions

This study underscores a deep understanding of illegal pedestrian crossing behavior at midblock locations in a dense urban built environment. In addition, identification of key factors affecting this behavior, such as pedestrian crossing time (s), total time (s), and age group (40–60 Y), are the main contributing factors pushing model prediction towards

illegal behavior. ML methods deeply infer hidden factor layers not clarified in traditional discrete choice models.

A comprehensive suite of ensemble and tree-based ML models, including random forest, extra trees, decision tree, AdaBoost, gradient boosting, and CatBoost, was systematically evaluated. Among these, CatBoost demonstrated clear superiority in predictive performance and generalization capability, achieving an accuracy of 0.98, an adjusted F1 Score of 0.975, a precision of 0.96, and a Recall of 0.989. The robustness of this model was further confirmed by AUC values of 0.99 for the training dataset and 0.97 for the test dataset, indicating minimal overfitting and strong discriminative power. This level of performance represents a substantial advancement over commonly applied statistical and econometric pedestrian behavior models reported in the prior literature.

Beyond prediction, a key methodological contribution of this research study is the application of sensitivity analysis and SHAP-based explainability to interpret model behavior. Feature importance results consistently identified pedestrian crossing time, total travel time, and the 40–60 year age group as the most influential predictors of illegal crossing behavior. SHAP analysis quantified these effects with values of approximately 3, 2, and 0.60, respectively, providing transparent attribution of model decisions. Importantly, SHAP sensitivity results revealed that the higher total time and the 40–60 year age group exert negative influence, increasing the likelihood of illegal crossing behavior. This behavioral asymmetry offers new empirical insights into time-pressure-driven risk taking that are not directly observable through aggregate statistical inference.

From an applied perspective, the findings demonstrate that CatBoost constitutes a highly effective and scalable analytical tool for diagnosing illegal pedestrian crossing behavior near midblock foot bridges. The results also expose critical deficiencies in current FOB design, particularly in addressing pedestrian comfort, accessibility, and directness of movement. The prevalence of illegal crossing behavior within short walking distances from FOBs highlights the disconnect between infrastructure provision and actual pedestrian needs. Accordingly, FOBs should be complemented with user-oriented design elements such as ramps, elevators, adequate lighting, and seamless inlet and outlet connectivity to public transport stops, major destinations, and urban landmarks.

Finally, this study advances policy discourse by emphasizing that pedestrian-friendly infrastructure alone is insufficient to maximize FOB utilization and safety outcomes. Effective mitigation of illegal crossing behavior requires an integrated strategy combining improved geometric design, targeted law enforcement, educational and behavioral awareness campaigns, and AI-driven monitoring and decision-support tools. Complementary enforcement measures, including physical sidewalk barriers, median island treatments, and controlled parking restrictions within the effective influence zone of FOBs, can further discourage unsafe crossing. Collectively, these interventions support safer pedestrian mobility, promote active travel, and contribute to resilient and sustainable urban environments aligned with smart city objectives.

While this study offers insights into the key behavioral factors influencing pedestrians' risky crossing behavior at midblock crossings supported by grade-separation pedestrian facilities, it is recommended to enrich the study outcomes by expanding the case studies to include different contexts, regions, income levels, road categories, etc.

On the other hand, it is recommended to develop choice models using traditional approaches (e.g., binary logit) and conduct a comparison with the ML method to quantify the superiority of ML over traditional modeling.

Author Contributions: Conceptualization and methodology, A.M.D.; software, A.M.D., S.S., and M.S.A.; validation, A.M.D., S.S., A.Q., T.O.A., and M.S.A.; formal analysis, A.M.D., S.S., M.Z., M.E., A.E., and M.S.A.; investigation, A.M.D.; data curation, A.M.D., M.Z., M.E., and A.E.; writing—original

draft preparation, A.M.D., S.S., and M.S.A.; writing—review and editing, A.M.D., S.S., M.Z., M.E., A.Q., T.O.A., A.E., and M.S.A.; manuscript revision after peer review, A.M.D., S.S., A.Q., T.O.A., and M.S.A.; visualization, supervision, and project administration, A.M.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research study received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript.

