Case Studies on Transport Policy 13 (2023) 101038

Contents lists available at ScienceDirect

Case Studies on Transport Policy

journal homepage: www.elsevier.com/locate/cstp

Modeling pedestrian behavior at urban signalised intersections using statistical-ANN hybrid approach – Case study of New Delhi

Faizanul Haque *, Farhan Ahmad Kidwai

Jamia Millia Islamia, Department of Civil Engineering, New Delhi 110025, India

ARTICLE INFO

Keywords:
Pedestrian signal violation
Artificial Neural Network (ANN)
Pedestrian Safety
Signalised intersection
Sensitivity analysis

ABSTRACT

Traffic crashes at intersections are major hazard all over the world. The victims of road crashes are mostly pedestrians. A large share of accidents can be related to unsafe crossing behavior by pedestrians. The accidents are mostly fatal for pedestrians especially if the unsafe act is signal violation. This study primarily focused on cognizing factors which are significantly affecting the signal violation behavior of pedestrians' at urban signalised intersections. Data is collected at 11 signalised intersections in New Delhi, India using video recording technique. Significant variables are identified using conventional statistics. The variables are then used as input neuron for the Artificial Neural Network (ANN) for modeling signal violation behavior. Sensitivity analysis is performed to identify the relative importance of factors related to the pedestrian signal violation. Host of different factors such as pedestrian demographic, behavioral attributes, crossing state and mobile usage are notably influencing signal violations. Pedestrian crossing speed, crossing path and waiting time are the top three predictors of violation behavior. The accuracy of ANN to predict signal violation behaviour is found to be about 85% and is considerably higher than conventional binary logistic regression (BLR) model. The area under the receiver operating characteristic (ROC) curve (AUC) is found to be 0.753, suggesting good model performance. The results suggest that ANN is a robust alternative to conventional regression, for studying and modeling pedestrian behavior. The findings from this study would assist engineers and policy makers to take proactive measures for designing pedestrian friendly facilities to reduce signal violations. This would ultimately improve pedestrian safety at urban signalised intersections.

1. Introduction

Road accidents have become a major safety concern all over the world. Almost 1.35 million people die every year due to road accidents (World Health Organization, 2018). The problem is distressing in developing countries such as India. To put into context a total of 464,910 road accidents were reported in India resulting in 147,913 fatalities in 2017(Transport Research Wing, 2018).In a study which included 22 major countries worldwide, India ranks second in terms of incidence of road accident deaths per million population(International Road Federation, 2017). The predicament is alarming in big Indian cities as they reckon for 17.7% of the total number of accidents and 11.5% of fatalities.

The preponderance victims and ultimate sufferers of road accidents are pedestrians. For instance more than 270,000 traffic related pedestrian fatalities happened which account for 22% of the total deaths globally(Zhang et al., 2016). India being an emerging economy is no exception to this global trend.In India 13.8% of the total road accident deaths consists of pedestrians during 2017 which is 3.4% more than the previous year. The odds of pedestrian crashes are high at intersections due to the merging and diverging movement of vehicles and pedestrians from different directions and frequent interactions between them within the shared space. Accident involving pedestrians at signalised intersections is found to be more than 60% and 50% in Japan and China respectively(Marisamynathan and Vedagiri, 2018). In Delhi alone 62% of the total accidents occurred at intersections(Transport Research Wing, 2018). Furthermore, many of these accidents can be attributed to unsafe pedestrian crossings at intersections such as signal violations and not crossing along the crosswalk(Zaki and Sayed, 2014). Most of the intersections in India do not have dedicated phases for pedestrians. Pedestrians are released along with right and free left turning vehicles in a shared-signal system leading to frequent interaction and conflicts. In the absence of adequate facilities, pedestrians are compelled to cross the road in an unsafe manner.

* Corresponding author at: Jamia Millia Islamia, Department of Civil Engineering, Jamia Nagar, Okhla, New Delhi 110025, India.
E-mail address: faizanul26@gmail.com (F. Haque).

https://doi.org/10.1016/j.cstp.2023.101038
Received 4 October 2022; Received in revised form 4 March 2023; Accepted 10 June 2023
Available online 15 June 2023

Researchers have identified various factors relating to pedestrian demographic, socio-economic, traffic conditions, road and intersection geometry and built environment factors which are affecting pedestrians' crossing behavior. Studies suggest that pedestrian demographics such as age and gender are significantly associated with signal violation behavior(Guo et al., 2011; Hamed, 2001; Haque and Kidwai, 2023). Male pedestrians have more inclination for violations as they have higher risk taking tendencies(Lipovac et al., 2013). Younger pedestrians commit more violations whereas elders are most law-abiding(Ren et al., 2011). Pedestrians' socio-economic such as education and employment is the key determinant of crossing behavior at signalised intersections and suggest that higher level of education is associated with less violations(Marisamynathan and Vedagiri, 2018).

Pedestrian crossing speed, waiting time and crossing alone or in groups are also included in a few studies. Crossing speed, which is an important parameter for designing of pedestrian facilities is found to vary between 1.2 and 1.4 m/s. Pedestrians having illegal crossing behavior has higher crossing speed(Zaki and Sayed, 2014). Further crossing speed is mostly related to age, gender and group size, indicating that males have higher and old aged pedestrians have lower crossing speed. Increase in waiting time before crossing upsurges the risk taking and violation tendencies(Keegan and O'Mahony, 2003; Tiwari et al., 2007).Single and smaller groups have more likelihood of illegal crossings as compared to larger groups(Marisamynathan and Perumal, 2014).

Pedestrian crossing behavior is highly subjective to signal system, road or intersection geometry and presence of warning signs. Properly designed pedestrian facilities encourage them to cross the road safely (Sisiopiku and Akin, 2003). In a study in Japan, it is found that poor pedestrian signal design with more waiting time increases the odds of pedestrian's unsafe behavior(Iryo-Asano and Alhajyaseen, 2014). In terms of road geometry low compliance is found where road width is comparatively shorter(Ren et al., 2011). Pedestrians' decisions on when and where to cross are highly idiosyncratic and are influenced by many factors such as comfort, convenience and safety(Al Bargi et al., 2017). Nevertheless many studies have established the fact that pedestrian violation is predominant cause of crashes at signalised intersections (SPF, 2013; Zaki and Sayed, 2014).

