ORIGINAL ARTICLE

Estimation of Pedestrian Safety Index Value at Signalized Intersections Under Mixed Traffic Conditions

S. Marisamynathan[1] · P. Vedagiri[1]

Abstract

The movement of pedestrians in the urban environment is a key factor in sustaining the social and economic relationships which are essential to the quality of life and maintaining a healthy life. To enhance pedestrian safety, there is a need to improve the pedestrian facilities at signalized intersections. The study objective is to develop pedestrian safety index model in crosswalks at signalized intersections under mixed traffic conditions. The data were collected from selected eight signalized intersections in Mumbai, India by performing video graphic and questionnaire surveys. The Pearson correlation test was performed to identify significant factors with respect to pedestrian perceived safety index score. Stepwise linear regression method was applied to develop a safety index model at 95% confidence interval and k-means clustering was used to define the threshold values for each safety index rating. The proposed model and threshold values were validated by using field data. The validation results showed that the proposed model and threshold values were estimated accurate safety levels of a pedestrian at a signalized intersection. Finally, the sensitivity of each model variable was analyzed by using Tornado diagram and improvement measures on pedestrian safety were applied and analyzed theoretically at selected signalized intersection. This study is helpful to improve the existing conditions of intersections and recommends guidelines for providing adequate pedestrian facilities to cross the crosswalk safely and comfortably at signalized intersections.

Keywords Pedestrian · Safety index · Signalized intersection · Regression · Clustering

Introduction

At signalized intersections, pedestrian traffic is very high in highly populous cities (for example, Mumbai, India) with the least amount of safety measures provided to them. Various types of pedestrian control strategies, such as fixed time control, dynamic control, coordinated control and adaptive control are provided at signalized intersections based on pedestrian flow, traffic flow, and geometric conditions. The fixed-time control signal is widely used at signalized intersections, especially in Mumbai, India. A fixed time control for highly populated areas needs to be reconsidered since the flow of pedestrian varies largely. Adaptive traffic control or optimization is required for better traffic control in such signalized intersections. The former needs complete rearrangement of traffic control signals while the latter can be utilized with existing design. To afford better pedestrian facilities at signalized intersections, the appropriate standard and control of the facilities need to be determined and maintained.

The traffic condition in India is characterized by mixed land-use pattern and mixed traffic conditions without proper lane discipline. The high pedestrian volumes at intersections in developing countries, like India, imply the need for providing suitable pedestrian facilities to improve the serviceability and safety of pedestrians. Due to high traffic and pedestrian volume at signalized intersections, pedestrians may experience conflicts with vehicles or even suffer severe accidents. Recent statistics show that more than 1,41,500 people have died and 4,88,731 people have been injured in road traffic accidents in the year 2014 alone, which is more than the past rates of Indian accidental deaths [1]. Mumbai is the most populous city in India with a population of 21 million as of 2014 and, 55% of the population are pedestrians. It has been reported that Mumbai has the maximum number

✉ S. Marisamynathan
marisamy1989@gmail.com

P. Vedagiri
vedagiri@civil.iitb.ac.in

1   Transportation Systems Engineering, Department of Civil Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India

of 'accidental deaths' including pedestrian fatalities accounting for 12.9% of the total accidental deaths as reported by 53 major cities in India [1, 2]. As per Mumbai traffic police record, out of the total 3040 fatal accidents recorded from 2007 to 2010 in Mumbai, 54% of the accidents took place at or close to intersections. The most vulnerable entity at intersections in India is the pedestrian. It is inferred that an increase in the percentage of pedestrian death rate is due to the decrease in pedestrian serviceability and safety. It is important to understand the needs of pedestrian for planning, designing, and development of pedestrian facilities, especially in crosswalks at signalized intersections. Hence, there is a need to investigate measures that can make pedestrians feel safe when crossing signalized intersections. Keeping this as the motivation, this research has been undertaken with the aim of modeling pedestrian safety index in crosswalks at signalized intersections.

Literature Review

The evaluation of pedestrian safety at intersections can be divided into three major categories: accident rate method, conflict method and level of service model. Accident rate method is the most referred method to evaluate pedestrian safety level, however, it has some limitations, such as it require larger data set, need secondary data of long periods in accident statistics, and it produces lesser evaluation results in findings [3]. Conflict method is simple to follow and the model has been developed by considering the interaction between pedestrians and vehicles in the crosswalk. Various factors are considered in model development such as volume, vehicle type, suitable gap, lane, crosswalk length, and speed. However, the accuracy levels of developed models fail to produce accurate pedestrian safety level at intersections [4, 5].

Most of the studies have addressed the pedestrian safety by developing a level of service model [6] and previous researchers have developed safety model based on conventional linear regression methods such as linear or multiple linear or stepwise or generalized models [7, 8]. Linear regression is simple to develop and is most widely used. The linear regression model is generally based on the following two assumptions: (1) the observations follow a normal distribution, and (2) means are varying with respect to independent variables. It has been found that the data are ordered (or ranked) in nature and cannot be defined by probability distributions [9]. Cumulative logistic regression method is more suitable and is also a useful technique to develop regression models when the observations are in ordered and qualitative data type. Very few studies adopted cumulative logistic regression method for modeling level of service [10]

and there are no studies for modeling pedestrian safety level at a signalized intersection.

Therefore, the objective of this paper is to identify the significant factors, develop pedestrian safety index model and define threshold values for each safety index rating that suit for Indian mixed traffic conditions. The required data are collected from selected eight signalized intersections by conducting video graphic and questionnaire surveys. Pearson's correlation test is performed to identify significant factors and stepwise regression is used to develop a safety index model. k-means clustering is performed to define threshold values and tornado diagram is plotted for sensitivity analysis with detailed discussions.

