



Pedestrian overpass utilization modeling based on mobility friction, safety and security, and connectivity using machine learning techniques

Arunabha Banerjee¹ · Rahul Raoniar¹ · Akhilesh Kumar Maurya¹

Published online: 29 August 2020

© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

The identification, understanding, and treatment of predictors that drive the use of pedestrian overpass or Foot Over Bridge (FOB) are very much essential for city planners and policymakers. There is paucity of studies that uses modern soft computing techniques to understand the factors driving the use of FOBs. The aim of the work presented here was to identify the important predictors that determine usability of FOBs. The study utilized both questionnaire survey (perception in terms of satisfaction/dissatisfaction) and field data collected across fourteen locations in six different Indian cities. The goal was to identify the essential features that drive the usability of FOBs under four different contexts, i.e., mobility friction, safety and security, and vertical connectivity and horizontal connectivity. Three soft computing algorithms such as generalized linear model (GLM), random forest (RF), and gradient boosting machine (GBM) were trained to predict the future usability of pedestrians. The modelling approach involved data collection from 14 FOB locations, preprocessing involving data input in spreadsheets, removing missing values and normalization for model training. The next stage involved splitting the data into training and testing set, followed by model training and hyper-parameter optimization using tenfold cross-validation. Finally, the developed models were evaluated for test dataset for generalization. The study results revealed that GBM algorithm showed highest classification accuracy on test dataset over the other two techniques at various scenarios. GBM helped in identifying the essential parameters that drive the usability of FOBs under the four different contexts. Sensitivity analysis supported the fact that gender and age had significant impact on the choice of pedestrians under different contexts. Further, the respondents' feedbacks regarding existing problems were used to validate the findings. The safety and security, walk environment, frequency of daily use, comfort, location type, length of travel, stairway dimensions and reduced walkable width affected the choice of using the FOBs. Therefore, provision of CCTV cameras and security personnel, removal of obstruction, provision of proper lighting and all-weather shade, and regular maintenance of the facilities will significantly improve the pedestrians' choice to use the FOBs. The identification of important variables not only provides better insight of factors that affects the choice of pedestrians using the elevated facilities but also provides a valuable source of information to researchers, planners and policymakers to construct a better-planned pedestrian friendly infrastructure.

Keywords Pedestrian overpass · Utilization · User perception · Machine learning

Communicated by V. Loia.

✉ Arunabha Banerjee
arunabhabanerjee77@gmail.com

Rahul Raoniar
rahul.raoniar@gmail.com

Akhilesh Kumar Maurya
akmaurya@gmail.com

¹ Department of Civil Engineering, IIT Guwahati, Guwahati, Assam 781039, India

Abbreviations

General terms

FOB	Foot over bridge
CCTV	Closed-circuit television
IRC	Indian roads congress
WHO	World Health Organization
CBD	Central business district
PM	Particulate matter
DLTSM	Development of the long short-term memory
PTT	Public transport terminal

Evaluation metrics

ANOVA	Analysis of variance
RMSE	Root-mean-square error
MSE	Mean-squared error
MAE	Mean absolute error
MAPE	Mean absolute percentage error
NRMSE	Normalized root-mean-squared error
AUROC	Area under receiver operating characteristics curve
AUPR	Area under precision recall curve
ROC	Receiver operating characteristic curve
SMAPE	Symmetric mean absolute percentage error
MAD	Mean absolute deviation

Modelling approaches

GLM	Generalized linear model
RF	Random forest
GBM	Gradient boosting machine
CART	Classification and regression trees
RFRC	Random forest regression and classification
BTCR	Boosted tree classifiers and regression
MARS	Multivariate adaptive regression splines
E-CHAID	Chi-squared automatic interaction detector
SVM	Support vector machine
ML	Machine learning
GBDT	Gradient boosting decision tree
MNL	Multi-nominal logit model
BPNN	Back propagation neural network
DCapsNet	Develop capsule network model
DNS-PS	Determined structure network-particle swarm
SAQPM	Smart air quality prediction model
DNN	Deep neural network
XGB	eXtreme gradient boosting
GBRT	Gradient-boosted regression tree
DT	Decision tree
NGSIM	Next-generation simulation
RT	Regression tree

1 Introduction

Walking is the most effective and efficient mode of transportation for short trips. Approximately 1–2 km length of urban trips are performed on foot daily (IRC: 103 2012). The physical structure of the street networks affects the quality of walking journeys (Kang et al. 2018). The absence of proper pedestrian facilities forces pedestrians to use the carriageway and thus come in direct contact with the motorized traffic. One of the most commonly adopted solutions is separating pedestrians from the traffic stream through grade-separated facilities (such as overpasses or underpasses) or providing at-grade signal-controlled pedestrian crosswalks. However, the provision of signalized crossing does not guarantee that the vehicular traffic would stop for the pedestrians, and provide a safe and comfortable crossing experience. Many drivers have a propensity to break traffic rules and even tend to drive their vehicles during red stop signals, which might lead to a pedestrian-vehicle collision (Herms 1972; Koepsell et al. 2002). Also, the pedestrians have a general tendency to avoid properly designated crosswalks and cross the roads through illegal median openings to save time (Golakiya et al. 2019).

According to the World Health Organization (Global Safety Report 2018), among all types of road users, 43% consists of vulnerable road users (with pedestrians accounting for 23%). Also, as per the World Health Organization (WHO) 2018 report, for children and adults (aged between 5 and 29 years) in low- and middle-income countries, road traffic injuries are the primary cause of deaths. The majority of these accidents occurred at the uncontrolled mid-block crossings, where no proper pedestrian crossing facilities were provided. As per the “Road Safety in India Status Report 2016” (Mohan et al. 2017), the reported pedestrian fatality rate was nearly 35–40% of the total fatalities. Metropolitan cities like New Delhi, Bangalore, and Kolkata, had pedestrian fatality share more than 40% (Mohan 2009). Thus, in order to save more lives, it is essential to separate the pedestrian movement from the traffic stream by providing overpasses (foot over bridges) or underpasses (subways). Besides, providing proper street connectivity leads to a better sustainable environment for safe and efficient travel (Zlatkovic et al. 2019).

Past studies showed that even when such grade-separated facilities were provided, pedestrians were reluctant to use them and crossed illegally using available at-grade median openings (Malik et al. 2017; Saha et al. 2011; Das and Barua 2015; Pasha et al. 2015; Sinclair and Zuidgeest 2016). Researchers across the globe had studied different crossing facilities to understand factors that led to the use or non-use of such grade-separated facilities. Studies

related to FOBs revealed that pedestrians preferred the shortest path while choosing the route (Methorst 2004), with time, distance, and extra effort versus safety benefit playing a significant part in the pedestrians' decision-making (Li 2013; Rankavat and Tiwari 2016). In a study at Ankara (Turkey), Räsänen et al. (2007) reported that depending on the time saving, safety, and familiarity with an area, the use of FOBs varied between 6 and 63%. Although the existence of escalators increased the usage rate, however, the existence of a traffic signal in the nearby vicinity of the FOB decreased the usability. Fear of height (Opdyke et al. 1995; Juan and Pérez 2009) and the absence of adequately designed stairways (Mutto et al. 2002; Rizati et al. 2013) significantly affected the motive of the pedestrians toward using elevated facilities. Pedestrians across different Asian cities usually preferred using at-grade facilities rather than using underpasses or overpasses due to security concerns, poor accessibility, and encroachment by hawkers (Saha et al. 2011; Pasha et al. 2015; Malik et al. 2017; Anciaes and Jones 2018).

Researchers had used different modeling approaches to obtain the critical parameters affecting the usability of the pedestrians (refer to Table 1). The regression model used by Abojaradeh (2013) predicted that in Jordan, the use of pedestrian bridges had a positive impact on pedestrians to reduce fatalities. Factors such as posted speed limit, traffic volume, the width of the crosswalk, condition of the bridge, and the existence of median barriers significantly affected the use of the bridges. Also, the likelihood of using FOBs increased when pedestrians had experience of injuries (Oviedo-Trespacios and Scott-Parker 2017). In China, Wu et al. (2014) used a binary logit model and identified that gender, age, career, level of education, license, detour wishes, detour distance, and crossing time played significant impact on the use of overpasses.

Moreover, a study based in Thailand (using logistic regression method) revealed that the proximity of bus stop to the FOB and self-experience of road accidents further influenced the choice of the pedestrians (Sangphong and Siridhara 2014). A Relative Importance Index measure highlighted that even structural design, angle of stairs, width, surface, and the existence of fence at-grade influenced the possibility of using the facility (Hasan and Napiah 2014). In addition to the above-stated factors, the role of parents in educating children about possible traffic risk of using the at-grade facilities influenced the utilization of the footbridges (Hasan and Napiah 2018). Moreover, the use or non-use of pedestrian facilities is a habit and not a coincidental behavior, based on safety and convenience perception (Rankavat and Tiwari 2016; Räsänen et al. 2007).

Table 1 shows that majority of the studies which were conducted in Asian countries, used logistic regression

(binary, multiple and mixed) in order to predict the pedestrian choice between the different crossing facilities. Also, questionnaire was the most preferred survey technique used by different researchers for modelling purposes. Apart from demographic characteristics, variables such as frequency of use, safety, security, hawker's presence, cleanliness, number of steps and type of facility available were the most used for questionnaire studies.

Apart from logistic regression, soft computing techniques are becoming quite famous among researchers globally. Table 2 shows the diverse empirical studies carried out across various domains using different soft computing approaches.

From Table 2, it is observed that different soft computing tools (generalized linear model, random forest, decision trees, neural networks, etc.) were used on survey data in the field of health care, environment, risk assessment, remote sensing and web app-security. Moreover, MAE, MAPE, RMSE, and R^2 square were the most common evaluation metrics used for model evaluation.

1.1 The study motivation and objective

From the previously mentioned literature (Table 1), it is evident that the perception of pedestrians in using the overpasses or FOBs has not been conducted extensively in a developing country, like India. Moreover, previous studies mainly used modeling approaches such as linear, binary logistic and ordered logistic regression models, to gain information on parameters influencing the use/non-use of FOBs. Alongside commonly used modeling approaches, in the past few decades many other high-performing machine learning algorithms, such as tree-based ensemble learners (i.e., random forest, bagging and boosting) and neural networks (deep learning) were introduced in various fields of research (Table 2), as they outperformed other traditional modeling approaches in terms of prediction accuracy (Couronné et al. 2018).

Hence, taking the model performance in consideration and to get more accurate insight on the most prominent factors affecting the usability of pedestrians (users), this study attempts to compare between three different modeling approaches (GLM: generalized linear model, RF: random forest and GBM: gradient boosting machines) in order to predict future usability of overpasses by pedestrians.

The main objective of the current study was to identify the potential parameters that drive the use of foot over bridges across various Indian cities. Both tangible (i.e., field measurements) and intangible (i.e., survey-based ratings of existing FOB's condition) parameters were used in the modeling process.

Unlike the past studies which estimated the usability of FOBs from a single context point of view (i.e. aggregate usability measurement in terms of safety and security), this

Table 1 Various empirical study details focused on the usability of different crossing facilities

Author and Country	Type of study	Type and number of facilities studied	Sample size	Variables used	Model/ method used
Rasanen (2007), Turkey	V & Q	FOB with two stairways (2), FOB with four stairways and with legal crossing at grade (1), FOB with six stairways (1), FOB with escalator and fencing at grade (1)	408	Familiarity with the area, convenience, license, time-saving, frequency of use, frequency of visiting CBD, bridge safety beliefs, age and sex	Binary Logistic Regression
Abojaradeh (2013), Jordan	Q	FOB (8) and underpass (2)	N.A.	Hourly traffic volume, posted speed limit, width of the crosswalk, type of facility, the existence of median and shade availability	Multiple Linear Regression
Rizati (Rizati et al. 2013), Malaysia	V & Q	FOB and at-grade (16)	287	Age, gender, occupation, frequency of use, walking ability, purpose, and status of walking, effective location, the reason of not using, group user, design criteria, type of facility and treatment	Multiple Linear Regression
Wu (2014), China	Q	FOB and at-grade (8)	873	Demographics, detour wish, security degree, overpass height, overpass width, overpass length, number of steps, detour distance, number of lanes of the street under the overpass, width of the street under the overpass	Binary Logistic Regression
Sangphong (2014), Thailand	Q	FOB and at-grade (6)	705	Gender, age, phone use, accidents on the road, the distance between FOB and bus stop, number of co-pedestrians and awareness about fines imposed due to jaywalking	Logistic Regression
Demiroz (2015), Turkey	O	FOB with illegal crossing (4)	231	Gender, age, carrying items, group size, vehicle speed and absence/presence of barriers	ANOVA
Das (2015), Bangladesh	Q	FOB with an improper at-grade opening (5)	143	Time and distance required to cross, height, hawker's presence, security, entry access, congestion over FOB, cleanliness, and location	Descriptive statistics
Rankavat (2016), India	Q	All three (1), Underpass and FOB (1), FOB and zebra crossing (1), Zebra crossing (1), Underpass (3)	500	Age, gender, safety perception, and convenience perception	Ordinal Logistic Regression
Hasan (2014, 2017), Malaysia	V & Q	FOB and at-grade (4)	191	Availability of escalator, the role of parents, the existence of fence, enforcement, aesthetics, alerting posters, signboards, tiredness while ascending stairways, imposing fine on jaywalker, the impact of the past accident, fear of heights, safety awareness	Relative Importance Index
Malik (2017), Pakistan	Q	FOB (12)	155	Safety, security, and awareness	Descriptive Statistics
Oviedo-Trespalcios (2017), Colombia	Q	FOB and at-grade (1)	210	Frequency of crossing, safety, and security of FOB, the danger of traffic in the area, the distance of FOB from the illegal crossing point, time to cross via FOB, time loss due to FOB, accessibility of FOB, illumination of FOB	Backward Stepwise Logistic Regression
Anciaes (2018), England	Q	Straight signalized crossing (9), staggered signalized crossing (12), FOB (1), underpass (3)	321	Age, gender, qualification, employment status, car ownership, location, perceived neighborhood, safety, convenience, crossing time, accessibility and security	Mixed Logistic Regression
Truong et al. (2019), Vietnam	O	FOB and at grade (10)	608	Gender, Age, Walking behavior, weather, overpass height and width, mean illegal crossing	Binomial Regression

