Australasian Transport Research Forum 2017 Proceedings
27 – 29 November 2017, Auckland, New Zealand
Publication website: http://www.atrf.info

Using crash severity to identify safe walking distance in vicinity of schools

Alireza Toran Pour[1], Sara Moridpour[2], Richard Tay[3], Abbas Rajabifard[4]

[1, 2] School of Engineering, RMIT University, Australia

[3]School of Business, IT and Logistics, RMIT University, Australia

[4]Department of Infrastructure Engineering, The University of Melbourne, Australia

Email for correspondence: alireza.toranpour@rmit.edu.au

Abstract

Walking to and from school can help children adopt physical activity habits for life. It can also assist in reducing traffic congestion, parking difficulties, and the associated environmental impacts. However, the factors influencing the crash severity of this group of pedestrians need to be analysed to decrease the risk, and severity of vehicle-pedestrian crashes. Pedestrians with less than 18 years of age or school-aged pedestrians are involved in about 18 per-cents of vehicle-pedestrian crashes in Melbourne metropolitan area. Providing a safe walking environment for school-aged pedestrians could encourage children and their parents to walk between home and school. This paper aims at identifying the vehicle-pedestrian crash severity contributing factors and introducing safe walking distance to schools. These findings could assist in applying effective countermeasures to reduce the vehicle-pedestrian crash severity and improve the safety of this group of pedestrians. In this research, Boosted Decision Tree (BDT) model has been applied to identify significant factors on vehicle-pedestrian crash severity. Results from BDT model show that the distance between home and school is the most significant factor in vehicle-pedestrian crash severity of school-aged pedestrians. In addition, applying Patrial Dependence Plot (PDP) identifies that the probability of crashes being fatal in 500 meters vicinity of schools is less than other shorter distances. Kernel Density Estimation (KDE) analysis in GIS shows that only 3.4% of crashes within the 500m walking distance from schools are in high and extreme crash severity. These results could assist in identifying safe walking distance for school-aged pedestrians and applying necessary safety programs.

1. Introduction

Physical activities during childhood play an important role in promoting physical, social and psychological health in the short and long terms. Similar to adults, children who are physically inactive are more likely to be overweight or obese and have poorer mental health and overall wellbeing (Biddle et al., 2004). It is recommended that children spend at least 60 minutes per day in moderate-to-vigorous activities such as organised sports (e.g. tennis or netball), active free-play (e.g. playing in the backyard, playing with the family dog) and active travel (e.g. walking and cycling to/from school) (Salmon and Shilton, 2004). Studies show that compared to children who are driven to school, children who walk or cycle tend to have higher levels of overall physical activity (Cooper et al., 2003).

Safety is an influencing factor on pedestrian walking and active commuting (Davison and Lawson, 2006). Thus, policies that address issues related to pedestrian safety may have a significant impact on active commuting patterns among young people. To address this issue and reduce the involvement of school children in crashes, we need to identify contributing factors on vehicle-pedestrian crashes for this age group and enhance safety in the school zones.

In recent years, many studies have been conducted to identify the contributing factors on vehicle-pedestrian crash frequency and severity (Tarawneh, 2001, Sze and Wong, 2007, Tay et al., 2011, Siddiqui et al., 2012, Tulu et al., 2015, Li et al., 2016, Toran Pour et al., 2016, Verzosa and Miles, 2016, Rifaat et al., 2017, Toran Pour et al., 2017b). For instance, Tay et al. (2011) identified that pedestrians' and drivers' age, and driving speed could influence vehicle-pedestrian crash severity in South Korea. Furthermore, in another study, Toran pour et al. (2017b) showed that neighbourhood social characteristics were as important as traffic and infrastructure variables in influencing the severity of pedestrian crashes.

However, there are relatively few studies that focused on school-aged vehicle-pedestrian crashes. Graham and Glaister (2003) found that the probability of child pedestrian casualty is higher in more deprived areas. In another study, Noland and Quddus (2004) found that more severe pedestrian injuries are associated with areas with lower income, higher percent of local roads, higher per capita expenditure on alcohol, and larger numbers of people. Abdel-Aty et al. (2007) developed a GIS base crash analysis and log-linear model for pedestrians under 19 years of age in Florida. In this research, they showed that majority of school-aged children crashes occurred in the areas near schools. Furthermore, age of drivers and pedestrians, road geometry, speed limit, and speed ratio were also found to be correlated with the frequency of crashes. Also, in another study, it is identified that child pedestrian crashes are strongly associated with built environment features (Rothman et al., 2014).