Data Collection and Analysis

Accurate and effective data on pedestrian behavior is necessary to improve safety, comfort and convenient movement of the pedestrian while crossing signalized intersections. The data collection method requires a careful procedure to ensure the accuracy of the data. Manual data collection is expensive, time-consuming, and is also error-prone. To overcome these limitations, a video data collection method is suggested to collect the required parameters at signalized intersections. Understanding of pedestrian, traffic, and geometric characteristics is very important for safety evaluation of pedestrian facilities at a signalized intersection. The data were collected and extracted during summer 2015 and summarized during the same period.

Site Selection

To fix the required number of crosswalks for this study, existing studies on pedestrian behavior and safety modeling at signalized intersections are reviewed and summarized here. Chen et al. analyzed pedestrian noncompliance and safety at a selected signalized intersection [11]. Huang and Ma analyzed pedestrian walking speed variations behaviors with the performance level at signalized intersections using data from two study locations [12]. Wasfi and Abu picked three locations and analyzed the pedestrian behaviors at crosswalks [13]. Ling et al. had chosen three locations for modeling interaction between pedestrian and vehicle at signalized intersections [14]. Muraleetharan et al. selected four crosswalks and developed a regression equation for pedestrian service level based on opinion survey and video survey [15]. Nagraj and Vedagiri selected four signalized intersections and developed pedestrian service level [16]. Also, Marisamynathan and Vedagiri selected four signalized intersections and developed a pedestrian delay models [17]. Gates et al. selected ten study locations and analyzed pedestrian crossing speed at signalized intersections [18].

Likewise, Zhang et al. analyzed pedestrian influence with right turning vehicles from selected ten signalized locations [19]. Kruszyna et al. conducted the survey at 14 study locations and developed an efficient model based on entry process [20]. Li et al. selected 15 locations and developed delay model for pedestrians [21]. Zhou et al. conducted the survey at 16 study locations and developed a probability model for pedestrian conflict probability at signalized intersections [22]. As per existing studies, the number of study locations varied from 1 to 16 for pedestrian crossing behavior studies, performance level study and safety model development at signalized intersections. In this study, eight signalized intersections were selected for Indian conditions based on the roadway, pedestrian and traffic conditions for analyzing pedestrian safety analysis and requiring model development.

Details of Selected Study Location

Before finalizing 8 signalized intersections, a reconnaissance survey was conducted and more than 20 intersections were selected to assess the suitability. The study locations are selected in such way that (1) the study sites selected were of typical four arm type signalized intersections with fixed traffic signal cycle lengths, (2) all approaches at the selected signalized intersections have two-way traffic and bi-directional pedestrian flow and (3) the study area covered the commercial area, institutional areas and a residential area with a high proportion of pedestrian volume. Eight signalized intersections were selected from a list of all signalized intersections in Mumbai suburban area, India and they are in the central part of the city with high pedestrian demand. The study locations are shown in Fig. 1. The information about the eight locations is given in Table 1.

Data Collection

Required data were collected from selected study locations by conducting field measurement, video graphic survey and questionnaire survey. Videographic survey and questionnaire survey were conducted simultaneously at each selected eight signalized intersections and the required data were extracted from video using ALL Capture video editor software. Data

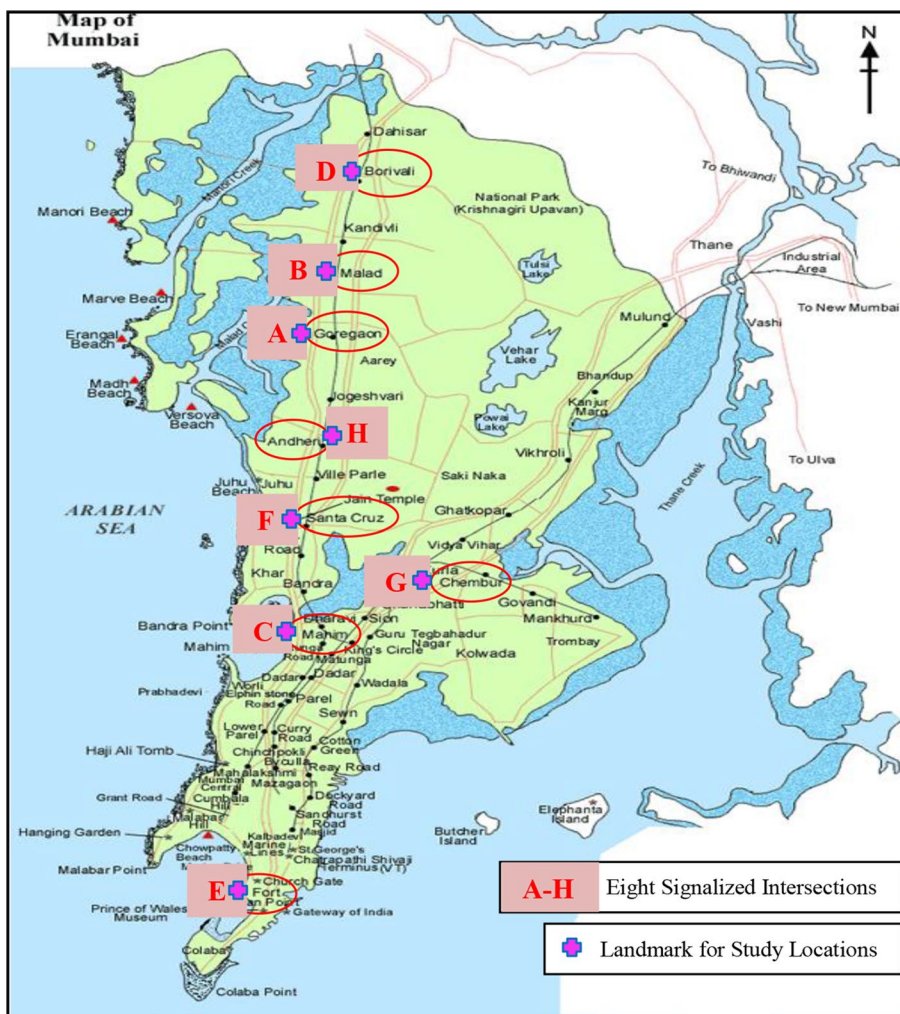**Fig. 1** Selected eight signalized intersections in Mumbai (Google Maps)

