Proceedings of 10th IOE Graduate Conference
Peer Reviewed
ISSN: 2350-8914 (Online), 2350-8906 (Print)
Year: 2021 Month: October Volume: 10

Prediction of Pedestrian Gap Acceptance Behavior in Urban Mid-Block Illegal Crossing under Mixed Traffic Condition

Krishna Chand [a], Anil Marsani [b]

[a] *Department of Civil Engineering, Pulchowk Campus, IOE, Tribhuvan University, Nepal*
Corresponding Email: [a] krchand555@gmail.com, [b] anilmarsani@ioe.edu.np

Abstract
The crossing movement of the pedestrian across the road at mid-block locations illegally leads to high risk of pedestrian- vehicle conflicts. This study is concentrated on the size of the vehicular gaps accepted by the pedestrian for crossing at mid-block section of the ring road. The middle four lanes which are free from the disturbance of the temporary parking of vehicles were only used for this study. Two popular machine learning models, Random Forest (RF) and Extreme Gradient Boosting (XG Boost) were used in addition to traditional Multiple Linear Regression (MLR) model to predict the size of the accepted gaps. In total, twenty one independent variables were extracted from the video recorded at site consisting of pedestrian demographic, behavioral and traffic related characteristics. The significant independent variables in MLR model was obtained based on the t-test value and p-value. Similarly, the importance score of independent variables for RF and XG Boost model was obtained using function feature_importances_ in Python Jupyter Notebook. Seven variables were found to be statistically significant in MLR model. To keep same number of variables in all three final models, seven corresponding important variables were chosen for fitting RF and XG Boost model. Two variables, safety distance and vehicle speed were found to be important in all three models. The fitness of model in terms of R-squared value with important variables only was found to be 89.20%, 99.66% and 99.92% for MLR, RF and XG Boost models respectively. Similarly, the performance of these trained models in terms of root mean squared error was obtained as 0.3358, 0.1114 and 0.0914. Both machine learning models performed better than MLR. XG Boost outperformed rest of the models in model fitness and gap size prediction.

Keywords
Multiple Linear Regression, Random Forest, Extreme Gradient Boosting, Accepted gap size, Rolling gap, Safety distance

1. Introduction

1.1 Background

One of the major mode of urban mobility is walking. It is the safest mode of travel if proper pedestrian facilities are provided. So, local transport authorities often promotes walking in urban areas.The demographic as well as socioeconomic characteristics of the pedestrian make them to decide whether to walk or not for short trips. Walking consists of two types of movements; along the road and across the road. Road crossing cannot be avoided in walk trips and pedestrians are the most exposed road users. While characterizing the road crossing behavior of pedestrian, the term 'critical gap' is encountered which is the duration in seconds below which a pedestrian will not try to start walking across the road [1].

Pedestrian's illegal crossing at overhead bridge locations for time saving and comfort [2] confirms that unless properly planned, the available pedestrian facilities cannot be utilized by the pedestrians..The motivations for illegal crossing at urban mid-blocks are habit (repeat prior goal oriented behavior or cue dependent automaticity), attitude (belief that is not wrong to do it) and efficacy (belief in one's self act) [3]. Pedestrians take calculated risks while crossing the road which may be different for different pedestrian. Their movement across the road intersect the vehicular path offering high risk of conflict with vehicles. The unprotected mid-block locations offer very few gaps to the crossing pedestrian under mixed traffic conditions. Safe gaps are rarely available in those few gaps. The availability of unsafe gaps compel the pedestrians to cross the road quickly by increasing crossing speed, changing crossing path or rolling over small gaps. The misjudgment of vehicular gaps by the pedestrian and inability of driver at high speed to yield to pedestrian may result severe pedestrian-vehicle conflict.