AI	artificial intelligence
AUC	area under the curve
ANN	artificial neural network
BCNNs	Bayesian combined neural networks
BGLM	Bayesian generalized linear model
CMF	crash modification factor
CNN	convolution neural network
DT	decision tree
DL	deep learning
ditARF	distributional auto-replicative random forest
ETs	extra trees
EFA	exploratory factor analysis
ERTs	extremely randomized trees
ELM	extreme learning machine
FHWA	Federal Highway Administration
FP	false positive
FN	false negative
FFNN	feedforward neural network
GB	gradient boosting
GMMs	Gaussian mixture models
GD	gradient descent
GCN	graph convolutional network
HDI	Human Development Index
IPCB	illegal pedestrian crossing behavior
IOT	Internet of Things
IRL	inverse reinforcement learning
KNN	k-nearest neighbor
KSONN	Kohonen self-organizing neural network
LMICs	low- and middle-income countries
LSTM	long short-term memory
MAAIRL	multiagent adversarial inverse reinforcement learning
ML	machine learning
MNN	modular neural network
MSE	mean square error
MARS	multivariate adaptive regression spline
MPS	midblock pedestrian signals
NN	neural network
NCPCB	non-compliant pedestrian crossing behavior
NB	naïve Bayes
piSVM	permutation-invariant support vector machines

PET	post-encroachment time
PHBs	pedestrian hybrid beacons
PCA	principal component analysis
RRFBs	rectangular rapid flashing beacons
RL	reinforcement learning
RNN	recurrent neural network
RBNN	radial basis function neural network
RF	random forest
RBM	restricted Boltzmann machine
SVM	support vector machine
SeSL	semi-supervised learning
SHAP	Shapley additive explanation
SSA	safety system approach
SEM	structural equation modeling
TxDOT	Texas Department of Transportation
TTC	time to collision
TP	true positive
TN	true negative
VSRI	vehicle-scaled risk indicator
XAI	explainable artificial intelligence

References

1. Dhoke, A.; Choudhary, P. Is there a relationship between time pressure and pedestrian non-compliance? A systematic review. *Transp. Res. Part F Traffic Psychol. Behav.* **2023**, *93*, 68–89. [\[CrossRef\]](#)
2. Shokry, S.; Alrashidi, A.; El-Bany, M.E.-S.; Darwish, A.M. Impact of road geometry on school-area traffic congestion using regression and machine learning analysis: Lessons from six Saudi cities. *Transp. Res. Interdiscip. Perspect.* **2025**, *34*, 101686. [\[CrossRef\]](#)
3. Sharif, H.O.; Joseph, J.; Dessouky, S.; Eid, E.; Weissmann, J. *Select High Risk Pedestrian Midblock Crossings and Perform Safety Evaluations for Developing Pedestrian Crossings*; University of Texas at San Antonio: San Antonio, TX, USA, 2025.
4. Zagow, M.; Elbany, M.; Darwish, A.M. Identifying urban, transportation, and socioeconomic characteristics across US zip codes affecting CO₂ emissions: A decision tree analysis. *Energy Built Environ.* **2025**, *6*, 484–494. [\[CrossRef\]](#)
5. Darwish, A.M.; Zagow, M.; Elkafoury, A. Impact of land use, travel behavior, and socio-economic characteristics on carbon emissions in cool-climate cities, USA. *Environ. Sci. Pollut. Res.* **2023**, *30*, 91108–91124. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Zagow, M.; Darwish, A.M.; Shokry, S. Modeling Health-Supportive Urban Environments: The Role of Mixed Land Use, Socioeconomic Factors, and Walkability in U.S. ZIP Codes. *Sustainability* **2025**, *17*, 10873. [\[CrossRef\]](#)
7. WHO. Road Traffic Injuries. Available online: <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries> (accessed on 22 November 2025).
8. Khuzam, E.A.; Lanzaro, G.; Sayed, T. Impact of jaywalking on pedestrian interaction behavior: A multiagent Markov Game-based analysis. *Accid. Anal. Prev.* **2025**, *220*, 108141. [\[CrossRef\]](#)
9. Sharma, S.N.; Dehalwar, K. A systematic literature review of pedestrian safety in urban transport systems. *J. Road Saf.* **2025**, *36*, 55–78. [\[CrossRef\]](#)
10. Eloglu, B. Developing a Decision-Support Tool for Footbridge Planning and Design Phases. Master's Thesis, Middle East Technical University, Ankara, Turkey, 2025.
11. CAPMAS. Central Agency for Public Mobilization and Statistics. Egypt. 2017. Available online: <https://www.capmas.gov.eg/> (accessed on 20 August 2025).
12. El-Din, E.-S.G. 24.5% Decrease in Road Fatalities in 2023 in Egypt: CAPMAS. Ahramonline. Available online: <https://english.ahram.org.eg/News/523833.aspx> (accessed on 19 December 2024).
13. Shaaban, K.; Muley, D.; Mohammed, A. Analysis of illegal pedestrian crossing behavior on a major divided arterial road. *Transp. Res. Part. F Traffic Psychol. Behav.* **2018**, *54*, 124–137. [\[CrossRef\]](#)
14. Zhang, W.; Guo, H.; Wang, C.; Wang, K.; Huang, W.; Xu, Q.; Tang, H.; Yang, B.; Yan, R. Analysis of pedestrian illegal crossing at unmarked segments: Environmental factors, pedestrian characteristics and crossing behaviours. *Transp. Res. Part. F Traffic Psychol. Behav.* **2023**, *99*, 339–355. [\[CrossRef\]](#)
15. Okail, M. Investigating Self-Report Pedestrian Safety-related Behavior in Developing Countries: Egypt as a Case Study. *Traffic Saf. Res.* **2022**, *3*, 000016. [\[CrossRef\]](#)