Table 1 Pedestrian flow and geometric information of the selected sites

C/W identity	Time of survey	C/W length (m)	Pedestrian flow (p/h)			Presence of C/W marking	Proper waiting area
			UtoD	DtoU	Total		
A	5.00–6.00 pm	27	101	74	175	No	Yes
B	5.30–6.30 pm	22.4	215	180	395	No	No
C	8.00–9.00 am	20	104	298	402	Yes	No
D	8.00–9.00 am	27.6	27	62	89	No	Yes
E	5.00–6.00 pm	31.5	164	173	337	Yes	No
F	8.00–9.00 am	19	74	74	148	Yes	No
G	8.00–9.00 am	27	137	166	303	No	Yes
H	5.00–6.00 pm	27	84	187	271	No	No

C/W crosswalk, *UtoD* upstream to downstream, *DtoU* downstream to upstream

were collected on weekdays during peak hours. The weather condition was sunny and warm during survey duration. Based on pedestrian flow and geometric characteristics, the major road crosswalks were selected for study purpose and most of the selected roads were urban arterial roads and sub-arterial roads.

Videographic Survey

The characteristics of the pedestrian and traffic conditions at signalized intersections were collected with a videographic survey. Cameras were setup in the direction of pedestrian upstream to downstream movement and downstream to upstream movement at the selected crosswalk in each intersection. The videographic survey covered the selected crosswalk and recorded pedestrian movements throughout the study area. The required data were extracted manually in the lab. This procedure takes a longer duration to extract data but has the advantage of providing more accurate data and long-lasting record of events. The positioning of the camera is shown in Fig. 2.

Questionnaire Survey

A questionnaire survey was designed to understand the pedestrians' perception level with respect to safety while crossing the crosswalks at signalized intersections. Several well-trained investigators were used to examine the pedestrians' perceptions towards level of safety when using the crosswalks at signalized intersections. Investigators were chosen based on their expertise in the local languages such as Hindi and Marathi along with English. Investigators stood on both sides of the crosswalks and questions were asked to pedestrians immediately after crossing the crosswalk. The meaning and importance of the question were explained to each pedestrian personally and asked them to rate from 1 to 5 with respect to pedestrian safety; where 1 represents excellence and 5 represents very poor. In addition, pedestrian age
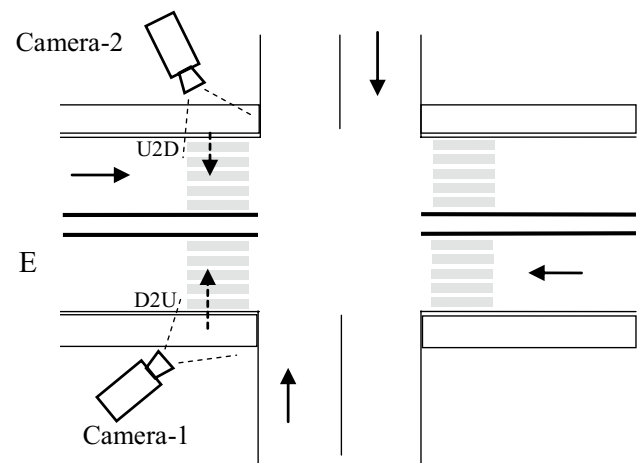

Fig. 2 Camera position set-up points for video graphic survey

was also collected. Investigators noted the time of the survey and dress color of each participant and these details were used to match the qualitative data with quantitative data.

Data Extraction

Study locations are operated by shared signal phase with the bidirectional pedestrian flow. Apart from geometric data, the required data for statistics and model development were extracted from collected video using ALL Capture video editor software. The software provided 20 numbers of images per 1 s interval and 72,000 images were extracted from the 1 h video. Two cameras were used and a total of 1,44,000 images were extracted. The required data were extracted from 1,44,000 images at one location for one direction (UtoD). Again, the procedure was repeated and data were extracted for the same location in another direction (DtoU). This process was repeated for all eight signalized intersections and all the required parameters were extracted.

The required model's variables are described in the following sections.

List of Extracted Variables

Based on earlier studies and field observations, several possible factors influencing pedestrian behavior at signalized intersections were identified for Indian conditions. The recorded video provided information about pedestrian crossing volumes, crossing time, crosswalk length, crossing locations, crossing behaviors (such as walking or running, alone or in groups and walking speed), pedestrian characteristics (like gender and age group) and pedestrian-vehicle interaction. The selected variables used in this study and model development are shown in Table 2 with encoded parameters based on existing literature and field conditions.