As per WHO (Global Status Report on Road Safety 2018)[4], road traffic injuries are the eighth leading cause of death for people of all ages. About 1.35 million people dies each year on world as road traffic deaths. The death rates in low income countries is three times higher than that of high income countries. Pedestrian, cyclists and motorcyclists constitute more than half of all traffic deaths. In the context of Nepal, the Fiscal year 2076/2077 and 2077/2078 suffered from 50 and 37 (at the end of eight month) pedestrian deaths in Kathmandu valley alone [5]. The economic burden of road traffic injuries (RTIs) in Nepal was USD 122.88 million as total costs of RTIs in 2017 which was equivalent to 1.52% of the gross national product [6].

Thus, pedestrian safety improvement in urban roads demands the understanding of pedestrian gap acceptance behaviors at mid-blocks. Once the pedestrian gap acceptance behavior is understood, it can be used to enhance the existing crosswalk facilities, installation of new pedestrian signal.

1.2 Research Objectives

The main purpose of this thesis is to develop pedestrian gap acceptance prediction model using statistical and machine learning approaches for the proposed study area. The detailed objectives are:

- To determine the pedestrian characteristics, pedestrian behavioral characteristics and traffic related characteristics that are influential to gap acceptance behavior of pedestrian.
- To predict the size of the accepted gap under the variables considered.
- To compare the performance of the model in predicting the accepted gaps.

2. Literature Review

Several researchers attempted to model the pedestrian gap acceptance behavior using different modeling techniques. The study were mainly focused on modeling two aspects of pedestrian behavior, vehicular gap size accepted and decision for crossing the road. The selection of variables for these studies were also different depending upon their site location, traffic characteristics and scope of their research work.

2.1 Modeling Techniques

Multiple linear regression (MLR) technique was widely used in existing literatures to model the size of gap accepted by the pedestrian [7, 8] as well as to understand whether a particular independent variable viz. presence of refuge island at mid-block crossing locations is significant to the size of the accepted gap or not [9]. The R-squared value of MLR model approximately ranges between 60% to 80% as reported by previous literatures. Even though the model performance was not so attractive, this technique is frequently used because of it's simplicity in model application and interpretation of the results. In addition to MLR technique,one study also applied Artificial Neural Network (ANN) to model the accepted gap size and obtained a better model performance with R-squared value of approximately 85% [10].

Similarly, lognormal regression technique was also extensively utilized in modeling the accepted vehicular gap size while crossing the road [11, 12, 13, 14, 15, 16]. This model is also a linear model with an output variable converted into their corresponding logarithmic value. The results of past studies depicts that the range of R-squared value for this model approximately lies between 45% to 85%. The model application and results interpretation is simpler as MLR model.

2.2 Significance of different variables

In case of critical distance, pedestrians crossing the road when the approaching vehicle is near will receive smaller gaps [7, 11]. However, one study did not found critical distance as significant variable[13].

Several studies reported that pedestrian behavior is significantly affected by the age of the pedestrian. They added, accepted gap size for older pedestrian is higher because of their low walking speed while crossing the road [12, 13, 14, 15]. However, pedestrian age was found insignificant to predict gap size by some studies ([7, 11].

Male selects larger vehicular gap showing less risky behavior[11]. On the other hand, some literatures reported that female are more possibly to accept larger gap size [14, 16]. In contrast to this finding, some other studies did not find gender as a significant variable [12, 13].

Many studies found that the gap size for pedestrian showing rolling behavior is small as the pedestrian changes their speed, crossing path to accept smaller gaps which indicates risky behavior [12, 13]. However, one study reported that larger gaps are accepted by the pedestrian showing rolling behavior [7].

The frequency of attempts for crossing affects the accepted gap size negatively. Pedestrians are more seemingly to cross the road in small gaps as the frequency of attempts for crossing increases [13, 14]. On the other hand, this feature was not found significant in some studies [12, 15].

Pedestrians prefer larger and safer gaps if the speed of the approaching vehicles to crossing zone is high [14, 15]. On the other hand, as the speed of approaching vehicles goes up, pedestrians accept smaller gaps. [16, 17].