Koopmans et al. (2015) investigated the vehicle-pedestrian injury crashes for pedestrians under 19 years of age in Chicago and found that environmental conditions such as weather condition, light, and location of crashes are contributing factors to crash injury severity of pedestrians. Lee et al. (2016) applied standard negative binomial and zero-inflated negative binomial models to identify the influencing environmental attributes of intersections on crashes involving children aged 10 to12 years of age near elementary schools in South Korea. They found that a higher number of student crossings, a wider road width, the presence of crosswalks, student-friendly facilities at the intersection, and four-leg intersections were significant and positively associated with perceived crash risk among school-aged children. Furthermore, Toran pour et al. (2017a) showed that the distribution of vehicle-pedestrian crashes for school-aged pedestrians is related to spatial distribution of schools.

Literature review shows that there are relatively few studies with emphasis on more refined spatial distribution of school-aged crashes, particularly in the areas surrounding schools. Furthermore, the existing studies mainly applied linear buffer around schools to identify crashes. Finally, safe walking distance is not considered in the existing studies.

The objectives of this study are:

(a) Identifying the contributing factors on vehicle-pedestrian crashes of school-aged pedestrians.

(b) Exploring the influence of distance from schools on school-aged pedestrian crash severity, and the safe walking distance from schools.

In this study, the Boosted Decision Tree (BDT) is developed to identify the influencing factors on this crash type. Partial Dependence Plot (PDP) in BDT is used in the model to assists in identifying the influence of distance on crash severity level. These results are then used in GIS to show safe walking distance. Finally, Kernel Density Estimation (KDE) is used to determine the most appropriate safe walking distance.

This paper is structured as follows. The next section of the paper presents the dataset and methodology of this research. The results are presented and discussed in Section 3. Finally, the outcomes are summarised in Section 4.

2. Dataset and methodology

2.1. Dataset

Data on pedestrian crashes are extracted from the Road Crash Information System (RCIS). RCIS is an online database providing crash data from Victorian road incidents dating back to 1987 and includes data on personal characteristics (e.g. driver/pedestrian age, gender), vehicle characteristics (e.g. vehicle type, weight), road and environment conditions (e.g. road surface, light and pavement conditions), and temporal parameters (e.g. date, day and time of the crash). In Victoria, only crashes resulting in injury of at least one road user involved in the crash are required to be reported to the police.

In Victoria, the severity of a crash is determined by the person with the most severe injury. A fatal crash refers to a crash in which at least one person dies within 30 days of a collision, while a serious injury crash refers to a crash in which at least one person is sent to hospital (VicRoads, 2016). It is noted that this classification is different from other schemes that use actual injury scale such as the Abbreviated Injury Scale (AIS) and might be an overestimation of the crash severity level because some of the road users who are sent to the hospital may only suffer minor injuries.

To investigate the variables influencing school aged vehicle-pedestrian crash severity, data for these crashes on public roadways of Melbourne metropolitan area from 2004 to 2013 are extracted from RCIS. From the 11,548 vehicle-pedestrian crashes, 2,161 are related to pedestrians less than 19 years of age. According to VicRoads severity classification, 0.6% of the school-aged vehicle-pedestrian crashes were fatal crashes, 37.5% were serious injury crashes, and 61.9% were minor injury crashes.

In addition to the crash data, data on the neighbourhood social and economic characteristics are extracted from the Australian Bureau of Statistics (ABS, 2013). ArcMap GIS 10.3 is used to extract the social and economic variables related to each suburb where the corresponding vehicle-pedestrian collision occurred. ArcMap GIS 10.3 is also used to extract the traffic volume data from the Melbourne road network database for each crash location. In addition, to identify the distance of crashes from schools, 1274 schools including primary, secondary, language and special schools which are located in Melbourne metropolitan area are mapped in GIS. Tables 1 and 2 show a summary of the continuous and categorical variables used in this study, respectively.

Table 1: Descriptive statistics for continuous variables applied in BDT model.