16. Sayed, I.; Abdelgawad, H.; Said, D. Studying driving behavior and risk perception: A road safety perspective in Egypt. *J. Eng. Appl. Sci.* **2022**, *69*, 22. [\[CrossRef\]](#)
17. Bayomi, A.; Shawky, M.; Osama, A. Investigating mid-block pedestrian crossing behaviour and safety at urban streets in Cairo. *Int. J. Inj. Contr. Saf. Promot.* **2024**, *31*, 72–85. [\[CrossRef\]](#)
18. Tom, A.; Granié, M.A. Gender differences in pedestrian rule compliance and visual search at signalized and unsignalized crossroads. *Accid. Anal. Prev.* **2011**, *43*, 1794–1801. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Araya-Porras, E.; Mora-Calderón, A.; Aguero-Valverde, J. Pedestrian crossing light violation in Costa Rica: Exploring factors affecting mid-block crossing behavior. *Ingeniería* **2022**, *32*, 111–128. [\[CrossRef\]](#)
20. Abdullah, M.; Dias, C.; Oguchi, T. Road Crossing at Unmarked Mid-Block Locations: Exploring Pedestrians' Perception and Behavior. *Iran. J. Sci. Technol. Trans. Civ. Eng.* **2021**, *46*, 1681–1698. [\[CrossRef\]](#)
21. Soathong, A.; Chowdhury, S.; Wilson, D.; Ranjitkar, P. Investigating the motivation for pedestrians' risky crossing behaviour at urban mid-block road sections. *Travel Behav. Soc.* **2021**, *22*, 155–165. [\[CrossRef\]](#)
22. Ravishankar, K.; Nair, P. Pedestrian risk analysis at uncontrolled midblock and unsignalised intersections. *J. Traffic Transp. Eng.* **2018**, *5*, 137–147. [\[CrossRef\]](#)
23. Xiao, Y.; Liu, Y.; Liang, Z. Study on road-crossing violations among young pedestrians based on the theory of planned behavior. *J. Adv. Transp.* **2021**, *2021*, 6893816. [\[CrossRef\]](#)
24. Elsayyad, M.; Muley, D.; Alhajyaseen, W. Determinants of road user behavior at marked midblock crosswalks. *Can. J. Civ. Eng.* **2025**, *52*, 1506–1522. [\[CrossRef\]](#)
25. Pechteep, P.; Luathep, P.; Jaensirisak, S. Factors Influencing the Violation Intentions of Pedestrians, Motorcycle Riders, and Car Drivers at Midblock Crosswalks. *J. Inf. Syst. Eng. Manag.* **2025**, *10*, 310–332. [\[CrossRef\]](#)
26. Tiwari, R.R.; Patel, S.; Soju, A.; Trivedi, P. Road use pattern and street crossing habits of schoolchildren in India. *Front. Public Health* **2021**, *9*, 628147. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Osei, K.K.; Obiri-Yeboah, A.A.; Adu-Gyamfi, L.; Ackaah, W. Road crossing behavior and preferences among pedestrians: From the lens of the theory of interpersonal behavior. *Traffic Inj. Prev.* **2024**, *25*, 91–100. [\[CrossRef\]](#)
28. Szagała, P.; Brzeziński, A.; Kieć, M.; Budzynski, M.; Wachnicka, J.; Pazdan, S. Pedestrian Safety at Midblock Crossings on Dual Carriageway Roads in Polish Cities. *Sustainability* **2022**, *14*, 5703. [\[CrossRef\]](#)
29. Pawar, D.; Yadav, A. Modelling the pedestrian dilemma zone at uncontrolled midblock sections. *J. Saf. Res.* **2021**, *80*, 87–96. [\[CrossRef\]](#)