Researchers have used several conventional behavioral and statistical techniques to study and model pedestrian behavior. Behavioral model such as the theory of planned behavior (TPB) is widely used method to model pedestrian violation behavior(Zhou et al., 2009). Various forms of regression are extensively used by several researchers to model pedestrian behavior. Discrete choice models such as binary and ordinal logit regression are used to model pedestrian violation and perception in a few studies(Haque and Kidwai, 2021; Marisamynathan and Vedagiri, 2018; Rankavat and Tiwari, 2016). Among various other techniques- principal component analysis, survival analysis, K-mean cluster analysis and mixed logit are also used by few researchers(Anciaes and Jones, 2018; Papadimitriou et al., 2017, 2012; Tiwari et al., 2007).

All of these conventionally used statistical techniques have their limitations. These methods are based on assumptions and precise attention is required for compliance. Data linearity and normality are other important constraints for the selection of these methods. These methods are data type specific and can perform only with the associated data types. Model fitting into the data is yet another challenge which needs to be addressed. The above limitations can be suitably tackled by using Artificial Neural Network (ANN). ANN is a soft computing technique which has gained popularity in recent years due to its simplicity and wide application. Although there are various statistical models that can handle the non-linearity in the data such as various forms of logistic regression(Rosenbloom, 2009; Schwebel et al., 2012; Thompson et al., 2013) and SEM-PLS(Díaz, 2002), ANN is highly robust in handling the non-linearity in the data. Further, it is not required to pre-select a model for our data and sufficient hidden nodes provide improved accuracy (Chakraborty et al., 2019). In spite of having several advantages of ANN, only a few researchers investigated the relevance of ANN in predicting pedestrian violation behavior (Zhang et al., 2022; Zhang et al., 2020).

The above appalling data indicates the poor state of road safety particularly the plight of vulnerable pedestrians in developing economies. Pedestrians' behavior is stochastic and is often challenging to develop universally accepted models related to their behavior. While several attempts have been made in developed countries, it requires more focus and attention in developing countries such as India. Several studies focused on crash data for safety assessment whereas quite a few researchers laid emphasis on pedestrians' violation behavior as a means of Surrogate Safety Measure (SSM) to assess pedestrian's safety at signalised intersections. In spite of wide application of ANN in various fields of traffic engineering, its implementation is limited in terms of pedestrian behavior analysis. Majority of researchers considered conventional statistical modeling techniques with little focus on ANN. Further it is difficult to model multiple parameters by regression techniques. So, to address the above mentioned research lacuna and requirement, the primary objective of this research is to understand the pedestrian signal violation behavior and the factors responsible, at urban signalised intersections. Further this study also aims at assessing the suitability of Artificial Neural Networks (ANN) to model pedestrians' signal violation behavior. Such studies have seldom been performed in developing economies such as India.

2. Research methodology

The steps involved in this research include selection of study site, data collection, data extraction-compilation, analysis and results. These steps are explained and discussed in the ensuing sections.

2.1. Study site selection

Initially a reconnaissance survey of several signalised intersections in New Delhi was performed to have a general understanding. After that few sites were filtered out considering various physical, vehicular and pedestrian factors. Finally after careful study of each filtered site, 11 signalised intersections with substantial pedestrian and vehicle volume were selected for this study. The selected intersections have varied pedestrian facilities, signal phases, physical and the built environment. Further all the intersections are spatially well parted with wide-ranging land uses in their surroundings. Details of study sites are given in Table 1.

2.2. Data collection

Road inventory surveys and video recordings were conducted to obtain the required intersection and pedestrian information. Data was collected on week days with normal weather conditions as during weekends sufficient traffic and pedestrian volume might not be available. An extensive format was prepared to record various physical and build environment features of the study sites. The format was used to record details about road geometry, signal phases, pedestrian facilities and land use in its surroundings.

Video graphic survey was used to record details about pedestrians such as demographics, crossing and violation behaviors. Two or more cameras were installed as per the site conditions to cover the entire section of study sites. The field of view and height of the cameras were adjusted so as to cover the ends of the carriageway including sidewalks, medians and crosswalks with few meters distance on both sides and signal phases. Two to three approaches were covered at each location. Video recordings were carried out during morning (9–10 AM) and evening (5–6 PM) hours without disturbing the normal traffic flow. Further the placement of cameras was not noticed by the pedestrians thus their naturalistic and actual behaviors were recorded. Data were extracted manually using AVS video editor software by trained volunteers. The entire duration (1 h morning and evening) of recorded videos was extracted. Video was played in ultra-slow motion at a rate of 2–3

Table 1
Study site details.

Intersection ID	Intersection Type	Pedestrian Volume (ped/h)	Vehicle Volume (veh/h)	Sample Size (N)	Carriageway Width (m)	Signal Violation (%)
A	3-Legged	289	7582	503	12.1	7.75
B	4-Legged	114	1472	225	9.65	17.33
C	3-Legged	517	3142	851	9.56	28.44
D	4-Legged	475	3250	747	12.42	2.68
E	4-Legged	383	1510	539	28.87	7.24
F	4-Legged	158	1602	266	8.23	14.28
G	4-Legged	391	1329	682	7.06	17.3
H	3-Legged	304	3180	409	10.97	21.27
I	3-Legged	401	4332	726	11.67	46.28
J	4-Legged	723	1964	959	12.12	26.8
K	3-Legged	198	2071	327	13.32	11.93

frames per second and required pedestrian data were extracted, coded and entered in preset excel formats. Pedestrian waiting and crossing times were recorded with an accuracy of milliseconds. Complete and useful information about 6234 pedestrians was extracted. Variables descriptions with coding are given in Table 2. Typical pedestrians' behaviors are shown in Fig. 1.

The exact age of the pedestrians could not be determined from video recordings, so it is approximately estimated by grouping them into young, middle and old age groups. The age group was judged based on their physical features. Factors such as facial features, physical appearances, hair color, walking style and clothing type were mostly used to estimate pedestrians' age group. Crossing alone is defined as single and if two pedestrians are crossing the road together then the group size is considered a pair. More than two pedestrians crossing together are defined as a larger group. Pedestrians observed to be crossing at a normal pace are defined as walking whereas those clearly observed to be running or walking at a high pace are classified as running. If a pedestrian crosses the road along a straight path then the crossing path is defined as a straight or else oblique crossing. Crossing direction is the near end if a pedestrian crosses the road from the sidewalk whereas it is far end if the crossing happens from median towards the sidewalk. If a pedestrian waits at the median or refuge island before crossing the road it is defined as a two stage crossing whereas crossing in one go without waiting at the median is defined as one stage crossing. The moment a pedestrian arrives at sidewalk or median is noted as arrival time and the moment a pedestrian steps on the carriageway is noted as departure time. After crossing the road the moment s/he sets foot on the sidewalk or median is noted as end time for road crossing. The difference between the departure time and arrival time is waiting time and the difference between the end time and departure time is defined as the crossing time. Crossing speed is defined as the width of the carriageway divided by crossing time.