Table 1 (continued)

Author and Country	Type of study	Type and number of facilities studied	Sample size	Variables used	Model/method used
Patra et al. (2020)	V	FOB and at-grade (2)	1134	Age, gender, luggage, number of adults travelling together, accompanied by a child, mobile phone use	Binary logit

V Video graphic survey, Q Questionnaire survey, O Observational survey

study estimates the relative importance of different parameters from four different users' perspectives (such as mobility, safety and security, vertical end connectivity and horizontal end connectivity) together using advanced machine learning tools (GLM, RF and GBM). A robust solution (using the advanced machine learning tools) was presented which might not be the quickest one, but shall be the best for policy making by highlighting the impact of different parameters altogether. The results of the study would provide useful information to planners and developers to either upgrade existing elevated pedestrian facilities or construct better facilities in the future.

2 Method

2.1 Survey location selection

Six major cities covering the different regions of India were visited, and all possible survey locations were observed prior to final data collection. Only those locations were selected where the FOBs had an adequate flow of pedestrians throughout the day and were connecting from one side of the road to the other side through a single entry and exit. To cover the variability among the FOBs, different land-use types ranging from commercial and public transport terminal to educational and residential areas were covered for this study. Figure 1 shows the six different Indian cities which were chosen for questionnaire survey along with the total sample response.

In total, 28 locations were visited over six different cities, out of which fourteen locations were finalized, where the final questionnaire surveys were conducted. The survey tried to cover various regions of the Indian sub-continent, covering metropolitan cities like Delhi, Mumbai, Kolkata and Bengaluru. In total 14 locations were covered and 552 completed questionnaire samples were collected. A detailed summary of the different foot over bridges selected for this study is illustrated in Table 3. All the foot over bridge locations considered for the study had different flow levels during peak and off-peak hours. The foot over bridges were provided with either stairway alone, or along with escalators, lifts, and ramps.

The measured length and walkable width of the FOBs across selected locations, varied between 21–88 m and 2.05–5.70 m, respectively, as illustrated in Table 3. It is consistent with the guidelines prescribed by IRC: 103 (2012), suggesting the minimum required walkable width of 1.8 m. Depending on the riser dimension, i.e., the vertical distance between two successive steps (Irvine et al. 1990), the number of steps keeps on varying across the different locations. The dimension of the tread, i.e., the horizontal top portion of a step where foot rests (Irvine et al. 1990), ranged between 26 and 32 cm.

Ideally, the suggested riser and tread dimensions as per IRC: 103 (2012) should be 15 cm and 30 cm, respectively. The locations in Delhi (i.e., ITO and Maharani Bagh) had stairways, which were similar to ramps with tread and riser dimensions of 30 cm and 2 cm respectively. Table 3 also indicates that most of the FOBs other than those in Bengaluru and Delhi lacked either ramps/lifts/escalators. Due to a lesser number of FOBs in Guwahati, only one FOB was found suitable for this study.

Similarly, in the case of Mumbai, the FOB situated outside the Indian Institute of Technology (IIT) Bombay campus was only considered. Due to time constraint and the commonality between the FOBs in Ghaziabad, only one of the overpasses was selected to represent the other FOBs.

Moreover, from Table 3, it was also observed that majority of the respondents across different locations were male pedestrians (67–76%). The pedestrian in the age category of 23–59 years were regular users of the overpasses. Also, pedestrians using luggage were found to be higher (5–17%) than the ones without luggage for all the different cities considered.

2.2 Questionnaire design

A questionnaire set was prepared, including three broad sections (A to C) representing demography, the current condition of FOBs, and future usability dependents (refer Table 4).

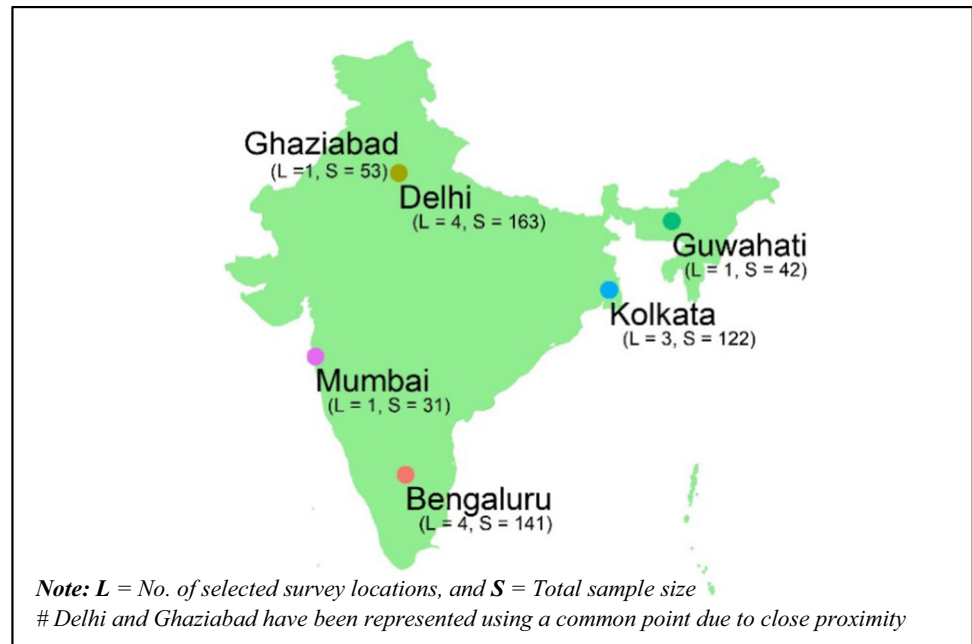
As per Table 4, Section A covered demographic characteristics including gender, age, profession, and frequency of daily use. Likewise, section B focused on capturing

Table 2 Application of soft computing approaches in various empirical studies across different domain

Authors	Dataset/source	Study type and description	Method used	Evaluation metrics	Findings
Al-Janabi et al. (2020a, b)	35 air quality monitoring stations	Environment: Prediction of six air pollutants (PM2.5, PM10, PM, NO2, CO and O3), Big Data Analysis	LSTM, SAQPM	SMAPE and tenfold CV	-DSN-PS could be used to identify best parameters for LSTM@ -DNS-PS and DLSTM reduces execution time but increases complexity
Al-Janabi et al. (2020a, b)	Five data sets from environment friendly database	Environment: Electrical energy generation from wind using multi-objective renewable energy generation (MORE-G) model	Deep learning techniques (DCapsNet and DCOM)	Error	Optimization method used to solve multi-objective function problems with different constraints
Al-Janabi and Andkaim (2020)	Six data sets	Intelligent data analysis: Novel tool development (DRFLLS) to estimate missing values estimation	Random forest and local least squares (LLS)	Pearson correlation (PC) and NRMSE	DRFLLS showed a higher performance and accuracy with regard to missing value problem
Mahdi and Al-Janabi (2020)	http://www.heritagehealthprize.com	Health Care: Predicting next year patient admission duration (days)	CART, MARS, RFRC, BTCR, CHAIR, E-CHAID, Bayesian Network, SVM, Self-Organization Network	RMSE, MSE, MAE, MAPE	HPC-based prediction algorithm is fast, accurate and could be used as alternative to other high performing algorithms.
Alam et al. (2019)	University of California at Irvine Machine Learning Repository	Epidemiology: Identifying predictors for 10 different diseases	Feature Ranking Algorithms, Random Forest	Accuracy, precision, recall, F-measure, AUROC, AUPR, RMSE	Feature ranking and selection-based approach increased model performance than without feature ranking-based selection
Zhou et al. (2018)	200 samples	Risk Assessment: Risk prediction of deep foundation pit in subway station	Random Forest	Accuracy	RF is useful for predicting risk classification and can be used with small sample and unbalanced datasets
Al-Janabi et al. (2018)	800 survey response	Recommender System: Built recommender system for improving time management in educational environments	Gaussian Mixture Model	Accuracy, Precision, Recall, F-Measure, RMSE, MSE, MAPE	The system provided better understanding of focus areas for better time management
Al-Janabi and Al-Shourbaji (2016)	760 survey responses	Educational: Understanding of information security awareness within educational environments	CART, Questionnaire Survey	Margin of error, ROC	Identified lack of knowledge of information security concepts, as well as a low level of awareness within the Educational Environments
Belgiu and Drăguț (6)	–	Remote Sensing: Review the utilization of RF classifier in remote sensing	Random Forest	–	The RF classifier performs better in classifying hyperspectral data and perform with higher dimensionality and highly correlated data.
Singh et al. (2016a, b)	502 experimental data sets	Environment: Predicting Vehicular noise level using soft computing techniques	Generalized Linear Model, Decision Trees, Random Forests and Neural Networks	R-square, Accuracy, MSE	RF algorithm most accurately predicted vehicles noise level compared to other algorithms

Table 2 continued

Authors	Dataset/source	Study type and description	Method used	Evaluation metrics	Findings
Patel et al. (2015)	–	Web App-Security: Developed a novel security framework to ensure secured against malicious and harmful application programs	Risk Filtering Data Mining algorithm (RFDM)	Accuracy, F-beta measure, precision, recall	RFDM algorithm enhances the quality of analyzing the risk by reducing false risks.

Fig. 1 Indian cities selected for the final survey

existing connectivity (i.e., the connection from one end of the stairway to the other), security (i.e., in the form of security personnel and CCTV cameras), comfort (i.e., shade, proper guardrails and cleanliness), walk environment (i.e., governed by the facility surroundings and whether they were pleasant or not) and obstruction (i.e., presence of hawkers/vendors, beggars, and standing pedestrians). The pedestrians were further asked whether the condition regarding obstruction, safety and security, vertical end connectivity (lift/escalator/ramp), and horizontal end connectivity improvement (refer Table 4, Section C) would govern their future usability. Further, a set of field measurements were noted down by the observers (see Table 4, section D).

2.3 Questionnaire survey

After finalizing the survey locations, the field observers (two in number for each location) gathered necessary

details of the FOBs including GPS coordinates and dimensions of the walkways and stairways. An interviewer-administered questionnaire survey (by two interviewers) of the pedestrians in the neighborhood of the facility was conducted at each survey location for weekdays between morning (8.30–11 am) and evening (5–7.30 pm) peak hours, respectively, to obtain a representative sample. During the survey, both set of participants (including users and non-users of FOBs) were randomly selected (using random sampling technique) and requested for the survey participation; and those willing to undergo the interview process were finally interviewed. Due to the massive rush in morning and evening peak hours, the participation rate was low (i.e., out of approximately twenty random pedestrians, only one participated when requested). Among all participants, only 552 respondents answered all the questions, thoroughly. Later, in the laboratory, these 552 questionnaire samples were manually entered into an excel sheet according to the final

Table 3 Site and sample characteristics

City	Location	Land-use type	Length (m)	Width (m)	Number of steps	Stair width (m)	Tread (cm)	Riser (cm)	Facility for vertical movement	Sample size	City wise characteristics of Respondents								
											Gender (%)		Age (%)			Luggage (%)			
											M	F	<12	13–22	23–45	46–59	>60	With	Without
Bengaluru	Tin factory bus stand	Commercial	49.8	2.6	50	2.09	25	12	S	141	69.5	31.5	2.1	26.9	40.5	21.9	8.6	54.6	45.4
	Yeshwantpur Station	Public Transport Terminal (PTT)	31.1	3.1	54	1.10	26	9	S, L										
	Marathahalli	Commercial	32.1	3.4	44	1.50	32	16	S, L										
	Christ University	Educational	21.2	2.8	39	3.10	29	14	S, L										
Mumbai	IIT main gate	Educational	46.5	3.6	52	1.60	32	17	S	31	70.9	21.1	0.0	54.8	19.4	19.4	6.4	58.1	41.9
Kolkata	Ultadanga	Commercial	42.4	3.1	41	2.40	26	20	S	122	75.4	24.6	3.2	19.7	41.8	25.4	9.9	54.9	45.1
	Lake Town	Residential	42.2	2.9	19	2.90	30	15	S, R										
	Sealdah	Commercial	36.9	2.4	53	2.10	30	15	S										
Delhi	Anand Vihar I.S.B.T.	PTT	61.5	5.7	48	2.30	30	16	S, E	163	73.6	26.4	2.4	17.7	38.6	33.1	8.2	52.7	47.3
	Akshardham	PTT	66.5	2.2	48	2.30	31	16	S, E										
	ITO	Commercial	39.2	2.6	117	2.20	31	2	R*, E										
	Maharani Bagh	Residential	41.3	2.6	115	2.80	30	2	R*, E										
Ghaziabad	Vaishali Metro	PTT	88.1	3.5	50	3.70	30	14	S, L	53	73.5	26.5	1.8	18.8	37.7	33.9	7.8	52.9	47.1
Guwahati	Maligaon	Commercial	33.8	2.05	55	1.22	28	8	S	42	66.7	33.3	2.3	14.3	35.7	33.3	14.4	54.7	45.3