Details of Extracted Data

A total of 2476 pedestrians were clearly observed from recorded video and detailed information on pedestrian crossing behavior was obtained. Pedestrians using the crosswalk during pedestrians' green phase were considered as compliance pedestrians while those who use them during non-green phase were considered as noncompliance pedestrians. The percentage of pedestrian compliance was defined as the ratio between the number of pedestrian using the crosswalk during the green phase and total number of pedestrian that arrived at crosswalks. Pedestrian characteristics and behavior information were presented in Table 3.

Table 3 Pedestrians classified by their characteristics and behavior

Characteristics	Variable	N (sample)	% of N
Gender	Male	1936	78.19
	Female	540	21.80
Age groups	Child	145	5.85
	Adult	1978	79.88
	Old	353	14.25
Platoon	Yes	742	29.96
	No	1734	70.03
Direction	UtoD	1083	43.74
	DtoU	1393	56.26
Departure signal	Green	1306	52.74
	Non green	1170	47.25
Crossing speed (m/s)	Less than 1	235	12.7
	1.0–1.2	551	29.8
	1.2–1.4	555	30.1
	1.4–1.6	286	15.5
	More than 1.6	222	12.0
Compliance with signal	Yes	1306	52.74
	No	1170	47.25

The statistics from the observed data shows that the proportion of male pedestrian is higher than female pedestrian during peak hours and comprises most of the adult pedestrians than children and elderly people. The data indicate that pedestrians are interested to walk while using the crosswalk rather than running, with 30% of pedestrians crossing the crosswalks at various crossing speeds ranging from 1.2 to

Table 2 Statistical results of factors influencing pedestrian safety level with variables description

Variable	Description	Pearson value	
		Value	Sign
Safety score	Obtained from questionnaire survey: SI A = 1, highly safe (excellent); SI B = 2, safe (normal); SI C = 3, average; SI D = 4, risk (danger); and SI E = 5, high risk (very danger)	1	–
Pedestrian crossing speed	Crosswalk length is divided by the actual travel time of pedestrian (expressed as m/s)	−0.050	0.396
Pedestrian volume	Counted pedestrian volume at intersections (expressed as ped/h)	0.024	0.688
Traffic volume	1. The volume of through movement vehicles and left turning vehicles from the street parallel to the crosswalk during pedestrian green phases, (expressed as vol/crosswalk/cycle)	0.254[a]	0.000
	2. The volume of moving vehicles through the crosswalk while pedestrian waiting in waiting area or median during pedestrian non- green phases (expressed as vol/crosswalk/cycle)	0.229[a]	0.000
	3. The numbers of vehicles occupied in the crosswalk at the upstream side for each pedestrian (expressed as vol/crosswalk/cycle)	0.201[a]	0.001
Number of interactions	A number of interactions between pedestrian and vehicle in a crosswalk. The interaction was counted for each pedestrian during their crossing	0.008	0.889
Conditions of crosswalk marking	0 for absent of pedestrian crosswalk marking and 1 for the presence of pedestrian crosswalk marking	−0.055	0.352
Median width	The width of the provided median (expressed as m)	−0.135[b]	0.220
Crosswalk length	Length of the crosswalk (expressed as m)	0.055	0.352

[a, b]Significant values at 99 and 95% confidence intervals, respectively

1.4 m/s. The average pedestrian crossing speed is found to be 1.29 m/s. Only 70% of pedestrians have been found to utilize the crosswalk. 53% of pedestrians are complying with traffic signals, indicating higher noncompliance being prevalent in highly populous regions during peak hours. The higher non-compliance rate in this study occurs at Samaj Junction with a noncompliance rate of 80%. Pedestrians do not comply with the traffic signal while a turning vehicle is an approach with the crosswalk compared to through movement vehicles. The percentage of pedestrians who receive median delay and interaction delay at crosswalk due to noncompliance are 33 and 14%, respectively. Pedestrian perceptions on safety were collected by conducting questionnaire survey and the sample size one each location were presented in Table 4.

The survey findings from the data collected for different pedestrians were analyzed and few of the important outcomes from the survey were discussed below. A total of 588 pedestrians participated in the questionnaire survey; 477 males and 111 females. About 7% of pedestrians were

less than age 18, 79% were 18–50 and 14% were aged 50 or older. The gender and age distribution of all locations are shown in Fig. 3.

Pedestrians were asked about the safety level while crossing the crosswalk. Only 38% of the pedestrians reported that felt safe when they cross the crosswalk. While 33% of pedestrians claimed that they felt risky and it shows that the high-level improvement is required to improve pedestrian safety (refer Fig. 4).

Identification of Significant Factors

Pearson correlation coefficient test was performed to identify the significant factors that influence the pedestrian perceived safety score in the crosswalk at signalized intersections. The test was performed in SPSS 16.0 software at 99% confidence interval and the results are presented in Table 2.

Table 4 Number of collected samples in the questionnaire survey

Location	A	B	C	D	E	F	G	H
Collected samples								
UtoD	21	20	60	29	83	45	32	27
DtoU	30	20	23	38	43	42	26	49
Total	51	40	83	67	126	87	58	76

Fig. 3 Pedestrian type by gender and age

Q) Gender: Male/Female?