Variable	Unit	Mean	Std. Deviation
Traffic Volume	Vehicle per day	11723.5	8995.8
Distance to school	Meters	594.6	467.1
Population Density	Person/Sq. kilometre	2145.0	1575.5
Median income	AUS Dollars	608.9	200.8
Use Public Transport	Percent	10.9	8.1
Use other type of transport	Percent	3.6	3.5
Use Private	Percent	61.9	15.4
Use Walk	Percent	4.1	7.8
Use Multimodal	Percent	5.6	5.9
Indigenous	Percent	0.4	0.3
Born in Australia	Percent	59.4	14.6
Born in other country	Percent	39.0	54.3

Table 2: Categorical explanatory variables applied in BDT model.

Variables		%	Variables		%
Severity	Fatal Crash	0.6	**Atmosphere Condition**	Clear	90.4
	Serious Injury Crash	37.5		Rainy	5.8
	Minor Injury Crash	61.9		Other	3.8
Day of Crash	Monday	18.3	**Pedestrian Gender**	Male	54.2
	Tuesday	23.1		Female	44.7
	Wednesday	18.4	**Driver Gender**	Male	47.5
	Thursday	19.7		Female	43.8
	Friday	20.5	**Month of Crash**	January	4.8
Time of Crash	Firs Peak (7:00-9:00 am)	32.7		February	8.8
	Off-Peak (10:00 am – 3:00 pm)	67.3		March	10.4
Light Condition	Day	98.1		April	7.9
	Dusk/Dawn	1.5		May	11.4
	Other	0.4		June	7.7
Node Type	Intersection	49.5		July	7.9
	Mid-Blocks	49.8		August	8.6
	Other	0.7		September	9.8
Surface Condition	Dry	87.6		October	8.2
	Wet	9.1		November	7.2
	Other	3.3		December	7.4

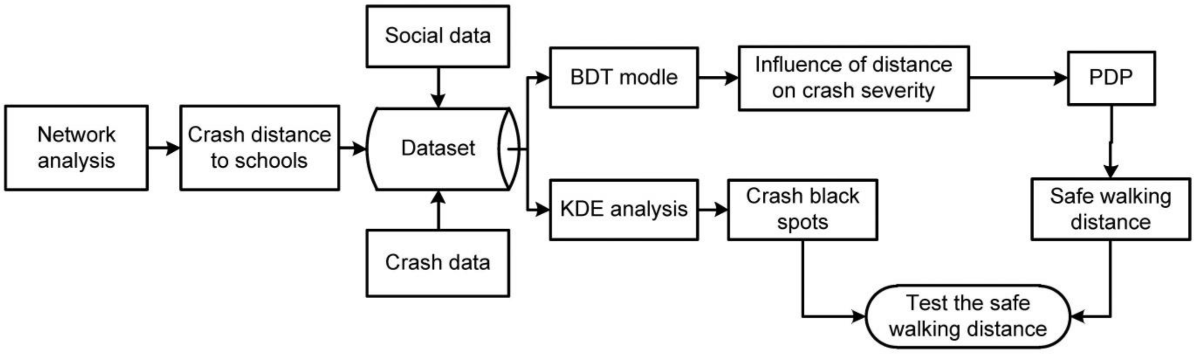
2.2. Methods

As mentioned before, the main objectives of this research include identifying the contributing factors on school-aged vehicle-pedestrian crash severity and exploring the influence of distance of crashes to schools on vehicle-pedestrian crash severity. In this research, network analysis in ArcMap 10.3 is used to identify the distance between crash locations and schools. Studies show that the majority of children live close to school but only a small minority of students walk the distance (Spallek et al., 2006, McDonald, 2008).Therefore, in this research it is assumed that the pedestrians who involved in crashes were attending to closest schools. Then, Boosted Decision Tree model is applied to identify the contributing factors including the distance to school on school-aged vehicle-pedestrian crash severity. The main advantage of using Decision Trees (DTs) approach is that there is no assumption to develop the model and, also finding correlation between the dependent and independent variables is not important in this technique. Furthermore, results from DT models are easy to interpret and justify. However, instability of DT models is known as the most important disadvantage of this model type.

To overcome this limitation, ensemble approaches such as BDT models are developed to reduce this problem. However, it can be difficult to understand the functional relations between predictors and an outcome when using ensemble approaches. One way to investigate these relations is applying Partial Dependence Plots (PDPs). These plots are graphical visualizations of the marginal effect of a given variable (or multiple variables) on an outcome. Therefore, PDP is applied to identify how the influencing factors contribute on severity level and find safe walking distance from schools.