3. Methodology

3.1 Study Area

A Ring road mid-block section was used in this study as shown in Figure 1. The section is located at Ekantakuna, Lalitpur in front of Department of Transport and Management office. The section has 8 lanes with two raised medians which separate the two lanes of either side from middle four lanes. The flow in outer two lanes of either side of the road is highly disturbed as it is used by passenger vehicles for short term parking and pedestrian loading and unloading. Hence, middle four lanes with 15m width is only used for this study.

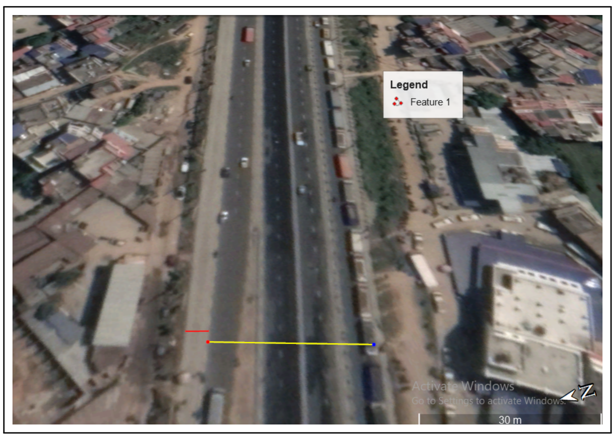

Figure 1: Site Location

The middle four lanes serve high volume of heterogeneous traffic flow in both directions. Though the centre of the middle four lane is median free, it is often used by pedestrian waiting for suitable time gap between vehicles to cross the road.

3.2 Data Collection and Extraction

Video-graphic survey was conducted on two typical weekdays in normal weather condition to collect necessary data. The recording time was selected based on previous day observation such that the pedestrian crossing the road was maximum. Video of study site was recorded from 10:00am to 12:15pm and from 1:00pm to 3:15pm in first and second survey day respectively. Data were extracted by playing the video repeatedly at lower speed. A total of 720 data points were extracted from the video in which pedestrian accepted vehicular gap to cross the road in presence of vehicle.

3.3 Variables Selection

The choice of variables for the prediction of pedestrian gap acceptance depends upon number of factors. It depends upon traffic conditions, geometrics of road crossing and other factors. Accepted gap is the minimum time gap between two consecutive vehicles at the pedestrian crossing point that is accepted by pedestrian for road crossing. It is measured in seconds from the video recorded. The demographic characteristics of the pedestrian viz. gender and age was obtained by visual perception from the recorded video. Age was only divided into two groups as age young (age below 35) and middle-aged (age above 35) because the number of childrens and old age people crossing the road at this location is significantly very less. In addition to this, pedestrians carrying bag or not, using mobile phone or not while crossing the road was also noted. Pedestrian group size in this study refers to number of road users walking together while crossing the road. This feature was also noted to understand the pedestrian behavior in group. Further, the influence of the pedestrian crossing opposite direction to subject pedestrian was also noted by counting the number of pedestrian passing from a close distance. Furthermore, waiting time of the pedestrian for obtaining suitable gap size was noted from the recorded video. The number of attempts made for crossing the road was recorded by counting total attempts to cross till gap was accepted. Moreover, whether the accepted gap occurred in near lane or far lane was also noted. Lag or gap was also noted based on whether the pedestrian crosses road before the arrival of the first vehicle or not. Similarly, rolling gaps, which occurs when pedestrian rolls over the small available gaps and running at the time of road crossing was obtained from video. Pedestrian speed was obtained from the ration of crossing distance and crossing time. Pedestrian path was divided into two groups as perpendicular or oblique based on visual inspection. The pedestrian speed change during road crossing is also obtained from video by visual inspection. While crossing the multi-lane road, pedestrian face vehicular gaps of different sizes in different lanes. In this study, the lane in which pedestrian gets minimum gap is termed as critical lane and the corresponding vehicle is termed as critical vehicle. The distance between this critical vehicle and pedestrian at the time of road crossing was computed from video with reference to the distance in the field and kept under variable safety distance. Further, the critical vehicle type was divided into three groups as two wheeler, four wheeler and other heavy vehicles, and noted. Furthermore, the speed of the critical vehicle was obtained from the ratio of trap length of 20m considered along approaching vehicles from pedestrian crossing point and the corresponding time taken to travel this distance. The yielding critical vehicle after observing pedestrian crossing the road was also recorded in order to detect the speed or lane change by the critical vehicle or not.