S Stairway, L lift/elevator, R ramp, R* ramp like stairway, E escalator, M male, F female

Table 4 Questionnaire survey format and field observations

Section	Title	Description
A	Demographic	Gender (0: Female, 1: Male), Age (0: < 12; 1: 13–22; 2: 23–45; 3: 46–59 and 4: > 60), Profession (0: Businessman, 1: Homemaker, 2: Others, 3: Retired, 4: Self-employed, 5: Service, 6: Student), Frequency of daily use (0: First time, 1: Occasionally, 2: Once, 3: Twice, 4: More than twice)
B	Existing conditions	Width, Surface, Connectivity, Safety and Security, Comfort, Walk Environment (0: Poor, 1: Satisfactory, 2: Good, 3: Very Good, 4: Excellent), Obstruction (0: None, 1: Some, 2: Many)
C	Future usability dependents	Obstruction removal or relocation (0: No/1: Yes), CCTV installation and security personnel deployment (0: No/1: Yes), Vertical end connectivity (lift/escalator/ramp) installation and maintenance (0: No/1: Yes), Horizontal end connectivity improvement (0: No/1: Yes)
D	Field observations	Length of FOB, Walkable width of FOB (indicated as width value), Number of steps, Stairway width, Tread dimension, Riser dimension, Land-use type (0: Commercial, 1: Residential, 2: Public Transport Terminal, 3: Educational), Type of facility for vertical movement (0: Stairway, 1: Escalator, 2: Lift, 3: Ramp)

analysis requirement and used for the final data analysis and modeling.

3 Data analysis

The demographic characteristics were obtained by performing exploratory data analysis on final prepared survey data containing 552 samples, using “tidyverse” package (Wickham 2017) in the R statistical programming environment. It was expected to get different behavior at various land-use types. Thus, instead of a city-wise analysis, the final analysis was carried out comparing the usability between different land-use types (commercial, educational, public transport terminal, and residential).

3.1 Demographic characteristics

The demographic parameters including gender, age, profession, and frequency of daily use are essential in understanding the existing usage pattern of pedestrians. Table 5 shows the demographic characteristics for different land-use types.

Table 5 shows that majority of the participants were male pedestrians (~ 65–75%). The variation of gender proportion within each gender group was found to be minimal. The highest proportion of male respondents were observed at residential locations (74.4%), while within the female gender group, the highest proportion of female pedestrians (34.8%) were observed at educational locations.

Majority (~ 55%) of the pedestrians using the overpass facility were at educational locations, between the age group of 13–22 years (refer Table 5). The finding is consistent with the study findings of Desriani and Komordjaja (2008) and Guo et al. (2014), which revealed that the utilization rate of FOBs was highest among young pedestrians

when the facilities were situated in educational locations. Similarly, for the other three land-use types, majority of the users (~ 40–42%) were in the age group of 23–45 years, which is also consistent with the report of Ministry of Statistics and Programme Implementation (2018) which states that India holds the highest proportion of the young population (i.e., around 242 million). It is further noticeable that the usage rate decreased with the increase in age, which is in accordance with the findings reported by Räsänen et al. (2007). The population of pedestrians below 12 years of age at educational locations was low, as parents tend to drop them using private vehicles or they travel by buses to school, instead of walking to their destinations using FOBs.

The statistics shown in Table 5 revealed that the majority of the pedestrians (~ 36–46%) were regular users, as they were using the FOBs daily twice or more than twice. The frequency of users using the facility daily once was low (~ 12–17%), as people using the FOBs generally used the facility for both ways of their trips (for example, from residence to workplace and back). There were very few first-timers (~ 2–6%) who used the facility for commuting.

Based on Table 5, it was also observed that majority of the regular users in educational and residential areas consisted of predominantly students (54.6% in educational location), while in commercial and PTTs, the proportion of servicemen (38–41%) and self-employed personnel (14–17%) were high. The results were similar to the previous findings reported by Saha et al. (2011) and Wu et al. (2014) that the tendency of using FOBs increased with higher education and better employment. The percentage of retired persons (age > 60 years) was significantly low throughout all the locations, but a small group of homemakers was noticeable in commercial (6.4%) and residential (8.1%) areas. The next subsection gives a vivid discussion about the respondents’ response ratings

Table 5 Demographic characteristics of respondents' under different land-use types

Demographic characteristics	Land-use type			
	Commercial	Educational	PTT	Residential
Gender (%)				
Male	73.20	65.20	72.80	74.40
Female	26.80	34.80	27.20	25.60
Age (%)				
≤ 12	2.70	0.00	1.10	5.80
13–22	15.00	54.50	21.10	19.80
23–45	40.00	22.70	40.60	41.90
46–59	31.80	16.70	28.90	24.40
≥ 60	10.50	6.10	8.30	8.10
Daily frequency (%)				
First time user	5.49	3.00	3.90	2.30
Occasional user	22.30	25.80	22.80	27.90
Daily once	16.40	12.10	15.60	12.80
Daily twice	38.20	45.50	36.10	37.20
More than twice	17.70	13.60	21.70	19.80
Profession (%)				
Business	10.50	9.11	11.70	9.30
Home maker	6.39	1.50	4.40	8.10
Retired	5.00	3.00	2.80	3.50
Self-employed	14.10	9.11	17.20	11.60
Service	40.50	21.20	38.30	27.90
Student	19.10	54.60	21.70	32.60
Others	4.50	1.50	3.90	7.00

regarding perceived satisfaction/dissatisfaction, on the existing condition of FOBs.

3.2 Satisfaction of existing FOBs

For understanding how much pedestrians were satisfied/dissatisfied with the available features in existing FOBs, seven quality assessment parameters were asked to the participants across different land-use types. These parameters were the comfort, connectivity, safety and security, surface condition, walk environment, width, and obstruction (as described in Sect. 2.2). The factors were selected from a list of physical and user characteristics provided by IRC: 103 (2012) for pedestrian facilities. The rating of these parameters were obtained from poor (0) to excellent (4) categories. The participants' responses from all land-use types were combined and analyzed together as shown in Fig. 2.

The response statistics from Fig. 2 revealed that safety and security was the most critical parameters, which the pedestrians perceived to be 'poor' in commercial (29.6%), residential (20.9%) and PTT (20.6%) locations. The finding was also consistent with the feedback (i.e., regarding

existing issues) provided by the respondents (presented in Table 6).

The users in commercial locations followed by the users in residential locations expressed the highest dissatisfaction related to comfort (19.1%), surface (21.4%) and walk environment (20.9%). Respondents' satisfaction regarding connectivity between FOBs with their desired destination was perceived to be extremely poor in the commercial locations (9.1%), while in educational, PTT and residential land-use types, the users expressed satisfaction with the existing connectivity.

Further respondents' felt that the walkable width and perceived obstructions (refer Figs. 2, 3) were not satisfactory for comfortable movement at commercial (17.7% and 23.6% respectively), PTT (9.4% and 4.4% respectively) and residential (5.8% and 29.1% respectively) locations. The dissatisfaction was mainly because during peak hours; these locations were crowded by standing pedestrians/vendors/beggars (see Table 6), reducing the effective walkway width and resulting in increased mobility friction.

The field measurement of FOBs indicated that the mean walkway width was minimum at commercial ($\bar{x} = 2.69$ m and $\sigma = 0.49$) and residential ($\bar{x} = 2.75$ m and $\sigma = 0.21$) locations, which could cause higher mobility friction.



Fig. 2 Perceived satisfaction/dissatisfaction on existing features of FOBs

Table 6 Respondents’ common feedback concerning different issues in existing FOBs

Features	Feedbacks	Feedback count across different land-use types			
		Commercial	Residential	Educational	PTT
Safety and security	Absence of security personnel/CCTV	26	8	3	17
	Antisocial activities and pickpocket	0	0	0	10
	Unsafe and uneasiness at night	3	1	0	0
Lift and escalator	Unavailable	34	8	6	12
	Improper functioning	7	6	3	9
Ramp	Not working	1	0	1	6
Ramp	Unavailable	9	4	0	6
Obstruction	Vendors & beggars	3	7	2	5
	Standing pedestrians	5	0	0	9
Cleanliness & Maintenance	Dirty pathways	10	4	0	5
	No regular scheduled maintenance	16	5	1	7
Lighting	Unavailable	10	0	0	0
	Improper functioning	0	0	0	2
	Not working	1	0	0	5
Shade	Unavailable	12	0	0	13
	Damaged and not repaired	5	5	0	1

Further, at PTT locations the mean width was found to be highest ($\bar{x} = 3.63$ and $\sigma = 1.49$), but due to the high peak hour flow and presence of mobility frictions (i.e., presence of standing pedestrians/beggars/vendors), the effective walkway width was fully utilized, and the space available to the pedestrians reduced substantially. These facts seemed to decrease the satisfaction regarding width and gave a sense of obstruction or acted as friction to a small group of pedestrians (4.4%).

4 Modeling approaches

An effort was made to understand the best-suited model (in terms of accuracy), which could predict the future usability of the foot over bridges under four different contexts (such as mobility, safety and security, vertical end connectivity and horizontal end connectivity). The different modeling approaches used in the current study were generalized linear modeling (GLM), random forest (RF), and gradient

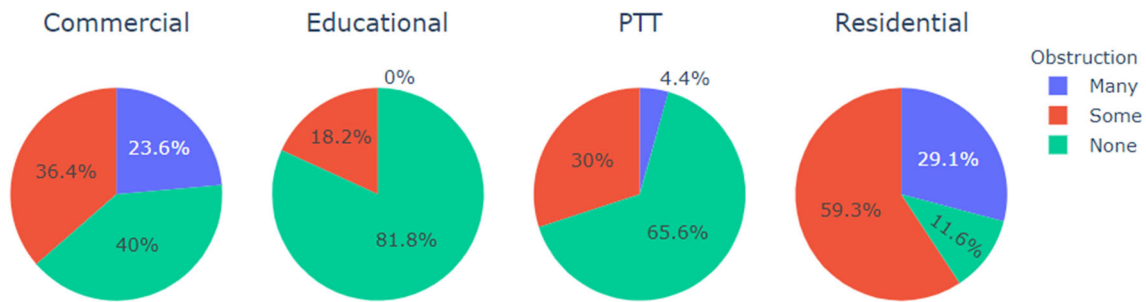


Fig. 3 Perceived obstruction across different land-use types

boosting machine (GBM). A short description of the models is provided in the following subsections.

4.1 Generalized linear modeling (GLM) framework

The GLM estimates regression models for outcomes following an exponential distribution. The structural form of the model describes the patterns of interactions and associations. The two categories of models produced by GLM are classification and regression. Binary logistic regression is the form of GLM which performs binary classification and estimates whether a probability characteristic is present (i.e., estimates the binary class probability). The form in which binary logistic regression is used is shown in Eq. 1.

$$P_i(Y_i = 1|X_i = x_i) = \frac{\text{EXP}[\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i]}{1 + \text{EXP}[\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i]} \quad (1)$$

where P_i is the probability of whether an occurrence happens or not; X is the observed value of explanatory variables, which can be discrete, continuous, or a combination of both; and β is the regression coefficient. The GLM-based algorithms are easy to train, though suffers from overfitting related issues. Thus, Lasso- or Ridge-based penalties are used to reduce the overfitting issues.

4.2 Random forest (RF) modeling framework

A decision tree is the simplest, perhaps most easily understandable algorithm used in Machine Learning (ML). A decision tree is similar to a flowchart, and Breiman (2017) first implemented this decision tree algorithm for classification and regression. Decision trees are mostly used for decision-making and predictive modeling.