30. Pechteep, P.; Luathep, P.; Jaensirisak, S.; Kronprasert, N. Analysis of Factors Influencing Driver Yielding Behavior at Midblock Crosswalks on Urban Arterial Roads in Thailand. *Sustainability* **2024**, *16*, 4118. [\[CrossRef\]](#)
31. Zou, F.; Ogle, J.; Jin, W.; Gérard, P.; Petty, D.; Robb, A. Pedestrian Behavior Interacting with Autonomous Vehicles during Unmarked Midblock Multilane Crossings: Role of Infrastructure Design, AV Operations and Signaling. *arXiv* **2023**. [\[CrossRef\]](#)
32. Theofilatos, A.; Ziakopoulos, A.; Oviedo-Trespalacios, O.; Timmis, A. To cross or not to cross? Review and meta-analysis of pedestrian gap acceptance decisions at midblock street crossings. *J. Transp. Health* **2021**, *22*, 101108. [\[CrossRef\]](#)
33. Tian, K.; Markkula, G.; Wei, C.; Lee, Y.M.; Madigan, R.; Merat, N.; Romano, R. Explaining unsafe pedestrian road crossing behaviours using a Psychophysics-based gap acceptance model. *Saf. Sci.* **2022**, *154*, 105837. [\[CrossRef\]](#)
34. Forrest, M.; Heydari, S.; Cherrett, T. Examining the impact of exposure, built environment and socio-demographics on pedestrian safety: A case study of Greater London. *Saf. Sci.* **2023**, *159*, 106015. [\[CrossRef\]](#)
35. Hasanat-E-Rabbi, S.; Hamim, O.F.; Debnath, M.; Hoque, S.; McIlroy, R.C.; Plant, K.L.; Stanton, N.A. Exploring the relationships between demographics, road safety attitudes, and self-reported pedestrian behaviours in Bangladesh. *Sustainability* **2021**, *13*, 10640. [\[CrossRef\]](#)
36. Mamun, S.; Caraballo, F.J.; Ivan, J.N.; Ravishanker, N.; Townsend, R.M.; Zhang, Y. Identifying association between pedestrian safety interventions and street-crossing behavior considering demographics and traffic context. *J. Transp. Saf. Secur.* **2020**, *12*, 441–462. [\[CrossRef\]](#)
37. Molyneaux, N.; Scarinci, R.; Bierlaire, M. Design and analysis of control strategies for pedestrian flows. *Transportation* **2021**, *48*, 1767–1807. [\[CrossRef\]](#)
38. Hossain, A.; Das, S.; Jafari, M.; Starewich, M.; Chakraborty, R.; Kutela, B. Behavioral and psychological determinants of pedestrian collisions on arterial roads with evidence from random parameter models. *Sci. Rep.* **2025**, *15*, 31684. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Verma, R.; Agarwal, M.M. A systematic review on road accident prediction: A special attention towards machine learning and deep learning approaches. *J. Circuits Syst. Comput.* **2025**, *35*, 2530009. [\[CrossRef\]](#)
40. Sun, Y.; Ortiz, J. Machine learning-driven pedestrian recognition and behavior prediction for enhancing public safety in smart cities. *J. Artif. Intell. Inf.* **2024**, *1*, 51–57.
41. Haque, F.; Kidwai, F.A.; Thapa, I.; Ghani, S.; Mtapure, L.M. Modeling and Evaluating the Impact of Mobile Usage on Pedestrian Behavior at Signalized Intersections: A Machine Learning Perspective. *Future Transp.* **2025**, *5*, 11. [\[CrossRef\]](#)