Table 2
Variables description and coding details.

Pedestrian Characteristics	Category	Description (Coding)	Value
Gender	Male Female	Male (0), Female (1)	76.92% 23.08%
Age Group	Young Middle Old	Young: Up to 30 (0), Middle: 31 to 55 (1), Old: Above 55 (2)	45.56% 46.60% 7.84%
Group Size	Single Pair > 2	Pedestrian group size while crossingSingle (0), Pair (1), More than 2 (2)	65.42% 20.34% 14.24%
Walking Pace	Walk Run	Pedestrian walking speedWalking (0), Running (1)	92.91% 7.09%
Mobile Usage	No Yes	Using mobile, talking over phone and using head phones Yes (1), No (0)	93.10% 6.90%
Crossing Path	Straight Oblique	Pedestrian crossing pathStraight (0), Oblique (1)	64.13% 35.87%
Crossing Direction	Near end Far end	Pedestrian starts crossing from sidewalkNear end (0) otherwise Far end (1)	52.31% 47.69%
Stage Crossing	One stage Two stage	Pedestrian crossing the road in single or two stages Single Stage (0), Two Stages (1)	74.69% 25.31%
Carrying Object	No Yes	Pedestrian carrying any luggage, backpack or hand bags Yes (1), No (0)	42.94% 57.06%
Waiting time	Mean waiting time	Waiting time of pedestrian before crossing (sec)	5.543 sec
Crossing speed	Mean crossing speed	Width of carriageway divided by crossing time of pedestrian (m/sec)	1.302 m/s

2.3. Analysis and modeling

In this research a two-step – conventional statistics and ANN based hybrid approach is implemented. In the first step significant variables associated with signal violation behavior are identified using various statistical hypothesis testing techniques. In the second step the significant variables are taken as input parameters for the ANN to model pedestrians' signal violations. The use of two-step approach complements each other as conventional statistics are suitable for hypothesis testing but have comparatively less predictive ability. While ANN has significant predictive power but is not suitable for hypothesis testing as it follows " black box" approach(Hew et al., 2016).

Initially descriptive statistics are performed for the variables, to have brief inference of sample distribution. Concise information about pedestrian characteristics and behavior related to signal violation is obtained. To identify significant variables, Chi-square test is performed for categorical variables. Further independent sample *t*-test and ANOVA are used to compare the means of metric variables such as waiting time and crossing speed. Subsequently the significant variables are used as input for ANN to predict signal violation behavior.

Neural networks are akin to the human brain in terms of working and information processing. It learns from instances by means of interconnected neurons or nodes. Multilayer Perceptron (MLP) is most widely used neural network typology. It works on the basis of Feed Forward Back Propagation (FFBP) algorithm. Determination of number of hidden layers is an important task. Use of one to two hidden layers is suggested by many researchers. Activation functions are a critical part of the design of neural network. They are some specific functions which link the weighted sums of units in a layer to the values of units in the succeeding layer. Several forms of activation functions are available and their use largely depends upon the data type and the associated problem. Sigmoid activation function is used in this study as it is most widely used because of its differentiability(Ceylan et al., 2014). Further a gradient-descent supervised learning paradigm is used. This algorithm trains the neural networks by iteratively updating the synaptic weight values

Fig. 1. Typical instances of **a)** Signal violation **b)** Signal compliance **c)** Group crossing **d)** Mobile usage.

until the error function reaches a local minimum. Sensitivity analysis is performed to assess the relative importance of input variables on pedestrian signal violation behaviour. The schematic of FFBP is shown in Fig. 2.

3. Results and discussion

Pedestrian signal violation at each location is given in Table 1, and it is found to vary across all the locations. Overall signal violation is found to be about 20%. Lowest violation is observed at location D (2.68%), whereas highest was at location I (46.28%). Low rates of violation at location D could be explained as follows: The road width is about 12.5 m with heavy traffic, so pedestrians might not find a suitable rolling gap to cross the road, hence wait until the traffic has stopped. Countdown signals are present for both vehicles and pedestrians, hence pedestrian knows the waiting duration and mentally prepare themselves. The crosswalks were properly marked with stop lines and all the vehicles stopped before the stop line as traffic surveillance cameras were also installed. Contrary to the above no such features are available at location I.

The effect of the built environment and land use features on the violation tendencies of pedestrian is also explored in some studies. The inconsistencies in violations can also be attributed to characteristics of locations. Location D is predominantly commercial whereas location I has residential dwellings. The results are contrary to previous findings, which reported that commercial, industrial and alcohol outlets might influence a pedestrian to cross against the signal(Cinnamon et al., 2011). The above results suggest that the violation behaviour of pedestrians is highly location specific and often changes according to surroundings and built environment. The result supplements the outcomes obtained in

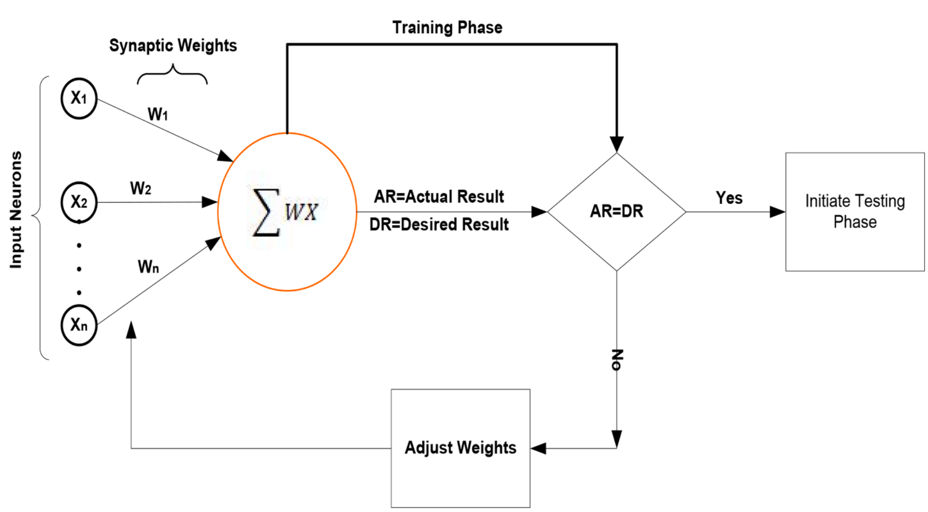

Fig. 2. Schematic of FFBP.

previous studies(Al Bargi et al., 2017).

