Study of Factors Impacting Safety-Security and Mobility Friction on the Choice of Pedestrians in Using Skywalk Facilities through Soft Computing Approaches

Arunabha BANERJEE [a], Rahul RAONIAR [b], Akhilesh Kumar MAURYA [c]

[a,b,c] *Department of Civil Engineering, IIT Guwahati, India*
[a] *E-mail: arunabhabanerjee77@gmail.com*
[b] *E-mail: rahul.raoniar@gmail.com*
[c] *E-mail: akmaurya@gmail.com*

Abstract: Preference of pedestrians in using elevated walkways (skywalks) is influenced by different factors. In the present study an interviewer-administered questionnaire survey was conducted to predict the factors impacting choice of pedestrians' using skywalks under two different contexts (mobility friction and safety-security) using advanced machine learning tools for Mumbai city (India). Light Gradient Boosting Classifier (LGBC) outperformed the other soft computing approaches in terms of accuracy. Feature importance plots (SHAP values) in terms of safety-security context revealed that the walk environment, gender (female) and frequency of daily use impacted the preference of the pedestrians in using the facilities. Similarly, mobility friction context model revealed that age and available walking width impacted the user preference. The outcome of the present study would provide detailed information to urban planners and designers on the importance of improving safety-security and mobility friction on the choice of pedestrians in using the elevated facilities.

Keywords: Pedestrians, questionnaire, skywalk, soft computing, safety-security, mobility friction

1. INTRODUCTION

IRC-103 (2012) reports that pedestrians walk between 1-2km road length in urban sections. Post the new millennium, the sudden growth in vehicular traffic led to curbing of pedestrian facilities for wider roads for vehicular traffic. This forced pedestrians to use the carriageway and thus come in direct contact with the motorized traffic. In order to segregate the pedestrians from the motorized traffic, the authorities generally tend to provide at-grade signal-controlled crosswalks or construct grade separated facilities (such as overpasses and underpasses). However, studies by Herms (1972) and Koepsell (2002) have shown that the drivers tend to break traffic rules and even drive during red stop signals. A study by Golakiya (2019) indicated that pedestrians also have a tendency to avoid designated signal-controlled crosswalks and cross through illegal openings to save time. WHO (2018) reported that among different road users, pedestrians comprise 23% of the 43% vulnerable road users. As per the report, the primary cause of death amongst child and adult pedestrians (between the age of 5-29 years) in lower and middle-income countries are deaths due to accidents. Thus, in order to prevent loss of life due to accidents, providing grade-separated facilities is of utmost importance.

The grade-separated facilities can be either in the form of overpasses (such as skywalks and footbridges: FOBs) or underpasses (or subways). The FOBs allow pedestrians to cross comfortably from one side of the road to the other, and avoid interaction with the vehicular

traffic. In India, the concept of elevated skywalk system was introduced in the city of Mumbai by the Mumbai Metropolitan Region Development Authority (MMRDA) under the Station Area Traffic Improvement Scheme (SATIS) in 2007. The aim of developing such elevated walkway facilities was to segregate the pedestrian movement completely from the vehicular traffic in form of both lateral and longitudinal movements over long distances. The skywalks constructed in Mumbai generally connect the busy railways stations to strategic locations and range from 149-1780m in length. Past studies (Malik, 2017; Saha, 2011; Das, 2015; Pasha, 2015; Sinclair, 2016) have shown that pedestrians tend to avoid grade-separated facilities and use median openings and carriageways for travelling. The reasons for avoiding such facilities were absence of proper safety-security measures, improper connectivity and encroachment concerns (Mutto, 2002; Rizati, 2013; Saha, 2011; Pasha, 2015; Malik, 2017; Anciaes, 2018; Banerjee, 2020). In general traditional modelling approaches (such as regression) were used to understand the pedestrians' choice of using the elevated facilities (refer to Table 1).

Table 1. Empirical studies related to pedestrian overpasses using questionnaire survey

Author	Data collection technique	Sample size	Method used	Significant variables
Banerjee (2020), India	Questionnaire	350	Forward stepwise binary logistic regression	***Questionnaire:*** Demographics, width, obstruction/ hawker's presence, safety-security, walk environment convenience, accessibility ***Videography:*** Demographics, safety, enforcement, tiredness due to stairways, fear of height, frequency of crossing, distance from illegal crossing, time to cross, facility type, location, design criteria safety, frequency of use
Anciaes (2018), England	Questionnaire	321	Mixed logistic regression	
Malik (2017), Pakistan	Questionnaire	155	Descriptive statistics	
Hasan (2017), Malaysia	Videography & Questionnaire	191	Relative importance index	
Oviedo-Trespalacios (2017), Colombia	Questionnaire	210	Backward stepwise logistic regression	
Rankavat (2016), India	Questionnaire	500	Ordinal logistic regression	
Demiroz (2015), Turkey	Questionnaire	231	ANOVA test	
Rizati (2014), Malaysia	Videography & Questionnaire	287	Multiple linear regression	
Wu (2014), China	Questionnaire	705	Binary logistic regression	
Rasanen (2007), Turkey	Videography & Questionnaire	408	Binary logistic regression	

Table 1 shows that questionnaire (qualitative) survey was the most preferred method across different studies related to overpass facilities and majority of such studies were conducted across different Asian countries only. However, studies by Rasanen (2007), Rizati (2014), Hasan (2017) and Patra (2020) used videography (quantitative) survey along with questionnaire survey as well. Questionnaire survey provides an in-depth source of information related to the preference of the overpass users and gives a detailed description about the current status of the facilities along with the requirements of the daily users. The perception analysis helps urban planners and designers construct better pedestrian facilities along with developing proper guidelines to improve the usability of the facilities.

From Table 1 it is also observed that majority of the studies related to questionnaire survey preferred to use demographic characteristics, safety-security, obstruction/ barrier in the path of travel, geometry of the facility, and accessibility in predicting the pedestrian inclination towards using the facility. In case of studies which used videographic data collection techniques, the

2

geometry of the facility, frequency of use and safety played crucial part in determining the choice of pedestrians in using overpass facilities.

In general (as per Table 1), researchers mainly focused on applying different traditional modelling approaches (such as binary or mixed logit) to predict the choice of pedestrians in using overpass facilities. Even though discrete models provide better interpretability, however, based on the prediction accuracy the machine learning tools outperform the traditional methods. The non-linearity, regularization methods and different bias-variance tradeoffs of the machine learning methods help in better prediction of the developed models (Paredes, 2017). Thus, recently shift has been made towards using soft computing approaches in the transportation sector to develop more robust models for prediction purposes.

Table 2 shows the different studies which have been conducted using soft computing techniques for pedestrian related studies.