42. Niture, N.; Abdellatif, I. A systematic review of factors, data sources, and prediction techniques for earlier prediction of traffic collision using AI and machine learning. *Multimed. Tools Appl.* **2025**, *84*, 19009–19037. [\[CrossRef\]](#)
43. Kaggle. “Kaggle”. Available online: <https://www.kaggle.com/> (accessed on 20 November 2025).
44. Singh, A.; Thakur, N.; Sharma, A. A review of supervised machine learning algorithms. In *Proceedings of the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*; IEEE: Piscataway, NJ, USA, 2016; pp. 1310–1315.
45. Shrestha, A.; Mahmood, A. Review of deep learning algorithms and architectures. *IEEE Access* **2019**, *7*, 53040–53065. [\[CrossRef\]](#)
46. Cai, J.; Wang, M.; Wu, Y. Research on pedestrian crossing decision models and predictions based on machine learning. *Sensors* **2024**, *24*, 258. [\[CrossRef\]](#)
47. Singh, D.; Das, P.; Ghosh, I. Prediction of pedestrian crossing behaviour at unsignalized intersections using machine learning algorithms: Analysis and comparison. *J. Multimodal User Interfaces* **2024**, *18*, 239–256. [\[CrossRef\]](#)
48. Taamneh, M.M.; Alomari, A.H.; Taamneh, S.M. Using machine learning to predict pedestrian compliance at crosswalks in Jordan. *Appl. Sci.* **2024**, *14*, 4945. [\[CrossRef\]](#)
49. Ishaque, M.M.; Noland, R.B. Behavioural issues in pedestrian speed choice and street crossing behaviour: A review. *Transp. Rev.* **2008**, *28*, 61–85. [\[CrossRef\]](#)
50. Banstola, A. Modelling Pedestrian-Vehicle Conflict and Severity at Uncontrolled Midblock Crossings Inside Kathmandu Valley. Ph.D. Thesis, Tribhuvan University, Lalitpur, Nepal, 2025.
51. Sulle, M.; Mwakalonge, J.; Comert, G.; Siuhi, S.; Gyimah, N.K. Unraveling Pedestrian Fatality Patterns: A Comparative Study with Explainable AI. *arXiv* **2025**, arXiv:2503.17623. [\[CrossRef\]](#)
52. Piyalungka, S.; Kanitpong, K.; Karoonsoontawong, A. Pedestrian gap acceptance behavior at unsignalized mid-block crossing under mixed traffic conditions. *IATSS Res.* **2025**, *49*, 105–113. [\[CrossRef\]](#)
53. Wang, Y.; Niu, Y.; Zhu, W.; Chen, W.; Li, Q.; Wang, T. Predicting pedestrian crossing behavior at unsignalized mid-block crosswalks using maximum entropy deep inverse reinforcement learning. *IEEE Trans. Intell. Transp. Syst.* **2023**, *25*, 3685–3698. [\[CrossRef\]](#)
54. Nasernejad, P.; Sayed, T.; Alsaleh, R. Multiagent modeling of pedestrian-vehicle conflicts using Adversarial Inverse Reinforcement Learning. *Transp. A Transp. Sci.* **2023**, *19*, 2061081. [\[CrossRef\]](#)
55. Kim, Y.; Abdel-Aty, M.; Choi, K.; Islam, Z.; Wang, D.; Zhai, S. Pedestrian Crossing Direction Prediction at Intersections for Pedestrian Safety. *IEEE Open J. Intell. Transp. Syst.* **2025**, *6*, 692–707. [\[CrossRef\]](#)
56. Liu, S.; Evdorides, H. Data Mining Applications for Pedestrian Behaviour Patterns at Unsignalized Crossings. *Sustainability* **2025**, *17*, 776. [\[CrossRef\]](#)
57. Prakash, S.; Karuppanagounder, K. Pedestrian path changing behaviour prediction models for urban mid blocks. *Transp. Res. Interdiscip. Perspect.* **2023**, *22*, 100973. [\[CrossRef\]](#)
58. Kim, S.-K.; Chan, I.C. Novel Machine Learning-Based Smart City Pedestrian Road Crossing Alerts. *Smart Cities* **2025**, *8*, 114. [\[CrossRef\]](#)
59. Bhagat, S.; Shihab, I.F.; Wood, J. Identification of Potentially Misclassified Crash Narratives using Machine Learning (ML) and Deep Learning (DL). *arXiv* **2025**, arXiv:2507.03066. [\[CrossRef\]](#)
60. Zokirjonovich, O.O. Pedestrian Safety in Developed Countries: Best Practices, Technologies, and Implementation Opportunities. *Mod. Am. J. Eng. Technol. Innov.* **2025**, *1*, 32–40.
61. UNDP. *A Matter of Choice: People and Possibilities in the Age of AI*; Human Development report 2025; United Nations Development Programme: New York, NY, USA, 2025.
62. Kwayu, K.M.; Kwigizile, V.; Oh, J.S. Evaluation of pedestrian crossing-related crashes at undesignated midblock locations using structured crash data and report narratives. *J. Transp. Saf. Secur.* **2022**, *14*, 1–23. [\[CrossRef\]](#)
63. Mukherjee, D. Analyzing Key Factors Influencing Pedestrian Non-utilization of Designated Crossings and Sidewalks in Urban Areas of Developing Countries. *Transp. Dev. Econ.* **2025**, *11*, 31. [\[CrossRef\]](#)
64. Mohammed, H. Assessment of distracted pedestrian crossing behavior at midblock crosswalks. *IATSS Res.* **2021**, *45*, 584–593. [\[CrossRef\]](#)
65. Sadeek, S.N.; Rahman, M.H.; Rifaat, S.M. Understanding pedestrian bridge usage considering perception and socio-demographic characteristics of the road users in Dhaka city. *Transp. Res. Interdiscip. Perspect.* **2025**, *31*, 101384. [\[CrossRef\]](#)
66. Pratiwi, P.C.; Suhardi, B.; Laksono, P.W. Analyzing safety factors of pedestrian bridge users in Asia: A systematic literature review. *Asian J. Soc. Sci. Manag. Technol.* **2025**, *7*, 1.
67. Green, O.; Ivan, J.N.; Auguste, M.E.; Wang, K.; Chacon-Hurtado, D. *Using Pedestrian Crossing Experiences, Physical Intersection Characteristics, and Socioeconomic Variables to Predict Pedestrian-Vehicle Conflicts and Signal Compliance*; Elsevier: Amsterdam, The Netherlands, 2025.
68. Raoniar, R.; Singh, S.; Pathak, A.; Maurya, A.K. Analysis of Pedestrian–Vehicle Interaction Dynamics at Signalized Urban Intersections: A Surrogate Safety Measure Approach. *Transp. Res. Rec.* **2025**, *2679*, 1042–1063. [\[CrossRef\]](#)