Finally, KDE is applied for all school-aged vehicle-pedestrian crashes and sever crashes to identify crash black spots. According to KDE results, crash blackspots are divided into 4 risk areas (levels) including low, moderate, high and extreme risk area. Then, the number of school-aged vehicle-pedestrian crashes that occurred in different risk area (level) counted to test the validity of safe walking distance. Figure 1, shows the methodology of this research.

Figure 1: The methodology of this research.



2.3. BDT model

Non-parametric techniques such as Decision Tree (DT) and support vector machine have been used in different traffic crash severity analysis (Chang and Wang, 2006, Li et al., 2008, Kashani and Mohaymany, 2011, Abellán et al., 2013, Chang and Chien, 2013, Jung et al., 2016). The DT approach is a simple and powerful method to solve classification problems and provides a graphical structure using a tree with many branches and leaves. These graphical features are useful in understanding and interpreting the results (Kashani and Mohaymany, 2011). In DT models, the root node on top of the tree which contains all objects is split into two homogeneous sets that are called child nodes. Then, DT splits child nodes until no further split can be made (i.e. all child nodes are homogenous or a user-defined minimum number of objects in the node is reached). These final nodes are called terminal nodes or leaves and they have no branches.

One disadvantage of the DT models is that it is often unstable and the results may change significantly with changes in the training and testing data (Lord et al., 2007). Therefore, ensemble models, such as Bagged Decision Tree (BgDT), Boosted Decision Tree (BDT) and Random Forest (RF), are applied in some studies to improve the reliability and accuracy of DT models (Pham et al., 2010, Chung, 2013, Xu et al., 2014, Lee and Li, 2015, Jiang et al., 2016, Toran Pour et al., 2017b). Since BDT has better performance in vehicle-pedestrian crash severity analysis than DT and BgDT models (Toran Pour et al., 2017b), BDT model is applied in this research to explore the factors contributing to vehicle-pedestrian crash severity.

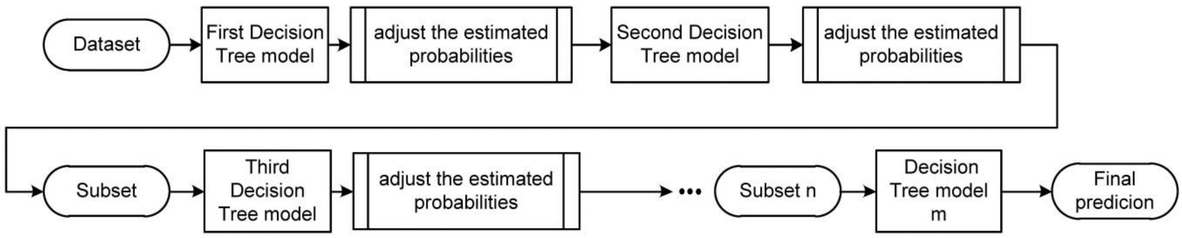
BDT combines regression trees and boosting technique to improve the performance of DT models. Boosting is a forward and stage-wise procedure in which a subset of data is randomly selected to iteratively produce new trees and improve the quality of prediction (Elith et al., 2008). This study uses a stochastic gradient boosting procedure that can improve model performance and reduce the risk of over-fitting (Friedman, 2001). In stochastic gradient boosting procedure, at each iteration, a subsample of the training data is drawn randomly (without replacement) from the full training dataset. To fit the base learner, the randomly selected subsample is then used, instead of the full sample. Figure 2 shows the BDT procedure flowchart.

In this approach, the residuals will be calculated after the first tree is fitted. In BDT, the difference between the target function (function in which the learning problem attempts to approximate) and the current predictions of an ensemble is called the residual. At the next stage, observations with high residual values are defined as observations with high prediction error. Then, BDT calculates the adjustment weights using Equation 1:

$$w(i) = \frac{(1+m(i)^4)}{\sum_{i=1}^{n}(1+m(i)^4)} \quad (1)$$

Where $0 \leq m(i) \leq n$ and n is the number of fitted decision tree models and m is the number of models that misclassifies case i in the previous step. The weights are then used to adjust the estimated probabilities and minimise the misclassification error rate. Hence, subsequent trees are fitted to the residual of the previous tree (Matignon, 2007). This procedure is repeated n times and on m models to adjust the estimated probabilities.

Figure 2: flowchart for the boosting DT method.