Table 2. Pedestrian related studies using soft computing approaches

Author	Facility	Type of data collection	Sample size	Modelling approach	Variables used
Banerjee (2020), India	FOB	Questionnaire	552	RF, GBM, GLM	Demographics, safety-security, obstruction, connectivity
Tordeux (2020), Germany	Corridor, bottleneck	Experimental	230	ANN	Geometry, density, flow
Bansal (2019), India	Signalized intersection	Videography	-	SLR	Demographics, geometry, group size
Guo (2019), China	Indoor	Experimental	101	DT, NN, SVM, RF	Demographics, walking pattern
Herrera-Angulo (2018), Peru	-	Experimental	100	ML tools	-
Rengarasu (2012), Sri-Lanka	-	Videography	50	CART, CHAID	Demographics
Park (2012), USA	Long corridor	Experimental	14	SVM	Walking characteristics

Note: *FOB- Foot over bridge, RF- Random Forest, GBM- Gradient Boosting Machine, GLM- Generalized Linear Model, ANN- Artificial Neural Network, SLR- Stepwise Linear Regression, DT- Decision Tree, SVM- Support Vector Machine, ML- Machine Learning, CART- Classification & Regression Trees, CHAID- Chi-square Automatic Interaction Detector.*

Table 2 shows that majority of the studies apart from Banerjee (2020), used videography or experimental based study approaches to predict the pedestrian walking behaviour using soft computing approaches.

1.1. Motivation and objective of the study

The previous literature shows the lack of research in understanding the pedestrian preference/ choice in using elevated walkways (such as skywalks). In such studies, researchers estimated the preference or choice of pedestrians in using an elevated facility for a single context (for example, safety-security). However, in the present study an attempt is made to understand the factors impacting the pedestrians' choice towards using skywalk facilities for two different contexts (safety-security and mobility friction) using different soft computing approaches.

2. METHODOLOGY OF DATA COLLECTION

2.1. Selection of survey locations

In total twelve skywalk location were visited in the city of Mumbai (India), and 7 locations were finalized for questionnaire data collection based on the flow level and usability of the pedestrians. In order to capture the variability amongst different pedestrians, the skywalks connecting public transport terminal (PTT) on one side to either commercial/ educational/ residential/ shopping location on the other side were selected for final data collection. Table 3 shows the details of the locations along with the geometric descriptions.

Table 3. Details of skywalk locations

Site	Type of land use	Geometric details						Sample size
		Mid-block width (m)	Total length (m)	No. of steps	Stairway width (m)	Tread (cm)	Riser (cm)	
Kalyan	Shopping	2.40	1287	54	1.90	32	17	55
Santa Cruz	Shopping	3.60	685	42	1.00	32	14	46
Vile Parle	Educational	4.00	460	57	1.43	28	14	54
Bandra	Commercial	4.50	970	51	1.46	28	14	48
Ghatkopar	Residential	3.70	315	42	1.40	30	14	47
Goregaon	Commercial	4.40	625	49	1.58	30	17	49
Andheri	Commercial	3.94	581	44	1.40	34	18	51
Total								350

From Table 3, it is observed that the total length of the skywalks across all 7 locations ranged between 315-1287m. The midblock and stairway widths varied between 2.4-4.5m and 1.0-1.9m respectively. The mid-block refers to the elevated part of the skywalk facility which the pedestrians use for traversing long distances after ascending the stairways. The dimension of the tread and riser ranged between 14-18cm and 28-34cm. Figure 1 shows the mid-block and stairway sections of the skywalks.

(a) Mid-block section (b) Stairway section

Figure 1: Skywalk geometric details

4

2.2. Questionnaire design

A set of questions were framed including demographic characteristics (age, gender, profession, and frequency of daily use), rating of existing condition of the skywalks (width, surface, safety-security, comfort, connectivity, walk environment, and obstruction) and geometric descriptions (length, mid-block width, stair width, tread and riser dimension, and location type). A 5-point Likert scale (1: Very poor to 5: Very good) was chosen for rating the different factors related to the existing condition. Table 4 shows the details of the designed questionnaire survey sheet used for data collection.

Table 4. Format of questionnaire survey used for data collection

Section	Title	Description of variable	
		Variable	***Sub-variable***
A	Demography characteristics	*Gender*	0: Female, 1: Male
		Age (in years)	0: <10, 1: 11-20, 2: 21-40, 3: 41-60, 4: >60
		Luggage	0: With, 1: Without
		Daily frequency	0: First time user, 1: Occasional user, 2: Daily once user, 3: Daily twice user, 4: More than twice user
		Profession	0: Self-employed, 1: Servicemen, 2: Student. 3: Businessmen, 4: Homemaker, 5: Retired, 6: Others
		Trip purpose	0: Work, 1: Education, 2: Shopping, 3: Mode change, 4: Returning home, 5: Jaywalking and others
B	Rating of existing conditions	*Width*	0: Very Poor, 1: Poor, 2: Satisfactory, 3: Good, 4: Very Good
		Surface	
		Connectivity	
		Safety concern	
		Comfort	
		Walk environment	
		Obstruction	0: Many, 1: Some, 2: None
C	Field observations	*Geometric conditions*	Length of facility, Width of mid-block section No. of steps, Stairway width, Tread dimension, Riser dimension, No. of entry-exit points
		Land use type	0: Commercial, 1: Educational, 2: Residential, 3: Shopping
D	Choice of using skywalk under available conditions	*Preference*	0: No, 1: Yes

2.3. Questionnaire survey

Before data collection, field observers (two at each location) gathered information on the geometric dimensions and GPS coordinates of each locations. Thereafter, interviewer-administered questionnaire survey was conducted in the neighborhood of the facility at each location during morning (8.30-11am) or evening (5-7.30pm) peak/ off-peak hours to collect sufficient representative samples. The participants were randomly chosen (using random sampling technique) and those willing to undergo the entire process of the survey were finally interviewed. The participation rate during rush hour at each location was low (1 out of 20 randomly chosen participants). In total 386 samples were acquired from the 7 selected locations and the 350 completely filled questionnaires were used for final analysis.

5

3. DATA ANALYSIS

3.1. Demographic characteristics

In order to understand the usage pattern of the elevated walkways, demographics of the pedestrians is essential. The age, gender, presence of luggage and frequency of daily use along with profession and trip purpose impact the pedestrian's choice towards using a particular pedestrian facility. Figures 2-7 shows the demographic characteristics of the pedestrians using the seven selected skywalk locations.