A simple example can be where a person is choosing between two available alternative infrastructures, i.e., FOB and another available route, for crossing a busy road, based on different available features, shown in Fig. 4. In a decision tree, the top node is called “root node” and the node at the bottom is called “terminal node”. The other

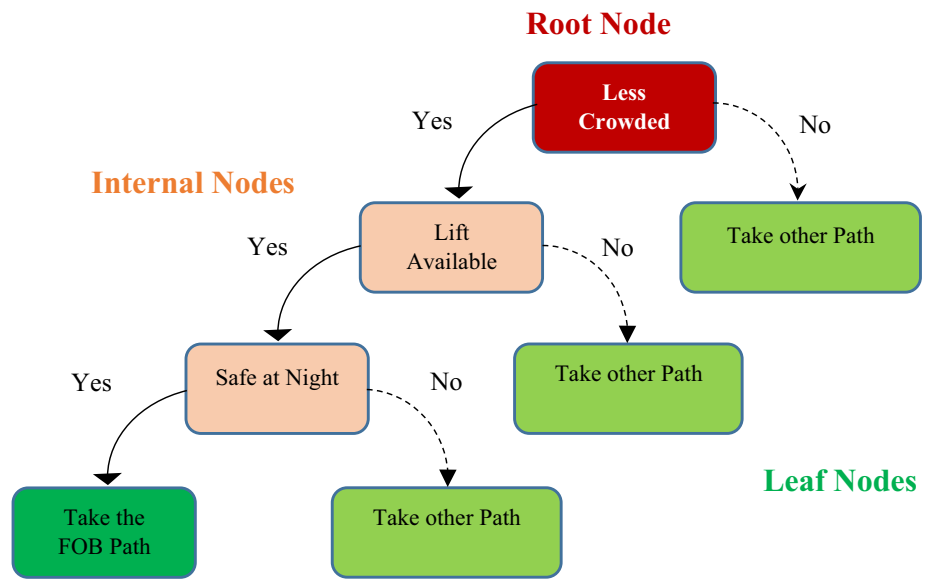
nodes except the root node and terminal node are called “internal nodes”. Each internal node includes a binary test condition, while each leaf node contains associated class labels (for binary choice example, yes/no).

A classification tree uses a split condition to predict a class label based on the supplied one or more input variables. The splitting process starts from the root node, and at each node, it checks whether supplied input values recursively continue to the right or left sub-branch as per a splitting condition. This process stops when a leaf or terminal node is reached. Most mathematical algorithms use an impurity measure as a splitting criterion, and one of the common impurity measures is the Gini Index (Breiman 2001). Lower the Gini value; higher will be the purity of split (refer to glossary section). The other splitting criteria are entropy (information gain) and misclassification rate.

Though a single decision tree-based model is easy to build and interpret, yet it suffers from various drawbacks, such as high variance (as often a small change in the training data results in very different splits) and overfitting due to deep grown fragile tree structure. To overcome these drawbacks, ensemble learners were introduced (Breiman 2001). Ensemble learning is a method where more than one model is built to gain higher prediction accuracy. A voting mechanism is used to aggregate the results from all models. For classification, the ‘majority voting technique’ is used, where each ensemble is asked to predict the class label. Once all the classifiers are queried, the class that received the highest number of votes returns as the final decision of the ensemble learner.

There are three popular ensemble learning algorithms such as Bagging (Boot Strap Aggregation), Random Forest and Boosting. Random forest is one of the widely used ensemble learners used in Ecology, Medicine, Astronomy, Autopsy, Agriculture, Bioinformatics, and Traffic and Transportation planning (Fawagreh et al. 2014). Though RF is widely popular, still limited researchers have used this approach for solving classification and regression problems in traffic and transport planning domains (see Table 7).

Fig. 4 A decision tree for choosing between two alternative routes



RF was mainly used by different researchers in crash prediction analysis, apart from travel demand forecasting and mode choice modelling (refer Table 7).

The random forest modeling approach uses a bootstrapping (means sampling, ‘m’ number of rows/observations at random from the ‘n’ size training dataset with replacement) analogy similar to bagging, but adds additional randomness to models at each split by randomly selecting input variables/features (often referred as “mtries”, where mtries < available number of variables) using a feature bagging method (Cook 2016; Yu-Wei 2015), refer Fig. 5. As there are fewer number of variables available to choose from, hence less information is available to the model during the training process. Thus, this additional randomness makes trained trees more different from each other, in other words, this makes less correlated trees which improves the prediction performance (Breiman 2001). Though RF provides accurate solution, still it is computationally very expensive and difficult to interpret.

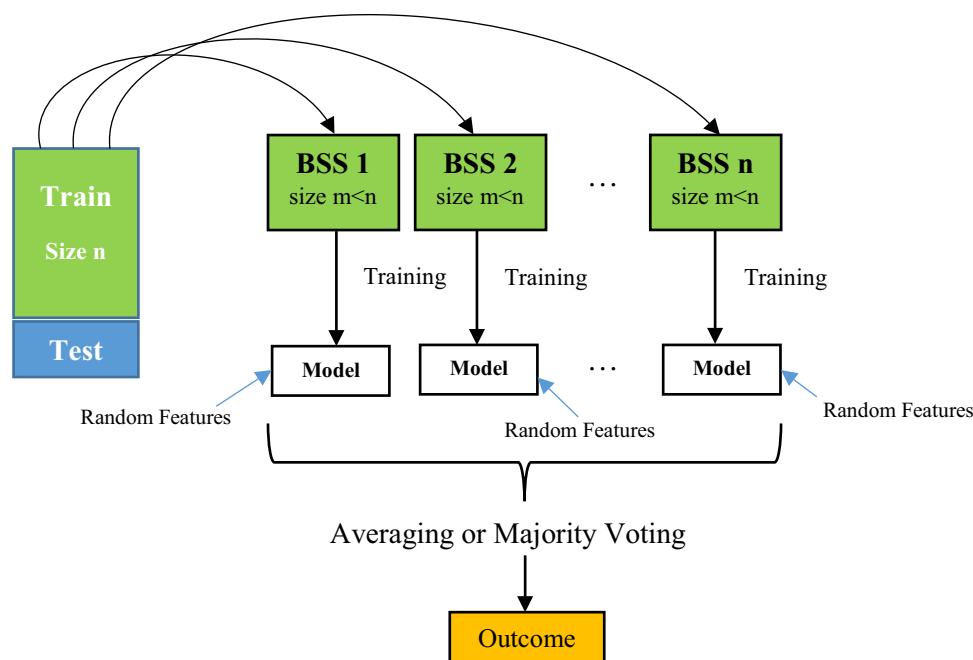
4.3 Gradient boosting machine (GBM) framework

GBM is another decision tree-based ensemble method (like RF) for regression and classification; and which primarily focusses on difficult rows of training (i.e., the ones that are hard to learn). GBM and RF differ in the way the trees are built and the way the results are combined. In RF, each tree is trained independently using random sample data; whereas in GBM one tree is built at a time, where each new tree helps to correct errors made by the previously trained tree. While in RF, mtries (i.e., number of variables to randomly choose as candidate at each split) and ntrees (number of trees to make) are the most important parameters which need tuning; in GBM, parameters such as ntrees, max_depth (how deep each tree will be allowed to grow) and learning rate (weighting factor applied for new trees when added to the model to slow down the learning) are the most prominent parameters which need modification. Studies have shown that GBM performs better than

Table 7 Studies in traffic and transportation planning related to random forest modeling approach

Author’s name	Application
Hossain and Muromachi (2011)	Used in understanding crash hazards on urban expressways
Pande et al. (2011)	Used to understand the importance of essential predictors of crash occurrence
Rasouli and Timmermans (2014)	Used model ensembles for predicting travel demand forecasting
Ermagun et al. (2015)	Used for predicting the decision of escorting and modes of transport for school trips
Chalumuri et al. (2016)	Used for modeling mode choice of commuters in Delhi (India), and results highlighted that RF-based models outperformed the multi-nominal logit model.

Fig. 5 The random forest modelling approach



Note: BSS: BootStrap Sample

RF, as GBM tries to add new trees that complement the already built ones. Similar to GLM and RF-based algorithms, the boosting algorithms has limitations too. GBM is more focused toward bias correction than variance, and is computationally more expensive.

Similar to RF, GBM has been less explored in the field of transportation and planning. Table 8 shows the studies which have been conducted on GBM in the related field.

Table 8 shows that majority of the studies used GBM to predict travel time and traffic patterns.

4.4 Study methodology

The step-by-step methodology adopted for modeling usability is illustrated in Algorithm 1. The study methodology involved literature survey, preliminary site inspection and questionnaire design, data collection and extraction, followed by modeling of usability from four

different contexts, and finally extracting the important features for policy decision.

As explained in Sect. 2, the data was collected from fourteen FOB locations across India. This data must be handled carefully before using them for prediction model. Incomplete or partial data sets were ignored from the analysis. Afterward, normalization was applied to each column of the data set using min–max scalar. Normalized dataset was randomly divided into 80:20 ratio for training and testing of the developed model. A hyper-parameter grid was selected based on the past test experiments. A tenfold cross-validation approach was applied on the 80% dataset for model and hyper-parameter tuning. To obtain faster solution, a random grid search was adapted. The Area Under Curve (AUC) metric was selected for model performance evaluation due to class imbalance in the outcome variable. For faster training an early AUC-based stopping criterion was adopted. If AUC did not improve by 0.1% for the ten successive models the model training and

Table 8 Studies in traffic and transportation planning related to gradient boosting machine approach

Author's name	Application
Regue and Recker (2014)	Used in predicting the bike-sharing station states
Li and Bai (2016)	Used to predict freight vehicle travel time
Yang et al. (2017)	Used in predicting short-term traffic
Ling et al. (2018)	Used for prediction of urban railways transit system (subway) passenger flows under different traffic conditions
Cheng et al. (2019a, b)	Used for prediction of travel time on freeways

hyper-parameter tuning stopped and next successive search started. The final best performing model then extracted and tested on the remaining 20% unseen/test dataset.

4.6 Study hypothesis and limitations

The present study predicts the pedestrian usability of FOBs under four different contexts. Majority of the studies,

Algorithm #1: FOB use or non-use prediction

Input: Demographic, frequency of use, existing condition, usability dependent, land use type and facility, walkway length and width, steps and stair characteristics.

Output: Predicting the use or non-use of FOB facility

```

// Pre-Processing Stage
1. For column in FOB dataset
2. | Call handle missing values
3. | Call normalize
4. End For
// Build the Models
4. For i in range (1: total samples)
5. | Split dataset into 80% training and 20% testing
6. | Split 80% dataset according to 10-Cross Validation Training and Testing dataset
7. End for
8. For each 10-CV training part
9. | For i in range (total number hyper-parameters combination)
10. | | for each i based model training
11. | | IF AUROC improves <0.1% for the ten successive models //stopping criteria
12. | | Stop & Go to step 9
13. | End for
14. | Call evaluate AUC
15. End for
// Evaluation stage
16. For each trained models
17. | Evaluate on the remaining 20% test dataset
18. | Sort model based on AUC criteria
19. | Save best model
20. | Compute and rank variable importance
21. End for

```

4.5 Future usability model development

To obtain the essential parameters that determine the use of FOBs, in the current study three different modeling approaches (GLM, RF, and GBM) were explored to predict the future usability determinants of FOBs (binary outcomes, Yes/No) using open-source statistical programming language R (version 3.4.3), under four different contexts. The four contexts include obstruction removal and relocation (Model 1), CCTV installation and security personnel deployment (Model 2), vertical end connectivity improvement in terms of lift/escalator/ramp installation and maintenance (Model 3) and horizontal end connectivity improvement (Model 4) corresponding to their selected predictors, presented in Table 9.

The samples split used was 80:20, and thus 443 samples were used for training the models and 109 for testing them (Table 9).

which were conducted to identify the pedestrian behavioral attributes (to improve the usability of elevated facilities), were mainly focused on understanding the relevant factors attributed to the use of elevated facilities under a single context only, and those limited studies emphasized on accurate solutions for reliable estimates for policy decision. Thus, the current study tried to fill this gap. It is hypothesized here that similar to other study domain, in pedestrian research; high-end machine learning (ML) algorithms could be used to obtain accurate solution that helps in policy decision-making. Therefore, the objective of the current study was to come up with reasonably accurate solution for FOB utilization modelling for policy makers, rather than real-time choice decision-making system for FOB use using four different contexts.

Due to the advancement in algorithms and hardware, researchers were able to develop and test various ML-based algorithms to solve similar problems, but in research trying and comparing all type of techniques is neither possible nor practical. Thus, in the current study, common algorithm (GLM) as well as advanced algorithms

Table 9 Model parameters for modeling the usability of FOBs

Model no.	Context	Future usability	Predictors	Sample split
Model 1	Obstruction (mobility friction)	0: No/1: Yes	Gender, age, daily frequency, safety and security, comfort, walk environment, location type, effective width, effective width value, stair width	<i>Training</i> = 80% ($n = 443$) and <i>Testing</i> 20% ($n = 109$)
Model 2	Safety and security	0: No/1: Yes	Gender, age, daily frequency, obstruction, walk environment, location type	
Model 3	Vertical end connectivity (lift/escalator/ramp)	0: No/1: Yes	Gender, age, presence or absence of luggage, daily frequency, connectivity, comfort, location type, number of steps, stair width, tread dimension, riser dimension	
Model 4	Horizontal end connectivity	0: No/1: Yes	Gender, age, luggage, daily frequency, comfort, location type, length of FOB	

(ensembles: RF and GBM) are used to get an accurate solution and thus identify the factors that influence the usability of FOBs.