Q) Age: Child/Adult/Elderly?
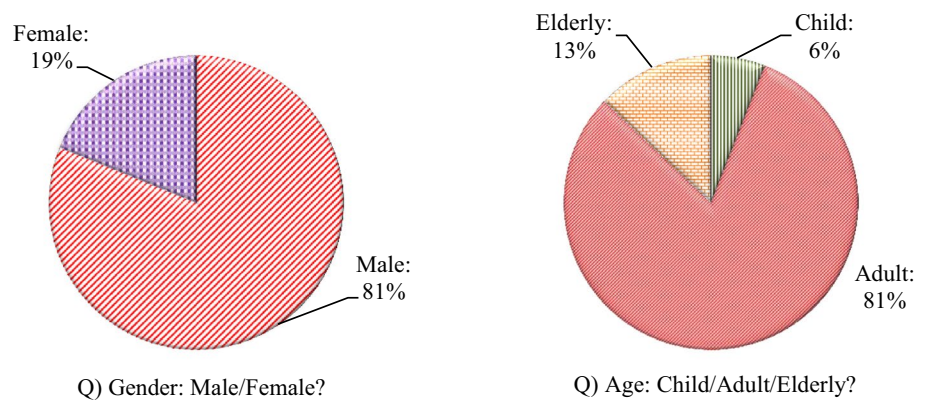
Fig. 4 Safety conditions of pedestrians at a crosswalk

- Highly Safe
- Safe
- Average
- Risk
- Highly Risk
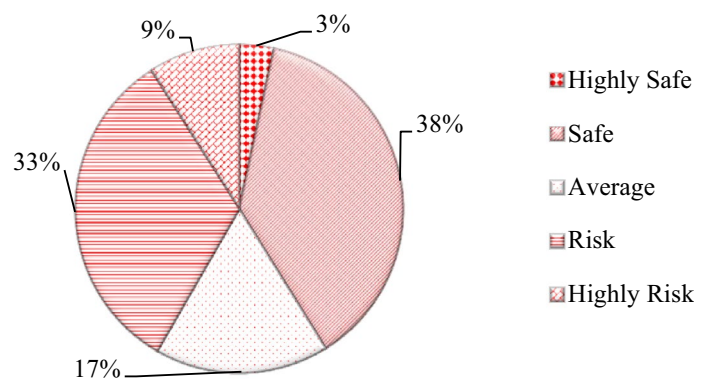
From Table 2, median width, the number of vehicles during pedestrian non-green phases, the number of parallel through movement vehicles and the number of vehicles occupied in the crosswalk area had the significant effect on the pedestrian perceived safety score at 0.01 significance level. Other variables were not considered further because of their poor correlation with the dependent variable. Significant factors which have been identified in the analysis are used to develop a safety index model in the following section.

Pedestrian Safety Index Model Development

In the past, a variety of deterministic and stochastic models have been developed to solve all kinds of complex transportation engineering problems. Safety or service level of existing facilities is recognized by qualitative data such as user response score and it is very difficult to quantify. Linear regression is an approach to modelling dependent variables and one or more independent variables which can be used for prediction or forecasting purpose and to quantify the strength of the given variables. Many researchers have adopted conventional linear regression methods to develop ordered data such as pedestrian level of service and bicycle level of service [10, 15, 16, 23]. The major reason is that linear regression is simple to develop and is one of the most widely used techniques. In addition, the application linear regression is easy to use in field application. The linear regression model can determine the relative influence of one or more predictor variables to the criterion value and able to identify outliers based on correlation statistic value. Therefore, multiple linear regression techniques were adopted to decide if a multiple linear relationship might occur that can calculate the mean rating obtained for each respondent in the questionnaire survey. The generalized form of the multiple linear regressions is given as,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n, \tag{1}$$

where Y = dependent variable, X_{1-n} = explanatory variables, β_{1-n} = estimated parameters from the model, β_0 = constant. The questionnaire survey rating was considered as dependent variable and the variables obtained from the Pearson correlation test was selected as independent variables for model development. The primary structure of the safety index model is expressed in the following mathematical expression,

$$SI_{score} = \beta_0 + \beta_1 \times X_1 + \beta_2 \times \log(X_2) + \beta_3 \times \log(X_3) + \beta_4 \times \log(X_4), \tag{2}$$

where SI_{score} = pedestrian safety index score through questionnaire survey (rating 1–5), X_1 = median width in m, X_2 = volume of motorized vehicle through crosswalk during pedestrian non-green phase in vol/crosswalk/cycle, X_3 = volume of through movement vehicle from the street parallel to crosswalk during pedestrian green phases in vol/crosswalk/cycle, and X_4 = volume of motorized vehicles occupied in crosswalk during pedestrian green phases in vol/crosswalk/cycle.

The stepwise regression technique was performed in SPSS 16.0 software at 95% confidence interval and the results were shown in Table 5. The adjusted R^2 value for the proposed model is 0.4752, which specifies that 47.52% of the variation in the predicted dependent variable has been explained by the explanatory variables and this denotes the moderate accuracy level of the proposed model prediction.

From Table 5, the calculated t values are greater than the critical value (± 2.326) and the p values are less than the p critical value (0.05). This represents that the model variables are significant at 95% confidence interval.

Threshold Values for Each Safety Index

Many existing studies have utilized clustering techniques to define threshold values for defining pedestrian or bicycle or motorized vehicle level of service and safety level [24–27]. The same technique is adopted in this study to define the threshold values for each safety index category. Therefore, this paper utilized k-means clustering method for defining threshold values for each safety index rating. The methods are applied to a data set consisting of pedestrian perceived safety scores, which were obtained from user perceptions survey and the developed model. In k-means clustering, the user perceptions score for safety is given as input variables and the threshold values were obtained as output using MATLAB 2014. In k-means clustering, the main objective is to find the k center for each cluster by minimizing an objective function known as squared error function and is given by:

$$min \sum_{i=1}^{K} \sum_{J=1}^{N_i} ||X_j - C_i||^2, \tag{3}$$

where K is the number of clusters in the data set; N_i is the number of points in the cluster i; X_j is the jth observation vector; and C_i is the centroid of cluster i. The following algorithm is used in k-means clustering.