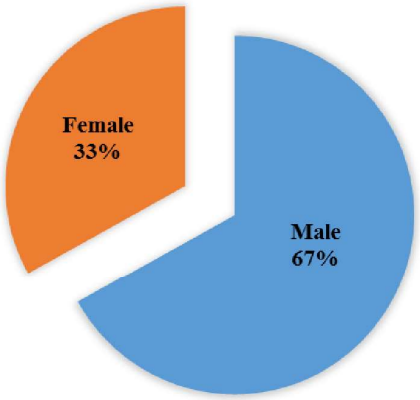

Figure 2. Gender distribution

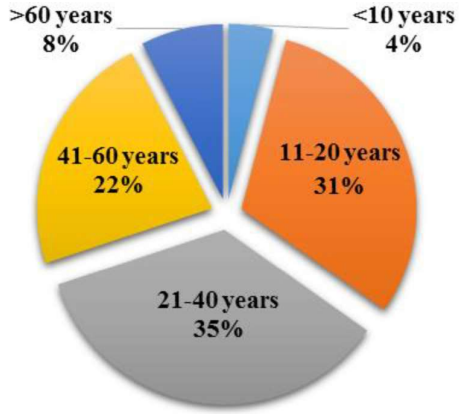

Figure 3. Age distribution

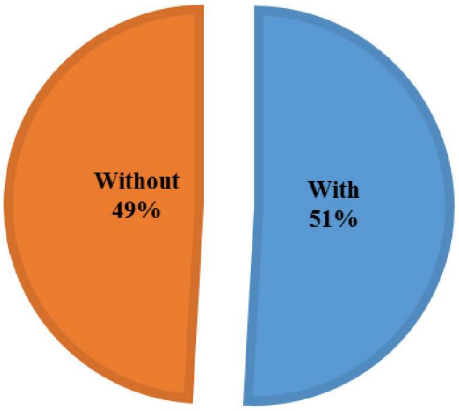

Figure 4. Luggage distribution

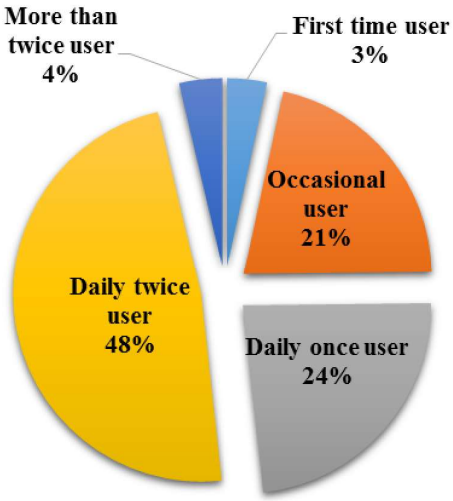

Figure 5. Facility use frequency

6

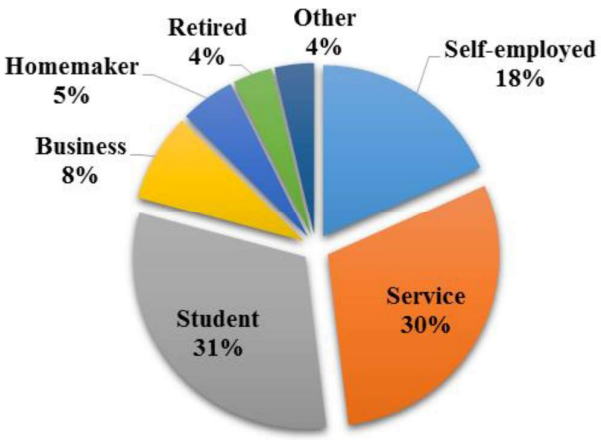

Figure 6. Profession distribution

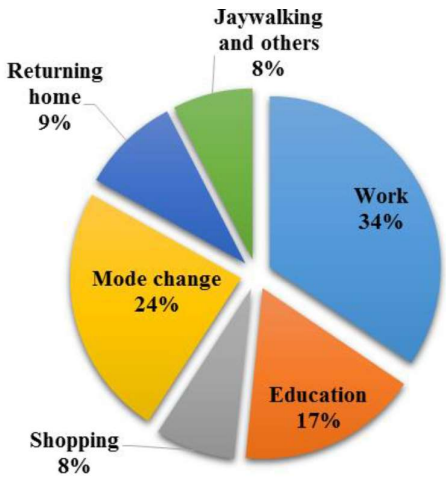

Figure 7. Purpose of trip distribution

Figures 2 and 3 show that majority of the respondents were male pedestrians (67%), in the age group of 11-40 years. The percentage of respondents with and without luggage were quite similar (50%), refer to Figure 4. The respondents were well aware of the existing situation of the skywalk facilities as they were using the facilities daily once or more than once (76%), as shown in Figure 5. Students (31%) and servicemen (30%) were the most frequent users of the different skywalk facilities as they used the facilities mostly for their work (33.7), education (17%) or mode change (24%) trips (refer to Figure 6 and 7).

3.2. Modelling approaches

In order to understand the factors impacting the safety-security and mobility friction of the skywalks users, different soft computing approaches (CBC: CatBoost Classifier, LGBM: Light Gradient Boosting Machine, XGM: Extreme Gradient Boosting, ETC: Extra Trees Classifier, ABC: Ada Boost Classifier, RFC: Random Forest Classifier, DTC: Decision Tree Classifier) were applied.

4. STUDY METHODOLOGY

The step-by-step method used for skywalk usability modelling is illustrated in Algorithm 1. The methodology involved literature review, data collection and extraction, followed by analysis and modelling the usability of the facility, and extraction of the meaningful features using Shaply values.

As explained in the previous sections, the process of the study starts with survey and geometric data collection across different Indian cities. Missing data were removed from the final sheet and normalized using min-max scaler approach. One-hot encoding was applied to categorical and ordered variables. The final data was split into 80% (for training) and 20% (for testing). Initially, a search was conducted using 80% training dataset across selected algorithms using a 10-fold cross validation with default hyper-parameters. After finding the best model a 10-fold cross validation was applied across 500 random set of hyper-parameters to identify the best hyper-parameters that maximizes the model classification power. Finally, the model performance was evaluated on the remaining 20% test dataset.

7

 # Algorithm for Skywalk Usability Modelling

Input: ***Context 1 (Mobility friction):*** gender, age, mid-block width, stairway width and different rating based ordinal variables (such as safety concerns, comfort, walk environment, obstruction, frequency of daily use and width)
Context 2 (Safety-Security): gender, age, location type and ordinal variables (such as obstruction, frequency of use and walk environment).