In order to identify the independent variables that have a significant impact on pedestrian signal violation behavior, Pearson chi-square, independent sample *t*-test and ANOVA are performed. Gender (G), crossing path (CP), crossing direction (CD), carrying object (CO), mobile usage (MU), crossing speed (CS) and waiting time before crossing (WT) are found to be significant predictors of signal violation. Descriptions of the significant variables are given in Table 3.

Subsequently these seven significant independent variables as obtained by statistical hypothesis tests are used as input neurons for the ANN model. The ANN analysis is performed using IBM SPSS neural network module. From the total data, 80% was randomly allocated for training and the remaining 20% was for testing the network. In order to minimize the bias associated with a random split of training and testing samples, k-fold cross validation is used. Entire data set is split into mutually exclusive k-subsets (folds) of nearly equal sizes. The network is then trained and tested k-times(Delen et al., 2006). Similar to Leong et al.(2020) 10-fold cross validation is used. Several combinations consisting of a different number of hidden layers and units were tried to obtain the best predictions. It was found that two hidden layers with 5 and 4 units in first and second hidden layers respectively gave the best results. The resulting ANN model is shown in Fig. 3. Root Mean Square Error (RMSE) associated with the network for training and testing data is shown in Table 4. The mean RMSE values are 0.192 and 0.195 for training and testing data respectively. The values are relatively small suggesting a good model fit.

To measure relative importance of each of the input neurons in predicting the signal violation behavior, sensitivity analysis is performed. Normalized importance value is obtained by dividing the relative importance by the maximum importance value and then converting it into percentage. Sensitivity analysis results are shown in Fig. 4. The results show that crossing speed, crossing path and waiting time are top three important predictors respectively, whereas crossing direction of pedestrian is the least important.

Crossing speed is a crucial parameter in designing of pedestrian facilities. Variation of mean crossing speed with respect to signal violation, gender, age group, group size and mobile usage is shown in Fig. 5. Mean crossing speed of signal violating pedestrian is more as compared to signal compliant. The result is in line with previous research(Zaki and Sayed, 2014). It is due to the fact that violating pedestrians increase their speed abruptly and run while crossing the road to avoid collisions with vehicles. It is observed that males, young and those who cross the road alone are having significantly higher crossing speed as compared to their counter parts. The results yet again complement the results obtained in earlier studies(Marisamynathan and Perumal, 2014; Muley et al., 2017). Pedestrians using mobile phone or any other electronic device have lesser crossing speed. Pedestrians using mobile are not much attentive while crossing and hence walk in a very casual manner without rushing.

Table 3
Significant variables with respect to signal violation.

Variables	Category	Signal Violation (SV)		p-value
		No	Yes	
Gender (G)	Male	79.40%	20.6%	0.067*
	Female	81.60%	18.4%	
Crossing Path (CP)	Straight	82.70%	17.3%	p < 0.01***
	Oblique	74.80%	25.2%	
Crossing Direction (CD)	Near end	81.10%	18.9%	0.011**
	Far end	78.50%	21.5%	
Carrying Object (CO)	No	78.70%	21.3%	0.038**
	Yes	80.80%	19.2%	
Mobile Usage (MU)	No	80.20%	19.8%	0.011**
	Yes	75.10%	24.9%	
Crossing Speed (CS)	Mean crossing speed	1.264 m/s	1.312 m/s	0.073*
Waiting Time (WT)	Mean waiting time	12.850 sec	3.703 sec	0.005***

Significant at: * 90%; ** 95%; *** 99% Confidence Level.

Pedestrian crossing path i.e. crossing along a straight or oblique path is the second most important predictor. Pedestrians crossing along straight path are more compliant than those crossing along an oblique path. Pedestrians crossing obliquely enter the intersection from random directions and also leave randomly. Due to the virtue of their random entry and exit at the intersection area, they might fail to notice or are not much concerned about signal. On the other hand straight crossing pedestrians cross the road at designated crossing locations along with signal compliance.

Waiting time of signal violating pedestrians is found to be 3.5 times less than compliant ones. The result is in line with previous studies in which waiting time was found to be positively associated with signal violation(Keegan and O'Mahony, 2003; Tiwari et al., 2007). This can be explained by considering pedestrians' psychology and behaviour towards law abidance and safety. The pedestrians who are more concerned about safety and law might prefer to wait more for a safer opportunity to cross the road. Once pedestrians reach their threshold value of waiting time they would become impatient and exhibit dangerous behavior by crossing during the red light. Variations of mean waiting time within various groups are shown in Fig. 6. Males and younger pedestrians prefer to wait for a longer duration before crossing. Disparity in results from previous studies can be witnessed in this regard (Asaithambi et al., 2016; Lobjois and Cavallo, 2007; Yang et al., 2015). The inconsistent results establish the fact that pedestrian psychology and behaviour are highly indifferent across different locations. It is interesting to note that among group sizes, larger groups have substantially more waiting time than single or pair. This can be explained by virtue of platoon behaviour. Pedestrians might feel safer to cross the road in larger groups hence they wait for longer durations for the formation of platoon. The larger is the size of the group, the more they wait and the less likely people infringe on traffic rules(Guo et al., 2011; Hamed, 2001). Mobile phones using pedestrians have considerably more waiting time as compared to their counterparts.

Mobile usage is the fourth important predictor and is significantly affecting the signal violation behavior. About a quarter of mobile users are observed to violate the signal. Mobile users are highly distracted and might not notice the signal thus resulting in violations. In general mobile usage and distraction increase the probability of signal violations. The results are in line with previous findings in which several researchers have established the fact that distraction causes certain cognitive and sensory impairments, reduces situational awareness, resulting in violation tendencies and unsafe behaviors(Aghabayk et al., 2021; Russo et al., 2018).

Further with respect to gender the percentage of signal violation is slightly more for males as compared to females. It can be attributed to the fact that male pedestrians have higher violation and risk taking tendencies than females. Contrary to this, females are more sensitive to risk and have less preference to cross the road during the red phase(Díaz, 2002; Guo et al., 2012, 2011; Rosenbloom, 2009). However the results are contradictory to results from China where males were more likely to comply with traffic rules (Ren et al., 2011).Several previous studies have reported that gender does not have a significant effect on pedestrian's compliance behavior(Demiroz et al., 2015; Dommes et al., 2015).