Table 5 Multiple linear regression model

Variables	Model estimate	Coefficients	Standard error	t value	Sign
Constant	β_0	2.202	0.073	4.652	0.000
X_1	β_1	− 0.029	0.095	− 3.099	0.021
X_2	β_2	0.223	0.095	2.355	0.019
X_3	β_3	0.346	0.001	2.723	0.046
X_4	β_4	0.305	0.003	2.952	0.003

Step 1: Randomly select cluster centers C

Step 2: Calculate the distance between each data point and cluster center value

Step 3: Assign the data points to the cluster center based on the minimum distance

Step 4: Recalculate the new center values for each cluster by using the equation

$$C_i = \frac{1}{N_i} \sum_{j=1}^{N_i} X_i, \tag{4}$$

Step 5: Again, calculate the distance between each point and the new cluster centers.

Step 6: Repeat again from step 3 until there is no change in cluster centers' values.

Finally, the threshold values for each safety index rating were calculated by using k means methods and the results are presented in Table 6.

The performance of proposed threshold value is validated by comparing the distributions of each SI rating field value and the results are presented in Table 7.

From Table 7, wider ranges are available for SL rating A–E. The proposed methods were compared by statistical analysis. The percentage of accurate distribution for the k-means method has been 84.85%. Results show that k-means clustering method delivers a reasonable threshold value of SI rating A through E for the mean pedestrian safety level score for the field data. In addition, the results obtained from cluster analysis and a regression model was compared and various statistical performance tests such as mean absolute percentage error (MAPE), root mean square error (RMSE), and R values were conducted to check the accuracy level of the results. MAPE, RMSE, and R values were 15.36%, 0.5318, and 0.8802, respectively. MAPE and RMSE values were very less with a good percentage of successful prediction (R value) of safety index category. Finally, the statistical performance test results indicate that the proposed threshold value for the pedestrian safety index is more precise and a reliable fit for Indian conditions. Further, k-means threshold values were

Table 6 Threshold values of pedestrian safety index rating at signalized intersection

SI rating	Numerical ranking	Threshold values
		Method: k-means clustering
A	1	Mean ≤ 2.00
B	2	2.00 < mean ≤ 2.75
C	3	2.75 < mean ≤ 3.55
D	4	3.55 < mean ≤ 4.55
E	5	4.55 < mean

Table 7 Comparisons of proposed techniques based on distributions

SI rating	Distribution of observed rating (%)				
	Method: k-means clustering				
	1	2	3	4	5
A	100	NA	NA	NA	NA
B	14	84	3	NA	NA
C	NA	3	89	8	NA
D	NA	6	6	81	6
E	NA	NA	NA	22	78

NA not available

used for identification of the safety level of pedestrians at signalized intersections in this study.

Applications

The proposed pedestrian safety index model is validated with another new four-arm signalized intersection. The evaluation was carried out with the data collected at the new crosswalk in the Santacruz–Juhu junction, Mumbai, India. A total of 66 pedestrians were interviewed with the same user perceptions survey and required parameters were extracted from the captured video graphic survey. The field observed median width, the number of vehicles during pedestrian non-green phases, the number of parallel through movement vehicles and the numbers of vehicles occupied in the crosswalk area are 0.7 m, 13, 9 and 11 veh/crosswalk/cycle.

Pedestrian safety index score

$$= 2.202 - (0.029 \times 0.7) + (0.223 \times \log(13))$$
$$+ (0.346 \times \log(9)) + (0.305 \times \log(11))$$
$$= 3.08$$
$$= \text{SI rating C.}$$

The field observed mean SI_{score} value is 3.3 which also follows the SL rating C. The individual percentage rating was also compared between predicted and observed values. Based on the application and validation, it can be concluded that the developed model estimates the pedestrian safety index accurately for crosswalks at a signalized intersection.

Sensitivity Analysis of Model Variables

Sensitivity analysis is a method followed to determine the impact of an independent variable with a particular dependent variable under a given set of assumption. A sensitivity analysis was performed between the dependent variable and independent variables of the safety index

model. The data were taken from Santacruz–Juhu junction and field values were considered as baseline values for sensitivity analysis. Independent variables such as median width, number of vehicles during pedestrian non-green phases, number of parallel through movement vehicles and numbers of vehicles occupied in the crosswalk area were changed by ± 5, 10, 15 and 20% and the effects on the dependent variable (SI score) were tabulated in Table 8.

Tornado diagram is a special type of bar chart, which can be used to compare the relative importance of variables. In this study, tornado diagram is plotted for understanding the positive and negative changes of 20% in independent variables and the result is shown in Fig. 5.

Figure 5 shows that if number of parallel through movement vehicles (X_3) goes down by 20%, then large negative changes in safety index score and same thing if the value increase by 20% that has a huge positive change in safety index score. Similarly, other variables such as number of vehicles during pedestrian non-green phases (X_2) and numbers of vehicles occupied in the crosswalk area (X_4) have the same effect with SI score. However, the median width (X_1) has reversed effect with the safety index score with comparatively lesser effect. Thus, the results concluded that variables are prioritized and the safety index can be improved by changing the most significant variables at Santacruz junction.