69. Hou, M.; Wang, C.; Easa, S.M.; Cheng, J. Eye Movement Evaluation of Pedestrians' Mobile Phone Usage at Street Crossings. *Transp. Res. Rec.* **2025**, 2679, 1495–1512. [\[CrossRef\]](#)
70. Sulle, M.; Mwakalonge, J.; Comert, G.; Siuhi, S.; Gyimah, N.K.; Roberts, J.; Ruganuza, D. Analysis of Distracted Pedestrians Crossing Behavior: An Immersive Virtual Reality Application. *arXiv* **2025**, arXiv:2503.16443.
71. Frej, D. Observational Study of Pedestrian Behavior at Signalized Crosswalks. *Komunikácie-Ved. Listy Žilinskej Univerzity V Žiline* **2025**, 27, 54–66. [\[CrossRef\]](#)
72. Demiroz, Y.I.; Onelcin, P.; Alver, Y. Illegal road crossing behavior of pedestrians at overpass locations: Factors affecting gap acceptance, crossing times and overpass use. *Accid. Anal. Prev.* **2015**, 80, 220–228. [\[CrossRef\]](#)
73. Fahami, F.A.; Daniel, B.D. Influence of Traffic Conditions on Pedestrian Crossing Decision at Urban Street Crosswalk. *Recent. Trends Civ. Eng. Built Environ.* **2025**, 6, 256–268.
74. Bandara, D.; Hewawasam, C. A comparative study on effectiveness of underpass and overpass among pedestrians in different urban contexts in Sri Lanka. *J. Serv. Sci. Manag.* **2020**, 13, 729–744. [\[CrossRef\]](#)
75. Räsänen, M.; Lajunen, T.; Alticafarbay, F.; Aydin, C. Pedestrian self-reports of factors influencing the use of pedestrian bridges. *Accid. Anal. Prev.* **2007**, 39, 969–973. [\[CrossRef\]](#) [\[PubMed\]](#)
76. Landa-Blanco, M.; Ávila, J. Factors related to the use of pedestrian bridges in university students of Honduras. *Transp. Res. Part F: Traffic Psychol. Behav.* **2020**, 71, 220–228. [\[CrossRef\]](#)
77. Shaaban, K.; Abdel-Warith, K. Agent-based modeling of pedestrian behavior at an unmarked midblock crossing. *Procedia Comput. Sci.* **2017**, 109, 26–33. [\[CrossRef\]](#)
78. Bayezid, M.; Abir, K.; Al Milhan, A.; Sakib, S. Influence of Various Factors on Jaywalking Behavior: A Behavioral Analysis. In Proceedings of the 1st International Conference on Core Engineering & Technology (IUT-ICCET 2024), Virtual, 22–24 February 2024.
79. Manoguid, A.; Tomas, J.P. Nighttime Detection of Illegal Crossing by Pedestrians and Pedestrian Lane Obstruction by Vehicles through Effective Deep Learning Model. In Proceedings of the 11th World Congress on Electrical Engineering and Computer Systems and Science, Paris, France, 17–19 August 2025.
80. Trafikverket. This is Vision Zero. Available online: https://bransch.trafikverket.se/en/startpage/operations/Operations-road/vision-zero-academy/This-is-Vision-Zero/?utm_source=smartcitysweden.com&utm_medium=link&utm_campaign=promotion&utm_source=smartcitysweden.com&utm_medium=link&utm_campaign=promotion (accessed on 19 November 2025).
81. Mostafa, A.M.; Aldughayfiq, B.; Tarek, M.; Alaerjan, A.S.; Allahem, H.; Elbashir, M.K.; Ezz, M.; Hamouda, E. AI-based prediction of traffic crash severity for improving road safety and transportation efficiency. *Sci. Rep.* **2025**, 15, 27468. [\[CrossRef\]](#)