Finally, two new variables coded as X_B and X_NB are introduced in this study based on the observation of recorded video. There is not any particular location such as median cut, for crossing the road. So, pedestrian crosses the road from multiple points at the same time in the similar fashion as if they are moving together in a single group. So, the number of pedestrians facing the critical vehicle before and after targeted pedestrian, within a distance of 20m on either side from targeted pedestrian was noted under variables X_B and X_NB respectively.

3.4 Multiple Linear Regression

The linear relationship between the independent and dependent variables can be established using multiple linear regression technique (MLR). MLR is widely used for developing the pedestrian gap acceptance model. The equation of the MLR is in the form:

$$Y = ß0 + ß1 * X1 + ß2 * X2 + ß3 * X3 + \ldots + ßn * Xn$$

Where, Y = dependent variable, ß0 = intercept, ß1, ß2, ß3,. . . , ßn = variable coefficients and X1, X2, X3, . . . , Xn = independent variables.

3.5 Random Forest

Random forest (RF) is widely used in crash prediction model. It can be used for both classification as well as regression problem [18, 19]. RF is an ensemble method under supervised machine learning technique. Ensemble method refers to the technique in which the final output of the model is generated by the aggregation of individual regression trees also known as weak learners. In supervised machine learning techniques both input and output variables are fed to the model. In order to form a forest, RF uses ensemble of regression trees. For regression problem, each tree predicts an output as shown in Figure 2. With all the given inputs, the final prediction will be based upon the mean of all predicted outputs. To reduce the bias, RF grows each tree to maximum depth. Variance reduction in RF is achieved by bootstrap technique and selection of subset of variables for each tree. Bootstrap technique allows us duplication of rows and columns while selecting subset of data to grow individual tree. It chooses the best splitting variable from a randomly selected variable subset at each node for each tree. Jupyter Notebook is employed for the execution of RF algorithm in this study which is an open source web-based application and provides an environment to edit and run python programming codes.

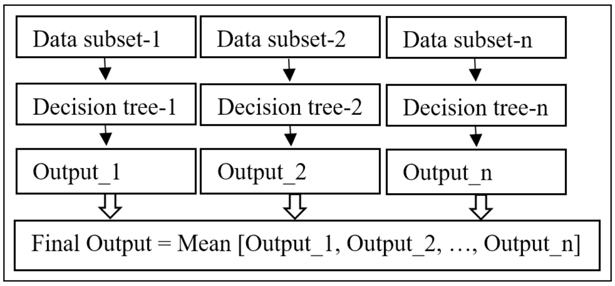

Figure 2: Random Forest Prediction

3.6 Extreme Gradient Boosting

Extreme Gradient Boosting (XG Boost) is also used in classification as well as regression problems such as crash severity prediction, taxi trip travel time prediction [20, 21]. XG Boost is also an ensemble method under supervised machine learning technique. XG Boost algorithms are the improvement over gradient boosting. Both XG Boost and gradient boosting algorithms fit residuals to regression trees but XG Boost algorithm consists of regularization parameter which reduces the overfitting problem. Overfitting refers to the problem in which the model performs good with the training dataset only resulting poor accuracy in prediction for testing dataset. Figure 3 shows the working of this model. At first, the predictor variables dataset are fed to each of the individual gradient boosting trees. Then, all gradient boosting trees produce their individual outputs. The final predicted output by the XG Boost model is obtained as: Prediction = Base model value + Learning rate*(Output_1 + ... + Output_n). Here, base model is a model created initially having a constant value for all predicted output values to obtain residuals as shown in Figure 3. And learning rate is a parameter that determines size of the step at each iteration while moving towards maximization of the gain function. It' value ranges from 0 to 1. Jupyter Notebook is employed for the execution of XG Boost algorithm in this study.