In this research, fitted BDT models are obtained using the “gbm” library (Ridgeway, 2007) in the R software (Team, 2014), “caret” package (Kuhn, 2008). To develop the BDT model, the repeated *k*-fold cross-validation technique is applied. The dataset is randomly divided into *k* blocks of roughly equal size instead of dividing the data into training and testing sub-sets. At each iteration, one block is left out and the other *k-1* blocks are used to train the model. Each k block is left out once and the left out block is used for prediction. These predictions are summarized in a performance measure (e.g. accuracy). This procedure is repeated *s* times to decrease the error and find the most robust model. The (*s x k)* estimates of performance are then averaged to obtain the overall re-sampled estimate.

To identify the contributing factors of crash severity, PDP is applied. PDP depicts the relationship between the response and one predictor variable, while accounting for the average effects of all other predictors (Friedman, 2001, Friedman and Meulman, 2003).

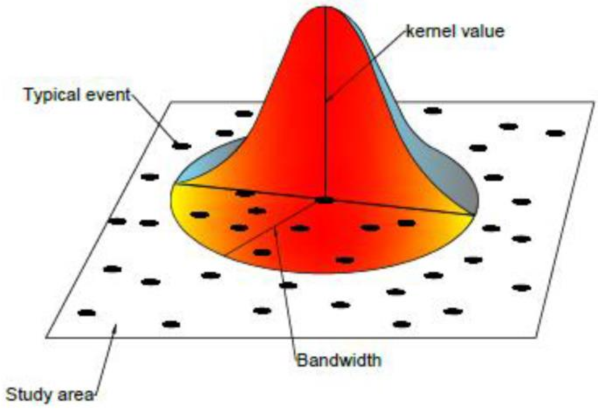
2.4. KDE analysis

In this research, KDE is applied to identify vehicle-pedestrian crash risk areas (black spots) around schools and the percent of crashes inside the crash black spots at various distances from the schools are computed to determine the most appropriate safe distance around schools.

KDE involves placing a symmetrical surface over each point, evaluating the distance from that point to a reference location (e.g. ???), and summing the values for all the surfaces for that reference location. This procedure is repeated for successive points. This provides the opportunity to place a kernel over each observation, and provides the density estimate of the distribution of collision points. The surface value is maximum at the reference point and decreases with an increase in distance from the reference point and reaches zero at the radius distance from reference point (Pulugurtha et al., 2007)(see Figure 3). One common mathematical function used for KDE is:

$$f(x,y) = \frac{1}{nh^2}\sum_{i=1}^{n} K\left(\frac{d_i}{h}\right) \quad (2)$$

Figure 3: Diagram of Kernel function.



Where $f(x,y)$ is the density estimation at location *(x,y)*, *n* is the number of observations, *h* is the bandwidth or kernel size or smooth parameter, *K* is the kernel function, and d_i is the distance between the location *(x,y)* and the location of i^{th} observation. In the simple density method, a circular neighbourhood is considered around each cell. However, in the kernel method the research area is divided into predetermined number of cells. Thus, the kernel method draws a circular neighbourhood around each feature point (here each vehicle-pedestrian crash). There are different types of kernel functions, such as Gaussian, Quartic, Conic, negative exponential, and epanichnekov (Levine and CrimeStat, 2002, Kuter et al., 2011). In this research, the Quartic kernel which is one of the three most common types of kernel functions, is applied (Schabenberger and Gotway, 2004, Xie and Yan, 2008). The specific form of the Quartic kernel function is:

$$K\left(\frac{d_i}{h}\right) = K\left(1-\frac{d_i^2}{h^2}\right) \qquad \text{When } 0 < d_i \le h \quad (3a)$$

$$K\left(\frac{d_i}{h}\right) = 0 \qquad \text{When } d_i > h \quad (3b)$$

In Equations (3a) and (3b), *K* is the kernel function, and d_i is the distance between the location *(x,y)* and the location of i^{th} observation. In Equation (3b), *K* is applied to ensure the total volume under Quartic curve is 1. The common values used for *K* include $\frac{3}{\pi}$ and $\frac{3}{4}$.

According to the literature, the kernel function type in KDE, *K,* is less important than the bandwidth *h* in determining model accuracy (O'Sullivan and Unwin, 2014, Loo et al., 2011). Many studies have shown that the selection of the bandwidth or smoothing parameter in KDE is subjective (Bíl et al., 2013, Plug et al., 2011). Different studies have selected different bandwidth values according to the area of study and size of the dataset. In general, according to Equation (3), selecting large bandwidth values $(h \to \infty)$ will decrease the density $(f(x,y) \to 0)$ and will show significant smoothing and low-density values (over smooth). In contrast, a small bandwidth value will result in less smoothing (under smooth), producing a map that depicts local variations in point densities (Chainey et al., 2002).