4.7 Model configuration and hyper-parameters

In order to train the models and to use early stopping criteria, an open-source R package named “H2O” was used (H2O 2017). For training and performance testing of the models, a total of 552 samples were randomly split into 80% ($n = 443$) train and 20% ($n = 109$) test dataset. To identify the best set of model parameters and to avoid unnecessary search, a randomized grid search approach (randomDiscrete) was used in the case of RF and GBM methods, as this helps in minimizing the computational time (Bergstra and Bengia 2012). The grid search parameters are illustrated in Table 10.

A random grid search was performed to get the best set of parameter combinations that would provide better prediction accuracy. In the random grid search, a number of trees were tried ranging from 100 to 500 for both RF and GBM. Instead of default mtries (H2O default, square root of the number of variables for classification), a range of mtries for each of the four models were used in RF, reported in Table 10. Further, the column sample rate at tree level and the sampling rate was varied from 0.5 to 1.0. The maximum tree depth was fixed at 40 for RF; while it was varied between 2 and 10 at an interval of 2 for GBM. In the case of GBM, a learning rate of 0.01 was used.

For minimizing the training time and to avoid model overfitting, among all available early stopping criteria offered by H2O package (such as misclassification, logloss, MSE, and AUC), the AUC-based early stopping criterion was selected. The AUC-based early stopping criteria was applied with the condition that if in the successive ten models, the AUROC (Area Under Receiver Operating Curve) does not improve by 0.1%, then H2O stops further

grid search. Simultaneously, per-model level early stopping criteria was applied with the condition that if training goes with ten scoring rounds without any improvement at all (stopping tolerance = 0) in the AUC, then it stops. Due to smaller sample size ($n = 552$), instead of creating a separate validation data set, tenfold cross-validation criteria with a random fold assignment were used to get a better and reliable estimate of the trained models.

4.8 Model training and performance testing

The models were trained in R environment with the training samples ($n = 443$) using H2O package. In total 180 models were generated for Model 1 (predicting FOBs use in the context of future improvement in mobility friction) using RF and 225 models using GBM. Consecutively, for Model 2 (predicting FOBs use in the context of enhancement in safety and security), Model 3 (predicting usability in the context of future improvement in lift/escalator/ramp) and Model 4 (predicting usability in the context of future horizontal end connectivity improvement), RF generated 90, 47 and 135 models; while GBM generated 225, 225 and 450 models respectively. Next, all the generated models were sorted in decreasing order according to the AUC value, and the models with the highest AUC value were selected as the final optimized model. The final model summary is illustrated in Table 11.

The optimized model results (from Table 11) revealed that both RF and GBM identified the accurate solution using less than 220 trees. RF modes used max depth ranging from 10 to 16, while GBM achieved the best optimized model within maximum depth of 4 to 10. Additionally, the model trained on different context utilized the whole range of column sample rate per tree and sample rate, thus no distinct pattern was observed.

The various statistical measures used to measure the statistical significance of the models are Accuracy,

Table 10 Grid search parameters and early stopping criteria

<i>Hyper-parameter search criteria</i>				
	Random forest		Gradient boosting machine	
ntrees	100, 200, 300, 400, 500		100, 200, 300, 400, 500	
mtries	Model1 = 4, 6, 8, 10	Model2 = 4, 6	–	
	Model3 = 4, 6, 8, 10	Model4 = 3, 5, 7		
Max Tree Depth	40		2, 4, 6, 8, 10	
Sample rate	0.5, 0.7, 0.9		0.5, 0.7, 0.9	
Column sample rate per tree	0.5, 0.8, 1.0		0.5, 0.8, 1.0	
Learning rate	–		0.01	
<i>Early Stopping Criteria</i>				
	Random forest		Gradient Boosting Machine	
	On the grid	On the models	On the grid	On the models
Stopping metric	AUC	AUC	AUC	AUC
Stopping tolerance	0.001	0	0.001	0
Stopping round	10	10	10	10
Score tree interval	–	5	–	5

Table 11 Summary of the best model hyper-parameters

Optimized Parameters	Random forest models				Gradient boosting machine models			
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
ntrees	100	151	147	167	217	205	131	138
mtries	4	6	10	3	–	–	–	–
max depth	16	10	13	14	6	10	6	4
Column sample rate per tree	0.5	0.5	0.8	0.5	1.0	0.8	0.5	0.5
Sample rate	0.7	0.5	0.7	0.5	0.5	0.5	0.9	0.7
Learning rate	–	–	–	–	0.01	0.01	0.01	0.01

LogLoss, AUC (Area Under Curve), MSE (Mean Squared Error) and F_{beta} measure (good performance measure used for unbalanced classes; in this study only estimated during model performance on test data set), which are described in detail in the glossary section. Further, model prediction accuracy was separately estimated and reported. The model performance on the training data is shown in Table 12.

The results from Table 12 shows that in case of all the four models, GBM reasonably performed best in comparison to GLM and RF. Also, the cross-validation estimates for each final model (illustrated in Table 13) for RF and GBM showed that AUC ranged between 0.72 and 0.98 (in case of RF) and between 0.72 and 0.97 (in case of GBM), which indicated an overall good model prediction in case of both the approaches.

Further, for obtaining model performance on unseen test dataset, the final models were tested on the remaining 20%

($n = 109$) test dataset. Table 14 shows the model performance summary on the test data set.

The performance summary revealed that the overall optimized models using GBM (based on F_{beta} measure: 0.61–0.97) performed best in comparison to GLM (F_{beta} measure: 0.51–0.93) and RF (F_{beta} measure: 0.54–0.97) methods on the same test data as illustrated in Table 14.

4.9 Applications of advanced soft computing techniques in transportation and its comparison with the current study

Application of different advanced soft computing techniques in the transportation-engineering domain is presented in summary format in Table 15. Results of the present study highlighted that GBM could be one of the best choices for modeling the usability of pedestrian FOBs.

Table 12 Summary of model performance on training data

Modeling technique	Evaluation parameters	Accuracy	LogLoss	AUC	MSE
GLM	Model 1	0.7156	0.5528	0.7771	0.1865
	Model 2	0.7720	0.4994	0.8388	0.1629
	Model 3	0.9278	0.2454	0.9604	0.0742
	Model 4	0.7743	0.4087	0.7896	0.1304
RF	Model 1	0.7043	0.5248	0.7883	0.1769
	Model 2	0.7404	0.5500	0.8174	0.1831
	Model 3	0.9594	0.1217	0.9847	0.0356
	Model 4	0.7156	0.4424	0.7328	0.1416
GBM	Model 1	0.8804	0.3858	0.9475	0.1154
	Model 2	0.7968	0.4555	0.8934	0.1431
	Model 3	0.9774	0.1670	0.9927	0.0420
	Model 4	0.7743	0.4271	0.8451	0.1355

Table 13 Summary of tenfold cross-validation mean estimates of final models

Modeling approach	Evaluation parameters	Accuracy	LogLoss	Precision	Recall	AUC	MSE
RF	Model 1	0.7708	0.5145	0.6395	0.8623	0.8088	0.1636
	Model 2	0.7894	0.5482	0.7722	0.8901	0.8286	0.1819
	Model 3	0.9745	0.1213	0.9864	0.9832	0.9842	0.0362
	Model 4	0.7346	0.4433	0.4259	0.7431	0.7298	0.1407
GBM	Model 1	0.7835	0.5343	0.6587	0.8381	0.7943	0.1794
	Model 2	0.7930	0.5234	0.7767	0.9014	0.8246	0.1723
	Model 3	0.9631	0.2042	0.9627	0.9955	0.9759	0.0572
	Model 4	0.7454	0.4704	0.4572	0.7218	0.7258	0.1512

Table 14 Summary of model performance estimated on the test dataset

Modeling technique	Evaluation parameters	Accuracy	AUC	MSE	Precision	Recall	F-measure
GLM	Model 1	0.8073	0.8344	0.1704	0.7750	0.7209	0.7470
	Model 2	0.7982	0.8101	0.1801	0.7273	0.8511	0.7843
	Model 3	0.8991	0.9325	0.0839	0.9011	0.9762	0.9371
	Model 4	0.7523	0.6942	0.1477	0.6667	0.4118	0.5091
RF	Model 1	0.7982	0.8300	0.1587	0.6750	0.7500	0.7105
	Model 2	0.7431	0.8000	0.1879	0.8909	0.6901	0.7749
	Model 3	0.9633	0.9700	0.0415	0.9780	0.9780	0.9780
	Model 4	0.8165	0.7300	0.1422	0.5714	0.5217	0.5455
GBM	Model 1	0.8349	0.8800	0.1424	0.7750	0.7750	0.7750
	Model 2	0.7431	0.8200	0.1775	0.8727	0.6957	0.7742
	Model 3	0.9633	0.9800	0.0499	0.9780	0.9780	0.9780
	Model 4	0.8349	0.8500	0.1246	0.6667	0.5600	0.6087

The current GBM model showed a classification power ranging from 77.42 to 97.80%. As per past studies (refer Table 15), better prediction accuracy was obtained using boosting-based algorithms on different study domain. For

example, study by Ha et al. (2019) revealed, GBM could predict travel mode choice behavior with 95.1% accuracy. Similarly, Mousa et al. (2018), modeled the lane changing behavior with an advanced version of boosting-based

algorithm called XGBoost and achieved an astonishing accurate model (99.7% accuracy). The other studies proved the effectiveness of boosting-based algorithms in different study domain, as mentioned in Table 15.

Similar to other domain, in pedestrian-based researches, the use of such advanced algorithms is limited. Hence, the current study results tried to fill this gap, and showed the effectiveness of such algorithms in pedestrian research,

Table 15 Different studies in transportation sector using advanced soft computing techniques

Author, Country	Type of study	Techniques used	Preprocessing techniques	Type of dataset used	Evaluation metrics	Advantage and disadvantages of the method used
Cheng et al. (2019a, b), China	Travel mode choice	RF, SVM, AdaBoost, MNL	Training (80%) and testing (20%)	Household survey for travel information, demographics and impact of built environment	Accuracy and MAPE	RF and SVM were the best models predicting travel mode with 85.36% and 83.44%.
Cheng et al. (2019a, b), China	Travel time prediction of freeway	GBDT, BPNN, SVM	Training (Days 27–133), Testing (Day 1–26)	133 days dataset with 11 VISSIM generated variables	MAD, MAPE, RE	Prediction accuracy of GBDT (both training and testing) with respect to MAPE was better, as it was least sensitive to prediction horizons
Ha et al. (2019), Indonesia	Travel mode choice	GBM, DNN, Hyper-parameter, CV	Trip duration, commuter travel characteristics and locations details	Data was collected in 2000 by Japan International Cooperation Agency (JICA)	Relative importance plot	GBM outperformed DNN in case of single mode (95.1%), two modes (84.1%) and three modes (75.95%) data testing
Alajali et al. (2018), Australia	Traffic prediction at Intersection	GBRT, RF and XGB	Date, time, traffic volume, day of week, day or night, peak or not and sensor locations obtained	Sensor data, accident data and roadwork data obtained from VicRoads	MSE and MAE	The MSE values were quite similar with respect to GBRT (0.6721), RF (0.6874) and XGB (0.6721)
Mousa et al. (2018), USA	Lane changing maneuvers	DT, RF, GBM and XGB, Grid search, tenfold CV	421 vehicle data with vehicle position, lanes, vehicle dimensions and lateral distances were prepared	NGSIM vehicle trajectory data	AUC	XGB outperformed other algorithms with 99.7% accuracy and was used to extract feature importance
Yang et al. (2017), China	Short-term traffic forecast	GBM, SVM and BPNN	Train (eight-week data) and test (last week)	Caltrans Performance Measurement System (PeMS) for 9 loop detectors	MAPE, MAE	GBM performed better than SVM and BPNN
Ding et al. (2016), China	Driver's stop-or-run behavior	GBM	Vehicles' behavior for three preceding vehicles collected using effects from preceding vehicle decision	Three-month data collected from detectors across different intersections	R ²	GBM is proposed as it could better handle different predictor variables and fit complex non-linear relationships
Zhang and Haghani (2015), USA	Travel time prediction	RT, GBM, RF	With only 1% missing data, the study simply replaced the missing values with the mean of the closest surrounding values	Two months training data and seven days testing data from Regional Integrated Transportation Information System (RITIS) website	MAPE	The prediction accuracy by the GBM model was significantly higher in 14 out of 18 cases

DT decision tree, *RF* random forest, *GBM* gradient boosting machine, *SVM* support vector machine, *XGB* extreme gradient boosting, *MNL* multinomial logit model, *DNN* deep neural networks

which could act as a better alternative when model quality (accurate model) is the main goal.

5 Variable importance analysis

In the modeling process, variable importance was further estimated for each optimized model using the GBM method. The importance was estimated by calculating the relative influence of each variable: whether the variable was selected during the splitting in the tree building process and how much the squared error decreased. The variable importance obtained from each final selected model is illustrated in Fig. 6. All scaled importance ranging low (0) to high (1) for each selected factors were arranged in descending order to get the most important factors influencing the FOBs use, as illustrated in Fig. 6.