82. Duong, H.N.; Chu, M.C.; Huynh, N. Understanding psychological factors behind motorcyclists crossing behavior on undivided roads in mixed traffic conditions: A case study of Hau Giang, Vietnam. *IATSS Res.* **2025**, 49, 114–126. [\[CrossRef\]](#)
83. Peng, T.; Liu, J. Analysis of conflict between right-turning vehicles and pedestrians at urban intersections using random parameter Logit model. *Traffic Inj. Prev.* **2025**, 1–9. [\[CrossRef\]](#)
84. Zhang, C.; Zhou, B.; Qiu, T.Z.; Liu, S. Pedestrian crossing behaviors at uncontrolled multi-lane mid-block crosswalks in developing world. *J. Saf. Res.* **2018**, 64, 145–154. [\[CrossRef\]](#)
85. Abushattal, M.; Alhomaidat, F.; El-Yabroudi, M. Effects of intersection control types on driver yielding behavior to cyclists using mixed logit modeling. *Sci. Rep.* **2025**, 15, 33928. [\[CrossRef\]](#)
86. Sutantaviboon, T.; Kanitpong, K. Modeling Factors Influencing Motorcycle Riders' decision When Facing Pedestrians at Unsignalized Mid-Block Crosswalks: A Case Study from Bangkok. *Civ. Environ. Eng.* **2026**. [\[CrossRef\]](#)
87. Qing, J.K.T.; Daniel, B.D.; Mudjanarko, S.W.; Azmi, A.E.M.N. Modelling Driver-Pedestrian Interaction at Raised Crosswalks. *Int. J. Integr. Eng.* **2025**, 17, 20–28. [\[CrossRef\]](#)
88. Datta, S.; Kadali, B.R. Analyzing Yielding Drivers' Dilemma Behavior toward Pedestrians at Semiurban Uncontrolled Intersections on Arterial Roads. *J. Transp. Eng. A Syst.* **2025**, 151, 04025072. [\[CrossRef\]](#)
89. Ahsan, M.J.; Abdel-Aty, M.; Abdelrahman, A. Can mid-block pedestrian signals (MPS) provide greater safety benefits than other mid-block pedestrian crossings? *Accid. Anal. Prev.* **2025**, 218, 108105. [\[CrossRef\]](#)
90. Foster, N.; Monsere, C.; Carlos, K. Evaluating Driver and Pedestrian Behaviors at Enhanced, Multilane, Midblock Pedestrian Crossings. *Transp. Res. Rec.* **2014**, 2464, 59–66. [\[CrossRef\]](#)
91. Anwari, N.; Abdel-Aty, M.; Goswamy, A.; Zheng, O. Investigating surrogate safety measures at midblock pedestrian crossings using multivariate models with roadside camera data. *Accid. Anal. Prev.* **2023**, 192, 107233. [\[CrossRef\]](#) [\[PubMed\]](#)
92. Hussain, Q.; Alhajyaseen, W.; Pirdavani, A.; Brijs, K.; Shaaban, K.; Brijs, T. Do detection-based warning strategies improve vehicle yielding behavior at uncontrolled midblock crosswalks? *Accid. Anal. Prev.* **2021**, 157, 106166. [\[CrossRef\]](#)
93. Khabiri, R.; Jahangiry, L.; Birgani, H.R.; Sadeghi-Bazargani, H. Interventions for Increasing Pedestrian Visibility to Prevent Injury and Death: A Systematic Review. *Health Soc. Care Community* **2025**, 2025, 2958743. [\[CrossRef\]](#)
94. Malenje, J.; Li, P.; Han, Y. Vehicle yielding probability estimation model at unsignalized midblock crosswalks in Shanghai, China. *PLoS ONE* **2019**, 14, e0213876. [\[CrossRef\]](#)