Existing conditions of pedestrian facilities at Santacruz–Juhu junction were evaluated and the results were presented in the previous section. The existing SI rating at Santacruz–Juhu junction was C and it represents that the existing pedestrian facilities perform below average and pedestrian safety level was rated as average. Also, the importance of each model variable was identified in Tornado diagram. Therefore, there is a need to apply immediate improvement measures that can improve the pedestrian safety level. In this section, various possible improvement measures are applied and analyzed theoretically by improving geometric characteristics and traffic characteristics at Santacruz–Juhu junction.

Measure 1: Provide exclusive pedestrian signal plan or stop moving vehicles and left turning vehicles from the street parallel to the crosswalk during the pedestrian green phase. Measure 1 is applied at Santacruz–Juhu junction and results are shown in Table 9.

Measure 2: Provide proper vehicle stop line marking with guidelines, raised crosswalk, and improvised traffic control management to reduce the number of vehicles occupied in the crosswalk. The application of this measure is shown in Table 9.

Measure 3: Increasing the median width or provision of the refuge-island can help to improve the convenience level of pedestrians and to reduce people waiting at the crosswalk. The results are shown in Table 9.

Table 8 Sensitivity analysis of safety index (SI) model variables

Changes in variables (%)	X_1	SI score	X_2	SI score	X_3	SI score	X_4	SI score
80	0.56	3.14	1.024	3.08	0.92	3.06	0.72	3.08
90	0.63	3.14	1.152	3.11	1.035	3.10	0.81	3.11
100[a]	0.7	3.14	1.28	3.14	1.15	3.14	0.9	3.14
110	0.77	3.14	1.408	3.17	1.265	3.18	0.99	3.17
120	0.84	3.14	1.536	3.20	1.38	3.22	1.08	3.19

[a]Existing base line values for Santacruz–Juhu junction (existing conditions); X_1, X_2, X_3 and X_4 represents median width, number of vehicles during pedestrian non-green phases, number of parallel through movement vehicles and numbers of vehicles occupied in crosswalk area, respectively

Fig. 5 Tornado diagram for sensitivity analysis of safety index (SI) model variables
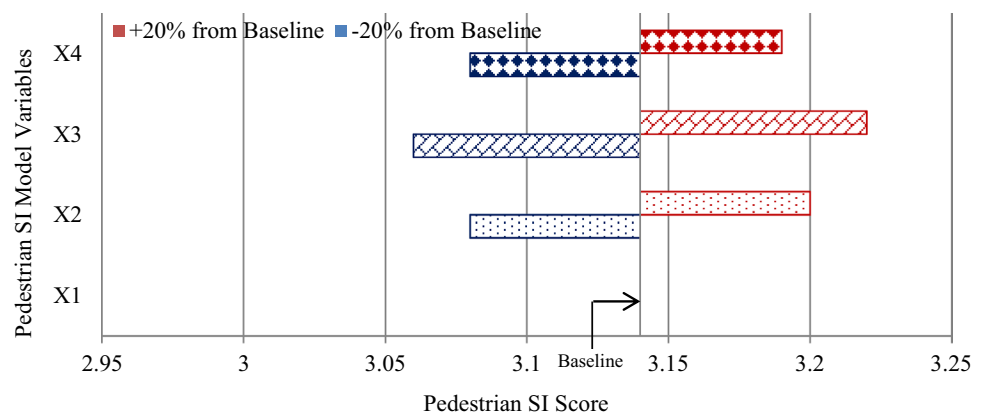
Table 9 Application of measures and results of SI changes

Measures	SL score		SL rating	
	Before	After	Before	After
1	3.1	3.10	C	C
2	3.1	2.74	C	B
3	3.1	2.87	C	C
1 and 2	3.1	2.70	C	B
1 and 3	3.1	2.83	C	C
2 and 3	3.1	2.47	C	B
1, 2 and 3	3.1	2.43	C	B

From Table 9, the application of individual measures 1 and 3; and combine measures of 1 and 3 do not allow any changes in SI rating. Apart from that SI rating are improved by applying other remedial measures. If the reduction of waiting time delay is considered with above-mentioned measures, then the overall SI may reach B or A.

Conclusions

Pedestrian environment is complicated because pedestrians are subjected to various parameters significantly affecting their perceptions of safety. Identification of these parameters is essential to assess pedestrian facilities, and assessment methods are needed to understand how well a facility accommodates pedestrians. The significant factors which influence pedestrian safety level score were identified by Pearson correlation analysis and the factors were observed as median width, number of vehicles during pedestrian non-green phases, number of parallel through movement vehicles and number of vehicles occupied in the crosswalk area. The linear regression technique was used to develop the safety level model by using significant factors. k-means clustering was utilized to define the threshold values for each safety index rating. The developed models and threshold values were validated with Santacruz–Juhu intersection data. The sensitivity of each model variable was performed with Tornado diagram and improvement measures were applied theoretically and analyzed with Santacruz–Juhu junction. The developed model helps intersection designers to understand the factors to enhance pedestrians safely at intersections and it will be useful to reduce pedestrian accidents in the crosswalk. By using the safety level rating, roadway designers can redefine existing conditions of pedestrian travel at intersection. Also, the model helps transport designers to decrease conflict and increase the safety level of the pedestrian at intersections. It can be used to find alternative intersection design standards to accommodate more pedestrians.

The limitation of this study is that the conventional linear regression method is used to develop the safety index model at signalized intersections. Because the linear regression method was used as a first attempt of the research work to quantify the pedestrian safety index value at signalized intersections which can give simple and workable procedure. Further, application of other methods such as ordered probit model and cumulative logistic regression could provide better prediction capabilities which can consider in future research scope. Due to unavailability of data on pedestrian-vehicle crash at signalized intersections, this study does not perform the comparison between the pedestrian perception of safety and actual safety performance. In the absence of crash data, surrogate safety measure can be considered to evaluate the actual safety performance of pedestrian facilities at signalized intersections. As a future study, the authors suggest comparing the pedestrian safety perception and actual safety performance by using surrogate safety measures for improving the conditions of pedestrian overall safety at the signalized crossing.