Output: Predicting the use or non-use of Skywalk facility (0: No; 1: Yes)

// Pre-Processing Stage
1. ***For*** *column in violation dataset*
2. | Call handle missing values
3. | Call normalise (for normalizing continuous variables)
4. End For
// Model Building and Ranking
5. ***For*** *i in range (1: total samples)*
6. | Split dataset into 80% training and 20% testing
7. | Split 80% dataset according to 10-fold CV Training and Validation dataset
8. ***End*** for
9. ***For*** *i in range (1: N model algorithms)*
10. | | ***For*** *each 10-fold CV training part*
11. | | Train model with 10 random hyper-parameter search space
12. | ***End*** for
13. | ***Call evaluate AUC on validation set for Model Ranking***
14. ***End*** *for*
// Hyper-parameter Tuning
15. ***For*** *best model's each 10-CV training part*
16 . | ***For*** *i in range (1 to 500 random hypermeters combination)*
17. | with each i train model
18. | ***Call evaluate AUC on validation set***
19. ***End*** for
// Evaluation stage
20. Evaluate the best model on the remaining 20% test dataset
21. Print test AUC score
22. Save best model
23. ***Compute Shaply Values and rank variables as per their importance***

5. MODEL DEVELOPMENT FOR PREDICTION OF FACTORS IMPACTING SAFETY-SECURITY AND MOBILITY FRICTION

The objective of the current study is to obtain an accurate classifier that can precisely predict the usability of skywalks. Two separate models were developed to understand the usability under two different contexts (**context 1:** mobility friction; **context 2:** safety-security). Mobility friction refers to the interruptions in the walking path of the pedestrians due to the standing pedestrians/ beggars/ vendors over the stairway or mid-block section of the facilities and the associated factors. Similarly, safety-security refers to the presence or absence of security personnel or CCTV cameras throughout the skywalk facilities and the associated factors. The models were developed using 350 collected survey samples. The features used in the usability

model related to mobility friction context were gender, age, mid-block width, stairway width and different rating based ordinal variables (such as safety concerns, comfort, walk environment, obstruction, frequency of daily use and width). Similarly, for safety and security context the utilized variables were gender, age, location type and ordinal variables (such as obstruction, frequency of use and walk environment).

As usability is binary in nature thus binary classification algorithms will be best suited for modelling usability. Thus, various cutting edge classification algorithms such as Gradient Boosting Classifier (GBC), Light Gradient Boosting Classifier (LGBC), Extreme Gradient Boosting Classifier (EGBC), Extra Trees Classifier (ETC), Adapting Boosting Classifier (ABC), Categorical Boosting Classifier (CBC), Random Forest Classifier (RFC), and Decision Tree Classifier (DTC) were used to train and estimate an accurate classifier for the present study. The model training was conducted using PyCaret 2.1 package through open-source Python programming language. A HP gaming notebook (16 GB Ram and i7 7th generation processor and 4 GB Nvidia GTx 1050Ti Graphics) was utilized for faster training and testing.

5.1. Model Training

As discussed in the methodological section the entire data was split into 80% for training and 20% for testing. As model comparison is both time and resource consuming process, thus, initially a 10-fold CV method was used with default hyper-parameters in order to get an overall idea of the models that performed well on the dataset. The 10-fold CV approach was adapted due to data limitation as well as to reduce overfitting. Thus, models trained using this approach will be much reliable.

The initial model training estimates are reported in Tables 5 and 6 using various classification metrics. Here, AUC has been utilized to minimize the classification error due to the class imbalance in outcome variable. Additionally, precision, recall, F1 score, kappa (Cohen, 1960) and Matthews correlation coefficients (MCC) evaluation metrics (Matthews, 1975) are also reported. Additionally, the training time (TT) was also estimated and reported in Tables 5 and 6 respectively. The 10-fold CV results revealed that for mobility friction and safety-security usability models, the light GBM classifier topped the list in terms of overall performance (AUC 0.847 & 0.668 respectively).

Table 5. Skywalk 10-fold CV mobility friction model comparison summary

Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
Light Gradient Boosting Machine	0.757	**0.847**	0.828	0.778	0.799	0.494	0.501	0.042
Ada Boost Classifier	0.786	0.845	0.894	0.776	0.828	0.548	0.564	0.028
Gradient Boosting Classifier	0.736	0.842	0.804	0.761	0.779	0.452	0.458	0.022
CatBoost Classifier	0.761	0.842	0.853	0.769	0.805	0.499	0.512	1.005
Random Forest Classifier	0.782	0.828	0.871	0.784	0.822	0.544	0.557	0.202
Extra Trees Classifier	0.786	0.807	0.828	0.809	0.816	0.560	0.565	0.177
Extreme Gradient Boosting	0.729	0.802	0.773	0.769	0.766	0.443	0.451	0.161
Decision Tree Classifier	0.739	0.736	0.760	0.783	0.769	0.47	0.473	0.385

Note: ***AUC***: Area Under the Curve; ***Prec.***: Precision; ***F1***: weighted harmonic mean of its precision and recall; ***Kappa***: Cohen Kappa; ***MCC***: Matthews Correlation Coefficient; ***TT***: Training Time

9

Table 6. Skywalk 10-fold CV safety-security model comparison summary

Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
Light Gradient Boosting Machine	0.682	**0.668**	0.802	0.726	0.760	0.291	0.297	0.033
Gradient Boosting Classifier	0.650	0.654	0.785	0.702	0.739	0.212	0.220	0.022
Ada Boost Classifier	0.657	0.641	0.862	0.681	0.760	0.188	0.201	0.027
Extreme Gradient Boosting	0.625	0.638	0.722	0.696	0.708	0.183	0.184	0.182
CatBoost Classifier	0.632	0.631	0.778	0.684	0.727	0.167	0.171	1.474
Random Forest Classifier	0.632	0.621	0.750	0.693	0.719	0.184	0.186	0.210
Extra Trees Classifier	0.589	0.594	0.647	0.681	0.659	0.136	0.138	0.180
Decision Tree Classifier	0.564	0.562	0.586	0.674	0.624	0.111	0.112	0.009

Note: ***AUC***: Area Under the Curve; ***Prec.***: Precision; ***F1***: weighted harmonic mean of its precision and recall; ***Kappa***: Cohen Kappa; ***MCC***: Matthews Correlation Coefficient; ***TT***: Training Time

5.2. Model Hyper-parameters optimization

By default, PyCaret 2.1 performs 10 random iterations over the hyper-parameters space. Thus, after getting the overall best performing models, they are trained using 500 random hyper-parameters using a 10-fold CV method (refer Table 7). This gives the idea of best hyper-parameters that would maximize the overall classifier performance.

Table 7. Hyper-parameters used for modelling

Model	Hyper-parameters	Definition of unique parameters
Light GBM	• **num_leaves:** [10,20,30,40,50,60,70,80,90,100,150,200] • **max_depth:** [int(x) for x in np.linspace (10, 110, num = 11)] • **learning_rate:** [0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1] • **n_estimators:** [10, 30, 50, 70, 90, 100, 120, 150, 170, 200] • **min_split_gain:** [0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9] • **reg_alpha:** [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9] • **reg_lambda:** [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]	• ***num_leaves:*** it controls the complexity of the tree based models. • ***max_depth:*** it defines how long a tree will be allowed to grow. • ***learning_rate:*** it is the process of adding weighting factor to new trees in the model to slow down the learning. • ***n_estimators:*** they represents the number of trees which needs to be built before maximum voting. • ***min_split_gain:*** it is the minimum loss reduction required to make further partition on leaf node of the tree • ***reg_alpha and reg_lambda:*** represent regularization terms based on weights

The hyper-parameter tuning using 10-fold CV showed an average AUC of 0.878 and 0.683 and standard deviation 0.070 and 0.088 for mobility friction (Table 8) and safety and security (Table 9) models respectively.