95. Li, C.-Y.; Liu, S.; Cen, X.-K. Safety and efficiency impact of pedestrian–vehicle conflicts at non signalized midblock crosswalks based on fuzzy cellular automata. *Phys. A-Stat. Mech. Its Appl.* **2021**, *572*, 125871. [CrossRef]
96. IBM. Available online: <https://www.ibm.com/think/topics/machine-learning-types> (accessed on 20 September 2025).
97. Darwish, A.M.; Almansour, M.; Salah, A.; Zagow, M.; Saeed, K.; Elkafoury, A. Sensitivity evaluation of machine learning-based calibrated transportation mode choice models: A case study of Alexandria City, Egypt. *Transp. Res. Interdiscip. Perspect.* **2024**, *24*, 101052. [CrossRef]
98. Nasteski, V. An overview of the supervised machine learning methods. *Horiz. Ser. B* **2017**, *4*, 56. [CrossRef]
99. Usama, M.; Qadir, J.; Raza, A.; Arif, H.; Yau, K.-L.A.; Elkhatab, Y.; Hussain, A.; Al-Fuqaha, A. Unsupervised machine learning for networking: Techniques, applications and research challenges. *IEEE Access* **2019**, *7*, 65579–65615. [CrossRef]
100. Naeem, S.; Ali, A.; Anam, S.; Ahmed, M.M. An unsupervised machine learning algorithms: Comprehensive review. *Int. J. Comput. Digit. Syst.* **2023**, *13*, 911–921. [CrossRef]
101. Rani, V.; Nabi, S.T.; Kumar, M.; Mittal, A.; Kumar, K. Self-supervised learning: A succinct review. *Arch. Comput. Methods Eng.* **2023**, *30*, 2761–2775. [CrossRef]
102. Banerjee, A.; Raoniar, R.; Maurya, A.K. Pedestrian overpass utilization modeling based on mobility friction, safety and security, and connectivity using machine learning techniques. *Soft Comput.* **2020**, *24*, 17467–17493. [CrossRef]
103. Das, S.; Le, M.; Dai, B. Application of machine learning tools in classifying pedestrian crash types: A case study. *Transp. Saf. Environ.* **2020**, *2*, 106–119. [CrossRef]
104. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [CrossRef]
105. Schapire, R.E. The boosting approach to machine learning: An overview. *Nonlinear Estim. Classif.* **2003**, *171*, 149–171.
106. Xu, F.; Huang, Y.; Wang, H.; Fan, Z. A novel heterogeneous data classification approach combining gradient boosting decision trees and hybrid structure model. *Pattern Recognit.* **2025**, *165*, 111614. [CrossRef]
107. Temlyakov, V. Brief introduction in greedy approximation. *arXiv* **2025**, arXiv:2502.13432. [CrossRef]
108. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
109. Kong, X.; Das, S.; Zhang, Y.; Wei, Z.; Yuan, C.-H. In-Depth Understanding of Pedestrian- Vehicle Near-Crash Events at Signalized Intersections: An Interpretable Machine Learning Approach. *Transp. Res. Rec. J. Transp. Res. Board* **2022**, *2677*, 747–759. [CrossRef]
110. Ibrahim, A.A.; Ridwan, R.L.; Muhammed, M.M.; Abdulaziz, R.O.; Saheed, G.A. Comparison of the CatBoost classifier with other machine learning methods. *Int. J. Adv. Comput. Sci. Appl.* **2020**, *11*, 738–748. [CrossRef]
111. Elkafoury, A.; Elboshy, B.; Darwish, A.M. Development of response surface method prediction model for traffic-related roadside noise levels based on traffic characteristics. *Environ. Sci. Pollut. Res.* **2023**, *30*, 94229–94241. [CrossRef]
112. CAPMAS. The Annual Bulletin Results of Cars and Train accidents 2024. Cairo. 2024. Available online: <https://www.censusinfo.capmas.gov.eg/metadata-en-v4.2/index.php/catalog/749/overview> (accessed on 10 September 2025).
113. Mohamed, A.F.A. A Study of Strategic Plans of Sustainable Urban Development for Alexandria, Egypt to Mitigate the Climate Change Phenomena. *Future Cities Environ.* **2023**, *9*, 14. [CrossRef]
114. Albert, S. Pedestrian Road Crossing Behavior and Optimal Selection of Pedestrian Facilities at Mid-Block Crossings. *Eur. Transp./Transp. Eur.* **2022**, *87*, 1–16. [CrossRef]
115. Cai, C.; Wong, S.-K.; Chen, T. Risk-Aware Pedestrian Behavior Using Reinforcement Learning in Mixed Traffic. *Comput. Animat. Virtual Worlds* **2025**, *36*, e70031. [CrossRef]
116. Zhang, C.; Sprenger, J.; Ni, Z.; Berger, C. Predicting Pedestrian Crossing Behavior in Germany and Japan: Insights into Model Transferability. *IEEE Trans. Intell. Veh.* **2024**, *10*, 4887–4902. [CrossRef]
117. Paul, D.; Moridpour, S.; Venkatesan, S.; Withanagamage, N. Evaluating the pedestrian level of service for varying trip purposes using machine learning algorithms. *Sci. Rep.* **2024**, *14*, 2813. [CrossRef] [PubMed]
118. Sakib, N.; Paul, T.; Ahmed, M.T.; Al Momin, K.; Barua, S. Investigating factors influencing pedestrian crosswalk usage behavior in Dhaka city using supervised machine learning techniques. *Multimodal Transp.* **2024**, *3*, 100108. [CrossRef]
119. Jin, C.J.; Luo, Y.; Wu, C.; Song, Y.; Li, D. Exploring the Pedestrian Route Choice Behaviors by Machine Learning Models. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 146. [CrossRef]
120. Sun, Q.; Wang, C.; Pan, Y.; Zhang, H.; Fu, R.; Guo, Y.; Yuan, W. Analysis of Pedestrian Crossing Behavior Characteristics and a Pedestrian Crossing Intention Recognition Model. *Automot. Innov.* **2025**, *8*, 935–948. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.