5.1 Future use concerning mobility friction improvement

In the future usability model corresponding to mobility friction (model 1: obstruction), the importance plot revealed that stair width and daily frequency of use were the most crucial factors which influenced the use of FOBs (illustrated in Fig. 6, Plot A). Usually, narrow stair width and high pedestrian flow increased the mobility friction. Even the presence of vendors/beggars/standing pedestrians reduced the effective walkway width, which was further responsible for the reduction in the use of the FOBs, and this was in accordance with the findings reported by Pasha et al. (2015). During the survey, a set of feedbacks were obtained from respondents (refer Table 6) which also confirmed that the mobility friction was land-use specific and occurred majorly at PTT, commercial and residential locations. A study by Saha et al. (2011), confirmed that in the Central Business District (CBD) and PTT locations, unwanted people's presence discouraged the use of overpass. It was further noticed that the type of location, age, existing walk environment, and security further also played a vital role in usability choice. Though the presence of people (vendors, standing pedestrians, and beggars) usually gave a sense of safety and security, yet this might not be true under all scenarios. During night time, the presence of vendors and beggars might not encourage pedestrians to use FOBs due to the prevalence of illegal activities and fear of victimization (see feedback Table 6). The study of Malik et al. (2017) also highlighted a similar concern, which made people feel insecure in the presence of many beggars and shops.

5.2 Usability concerning safety and security

Perceived safety and security is one of the most prominent factors related to overpass use (Räsänen et al. 2007). The future usability prediction model (Model 2) concerning perceived safety and security (i.e. CCTV installation and security personnel deployment) revealed walk environment (i.e., the facility surroundings and whether they are pleasant or not) and gender as the most crucial predictors that influenced the use of FOBs across Indian cities (see Fig. 6, plot B). The significant determinants of safety and security differed based on the time when pedestrians used the facility. At daytime, safety and security-related concerns arose among people in highly crowded areas (such as PTT), where the pickpocket and theft-related issues were frequently noticeable (see feedback Table 6). At night time, the security-related issues (such as walk environment being uncomfortable due to illegal activities) were primarily gender-specific (related to female pedestrians), which debarred them from using the FOBs (Malik et al. 2017; Pasha et al. 2015). Moreover, the absence of CCTV and security personnel at PTT and commercial locations (see feedback Table 6) demotivated users and increased the perceived fear of victimization. Insufficient security regarding the absence of lighting (Malik et al. 2017; Pasha et al. 2015), perceived insecurity of getting robbed (Hasan and Napiyah 2014; Malik et al. 2017; Villaveces et al. 2012) and criminal activities (Villaveces et al. 2012, Saha et al. 2011) were previously found to deter the use of FOBs among pedestrians. Past studies, as well as current respondents' expressed concerns also supports the fact that to enhance safety and security, most common measures are the provision of proper lighting (Hasan and Napiyah 2014; Malik et al. 2017; Villaveces et al. 2012; Pasha et al. 2015) and removal of advertisement banners (Koepsell et al. 2002; Malik et al. 2017; Oviedo-Trespalacios and Scott-Parker 2017). Moreover, the provision of surveillance systems such as CCTV cameras along with proper placement of security personnel (Gallegos 2012) and stricter laws (Hidalgo-Solórzano et al. 2010; Sabet 2013) also strengthen the safety perception among pedestrians and motivate them to use the facility frequently.

5.3 Usability concerning vertical end connectivity (lift/escalator/ramp)

The variable importance obtained from Model 3 (predicting usability concerning vertical end connectivity) illustrated future usability was highly related to the design-related parameters such as steepness of stairs (Mutto et al. 2002; Sabet 2013; Saha et al. 2011), narrow width of stairs (Hasan and Napiyah 2014) and absence of escalators/ramps

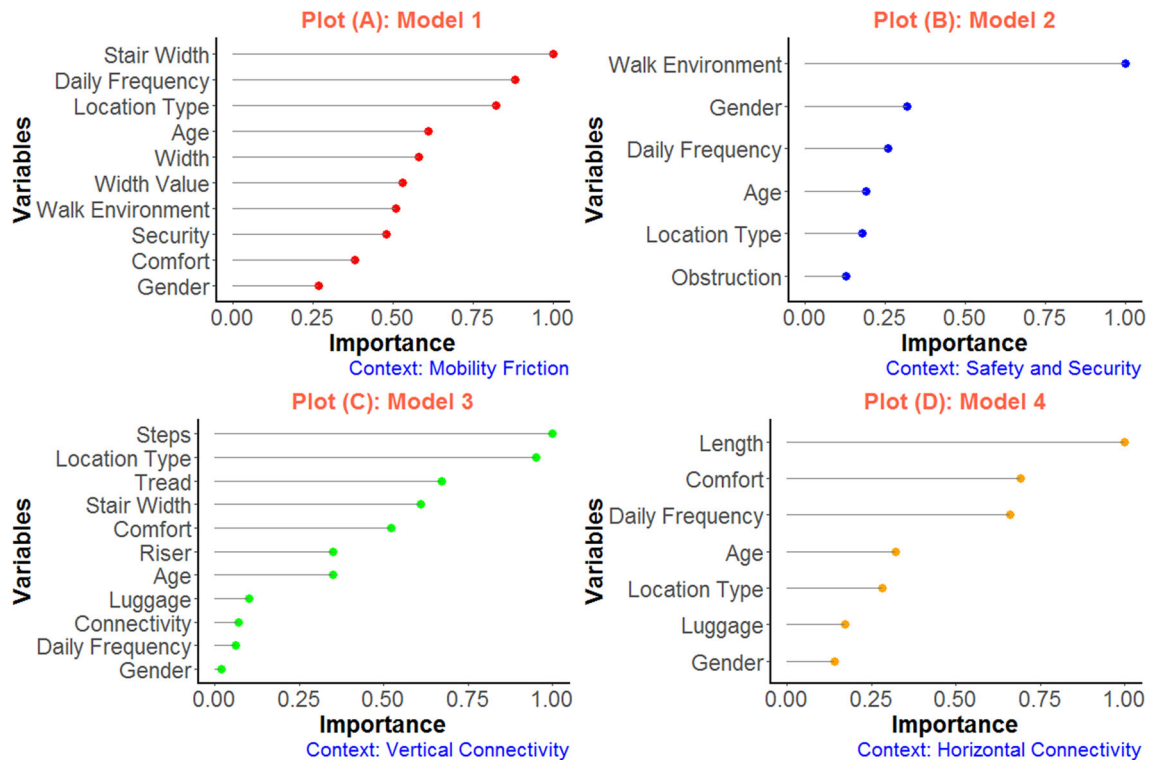


Fig. 6 Variable importance plot (scaled importance)

(Desriani and Komardjaja 2008; Hasan and Napiah 2014; Räsänen et al. 2007; Rizati et al. 2013) which caused discomfort to the users (Sharples and Fletcher 2001; Hasan and Napiah 2014, 2018; Saha et al. 2011). These reported facts also support the findings obtained from Model 3 (refer Fig. 6, plot C), where the dimension of the stairway (i.e., number of steps, tread dimension, width, and riser dimension) and comfort played a significant role in deciding the future use of the FOBs. Further, location type, i.e., where the facility is situated (see Fig. 6, plot C) also played a pivotal role in the choice of use. This finding is consistent with the previous studies, where researchers reported that the usability of FOBs in commercial and educational areas was significantly higher than residential and shopping areas (Desriani and Komardjaja 2008; Hasan and Napiah 2017; Räsänen et al. 2007; Rizati et al. 2013). Further, the age of the respondent was found out to be a prevalent determinant, as with the increase in age pedestrians felt uncomfortable in putting extra effort to climb stairs (Rankavat and Tiwari 2016). Provision of short stairs, ramps (Desriani and Komardjaja 2008) and lift/escalator (Demiroz et al. 2015; Hasan and Napiah 2018; Räsänen et al. 2007), while designing new elevated facilities and on existing FOBs would reduce extra effort and enhance comfort, leading to an increase in the use of FOBs.

5.4 Usability concerning horizontal end connectivity

The future usability concerning the horizontal end connectivity (Model 4) was majorly influenced by the length of travel, comfort, daily frequency of use and age of the pedestrian (see Fig. 6, plot D). The traveled length was found out to be the most influencing factor in deciding the use of the FOBs. Mutto et al. (2002) in their study highlighted the fact that extra traveled distance negatively influenced pedestrian behavior while choosing a crossing facility. Other studies also used similar predictors, but instead of traveled distance, traveled time was used as a function of covered distance, which was also directly related to the perceived comfort. Past studies identified that when the traveled length increased significantly as compared to at-grade facilities and the time needed to cover the distance was more than 50% (Anciaes and Jones 2018; Hasan and Napiah 2014; Malik et al. 2017; Rankavat and Tiwari 2016; Wu et al. 2014) people tried to avoid FOBs and used nearest illegal exits available by attempting to cross through median openings or jumping over fences (Demiroz et al. 2015; Desriani and Komardjaja 2008). This indicated that providing additional length and detour distances would discourage pedestrians from FOB use. The current model finding also revealed that pedestrians' perceived comfort was another important factor. The comfort

was derived based on whether proper horizontal end connectivity gave the pedestrians easy and direct access toward their destination or not. In this regard, proper signboards mentioning the connecting locations (i.e., entry and exit location name) would likely give initial information to pedestrians about their destination and provide comfort to existing as well as new users (Desriani and Komardjaja 2008).

6 Sensitivity analysis

In order to understand the robustness of the developed model, sensitivity analysis was carried out. In the sensitivity analysis the top parameter of context 1 (refer Fig. 7), i.e., stair width was varied between 0.66 and 3.74 m with a base value of 2.22 m (i.e., varied between -70 and $+70\%$). The other parameters from daily frequency to gender (refer Fig. 6, plot A, model 1), were fixed as per the frequency of the questionnaire survey.

The result of the sensitivity of Model 1 (as per Fig. 7) showed that for both gender and age groups (23–59 years), the usability starts increasing when the stair width is above 1.5 m, which is mainly due to the fact that in wider FOBs (above 1.5 meters) pedestrians feel more comfortable to use the elevated facility. As most stairways are open for bi-directional movement, and under Indian scenario average shoulder depth is considered to be 60 cm (Singh et al. 2016a, b), hence when two pedestrians are moving in opposite direction, a stair width above 1.5 m is preferable. Further the usability of FOB increases most for young male.

Similarly, as per context 2 (i.e. safety and security) sensitivity analysis, two perception scenarios (i.e., good and bad) were tested to understand the sensitivity of gender category, which is reported as one of the important predictors. The first scenario considered was good, where pedestrians use the facility twice a day, believe the safety and security is satisfactory, and are subjected to few obstructions. Similarly, the alternative scenario is bad, where pedestrians use the facility occasionally, believe the safety and security is poor and there are presence of many obstructions. Both the scenarios represent FOBs situated in commercial zone.

The result of the sensitivity test for Model 2 (refer Fig. 8) showed that compared to good perception when pedestrians had bad perception/experience regarding the facility, the preference or usability dropped and the magnitude of reduction in usability is more for old age people.

Sensitivity test for Model 3 (vertical connectivity) and Model 4 (horizontal connectivity) were also conducted. In the vertical connectivity, it was found that as number of steps increases pedestrians need to put more effort during

climbing up as well as down which eventually demotivates pedestrians to use the facility, and the highest impact is observed on the choice of aged pedestrians. Similarly, sensitivity analysis of horizontal connectivity revealed a negative relationship between length and facility use, this is because when travel length becomes significantly longer compared to at-grade traverse distance, pedestrians might feel negligent towards using the grade-separated facilities.

7 Conclusion

In the current study, information was obtained through interviewer-administered questionnaire survey and field measurement sessions near fourteen overpasses or Foot Over Bridges (FOBs) locations under different land-use types (commercial, residential, educational, and public transport terminal) across six Indian cities. In total, 552 valid survey samples were collected from the respondents. Analysis results revealed that the majority of pedestrians who were using the FOBs were young (13–45 years), regular users (used twice or more than twice daily) and mostly comprised of male (~ 65 – 75%) pedestrians. The career of the users further influenced the preference of use. The students and working professionals were more likely to use the FOBs than other profession types.