To evaluate the predictive performance of ANN model, metrics such as sensitivity, specificity, false positivity and accuracy are used. These metrics are calculated by using Eqn 1, 2, 3 and 4. The matrix for these calculations is shown in Table 5. Sensitivity measures the proportion of true positive or true signal violations. It tells the probability of truly predicted signal violations from all the predicted violations. Sensitivity value close to unity suggests that our model is highly sensitive in detecting signal violations. Specificity measures true negative or true signal compliance predictions. It tells the probability of truly predicted signal compliances from all the predicted compliances. Specificity value close to unity suggests that our model is capable of predicting signal

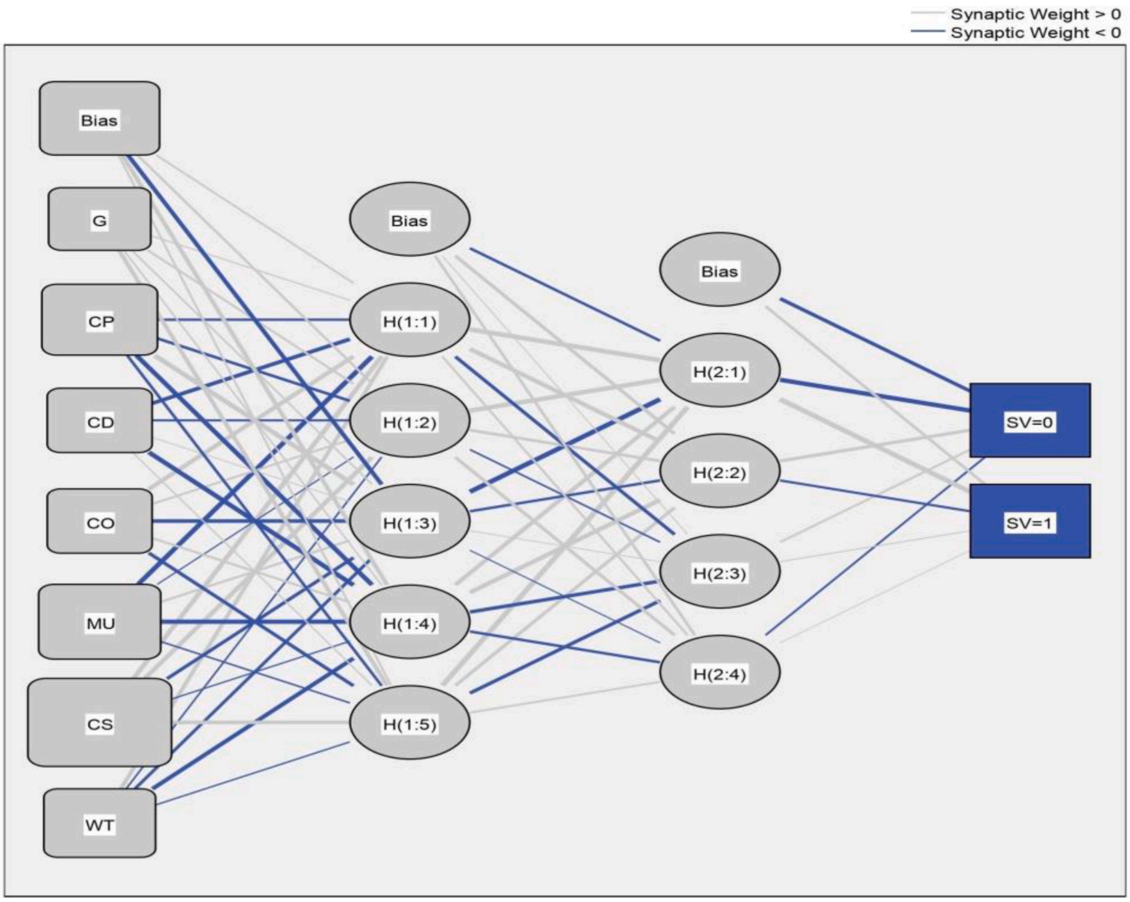

Fig. 3. ANN for signal violation.

Table 4
RMSE values for training and testing data sets.

Network	Training			Testing			Total Samples
	N	SSE	RMSE	N	SSE	RMSE	
1	5025	209.757	0.204	1209	43.474	0.190	6234
2	4941	179.979	0.191	1293	42.733	0.182	6234
3	4939	192.208	0.197	1295	45.530	0.188	6234
4	4948	194.974	0.199	1286	47.808	0.193	6234
5	4984	187.753	0.194	1250	42.589	0.185	6234
6	5025	199.725	0.199	1209	57.651	0.218	6234
7	4968	193.498	0.197	1266	48.246	0.195	6234
8	4959	193.171	0.197	1275	46.195	0.190	6234
9	5006	109.561	0.148	1228	53.088	0.208	6234
10	4956	187.561	0.195	1278	50.776	0.199	6234
Mean		184.819	**0.192**		47.809	**0.195**	
Standard Deviation		26.173	0.015		4.619	0.011	

Note: SSE = Sum of square error; RMSE = Root mean square error; N = Sample size.

compliance as well. The difference of unity and specificity is defined as a false positive rate. It is desirable to have lower values of false positivity. Accuracy value tells us about the overall correct predictions i.e. the final predictive ability of the model. Higher the value of accuracy, greater is the ability of the model to make correct predictions.

$$Sensitivity = \frac{TP}{TP + FN} \tag{1}$$

$$Specificity = \frac{TN}{TN + FP} \tag{2}$$

$$FalsePositivity = \frac{FP}{FP + TN} \tag{3}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

The accuracy of ANN to predict violation behaviour for training and testing data is found 85.88% and 85.19% respectively. Sensitivity, specificity and false positivity for training are 0.998, 0.607 and 0.393 respectively whereas for testing these values are respectively 0.999, 0.609 and 0.391. The receiver operating characteristic (ROC) is also used as an inclusive parameter to judge a model's functioning. The curve is plotted using sensitivity against false positivity (1-Specificity). The area under the ROC curve (AUC) is found to be 0.753, suggesting good model performance. All the above metrics indicate that ANN model can be significantly used to predict pedestrian signal violation behaviour.