10

Table 8. Summary of tuned 10-fold CV mobility friction model performance

CV-Fold	Accuracy	AUC	Recall	Precision	F1	Kappa	MCC
0	0.857	0.885	0.937	0.833	0.882	0.702	0.710
1	0.750	0.859	0.937	0.714	0.810	0.461	0.500
2	0.678	0.729	0.812	0.684	0.742	0.322	0.331
3	0.678	0.807	0.875	0.666	0.756	0.307	0.333
4	0.714	0.833	0.812	0.722	0.764	0.404	0.408
5	0.7857	0.916	0.937	0.750	0.833	0.543	0.570
6	0.821	0.937	1.000	0.761	0.864	0.615	0.666
7	0.857	0.893	0.875	0.875	0.875	0.708	0.708
8	0.892	0.946	0.941	0.888	0.914	0.771	0.774
9	0.857	0.973	0.941	0.842	0.888	0.690	0.699
Mean	0.789	**0.878**	0.907	0.773	0.833	0.552	0.570
SD	0.075	**0.069**	0.058	0.076	0.058	0.161	0.158

Table 9. Summary of tuned 10-fold CV Safety-Security model performance

CV-Fold	Accuracy	AUC	Recall	Precision	F1	Kappa	MCC
0	0.714	0.655	0.882	0.714	0.789	0.360	0.380
1	0.678	0.660	0.823	0.700	0.756	0.292	0.300
2	0.607	0.668	0.823	0.636	0.717	0.104	0.114
3	0.607	0.631	0.764	0.650	0.702	0.134	0.138
4	0.750	0.822	0.722	0.866	0.787	0.489	0.501
5	0.571	0.633	0.722	0.650	0.684	0.023	0.023
6	0.678	0.697	0.833	0.714	0.769	0.250	0.258
7	0.571	0.605	0.722	0.650	0.684	0.023	0.023
8	0.785	0.872	0.833	0.833	0.833	0.533	0.533
9	0.607	0.586	0.777	0.666	0.717	0.083	0.086
Mean	0.657	**0.683**	0.790	0.708	0.744	0.229	0.236
SD	0.071	**0.088**	0.053	0.076	0.047	0.176	0.179

The tuned Light Gradient Boosting classifier's AUC based performance can be obtained using an Area Under Receiver Operating Curve (False Positive Rate vs False Negative Rate), illustrated in Figures 8 and 9. It can be observed that for both 0 and 1 class the AUC value are 0.92 (mobility friction) and 0.66 (safety and security) respectively, which shows an overall good class distinction capability for both the models. The optimized model hyper-parameters that provide the best performance in 10-fold CV are presented in Table 10.

11

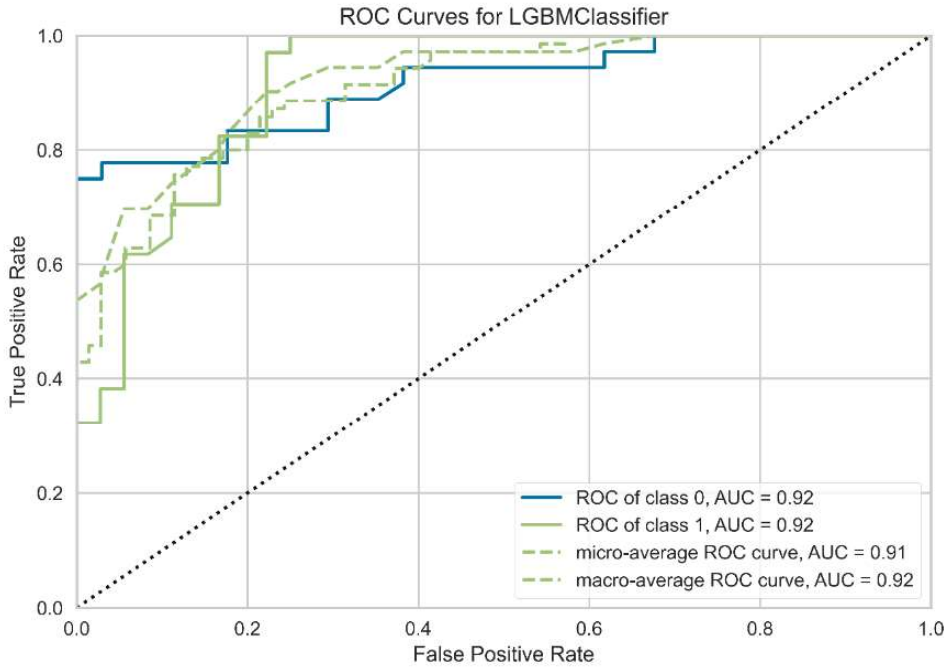

Figure 8. Mobility friction model ROC curve

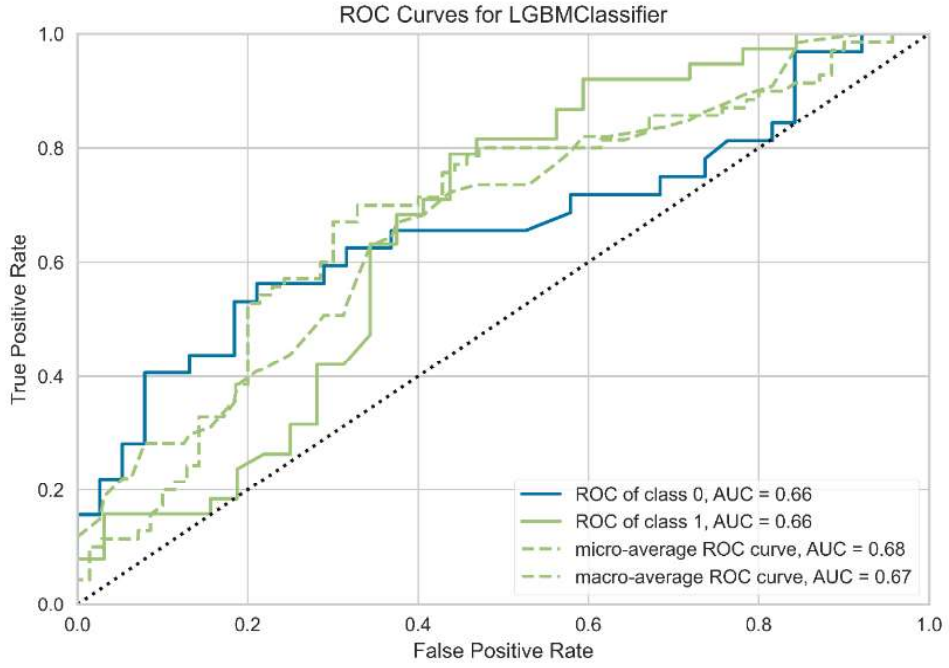

Figure 9. Safety-security model ROC curve

12

Table 10. Tuned best performing model hyper-parameters.