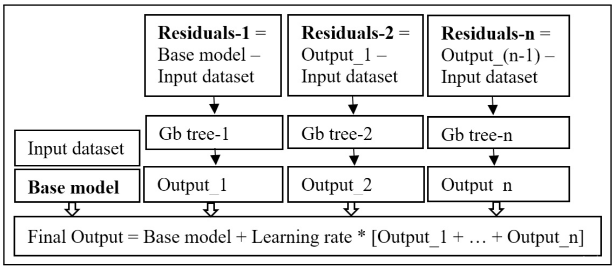

Figure 3: XG Boost Prediction

4. Data Analysis and Results

4.1 Overview of Data

The minimum accepted gap was 0.5s which was the case of pedestrian entering into the lane for adopting rolling behavior while the maximum accepted gap was 7.75s. Similarly, minimum and maximum waiting time was 0s and 100.24s respectively. The pedestrian accepted minimum safety distance of 6.5m and maximum safety distance of 72.39m. The minimum and maximum speed of the vehicle accepted by the crossing pedestrian was computed to be 7.31kmph and 92.74kmph respectively. The minimum and maximum speed of pedestrian while crossing the road was computed as 0.28m/s and 3.2m/s respectively.

The 720 datasets used in this study consists of 661 male and 59 female pedestrians. By visual inspection 329 pedestrian were categorized under below 35 years age group and rest were kept under above 35 years age group. The number of pedestrian crossing the road with bags or files on hand was significant. A total of 317 pedestrian crossed road with bag or files on hand. Few pedestrian also used mobile phone while crossing the road. A total of 39 pedestrian were observed walking across the road using mobile phone. Only 33 pedestrians were observed to have followed paths that are very close with that of pedestrians heading towards opposite direction at the same time. The number of pedestrians crossing the road in more than one attempts was 15 only. Out of 720 pedestrians, 222 accepted gap in far lane. Similarly 213 pedestrians accepted lag while rest accepted gap. 372 pedestrians rolls over small available gap while 47 pedestrians ran to cross the road quickly. Oblique path was chosen by 202 while 261 pedestrians changed their speed at the time of crossing the road. Only 91 vehicles yielded to pedestrians. Most of the pedestrians accepted the gap of two wheelers then four wheelers and finally others. A total of 506 pedestrians were observed to have accepted the gap of two wheelers while 142 accepted the gap of four wheelers.

4.2 Some Terminologies

Some of the functions that are used during the course of model development and validation include train_test_split, feature_importances_ and GridSearchCV. These fuctions are defined under a machine learning library known as scikit-learn or sklearn library.

4.3 Model Development

Randomly selected 75% of datasets by train_test_split function was used for training the models. The fitness of the model with the training dataset are described below.

4.3.1 Multiple Linear Regression

The t-test value and p-value were used to determine the significance of independent variables in predicting the accepted gap size. Out of twenty one, seven features were found to be statistically significant viz. group size, gap type (gap or lag), rolling gap, safety distance, vehicle yield, X_NB and vehicle speed.Cosidering only significant variables, the relationship between the accepted gap size and the independent variables was obtained as:

$$\begin{aligned}Acc\_gap &= 2.8833 - 0.0538 * Group\_size \\ &- 0.1094 * Gap\_or\_lag + 0.1104 * Rolling\_gap \\ &+ 0.0837 * Safety\_distance + 0.1616 * Veh\_yield \\ &- 0.0478 * X\_NB - 0.0607 * Veh\_speed \quad (1)\end{aligned}$$

The fitness of the MLR model with training dataset is obtained as: Rsquared value = 0.893 and adjusted Rsquared value=0.892 .