In the present study, machine learning techniques such as generalized linear model (GLM), random forest (RF) and gradient boosting machine (GBM) learning algorithms were compared to find the optimal solution that accurately predicts primary factors affecting the use of FOBs. Among different machine learning techniques, GBM outperformed the other two in terms of prediction performance on unknown (test) data set for identifying the essential parameters affecting pedestrians' choice of using the elevated facility under four different contexts (i.e., mobility friction, safety and security, vertical end connectivity and horizontal end connectivity). The major conclusions drawn from this study were described as follows:

- (a) One of the most crucial factors that decided the usability was safety and security, which was gender-specific and depended on the existing walk environment. The feedback provided by the respondents revealed that the unavailability of CCTV cameras and security personnel, along with the prevalence of antisocial activities were significant concerns that influenced the usability choice.
- (b) The age of the pedestrian played a significant role in the choice of FOBs when the decision was derived regarding ease in vertical movement (climbing stairs)

Fig. 7 Sensitivity analysis for Model 1 (Context: Mobility Friction)

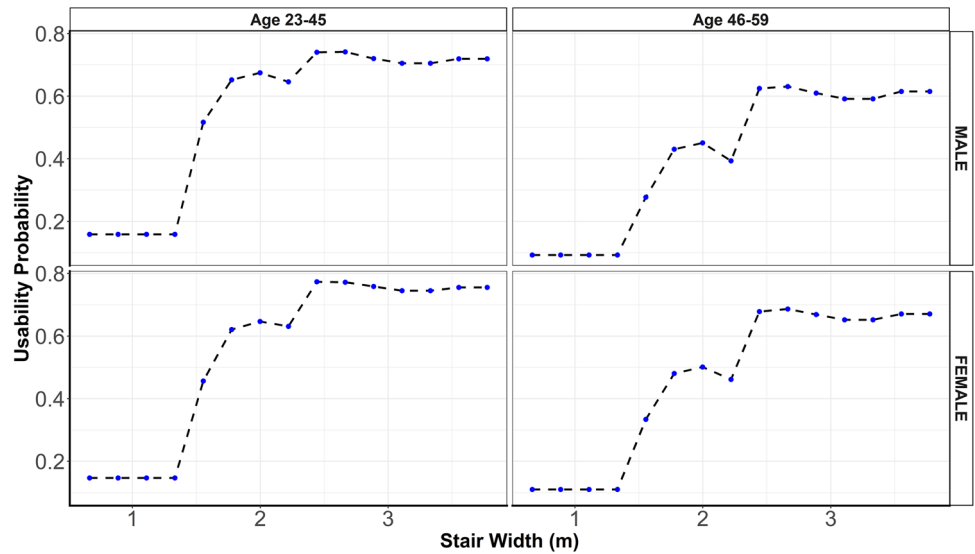
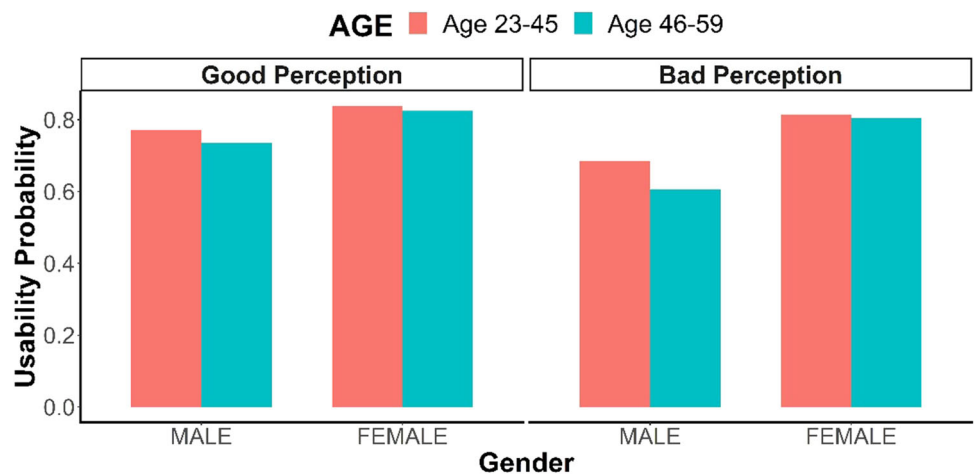


Fig. 8 Sensitivity analysis for Model 2 (Context: Safety and Security)



- and when longer travel distance (length) was a matter of concern.
- (c) Gender acted as a significant influencer when the usability choice was solely dependent on perceived safety & security.
- (d) Sensitivity analysis showed that stair width above 1.5 m increased the usability preference of pedestrians.
- (e) Overall good perception/experience regarding the facility, increased the usability preference.
- (f) As the number of steps to climb increased, the usability preference decreased among aged pedestrians.
- (g) The design parameters of FOB such as the number of stairs to climb, stair width, width of the walkway and length of the FOB, were associated with the perceived comfort and also determined whether a pedestrian would choose the facility. The provision of stairs with short risers and escalator/lift/ramp

- would enhance the comfort of pedestrians and motivate them to use FOBs more frequently.
- (h) The feedback provided by the users revealed that apart from improving security, removing obstructions and installing lifts/escalators also affect the decision of FOB use. The respondents further expressed their concern regarding an immediate need for proper lighting and shade, along with

Table 16 Binary contingency table

<i>Reference</i>		
Predicted	Relevant	Irrelevant
Relevant	A	B
Irrelevant	C	D

regular maintenance of the facilities to increase the future usability.

8 Advantages of boosting ML model in FOB research

The use of advanced soft computing models (boosting-based algorithms) reduces the prediction uncertainty in pedestrian research. The boosting-based algorithm works well compared to RF- or GLM-based algorithms, as the nodes in every tree take different subset of features for making best split which makes them uncorrelated. Additionally, each new tree considers and corrects the errors or mistakes made by the previous trees which helps in achieving better prediction accuracy.

The finding of these models would ultimately provide new insight into factors associated with the use of FOBs for researchers, planners, and policymakers. The outcomes of this study could be used to improve the existing users' experience (by improving the existing facilities or constructing better elevated facilities in the future), which could encourage them to use the facility more often and attract new pedestrians to use FOBs. To achieve these goals, it is also essential to make the facilities more attractive and user friendly, which provides the pedestrians a safe and comfortable crossing experience. By attracting new pedestrians and enabling existing users to use the facility continuously, could indirectly reduce at-grade road crossing risk. Further, awareness campaigns and strict enforcements come as handy tools to eradicate accidents due to the illegal road crossing.

9 Study limitations and future scope

Some of the significant challenges in the current study were low response rate, duration of data collection (i.e., restricted to a single day), language diversity, and the number of locations covered.

In future, research can be further extended by considering the broader population and covering more types of locations (e.g., institutional and recreational areas) across other Indian cities under different climatic conditions. Additionally, studies can be carried out by comparing both at-grade and grade-separated pedestrian facilities (FOBs or subways), considering both questionnaire and videography surveying techniques.

9.1 Glossary

Area Under Curve (AUC) Performance measure used for binary classifiers. Typically range from 0.5 to 1.0, higher the value better the model performance. Obtained from the plot between true positive rate and true negative rate.

Cross-Validation (CV) In the cross-validation, the total training sample is divided into k blocks, and each of the k blocks used as a validation dataset and rest used for training. The process repeats k times with different parts of the training dataset being the validation set each time. The error of the final model is obtained regarding the average value of all k models.

Pruning Pruning is a technique to chop off sections that are not very powerful in classifying examples to avoid overfitting and helps in improving model prediction accuracy.

Gini Index In tree-based models, Gini Index used as one of the common impurity measures in features splitting (Fawagreh et al. 2014) shown in Eq. 2. Lower the Gini value; higher will be the purity of split.

$$\text{Gini}(t) = 1 - \sum_{i=1}^N P(C_i/t)^2 \quad (2)$$

where t is a condition, N the number of classes in the data set, and C_i is the i th class label in the data set.

F-measure The *F-measure* is defined as the weighted harmonic mean of its precision and recall (Buckland and Gey 1994; Powers 2007; Yu-Wei 2015; Golakiya et al. 2019), shown in Eq. 3. A value closer to one indicates a better performing classification model. The measure for precision and recall is illustrated as a binary contingency table (refer to Table 16).

where Precision and Recall can be defined as:

$$\text{Precision} = \frac{A}{A+B} \text{ and, } \text{Recall} = \frac{A}{A+C}$$

Similarly,

$$F_{\text{measure}} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

Acknowledgements This research work was funded by CSIR - Central Road Research Institute (CRRI), New Delhi under the Supra Institutional Network Project (SINP) category as part of the 12th Five Year Plan Period (FYP).

Compliance with ethical standards

Conflict of interest The authors declare that there is no actual or potential conflict of interest in relation to this article.

References

- Abojaradeh M (2013) Evaluation of pedestrian bridges and pedestrian safety in Jordan. *Civ Environ Res* 3(1):66–79
- Alajali W, Zhou W, Wen S, Wang Y (2018) Intersection traffic prediction using decision tree models. *Symmetry* 10(9):386. <https://doi.org/10.3390/sym10090386>
- Alam MD, Rahman MS, Rahman MS (2019) A random forest based predictor for medical data classification using feature ranking. *Inform Med Unlocked*. <https://doi.org/10.1016/j.imu.2019.100180>
- Al-Janabi S, Alkaim AF (2020) A nifty collaborative analysis to predicting a novel tool (DRFLLS) for missing values estimation. *Soft Comput* 24:555–569. <https://doi.org/10.1007/s00500-019-03972-x>
- Al-Janabi S, Al-Shourbaji I (2016) A study of cyber security awareness in educational environment in the middle east. *J Inf Knowl Manag*. <https://doi.org/10.1142/s0219649216500076>
- Al-Janabi S, Salman MS, Fanfakh A (2018) Recommendation system to improve time management for people in education environments. *J Eng Appl Sci*. <https://doi.org/10.3923/jeasci.2018.10182.10193>
- Al-Janabi S, Alkaim AF, Adel Z (2020a) An Innovative synthesis of deep learning techniques (DCapsNet & DCOM) for generation electrical renewable energy from wind energy. *Soft Comput* 24:10943–10962. <https://doi.org/10.1007/s00500-020-04905-9>
- Al-Janabi S, Mohammad M, Al-Sultan A (2020b) A new method for prediction of air pollution based on intelligent computation. *Soft Comput* 24(1):661–680. <https://doi.org/10.1007/s00500-019-04495-1>
- Ancaies PR, Jones P (2018) Estimating preferences for different types of pedestrian crossing facilities. *Transp Res Part F Traffic Psychol Behav* 52:222–237. <https://doi.org/10.1016/j.trf.2017.11.025>
- Belgiu M, Dragut L (2016) Random forest in remote sensing: a review of applications and future directions. *ISPRS J Photogram Remote Sens*. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
- Bergstra J, Bengia Y (2012) Random search for hyper-parameter optimization. *J Mach Learn Res* 13:281–305
- Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32. <https://doi.org/10.1023/A1010933404324>
- Breiman L (2017) *Classification and regression trees*. Routledge, Abingdon
- Buckland M, Gey F (1994) The relationship between Recall and Precision. *J Am Soc Inf Sci* 45(1):12–19. [https://doi.org/10.1002/\(SICI\)1097-4571\(199401\)45:1%3c12:AID-ASI2%3e3.0.CO;2-L](https://doi.org/10.1002/(SICI)1097-4571(199401)45:1%3c12:AID-ASI2%3e3.0.CO;2-L)
- Central Statistics Office, Ministry of Statistics and Programme Implementation, Government of India (2018) *Women and Men in India: a statistical compilation of gender related indicators in India*. http://www.mospi.gov.in/sites/default/files/publication_reports/Women%20and%20Men%20in%20India%202018.pdf
- Chalumuri RS, Minal S, Madhu E (2016) Multimodal choice modeling using random forest decision trees. *Int J Traffic Transp Eng*. [https://doi.org/10.7708/ijtte.2016.6\(3\).10](https://doi.org/10.7708/ijtte.2016.6(3).10)
- Cheng J, Li G, Chen X (2019a) Research on travel time prediction model of freeway based on gradient boosting decision tree. *IEEE Access* 7:7466–7480. <https://doi.org/10.1109/ACCESS.2018.2886549>
- Cheng L, Chen X, De Vos J, Lai X, Witlox F (2019b) Applying a random forest method approach to model travel mode choice behavior. *Travel Behav Soc* 14:1–10. <https://doi.org/10.1016/j.tbs.2018.09.002>
- Cook D (2016) *Practical machine learning with H2O: powerful, scalable techniques for deep learning and AI*. (First Edit). O'Reilly Media, Incorporated
- Couronné R, Probst P, Boulesteix AL (2018) Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinform* 19(1):1–14. <https://doi.org/10.1186/s12859-018-2264-5>
- Das A, Barua S (2015) A survey study for user attributes on foot over bridges in perspective of Dhaka City. In: *International conference on recent innovation in civil engineering for sustainable development (IICSD-2015)*, pp 1–6. <https://www.slideshare.net/SauravBarua6/te-009a-survey-study-for-user-attributes-on-foot-over-bridges-in-perspective-of-dhaka-city>
- Demiroz YI, Onelcin P, Alver Y (2015) Illegal road crossing behavior of pedestrians at overpass locations: factors affecting gap acceptance, crossing times and overpass use. *Accid Anal Prev* 80:220–228. <https://doi.org/10.1016/j.aap.2015.04.018>
- Desriani RW, Komardjaja I (2008) Assessing pedestrian facilities on crossing bridges. In: *2nd international conference on built environment in developing countries*, pp 1617–1626. <http://eprints.usm.my/34884/1/HBP21.pdf>
- Ding C, Wu X, Yu G, Wang Y (2016) A gradient boosting logit model to investigate driver's stop-or-run behavior at signalized intersections using high-resolution traffic data. *Transp Res Part C Emerg Technol* 72:225–238. <https://doi.org/10.1016/j.trc.2016.09.016>
- Ermagun A, Rashidi TH, Lari ZA (2015) Mode choice for school trips long-term planning and impact of modal specification on policy assessments. *Transp Res Record* 2513:97–105. <https://doi.org/10.3141/2513-12>
- Fawagreh K, Gaber MM, Elyan E (2014) Random forests: from early developments to recent advancements. *Syst Sci Control Eng* 2(1):602–609. <https://doi.org/10.1080/21642583.2014.956265>
- Gallegos WLA (2012) Motives of disuse of pedestrian bridges in Arequipa. *Revista Cubana de Salud Pública* 38(1):84–97
- Golakiya HD, Patkar M, Dhamaniya A (2019) Impact of midblock pedestrian crossing on speed characteristics and capacity of urban arterials. *Arab J Sci Eng*. <https://doi.org/10.1007/s13369-019-03786-x>
- Guo H, Zhao F, Wang W, Zhou Y, Zhang Y, Wets G (2014) Modeling the perceptions and preferences of pedestrians on crossing facilities. *Discrete Dyn Nat Soc*. <https://doi.org/10.1155/2014/949475>
- Ha TV, Asada T, Arimura M (2019) The application of gradient boost machine and local interpretable modelagnostic explanations in examining the travel multi-mode choice, the case study of Jakarta City, Indonesia. *J Eastern Asia Soc Transp Stud* 13:503–522. <https://doi.org/10.11175/easts.13.503>
- Hasan R, Napiah M (2014) The effect of structure and street characteristics on the footbridge usage. *J Appl Sci Agric* 9(21):52–59
- Hasan R, Napiah M (2017) Pedestrians' behavior towards the use of footbridges under the impact of motivational alerting posters: the case of Ipoh city. *Adv Transp Stud, Malaysia*. <https://doi.org/10.4399/978882550352458>
- Hasan R, Napiah M (2018) The perception of Malaysian pedestrians toward the use of footbridges. *Traffic Injury Prev* 19(3):292–297. <https://doi.org/10.1080/15389588.2017.1373768>
- Herms BF (1972) Pedestrian crosswalk study: accidents in painted and unpainted crosswalks. *Highway Res Rec* 406:1–13
- Hidalgo-Solórzano E, Campuzano-Rincón J, Rodríguez-Hernández JM, Chías-Becerril L, Reséndiz-López H, Sánchez-Restrepo H, Baranda-Sepúlveda B, Franco-Arias C, Híjar M (2010) Use and non-use of pedestrian bridges in Mexico City: the pedestrian perspective. *Salud Publica Mex* 52(6):502–510. <https://doi.org/10.1590/s0036-36342010000600004>