Models	Model Hyperparameter
Light GBM Classifier (Mobility friction)	*num_leaves:* 60, *max_depth:* 20, *learning_rate:* 0.44, *n_estimators:* 200, *min_split_gain:* 0.3, *reg_alpha:* 0.5, *reg_lambda:* 0.3
Light GBM Classifier (Safety- Security)	*num_leaves:* 50, *max_depth:* 20, *learning_rate:* 0.20, *n_estimators:* 230, *min_split_gain:* 0.1, *reg_alpha:* 0.01, *reg_lambda:* 0.01

5.3. Model Evaluation on Test Data

After obtaining the best classifiers based on the 10-fold CV performance, Light Gradient Boosting classifiers are tested to check their performance on unseen/ test dataset. The test evaluation statistics showed exceptionally good predictive accuracy based on AUC score (0.916) and F1 score (0.829) for the model presenting mobility friction context (refer Table 11). Similarly, the LGBM for safety-security context showed an AUC of 0.664 and F1 score of 0.723, indicating overall good model performance.

Table 11. Summary of mobility friction model performance on the test dataset

Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
Light Gradient Boosting Machine *(Context: Mobility friction)*	0.800	0.916	1.000	0.708	0.829	0.604	0.658
Light Gradient Boosting Machine *(Context: Safety and Security)*	0.671	0.664	0.789	0.667	0.723	0.326	0.333

5.4. Variable importance analysis

After finalizing the best performing classifiers for both the contexts, a variable importance analysis was further carried out. The relative importance is usually computed based on whether the variable is selected during the splitting in tree building process. However global importance lacks detailed information for interpretation such as its impact direction (positive/negative) and trend (i.e., whether linear/non-linear). The traditional methods such as drop column or permutation methods are computationally expensive and have limitations. Thus, in the current analysis to cope up with the limitations, an advanced approach known as Shapley Additive Explanation (SHAP) technique was utilized (Lundberg and Lee, 2017). The goal of SHAP is to explain the prediction of an instance x by computing the contribution of each feature to the prediction. It computes shaply values using coalitional game theory, where it requires two entities; for example, “game” and “players”, where “game” is reproducing the outcome of the trained model and “players” are the features incorporated in the model. SHAP computes the shaply values that are based on the idea that the outcome of each possible combination (coalition) of players should be considered in determining the importance of a single player.

In the current study analysis, global importance for both classifiers were obtained using the variable importance plot. The importance of each variable was obtained in the model based on whether the variable was used in the splitting process during tree generation. For mobility friction model the top five variables obtained are obstruction, stair width, safety concerns, width value and daily frequency of use. Similarly, for the context of safety-security, the top five important contributors are frequency of daily use, obstruction, walk environment, gender (if the gender is especially female) and location type. The problem with the global importance plot is that it does not provide the direction of impact and also suffers from bias and produces unrealistic results when two or more correlated features exist in data.

Thus, in the current study, the Shaply values are computed for both LGBM classifiers and illustrated using summary plots (refer to Figures 10 and 11). The main advantage of summary plots is that it ranks the variables in descending order and also illustrates the impact on both the classes of the outcome variable (*0: No; 1: Yes*).

In the usability model related to mobility friction (Model 1), the importance plot revealed that obstruction rating, age (<20 years), stair width, perceived comfort rating and mid-block width were the top five factors impacting the use of skywalk facilities under the existing conditions (illustrated in Figure 10). The pre-existing knowledge of a particular skywalk facility being obstructed by presence of standing pedestrians/ vendors/ beggars usually impacts the choice of using the facility. Young (especially female) pedestrians are greatly affected by the presence of obstructions (related to standing pedestrians) and they tend to avoid skywalk facilities especially at night times due to prevalence of illegal activities and fear of victimization (Malik, 2017). Higher flow and presence of obstructions, tend to reduce the effective width of the stairways and mid-block sections and thus reduce the overall usage of the skywalk facilities (Pasha, 2015).

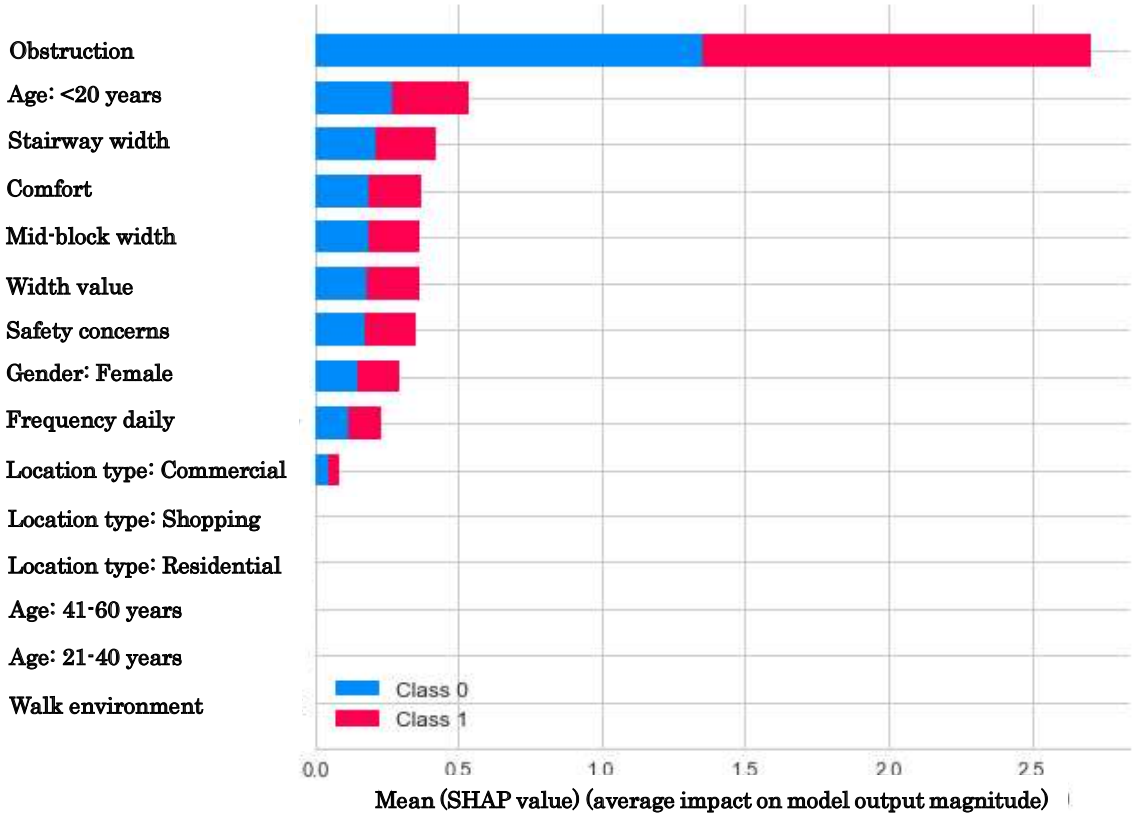

Figure 10. SHAP values plot indicating feature importance for the mobility friction model