4.3.2 Random Forest

The feature_importances_ function was used to determine the importance of independent variables in predicting the accepted gap size. The importance score is given to the features based on their ability to split the dependent variable values such that the sum of mean squared error while building decision trees is minimized. The seven important features were group size, waiting time, pedestrian speed, safety distance, vehicle type, X_NB and vehicle speed.The optimal value of hyperparameters for training dataset was obtained using function GridSearchCV. This function takes the multiple inputs of the selected hyper-parameters such as n_estimators, max_features, bootstrap and, selects the value of these hyperparameters such that the training dataset fit in the best way with the model. Considering important variables only, the optimal value of hyperparameters were obtained as: n_estimators = 1000, max_features = 7 and bootstrap = True . The fitness of the RF model with training dataset is obtained as: R-squared=0.9966 .

4.3.3 Extreme Gradient Boosting

In this model also, the importance of independent variables in estimating the accepted vehicular gap was determined using feature_importances_ function.The importance score is given to the features based on their ability to split the dependent variable values such that the gain value while building gradient boosting trees is minimized. The seven important features were waiting time, rolling gap, pedestrian path, safety distance, vehicle yield, vehicle type and vehicle speed.The optimal value of hyperparameters for training dataset was obtained using function GridSearchCV. This function takes the multiple inputs of the selected hyper-parameters such as max_depth, n_estimators, learning_rate and, selects the value of these hyperparameters such that the training dataset fit in the best way with the model. Considering important variables only, the optimal value of hyperparameters were obtained as: max_depth = 2, n_estimators = 1000 and learning_rate = 0.1 . The fitness of the XG Boost model with training dataset is obtained as: Rsquared=0.9992 .

5. Validation of the Model

Randomly selected 25% of datasets by train_test_split function was used for validation of the model.The scatterplot with the best fit line between predicted vs observed value of accepted gap size considering significant or important independent variables only for MLR, RF and XG Boost model is shown by Figure 2, Figure 4 and Figure 5 respectively.

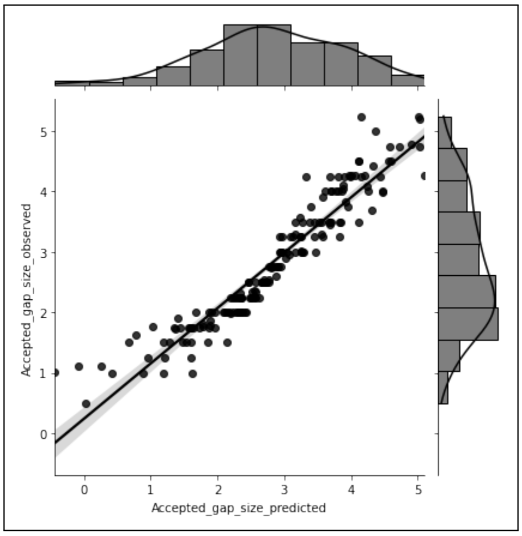

Figure 4: Observed Gap(Y) vs Predicted Gap(X)

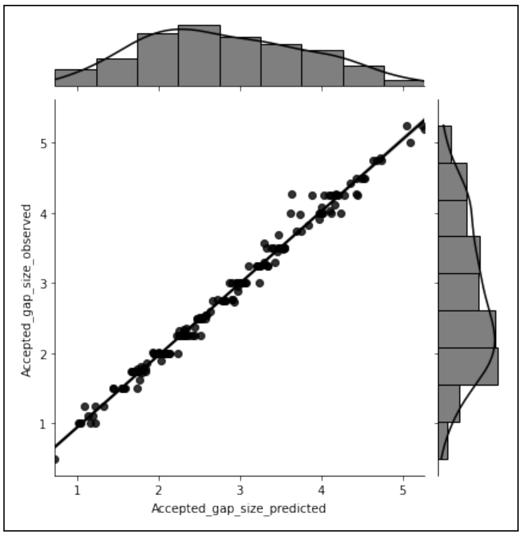

Figure 5: Observed Gap(Y) vs Predicted Gap(X)