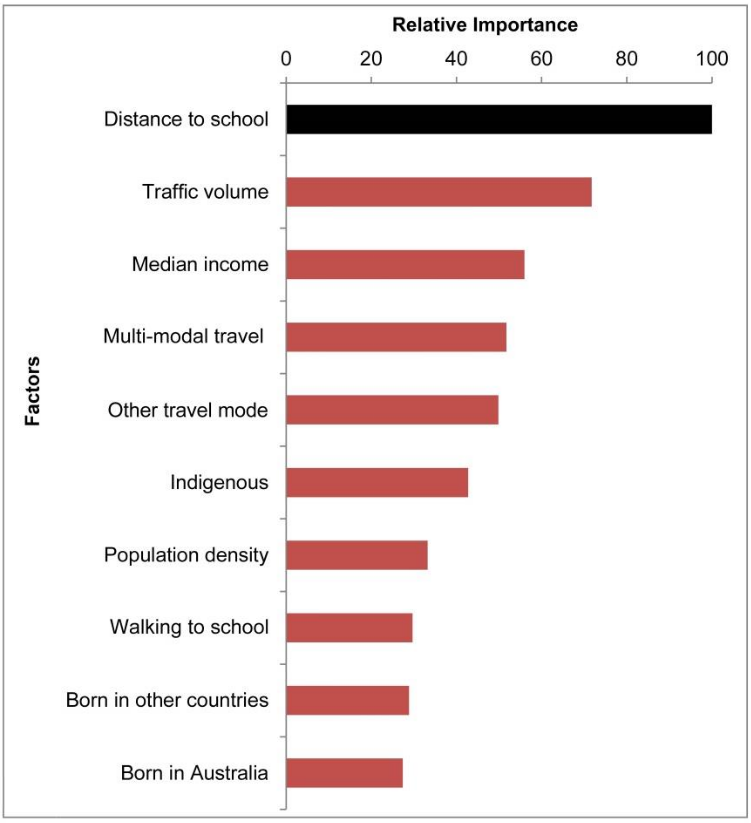
3. Results and discussions

3.1. Model performance and contributing factors

Accuracy (ACC) of model is the most widely used performance measure in machine learning methods. ACC is defined as the proportion of correct predictions made by the classifier relative to the size of the dataset. The ACC for the model that is developed in this research is 79% which shows an accurate performance in term of accuracy.