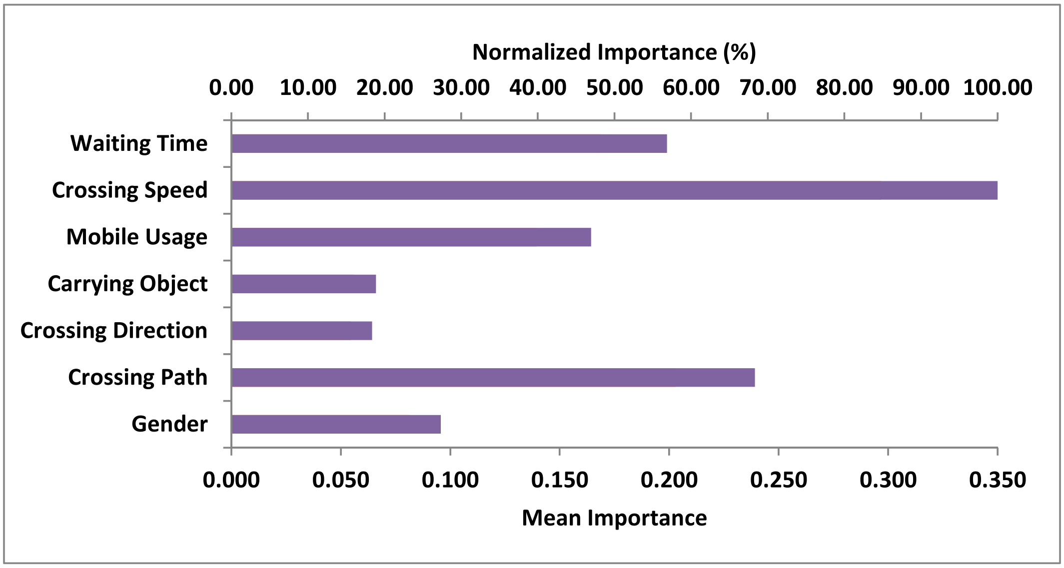

Fig. 4. Sensitivity analysis.

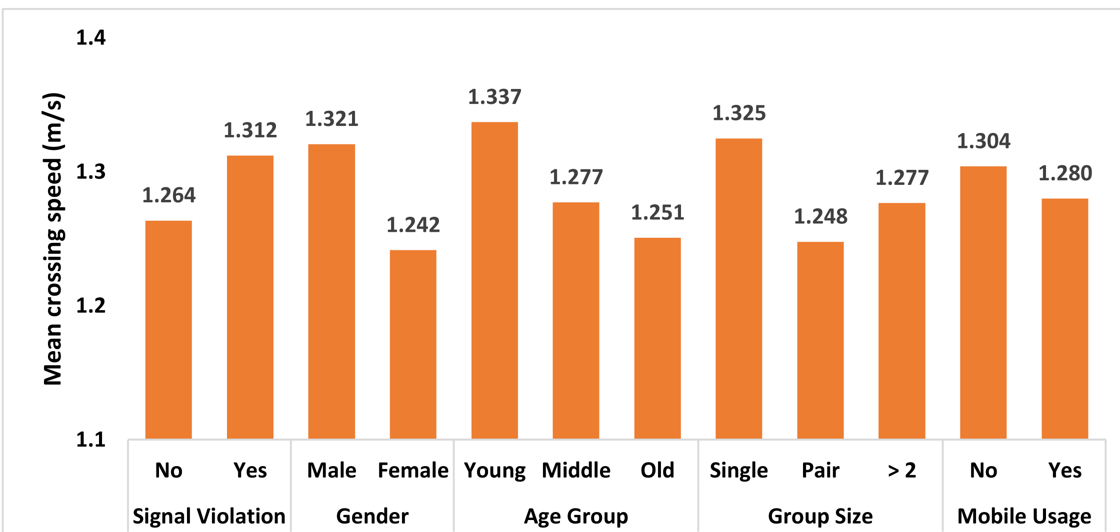

Fig. 5. Crossing speed variations.

Finally the suitability of ANN against conventional regression methods is evaluated. The results from the ANN are compared alongside the Binary Logistic Regression (BLR) model. The detailed process pertaining to the development of BLR can be found elsewhere(Haque and Kidwai, 2022). The BL model is developed using 80% of the data and the remaining 20% for validation. Results show that BL model accuracy is 79.8% and 80.3% for training and validation data sets respectively. Hosmer and Lemeshow test is used to assess the model performance. An insignificant p-value (0.181) is found at 95% confidence level suggesting that the BL model has significant predictive abilities. The accuracy of the ANN is higher than the conventional statistical BLR model. Hence it can be concluded that ANN model to predict pedestrian signal violation behavior is significantly better than the conventional BL regression model. As regression models have considerable drawbacks in terms of normality and linearity assumptions, ANNs can be a favorable choice.

4. Conclusions

Safety of pedestrians in urban Indian cities is a matter of serious concern and apprehension. Accidents involving pedestrians are being reported frequently at the intersection. Several of the accidents are mostly due to signal violations by the pedestrians. Therefore this research is focused on gaining plausible insight into factors affecting pedestrian's signal violation behavior at urban signalised intersections. The following important conclusions are made:

- Pedestrian behaviour and psychology greatly vary across different locations hence site specific understanding of people's attributes needs to be developed.
- Gender, crossing path, crossing direction, carrying an object, mobile usage, crossing speed and waiting time before crossing are found to have a significant association with pedestrian signal violation behaviour.
- Sensitivity analysis result reveals that crossing speed, crossing path and waiting time are the top three important factors affecting signal violation.
- This research successfully used ANN in modelling and predicting pedestrian's signal violation behaviour thus suggesting an

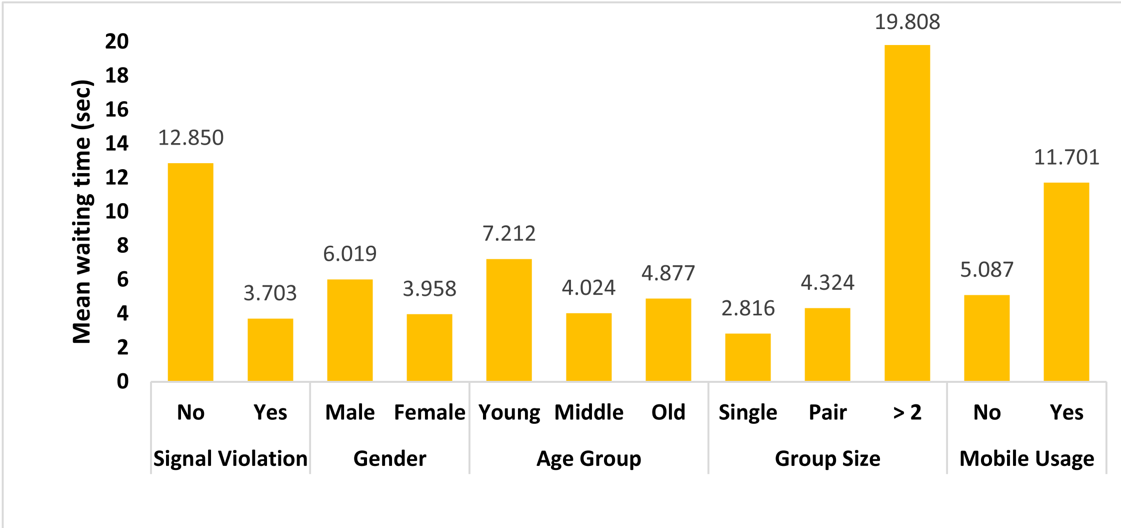

Fig. 6. Waiting time variations.