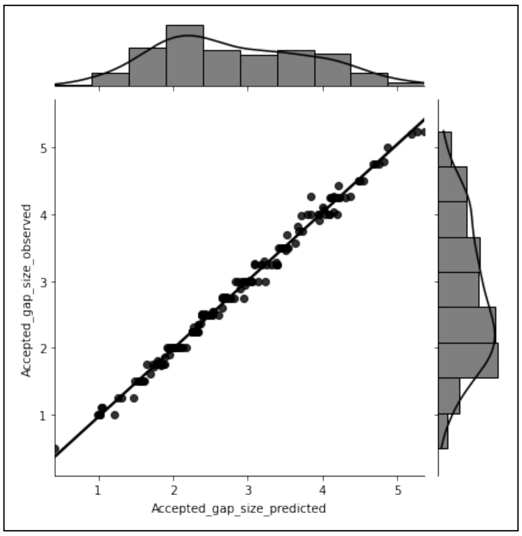

Figure 6: Observed Gap(Y) vs Predicted Gap(X)

Table 1 presents the results of regression analysis shown by Figure 4, Figure 5 and Figure 6 respectively.

Table 1: Validation Results

Multiple Linear Regression (MLR)
R-squared value: 89.60%
Significance (F-statistic): 2.42E-92
Root Mean Squared Error: 0.3358
Regession Equation: Y = 0.2388 + 0.9164*X
Random Forest (RF)
R-squared value: 98.85%
Significance (F-statistic): 1.48E-177
Root Mean Squared Error: 0.1114
Regession Equation: Y = -0.0877 + 1.0272*X
Extreme Gradient Boosting (XG Boost)
R-squared value: 99.22%
Significance (F-statistic): 3.30E-192
Root Mean Squared Error: 0.0914
Regession Equation: Y = -0.0570 + 1.0211*X

6. Conclusion and Recommendations

This study is focused on developing models that can predict the size of the vehicular gap accepted by the pedestrians while crossing the road. For this, video-graphic survey was conducted at mid-block section of the eight-lane ring road. A total of twenty one independent variables were extracted from the video recorded at site consisting of pedestrian demographic, behavioral and traffic related characteristics. The extracted data was used for developing three models viz. MLR, RF and XG Boost model.

As suggested by MLR model, independent variables including group size, gap type, X_NB and vehicle speed were found to be negatively related to the accepted gap size of the pedestrians. On the other hand, rolling gap, safety distance and vehicle yielding were found to be positively related to the vehicular gap size accepted by the pedestrians while crossing the road. The model fitness of MLR model, RF model and XG Boost model considering significant or important variables only in terms of R-squared value was found to be 89.30%,99.66% and 99.92% respectively. Similarly, the performance of the trained model was computed in terms of root mean squared error whose value was 0.3358, 0.1114 and 0.0914 for MLR, RF and XG Boost model respectively .

Out of seven important variables of RF model, four variables viz. group size, safety distance, X_NB and vehicle speed were also found significant in multiple linear regression model. Four important variables of XG Boost model viz. rolling gap, safety distance, vehicle yield and vehicle speed were also found significant in multiple linear regression model. Similarly, four important variables viz. waiting time, safety distance, vehicle type and vehicle speed, of XG Boost model were also in the list of seven most important variable of RF model. Two variables viz. safety distance and vehicle speed were found to be important in all three models. Both machine learning model performed better than statistical model in predicting the size of the accepted gap. XG Boost outperformed rest of the models in model fitness and gap size prediction of with R-squared value of 99.92% and least root mean squared error value of 0.0914 respectively.

The recommendations for further studies are as follows:

- Similar study can be carried out considering multiple road sections with different number of lanes.