14

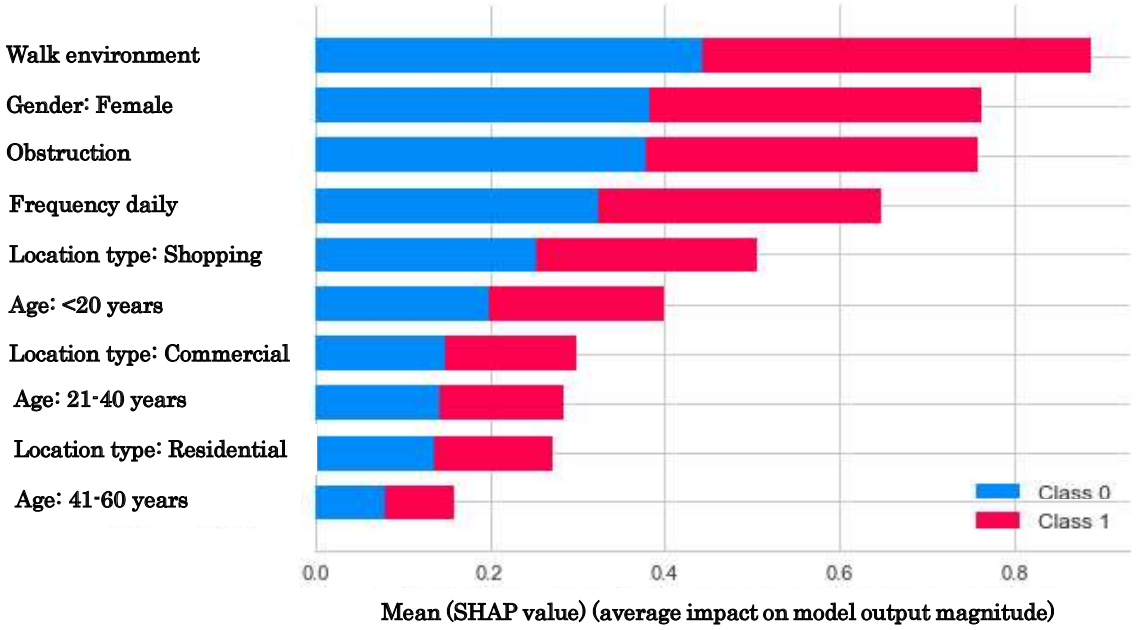

Figure 11. SHAP values plot indicating feature importance for the safety and security model

The usability concerning safety-security (refer to Model 2, Figure 11) shows that perceived safety-security is one of the most significant factors which impacts the choice of pedestrians towards using the skywalks. Figure 4 shows that with respect to safety-security; the walk environment, gender (female), obstruction, daily frequency of use, the location of the skywalk and age (<20 years) are the most significant parameters affecting the choice of the pedestrians. As the walk environment is dependent on the time of the day, thus during the daytime safety-security is mainly related to pickpocket and theft at crowded walkways; however, at night time the safety-security is related to the gender of the pedestrians (females) due to the fear victimization. Moreover, model 2 also showed that the frequency of daily use and the type of location where the skywalk was located impacted the choice of the pedestrians. For first time users or occasional users, the safety-security might not be a concern as they are not fully familiar with the existing environment; however, the regular users who are already aware about the surroundings would be more cautious while using the facility regularly.

6. CONCLUSIONS

In the present study, an interviewer-administered questionnaire survey was conducted in the neighborhood of seven skywalk locations under different land-use types (shopping, commercial, residential, and educational) across the city of Mumbai (India). 350 valid survey samples were collected from the respondents across all the seven locations. The demographic statistic results showed that majority of the pedestrians who were using the skywalk facilities were young adults (in the age group of 21–40 years), regular users (used twice daily), servicemen/ students (61%) and mostly comprised of male (67%) pedestrians. The users preferred to use the skywalks for mode change (train station to metro station/ bus terminals/ auto stands).

In the present study, different tree based machine learning algorithms (*such as GBC: Gradient Boosting Classifier, LGBC: Light Gradient Boosting Classifier, EGBC: Extreme Gradient Boosting Classifier, ETC: Extra Trees Classifier, ABC: Adapting Boosting Classifier,*

15

CBC: Categorical Boosting Classifier, RFC: Random Forest Classifier, and DTC: Decision Tree Classifier) were compared to find the optimal solution to accurately predict the factors affecting the use of skywalks under the contexts of mobility friction and safety-security concerns. Among different soft computing approaches, LGBC outperformed the other tree-based machine learning algorithms in terms of prediction accuracy on test dataset.

The important conclusions drawn from this study are discussed below:

i. LGBC showed an exceptionally good performance on test dataset for mobility friction (AUC: 0.916, F1 score: 0.829), and for safety-security model (AUC: 0.664, F1 score: 0.723).
ii. The impact of gender (female) variable on the usability choice was most significant in respect to both mobility friction and safety-security concerns.
iii. In case of mobility friction model, age (<20 years) and available effective width (stairway section and mid-block section) significantly impacted the choice of pedestrians using the skywalks.
iv. Walk environment (pickpocket/ theft during daytime and victimization during nighttime), frequency of daily use and the location of the facility impacted the preference of the users depending on safety-safety context.

7. APPLICATIONS AND RECOMMENDATIONS

The outcomes of the developed models would be useful to policy makers, planners and researchers to have an insight into the parameters affecting the pedestrians' choice towards using elevated facilities such as skywalks. Implementation of the model outcomes would attract new pedestrians to use the facilities and improve the walking experience of the existing users. This would also reduce the interactions between pedestrians-motorized traffic and improve safety to the pedestrians.

Improving the overall safety-security measures by setting up CCTV cameras as well as deploying security personnel would help in curbing the theft/ robbery/ victimization of pedestrians, especially during the night time. Moreover, to improve the frictions faced by pedestrians during their travel. restricting vendors/ beggars/ standing pedestrians could enhance the walking experience of the daily users, and provide them with a safe and comfortable travel.