References

1. NCRB (2015) Accidental deaths and suicides in India. National Crime Records Bureau Ministry of Home Affairs, New Delhi
2. Mohan D, Tiwari G, Bhalla K (2015) Road safety in India. Transportation Research and Injury Prevention Programme, Indian Institute of Technology Delhi, New Delhi
3. Brian B, Associate LB, Vecellio RL (1995) Vehicle and pedestrian accident models for median locations. J Transp Eng 3:531–537
4. Wang X, Abdel Aty M (2008) Modeling left-turn crash occurrence at signalized intersections by conflicting patterns. Accid Anal Prev 40:76–88. https://doi.org/10.1016/j.aap.2007.04.006
5. Wang X, Aty MA (2008) Analysis of left-turn crash injury severity by conflicting pattern using partial proportional odds models. Accid Anal Prev 40:1674–1682. https://doi.org/10.1016/j.aap.2008.06.001
6. Zhang L, Prevedouros PD (2000) Signalized intersection level of service incorporating safety risk. Transp Res Rec J Transp Res Board 3288:77–86
7. Leden L (2002) Pedestrian risk decrease with pedestrian flow. A case study based on data from signalized intersections in Hamilton, Ontario. Accid Anal Prev 34:457–464
8. Qi Y, Yuan P (2012) Pedestrian safety at intersections under control of permissive left-turn signal. Transp Res Rec J Transp Res Board 2299:91–99. https://doi.org/10.3141/2299-10
9. Kao C, Chyu C-L (2003) Least-squares estimates in fuzzy regression analysis. Eur J Oper Res 148:426–435. https://doi.org/10.1016/S0377-2217(02)00423-X
10. Ye X, Chen J, Jiang G, Yan X (2015) modeling pedestrian level of service at signalized intersection crosswalks under mixed traffic conditions. Transp Res Rec J Transp Res Board 2512:46–55. https://doi.org/10.3141/2512-06

11. Chen J, SHI JJ, LI XL, Qing Z (2011) Pedestrian behaviour and traffic violations at signalized intersections. In: International conference of chinese transportation professionals, pp 1103–1110

12. Huang C, Ma W (2010) A statistical analysis of pedestrian speed on signalized intersection crosswalk. In: International conference of chinese transportation professionals, ASCE, China, pp 1401–1407

13. Wasfi Z, Abu T (2007) Modeling pedestrian behavior on pedestrian crosswalks. An-Najah National University, Palestine

14. Ling Z, Ni Y, Keping L (2012) Modeling interactions between pedestrians and right turn vehicles at signalized intersections. In: COTA International conference of transportation professionals, pp 2072–2084

15. Muraleetharan T, Adachi T, Hagiwara T, Kagaya S (2005) Method to determine pedestrian level-of-service for crosswalks at urban intersections. J East Asia Soc Transp Stud 6:127–136

16. Nagraj R, Vedagiri P (2013) Modeling pedestrian delay and level of service at signalized intersections crosswalks under mixed traffic condition. J Transp Res Board 2394:70–76

17. Marisamynathan S, Vedagiri P (2016) A new approach to estimate pedestrian delay at signalized intersections. Transport. https://doi.org/10.3846/16484142.2016.1158208

18. Gates TJ, Noyce DA, Bill AR, Ee N, Van (2006) Recommended walking speeds for pedestrian clearance timing based on pedestrian characteristics. In: 85th Transportation Research Board annual meeting, Washington, D.C.

19. Zhang W, Yang X, Xu Z (2009) Study on the capacity of right-turn movement under pedestrians influence at signalized intersections. In: ICCTP critical issues in transportation systems planning, development, and management, ASCE, pp 2976–2982

20. Kruszyna M, Mackiewicz P, Szydlo A (2006) Influence of pedestrians' entry process on pedestrian delays at signal-controlled crosswalks. J Transp Eng 132:855–861

21. Li Q, Wang Z, Yang J, Wang J (2005) Pedestrian delay estimation at signalized intersections in developing cities. Transp Res Part A Policy Pract 39:61–73. https://doi.org/10.1016/j.tra.2004.11.002

22. Zhou Z ping, Ren G, Wang W (2011) Modeling violations of pedestrian road-crossing behavior at signalized intersections. In: International conference of chinese transportation professionals, ASCE, pp 2490–2499

23. Hubbard SML, Bullock DM, Mannering FL (2009) Right turns on green and pedestrian level of service: statistical assessment. J Transp Eng 135(4):153–159

24. Fang FC, Elefteriadou L, Pecheux KK, Pietrucha MT (2004) Using fuzzy clustering of user perception to define levels of service at signalized intersections. J Transp Eng 129:657–663

25. Mohapatra SS, Bhuyan PK, Rao KVK (2012) Genetic algorithm fuzzy clustering using GPS data for defining level of service criteria of urban streets. Eur Transp Transp Eur 52:1–18

26. Bhuyan PK, Rao KVK (2011) Defining level of service criteria of urban streets in Indian context. Eur Transp Transp Eur 49:38–52

27. Azimi M, Zhang Y (2010) Categorizing freeway flow conditions by using clustering methods. Transp Res Rec J Transp Res Board 2173:105–114. https://doi.org/10.3141/2173-13