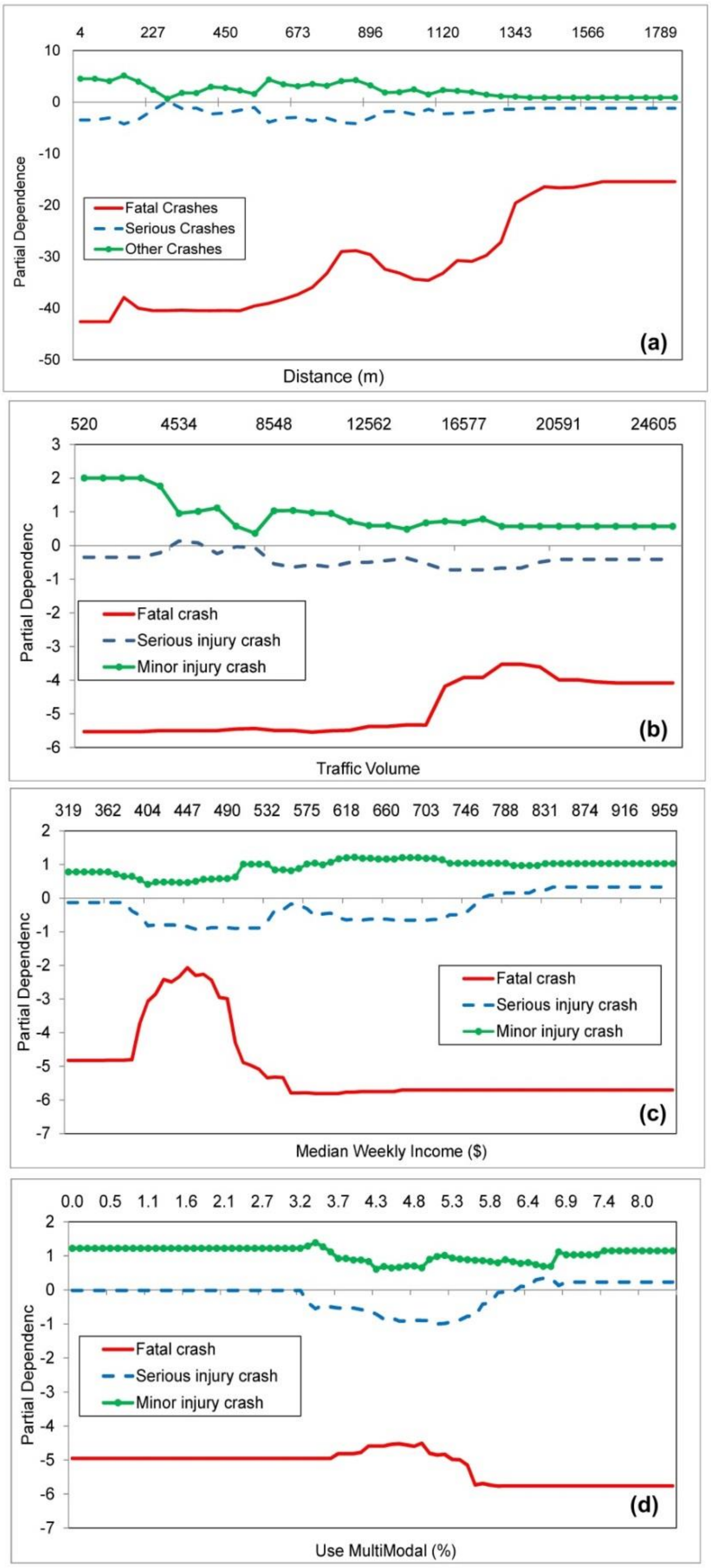
Figure 4 shows the results of BDT model and the top 10 most important predictor variables for school-aged vehicle-pedestrian crash severity. Also, Figure 5 shows the PDP for first top 4 significant factors. As shown in Figure 4, 'Distance to School' is the most important contributing factor to the severity of school-aged vehicle-pedestrian crashes, showing that this variable is a significant contributing variable to vehicle-pedestrian crash severity of this age group. The results from this study show that distance of crashes to/from schools needs to be considered in vehicle-pedestrian crash severity studies around schools and for school-aged pedestrians.

Figure 4: Top 10 relative importance of predictor variables for school-aged vehicle-pedestrian crashes in BDT model.



Traffic volume is the next contributing factor on the vehicle-pedestrian crash severity of this age group. This result is consistent with the results from other studies that found traffic volume could associate with vehicle-pedestrian crashes (Assailly, 1997, Zegeer et al., 2001, Morency et al., 2012, Toran Pour et al., 2017b). Figure 5 (b) illustrates that increase in traffic volume from about 15,000 to 19,000 vehicles per day in roads around schools could increase the risk of fatal crashes. According to this figure, the risk of fatal crash decreased and then remained stable after 21,000 vehicles per day. The results of the present research are consistent with the results of other studies showing that increase in traffic volume can increase pedestrian crash frequency and the probability of pedestrian crash severity (Zegeer et al., 2001, Miranda-Moreno et al., 2011, Pulugurtha and Sambhara, 2011, Morency et al., 2012, Toran Pour et al., 2017b). These results suggest that transport engineers and planners may need to target roads with more than 15,000 vehicles per day to improve the safety of these vulnerable road users. More pedestrian crossings, crossing supervision, pedestrian signals and flashing lights on these roads may assist in improving the safety of school-aged pedestrians.

Figure 5: PDP for first 4 top contributing factors on vehicle-pedestrian crashes severity of school-aged pedestrian.



According to Figure 4, median income of neighbourhoods in which crash occurred is another contributing factor on school-aged vehicle-pedestrian crash severity. This result is similar to the other research that found income could influence the vehicle-pedestrian crashes around schools (LaScala et al., 2004, Noland and Quddus, 2004). Furthermore, Figure 5(c) shows that the risk of fatal crashes that occurs in suburbs with median weekly income between $370.00 and $450.00 (low income) could be more than other suburbs. This result is consistent with the results from other studies that showed the probability of vehicle-pedestrian crashes to be fatal is more in low-income suburbs (Noland and Quddus, 2004, Zhu and Lee, 2008, Siddiqui et al., 2012). As highlighted in Figure 5(c), our results suggested that suburbs with low median income could be targeted for school-aged pedestrian safety educational programs or campaigns. These programs could increase the traffic safety knowledge, especially safe walking knowledge, and improve pedestrian safety for children living in these targeted suburbs.