8. GLOSSARY

Hyperparameters: These are the default model parameter which are tuned to attain the best calibrated model. Different modelling approaches have various hyper-parameters (for example in case of Light Gradient Boosting algorithm some of the model hyper-parameters which require tuning are: num_leaves, max_depth, learning_rate, n_estimators, min_split_gain, reg_alpha, reg_lambda). Moreover, the hyper-parameters have a range of values within which they can be tuned for obtaining the best performing model.

10-fold Cross Validation (CV): When sample size for model training is small, in such situations splitting data into train, test and validation reduces the training sample size which also increases the model overfitting issues. To increase the model performance and to reduce the overfitting issues, one of the popular approaches is fitting models using 10-fold CV. In this method the entire data set is divided into 10 different parts randomly, keeping 9 parts for training and 1 part for testing. The procedure is repeated 10 times and average of the performance is reported as final performance metric.

16

REFERENCES

IRC: 103 (2012) Guidelines for pedestrian facility. Indian Roads Congress, India.

Herms, B.F. (1972) Pedestrian crosswalk study: accidents in painted and unpainted crosswalks. *Highway Research Record*, 406, 1–13.

Koepsell, T., McCloskey, L., Wolf, M., Moudon, A.V., 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.

Golakiya, H. D., Patkar, M., & Dhamaniya, A. (2019). Impact of midblock pedestrian crossing on speed characteristics and capacity of urban arterials. *Arabian Journal for Science and Engineering*, *44*(10), 8675-8689.

World Health Organization, Global status report on road safety 2018. World Health Organization (2018).

Malik, I.A., Alwi, S.K.K., 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.

Saha, M.K., Tishi, T.R., Islam, M.S., Mitra, S.K. (2011) Pedestrian behavioral pattern and preferences in different road crossing systems of Dhaka city. *J Bangladesh Inst Plan*, ISSN. 2075, 9363.

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),* 1–6.

Pasha, M. M., Rifaat, S. M., Hasnat, A., & Rahman, I. (2015). Pedestrian's Behaviour on Road Crossing Facilities. *Jurnal Teknologi*, *73*(4).

Sinclair, M., & Zuidgeest, M. (2016). Investigations into pedestrian crossing choices on Cape Town freeways. *Transportation research part F: traffic psychology and behaviour*, *42*, 479-494.

Mutto, M., Kobusingye, O. C., & Lett, R. R. (2002). The effect of an overpass on pedestrian injuries on a major highway in Kampala–Uganda. *African health sciences*, *2*(3), 89-93.

Rizati, H., Ishak, S. Z., & Endut, I. R. (2013, April). The utilization rates of pedestrian bridges in Malaysia. In *2013 IEEE business engineering and industrial applications colloquium (BEIAC)* (pp. 646-650). IEEE.

Anciaes, P. R., & Jones, P. (2018). Estimating preferences for different types of pedestrian crossing facilities. *Transportation research part F: traffic psychology and behaviour*, *52*, 222-237.

Banerjee, A., & Maurya, A. K. (2020). Planning for Better Skywalk Systems Using Perception of Pedestrians: Case Study of Mumbai, India. *Journal of Urban Planning and Development*, *146*(2), 05020003.

Oviedo-Trespalacios, O., & Scott-Parker, B. (2017). Footbridge usage in high-traffic flow highways: The intersection of safety and security in pedestrian decision-making . *Transportation research part F: traffic psychology and behaviour*, *49*, 177-187.

Hasan, R., & Napiah, M. (2017). Pedestrians' behavior towards the use of footbridges under the impact of motivational alerting posters: the case of Ipoh city, Malaysia. *Advances in transportation studies*, *42*.

Rankavat, S., & Tiwari, G. (2016). Pedestrians perceptions for utilization of pedestrian facilities–Delhi, India. *Transportation research part F: traffic psychology and behaviour*, *42*, 495-499.

Demiroz, Y. I., Onelcin, P., & Alver, Y. A. L. Ç. I. N. (2015). Illegal road crossing behavior of pedestrians at overpass locations: factors affecting gap acceptance, crossing times and overpass use. *Accident Analysis & Prevention*, *80*, 220-228.

Wu, Y., Lu, J., Chen, H., & Wu, L. (2014). Identification of contributing factors to pedestrian overpass selection. *Journal of traffic and transportation engineering (English edition)*, *1*(6), 415-423.

Räsänen, M., Lajunen, T., Alticafarbay, F., & Aydin, C. (2007). Pedestrian self-reports of factors influencing the use of pedestrian bridges. *Accident analysis & prevention*, *39*(5), 969-973.

Patra, M., Perumal, V., & Rao, K. K. (2020). Modelling the effects of risk factor and time savings on pedestrians' choice of crossing facilities at signalised intersections. *Case studies on transport policy*, *8*(2), 460-470.

Banerjee, A., Raoniar, R., & Maurya, A. K. (2020). Pedestrian overpass utilization

modeling based on mobility friction, safety and security, and connectivity using machine learning techniques. *Soft Computing*, *24*(22), 17467-17493.
Tordeux, A., Chraibi, M., Seyfried, A., & Schadschneider, A. (2020). Prediction of pedestrian dynamics in complex architectures with artificial neural networks. *Journal of intelligent transportation systems*, *24*(6), 556-568.
Bansal, A., Goyal, T., & Sharma, U. (2019). Modelling the Pedestrian Speed at Signalised Intersection Crosswalks for Heterogeneous Traffic Conditions. *Promet-Traffic&Transportation*, *31*(6), 681-692.
Guo, G., Chen, R., Ye, F., Chen, L., Pan, Y., Liu, M., & Cao, Z. (2019). A pose awareness solution for estimating pedestrian walking speed. *Remote Sensing*, *11*(1), 55.
Herrera-Angulo, A., & Zenteno-Bolanos, E. (2018, December). A Low-Cost Microwave System for Pedestrian Speed Estimation. In *2018 IEEE MTT-S Latin America Microwave Conference (LAMC 2018)* (pp. 1-3). IEEE.
Rengarasu, T. M., Jayawansa, H. N., & Perera, G. P. W. (2012). Estimation of Pedestrian walking speeds at controlled cross walks in Sri Lanka-a pilot study. Presented at International Symposium on Advances in Civil and Environmental Engineering Practices for Sustainable Development (ACEPS 2012).
Park, J. G., Patel, A., Curtis, D., Teller, S., & Ledlie, J. (2012, September). Online pose classification and walking speed estimation using handheld devices. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (pp. 113-122).
Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, *20*(1), 37-46.
Lundberg, S., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874*.
Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, *405*(2), 442-451.
Paredes, M., Hemberg, E., O'Reilly, U.M. and Zegras, C., 2017, June. Machine learning or discrete choice models for car ownership demand estimation and prediction?. In 2017 *5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)* (pp. 780-785). IEEE.

18