- Similar study can be carried out at overhead bridge locations to evaluate the existing crosswalk facility.

- It is recommended to use machine learning approaches for modeling the complex relationships between dependent and predictor variables over traditional statistical models.

Acknowledgement

The authors acknowledge Roads Board Nepal under Government of Nepal and Center for Infrastructure Development and Studies (CIDS) under Institute of Engineering for the financial and other assistance provided for conducting this research.

References

[1] Highway capacity manual (2010).

[2] Y. I. Demiroz, P. Onelcin, and Y. Alver. Illegal road crossing behavior of pedestrians at overpass locations: Factors affecting gap acceptance, crossing times and overpass use. 2015.

[3] A. Soathong, S. Chowdhury snf D. Wilson, and P. Ranjitkar. Investigating the motivation for pedestrians' risky crossing behaviour at urban mid-block road sections. 2021.

[4] World Health Organization Technical report. Global status report on road safety. 2018.

[5] Nepal Police Traffic Directorate. Annual accidental description, 2020 (2076/2077). 2020.

[6] A. Banstola, J. Kigozi, P. Barton, and J. Mytton. Economic burden of road traffic injuries in nepal. 2020.

[7] K. Shaaban, D. Muley, and A. Mohammed. Modeling pedestrian gap acceptance behavior at a six-lane urban road. 2019.

[8] A. Chaudhari, J. Shah, S. Arkatkar, G. Joshi, and M. Parida. Evaluation of pedestrian safety margin at mid-block crosswalks in india. 2019.

[9] W. Saleh, M. Grigorova, and S. Elattar. Pedestrian road crossing at uncontrolled mid-block locations:does the refuge island increase risk? 2020.

[10] B. R. Kadali, P. Vedagiri, and N. Rathi. Models for pedestrian gap acceptance behaviour analysis at unprotected mid-block crosswalks under mixed traffic conditions. 2015.

[11] G. Yannis, E. Papadimitriou, and A. Theofilatos. Pedestrian gap acceptance for mid-block street crossing. 2013.

[12] M. M. Nassr, A. Zulkiple, W. A. Albargi, and N. A. Khalifa. Modeling pedestrian gap crossing index under mixed traffic condition. 2017.

[13] M. S. Serag. Modelling pedestrian road crossing at uncontrolled mid-block locations in developing countries. 2014.

[14] W. A. Al Bargi, B. D. Daniel, J. Prasetijo, M. M. Rohani, and S. N. Mohamad Nor. Crossing behaviour of pedestrians along urban streets in malaysia. 2017.

[15] W. A. Al Bargi, B. D. Daniel, and M. M. Muftah. Mid-block crossing behavior: A study of pedestrians and vehicles interaction along urban streets in malaysia. 2017.

[16] S. N. Mohamad Nor, B. D. Daniel, R. Hamidun, W. A. Al Bargi, M. M. Rohani, J. Prasetijo, M. Y. Aman, and A. Ambak. Analysis of pedestrian gap acceptance and crossing decision in kuala lumpur. 2017.

[17] I. Omran m.f alajnaf, K. Mohammed, A.Emhamed, and M. M. Almadani. Pedestrian gap acceptance and crossing decision outside crossing facilities along urban streets in malaysia: A case study of rughaya street, batu pahat, johor, malaysia. 2016.

[18] S. Elyassami, Y. Hamid, and T. Habuza. Road crashes analysis and prediction using gradient boosted and random forest trees. 2020.

[19] R. Cheng, M. Zhang, and X. Yu. Prediction model for road traffic accident based on random forest. 2019.

[20] H. Chen, Z. Liu, X. Sun, and R. Zhou. Analysis of factors affecting the severity of automated vehicle crashes using xgboost model combining poi data. 2020.

[21] K. D. Kankanamge, Y. R. Witharanage, C. S. Withanage, and M. Hansini. Taxi trip travel time prediction with isolated xgboost regression. 2019.