Table 5
Matrix to calculate metrics.

Observed	Predicted	
	No	Yes
No	True Negative (TN)	False Positive (FP)
Yes	False Negative (FN)	True Positive (TP)

alternative approach free from the assumptions and limitations of conventional methods.

The findings from this research opened a whole new approach to study and model pedestrian behaviours. It would help researchers, designers, policy makers and other stakeholders to better understand pedestrian's crossing behavior for the improvement of pedestrian's facilities so as to reduce signal violations. This would ultimately increase the overall safety of pedestrians at urban signalized intersections. Thus the findings from this research have wider practical as well as theoretical contributions. Special attention should be given while designing pedestrian facilities at signalized intersections.

The following recommendations are also made based upon the various findings so as to reduce violations:

- Special attention should be given while designing pedestrian facilities at signalized intersections. As pedestrian behaviour varies across intersections, site specific measures should be implemented to reduce violations and increase safety.
- Since violating pedestrians do not prefer to wait longer so short and dedicated pedestrian signals would be a crucial step in decreasing signal violations.
- Finally awareness regarding the significance of using crosswalk and following signal should be created at an early stage by means of education at the school and college levels.

In this study, efforts are made to identify significant factors affecting pedestrian signal violation behavior. However, only pedestrian personal attributes and crossing behaviors are considered. Road and intersection geometry, built environment, signals and traffic characteristics were not incorporated in this research. Further pedestrian-vehicular interaction could also be an important parameter influencing signal violations. Sample size can be further increased by including more intersections.

CRediT authorship contribution statement

Faizanul Haque: Methodology, Investigation, Formal analysis, Writing – original draft. **Farhan Ahmad Kidwai:** Conceptualization, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Aghabayk, K., Esmailpour, J., Jafari, A., Shiwakoti, N., 2021. Observational-based study to explore pedestrian crossing behaviors at signalized and unsignalized crosswalks. Accid. Anal. Prev. 151, 105990 https://doi.org/10.1016/j.aap.2021.105990.

Al Bargi, W.A., David Daniel, B., Prasetijo, J., Rohani, M.M., Mohamad Nor, S.N., Zainorizuan, M.J., Yee Yong, L., Alvin John Meng Siang, L., Mohamad Hanifi, O., Siti Nazahiyah, R., Mohd Shalahuddin, A., 2017. Crossing Behaviour of Pedestrians Along Urban Streets in Malaysia. MATEC Web of Conferences. 103, 08003.

Anciaes, P.R., 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.

Asaithambi, G., Kuttan, M.O., Chandra, S., 2016. Pedestrian Road Crossing Behavior Under Mixed Traffic Conditions: A Comparative Study of an Intersection Before and After Implementing Control Measures. Transp. Dev. Econ. 2, 1–12. https://doi.org/10.1007/s40890-016-0018-5.

Ceylan, H., Bayrak, M.B., Gopalakrishnan, K., 2014. Neural Networks Applications in Pavement Engineering : A Recent Survey. Int. J. Pavement Res. Technol. 7, 434–444. https://doi.org/10.6135/ijprt.org.tw/2014.7(6).434.

Chakraborty, A., Mukherjee, D., Mitra, S., 2019. Development of pedestrian crash prediction model for a developing country using artificial neural network. Int. J. Inj. Contr. Saf. Promot. 26, 283–293. https://doi.org/10.1080/17457300.2019.1627463.

Cinnamon, J., Schuurman, N., Hameed, S.M., 2011. Pedestrian Injury and Human Behaviour: Observing Road-Rule Violations at High-Incident Intersections. PLoS One. 6. e21063. https://doi.org/10.1371/journal.pone.0021063.

Delen, D., Sharda, R., Bessonov, M., 2006. Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. Accid. Anal. Prev. 38, 434–444. https://doi.org/10.1016/j.aap.2005.06.024.

Demiroz, Y.I., 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.

Díaz, E.M., 2002. Theory of planned behavior and pedestrians' intentions to violate traffic regulations. Transp. Res. Part F Traffic Psychol. Behav. 5, 169–175. https://doi.org/10.1016/S1369-8478(02)00015-3.

Dommes, A., Granié, M.A., Cloutier, M.S., Coquelet, C., Huguenin-Richard, F., 2015. Red light violations by adult pedestrians and other safety-related behaviors at signalized crosswalks. Accid. Anal. Prev. 80, 67–75. https://doi.org/10.1016/j.aap.2015.04.002.

Guo, H., Gao, Z., Yang, X., Jiang, X., 2011. Modeling pedestrian violation behavior at signalized crosswalks in China: A hazards-based duration approach. Traffic Inj. Prev. 12, 96–103. https://doi.org/10.1080/15389588.2010.518652.

Guo, H., Wang, W., Guo, W., Jiang, X., Bubb, H., 2012. Reliability analysis of pedestrian safety crossing in urban traffic environment. Saf. Sci. 50, 968–973. https://doi.org/10.1016/J.SSCI.2011.12.027.

Hamed, M.M., 2001. Analysis of pedestrians' behavior at pedestrian crossings. Saf. Sci. 38, 63–82. https://doi.org/10.1016/S0925-7535(00)00058-8.

Haque, F., Kidwai, F.A., 2023. Traffic Signal Violation Behavior of Pedestrian at Urban Intersections—A Case Study of New Delhi, in: Marano, G.C., Rahul, A. V, Antony, J., Unni Kartha, G., Kavitha, P.E., Preethi, M. (Eds.), Proceedings of SECON'22.Lecture Notes in Civil Engineering. Springer International Publishing, Cham, pp. 963–971. https://doi.org/10.1007/978-3-031-12011-4_81.

Haque, F., Kidwai, F.A., 2021. Modeling Pedestrian Compliance Behavior at an Urban Signalized Intersection. J. East. Asia Soc. Transp. Stud. 14, 1334–1348. https://doi.org/10.11175/easts.14.1334.

Haque, F., Kidwai, F.A., 2022. Crosswalk Utilization by Pedestrian : Perception versus Reality - Case Study of New Delhi. Int. J. Eng. Trans. C Asp. 35, 1123–1132. https://doi.org/10.5829/ije.2022.35.12c.08.

Hew, T.-S., Leong, L.-Y., Ooi, K.-B., Chong, A.-Y.-L., 2016. Predicting Drivers of Mobile Entertainment Adoption: A Two-Stage SEM-Artificial-Neural-Network Analysis. J. Comput. Inf. Syst. 56, 352–370. https://doi.org/10.1080/08874417.2016.1164497.

International Road Federation, 2017. World Road Statistics 2017. Geneva.