- Hossain M, Muromachi Y (2011) Understanding crash mechanisms and selecting interventions to mitigate real-time hazards on urban expressways. *Transp Res Rec* 2213(1):53–62. <https://doi.org/10.3141/2213-08>
- IRC: 103 (2012) Guidelines for pedestrian facility. Indian Roads Congress, India
- Irvine CH, Snook SH, Sparshatt JH (1990) Stairway risers and treads: acceptable and preferred dimensions. *Appl Ergon* 21(3):215–225. [https://doi.org/10.1016/0003-6870\(90\)90005-1](https://doi.org/10.1016/0003-6870(90)90005-1)
- Juan MC, Pérez D (2009) Comparison of the Levels of presence and anxiety in an acrophobic environment viewed via HMD or CAVE. *Presence Teleoper Virt Environ* 18(3):232–248. <https://doi.org/10.1162/pres.18.3.232>
- Kang Y, Fukahori K, Kubota Y (2018) Evaluation of the influence of roadside non-walking spaces on the pedestrian environment of a Japanese urban street. *Sustain Cities Soc* 43:21–31. <https://doi.org/10.1016/j.scs.2018.07.009>
- Koepsell T, McCloskey L, Wolf M, Moudon AV, Buchner D, Kraus J, Patterson M (2002) Crosswalk markings and the risk of pedestrian–motor vehicle collisions in older pedestrians. *JAMA* 288(17):2136–2143. <https://doi.org/10.1001/jama.288.17.2136>
- Li B (2013) A model of pedestrians' intended waiting times for street crossings at signalized intersections. *Transp Res Part B Methodol* 51:17–28. <https://doi.org/10.1016/j.trb.2013.02.002>
- Li X, Bai R (2016) Freight vehicle travel time prediction using gradient boosting regression tree. In: 15th IEEE international conference on machine learning and applications (ICMLA), pp 1010–1015. <https://doi.org/10.1109/icmla.2016.0182>
- Ling X, Huang Z, Wang C, Zhang F, Wang P (2018) Predicting subway passenger flows under different traffic conditions. *PLoS ONE* 13(8):e0202707. <https://doi.org/10.1371/journal.pone.0202707>
- Mahdi MA, Al-Janabi S (2020) A novel software to improve healthcare base on predictive analytics and mobile services for cloud data centers. In: Farhaoui Y (ed) *Big Data and networks technologies: BDNT 2019*, vol 81. Lecture notes in networks and systems. Springer, Cham
- Malik IA, Alwi SKK, Gul N (2017) A survey to understand people perception of pedestrian bridges usage on Shahrah-e-Faisal road, Karachi-Pakistan. *New Horizons* 11(1):111
- Methorst R (2004) *Vulnerable road users: report on the knowledge base for an effective policy to promote the safe mobility of vulnerable road users*. Rotterdam, the Netherlands, AVV Transport Research Centre. <http://publicaties.minienm.nl/documenten/vulnerable-road-users-report-on-the-knowledge-base-for-an-effect>
- Mohan D (2009) Road accidents in India. *IATSS Res* 33(1):75. [https://doi.org/10.1016/S0386-1112\(14\)60239-9](https://doi.org/10.1016/S0386-1112(14)60239-9)
- Mohan D, Tiwari G, Bhalla K (2017) Road safety in India: status report 2016. *Transportation Research & Injury Prevention Programme*, Indian Institute of Technology Delhi. http://tripp.iitd.ernet.in/assets/publication/2016_India_Safety_status1.pdf
- Mousa SR, Bakhit PR, Osman OA, Ishak S (2018) A comparative analysis of tree-based ensemble methods for detecting imminent lane change maneuvers in connected vehicle environments. *Transp Res Record* 2672(42):268–279. <https://doi.org/10.1177/361198118780204>
- Mutto M, Kobusingye OC, Lett RR (2002) The effect of an overpass on pedestrian injuries on a major highway in Kampala-Uganda. *African Health Sci* 2(3):89–93
- Opdyke D, Williford JS, North M (1995) Effectiveness of computer-generated (virtual reality) graded exposure in the treatment of acrophobia. *Am J Psychiatry* 1(152):626–628. <https://doi.org/10.1176/ajp.152.4.626>
- Oviedo-Trespalacios O, Scott-Parker B (2017) Footbridge usage in high-traffic flow highways: the intersection of safety and security in pedestrian decision-making. *Transp Res Part F Traffic Psychol Behav* 49:177–187. <https://doi.org/10.1016/j.trf.2017.06.010>
- Pande A, Das A, Abdel-Aty M, Hassan H (2011) Estimation of real-time crash risk: are all freeways created equal? *Transp Res Rec* 2237(1):60–66. <https://doi.org/10.3141/2237-07>
- Pasha MM, Rifaat SM, Hasnat A, Rahman I (2015) Pedestrian's behaviour on road crossing facilities. *J Teknol*. <https://doi.org/10.11113/jt.v73.4292>
- Patel A, Al-Janabi S, AlShourbaji I, Pedersen J (2015) A novel methodology towards a trusted environment in mashup web applications. *Comput Secur* 49:107–122. <https://doi.org/10.1016/j.cose.2014.10.009>
- Patra M, Perumal V, Rao KK (2020) Modelling the effects of risk factor and time savings on pedestrians' choice of crossing facilities at signalised intersections. *Case Stud Transp Policy* 8(2):460–470. <https://doi.org/10.1016/j.cstp.2019.10.010>
- Powers DMW (2007) Evaluation: from precision, recall and F factor to ROC, informedness, markedness and correlation, Technical Report SIE-07-001, Flinders University. https://bioinfopublication.org/files/articles/2_1_1_JMLT.pdf
- Rankavat S, Tiwari G (2016) Pedestrians perceptions for utilization of pedestrian facilities—Delhi, India. *Transp Res Part F Traffic Psychol Behav* 42:495–499. <https://doi.org/10.1016/j.trf.2016.02.005>
- Räsänen M, Lajunen T, Alticafarbay F, Aydin C (2007) Pedestrian self-reports of factors influencing the use of pedestrian bridges. *Accid Anal Prev* 39(5):969–973. <https://doi.org/10.1016/j.aap.2007.01.004>
- Rasouli S, Timmermans HJ (2014) Using ensembles of decision trees to predict transport mode choice decisions: effects on predictive success and uncertainty estimates. *Eur J Transp Infrastruct Res* 14(4). <https://www.semanticscholar.org/paper/Using-ensembles-of-decision-trees-to-predict-mode-%3A-Rasouli-Timmermans/f35371664813ee32a045b030f3efc5202235110b>
- Regue R, Recker W (2014) Using gradient boosting machines to predict bike sharing station States, Transportation Research Board. No. 14-1261. <https://trid.trb.org/view/1287729>
- Rizati H, Ishak SZ, Endut IR (2013) The utilization rates of pedestrian bridges in Malaysia. In: 2013 IEEE business engineering and industrial applications colloquium (BEIAC), pp 646–650. <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6560210>
- Sabet DM (2013) Fear is not enough: testing the impact of risk on pedestrian behavior in Dhaka, Bangladesh. University of Liberal Arts Bangladesh. <https://ces.ulab.edu.bd/wp-content/uploads/sites/18/2015/07/CES-Working-Paper-Fear-is-not-enough.pdf>
- Saha MK, Tishi TR, Islam MS, Mitra SK (2011) Pedestrian behavioral pattern and preferences in different road crossing systems of Dhaka city. *J Bangladesh Inst Plan*, ISSN. 2075:9363. <http://www.bip.org.bd/SharingFiles/201311191819541.pdf>
- Sangphong O, Siridhara S (2014) A study of footbridge utilization behavior in Nakhon Ratchasima. *Rev Integr Bus Econ Res* 3:93
- Sharples JM, Fletcher JP (2001) Pedestrian perceptions of road crossing facilities. Scottish Executive Central Research Unit. <https://www2.gov.scot/resource/doc/156816/0042173.pdf>
- Sinclair M, Zuidgeest M (2016) Investigations into pedestrian crossing choices on Cape Town freeways. *Transp Res Part F Traffic Psychol Behav* 42:479–494. <https://doi.org/10.1016/j.trf.2015.07.006>
- Singh D, Nigam SP, Agrawal VP, Kumar M (2016a) Vehicular traffic noise prediction using soft computing approach. *J Environ Manage* 183:59–66. <https://doi.org/10.1016/j.jenvman.2016.08.053>
- Singh N, Parida P, Advani M, Gujar R (2016b) Human ellipse of Indian pedestrians. *Transp Res Procedia* 15:150–160. <https://doi.org/10.1016/j.tpro.2016.06.013>

- The H2O.ai team. H2O (2017) R Interface for H2O. <https://github.com/h2oai/h2o-3>
- Truong LT, Nguyen HT, Nguyen HD, Vu HV (2019) Pedestrian overpass use and its relationships with digital and social distractions, and overpass characteristics. *Accid Anal Prev* 131:234–238. <https://doi.org/10.1016/j.aap.2019.07.004>
- Villaveces A, Nieto LA, Ortega D, Ríos JF, Medina JJ, Gutiérrez MI, Rodríguez D (2012) Pedestrians' perceptions of walkability and safety in relation to the built environment in Cali, Colombia, 2009–10. *Injury Prev* 18(5):291–297. <https://doi.org/10.1136/injuryprev-2011-040223>
- Wickham H (2017) Tidyverse: easily install and load “Tidyverse” Packages. <https://CRAN.R-project.org/package=tidyverse>. R package version, 1(1) 51
- World Health Organization, Global status report on road safety 2018. World Health Organization (2018) https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/
- Wu Y, Lu J, Chen H, Wu L (2014) Identification of contributing factors to pedestrian overpass selection. *J Traffic Transp Eng (Engl Ed)* 1(6):415–423. [https://doi.org/10.1016/S2095-7564\(15\)30291-9](https://doi.org/10.1016/S2095-7564(15)30291-9)
- Yang S, Wu J, Du Y, He Y, Chen X (2017) Ensemble learning for short-Term traffic prediction based on gradient boosting machine. *J Sens*. <https://doi.org/10.1155/2017/7074143>
- Yu-Wei CDC (2015) Machine learning with R cookbook. Packt Publishing Ltd
- Zhang Y, Haghani A (2015) A gradient boosting method to improve travel time prediction. *Transp Res Part C Emerg Technol* 58:308–324. <https://doi.org/10.1016/j.trc.2015.02.019>
- Zhou Y, Li S, Zhou C, Kuo H (2018) Intelligent approach based on random forest for safety risk prediction of deep foundation pit in subway stations. *J Comput Civ Eng*. [https://doi.org/10.1061/\(asce\)cp.1943-5487.0000796](https://doi.org/10.1061/(asce)cp.1943-5487.0000796)
- Zlatkovic M, Zlatkovic S, Sullivan T, Bjornstad J, Shahandashti SKF (2019) Assessment of effects of street connectivity on traffic performance and sustainability within communities and neighborhoods through traffic simulation. *Sustain Cities Soc* 46:101409. <https://doi.org/10.1016/j.scs.2018.12.037>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.