Iryo-Asano, M., Alhajyaseen, W.K.M., 2014. Analysis of Pedestrian Clearance Time at Signalized Crosswalks in Japan. Procedia Comput. Sci. 32, 301–308. https://doi.org/10.1016/j.procs.2014.05.428.

Keegan, O., O'Mahony, M., 2003. Modifying pedestrian behaviour. Transp. Res. Part A Policy Pract. 37, 889–901. https://doi.org/10.1016/S0965-8564(03)00061-2.

Leong, L., Hew, T., Ooi, K., Wei, J., 2020. Predicting mobile wallet resistance: A two-staged structural equation modeling-artificial neural network approach. Int. J. Inf. Manage. 51, 102047 https://doi.org/10.1016/j.ijinfomgt.2019.102047.

Lipovac, K., Vujanic, M., Maric, B., Nesic, M., 2013. Pedestrian Behavior at Signalized Pedestrian Crossings. J. Transp. Eng. 139, 165–172. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000491.

Lobjois, R., Cavallo, V., 2007. Age-related differences in street-crossing decisions: The effects of vehicle speed and time constraints on gap selection in an estimation task. Accid. Anal. Prev. 39, 934–943. https://doi.org/10.1016/j.aap.2006.12.013.

Marisamynathan, Perumal, V., 2014. Study on pedestrian crossing behavior at signalized intersections. J. Traffic Transp. Eng. (English Ed. 1, 103–110. https://doi.org/10.1016/S2095-7564(15)30094-5.

Marisamynathan, S., Vedagiri, P., 2018. Modeling pedestrian crossing behavior and safety at signalized intersections. J. Transp. Res. Board 2672, 76–86. https://doi.org/10.1177/0361198118759075.

Muley, D., Kharbeche, M., Alhajyaseen, W., Al-Salem, M., 2017. Pedestrians' Crossing Behavior at Marked Crosswalks on Channelized Right-Turn Lanes at Intersections, in: Procedia Computer Science. Elsevier B.V., pp. 233–240. https://doi.org/10.1016/j.procs.2017.05.339.

Papadimitriou, E., Theofilatos, A., Yannis, G., Sardi, G.M., Freeman, R.P.J., 2012. Road Safety Attitudes and Perceptions of Pedestrians in Europe. Procedia - Soc. Behav. Sci. 48, 2490–2500. https://doi.org/10.1016/j.sbspro.2012.06.1220.

Papadimitriou, E., Lassarre, S., Yannis, G., 2017. Human factors of pedestrian walking and crossing behaviour. In: Transportation Research Procedia. Elsevier B.V, pp. 2002–2015. https://doi.org/10.1016/j.trpro.2017.05.396.

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.

Ren, G., Zhou, Z., Wang, W., Zhang, Y., Wang, W., 2011. Crossing behaviors of pedestrians at signalized intersections: Observational study and survey in China. Transp. Res. Rec. 2264, 65–73. https://doi.org/10.3141/2264-08.

Rosenbloom, T., 2009. Crossing at a red light: Behaviour of individuals and groups. Transp. Res. Part F Traffic Psychol. Behav. 12, 389–394. https://doi.org/10.1016/j.trf.2009.05.002.

Russo, B.J., James, E., Aguilar, C.Y., Smaglik, E.J., 2018. Pedestrian behavior at signalized intersection crosswalks: Observational study of factors associated with distracted walking, pedestrian violations, and walking speed. Transp. Res. Rec. 2672, 1–12. https://doi.org/10.1177/0361198118759949.

Schwebel, D.C., Stavrinos, D., Byington, K.W., Davis, T., O'Neal, E.E., De Jong, D., 2012. Distraction and pedestrian safety: How talking on the phone, texting, and listening to music impact crossing the street. Accid. Anal. Prev. 45, 266–271. https://doi.org/10.1016/j.aap.2011.07.011.

Sisiopiku, V.P., Akin, D., 2003. Pedestrian behaviors at and perceptions towards various pedestrian facilities: An examination based on observation and survey data. Transp. Res. Part F Traffic Psychol. Behav. 6, 249–274. https://doi.org/10.1016/j.trf.2003.06.001.

SPF, 2013. Road traffic accident database 2009-2011. Singapore.

Thompson, L.L., Rivara, F.P., Ayyagari, R.C., Ebel, B.E., 2013. Impact of social and technological distraction on pedestrian crossing behaviour: An observational study. Inj. Prev. 19, 232–237. https://doi.org/10.1136/injuryprev-2012-040601.

Tiwari, G., Bangdiwala, S., Saraswat, A., Gaurav, S., 2007. Survival analysis: Pedestrian risk exposure at signalized intersections. Transp. Res. Part F Traffic Psychol. Behav. 10, 77–89. https://doi.org/10.1016/j.trf.2006.06.002.

Transport Research Wing, 2018. Road accidents in India-2017, MoRTH. Government of India, New Delhi, India.

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

Yang, X., Abdel-Aty, M., Huan, M., Peng, Y., Gao, Z., 2015. An accelerated failure time model for investigating pedestrian crossing behavior and waiting times at signalized intersections. Accid. Anal. Prev. 82, 154–162. https://doi.org/10.1016/j.aap.2015.04.036.

Zaki, M.H., Sayed, T., 2014. Automated Analysis of Pedestrians' Nonconforming Behavior and Data Collection at an Urban Crossing. Transp. Res. Rec. J. Transp. Res. Board 2443, 123–133. https://doi.org/10.3141/2443-14.

Zhang, S., Abdel-Aty, M., Yuan, J., Li, P., 2020. Prediction of Pedestrian Crossing Intentions at Intersections Based on Long Short-Term Memory Recurrent Neural Network. Transp. Res. Rec. J. Transp. Res. Board 2674, 57–65. https://doi.org/10.1177/0361198120912422.

Zhang, S., Abdel-Aty, M., Wu, Y., Zheng, O.u., 2022. Pedestrian Crossing Intention Prediction at Red-Light Using Pose Estimation. IEEE Trans. Intell. Transp. Syst. 23 (3), 2331–2339.

Zhang, W., Wang, K., Wang, L., Feng, Z., Du, Y., 2016. Exploring Factors Affecting Pedestrians' Red-Light Running Behaviors at Intersections in China. Accid. Anal. Prev. 96, 71–78. https://doi.org/10.1016/j.aap.2016.07.038.

Zhou, R., Horrey, W.J., Yu, R., 2009. The effect of conformity tendency on pedestrians' road-crossing intentions in China: An application of the theory of planned behavior. Accid. Anal. Prev. 41, 491–497. https://doi.org/10.1016/j.aap.2009.01.007.