Furthermore, Figure 4 and Figure 5(c) show that type of travel mode to/from school is a contributing factor to school-aged vehicle-pedestrian crash severity. School-home travel mode could influence walking distance and exposure of school-aged pedestrians to vehicular traffic. Finally, Figure 4 shows that other social characteristics at the location of crashes including ethnicity, place of birth and population density are other factors that influence school-aged vehicle-pedestrian crash severity.

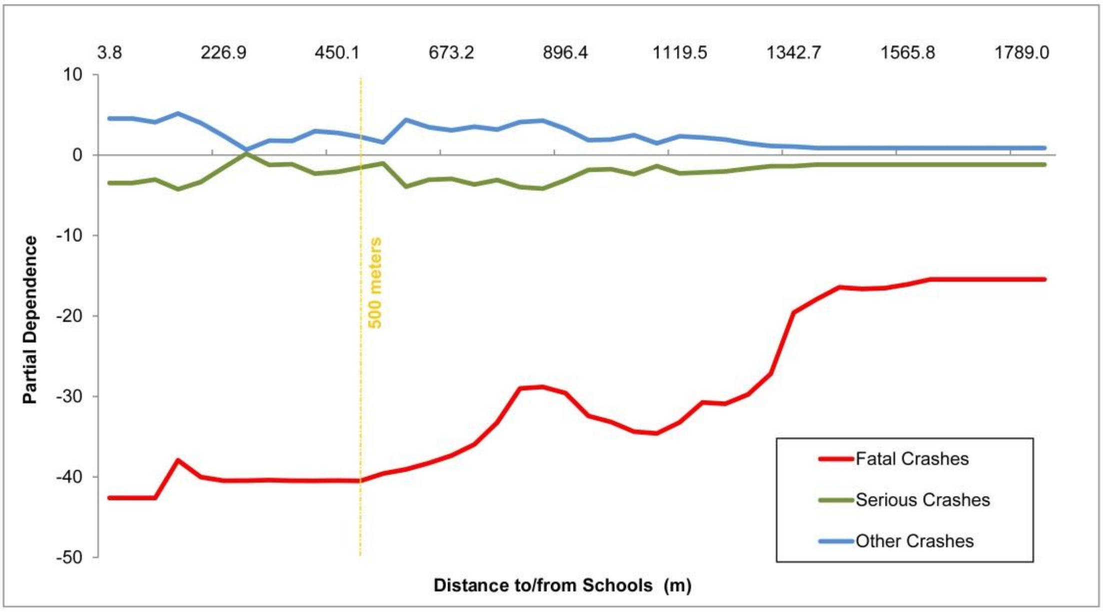
3.2. Safe walking distance

As mentioned, crash distance to school is the most significant factor in school-aged vehicle-pedestrian crash severity. Figure 6 shows the PDP for distance of crashes to/from schools for different levels of vehicle-pedestrian crash severity. This figure identifies the influence of this factor on vehicle-pedestrian crash severity levels. Increasing the distance of crash from/to schools beyond about 500 meters will significantly increase the probability of fatal crashes. Therefore, a distance of less than 500 meters around schools can be defined as a safe walk-to-school distance for school-aged pedestrians.

From both parents and children points of view, traffic safety is an important factor when considering walking to school (Dellinger et al., 2002, Timperio et al., 2004, Rossen et al., 2011, Collins and Kearns, 2001). As found by this research, 500 meters is a safe walk-to-school distance for school-aged pedestrians. However, this distance is much less than what has been found by previous research as a reasonable walking distance. A reasonable walking distance for parents and child is between 800 meters and 1.5 kilometres (Timperio et al., 2004, Bejleri et al., 2011, D'Haese et al., 2011). Therefore, a specific safety program is required to increase the safe walking distance t encourage parents and children to walk to school.

Therefore, results from this research can assist road safety professionals in identifying road segments to apply road safety strategies and plans for school-aged pedestrians including speed calming and warning systems. Improving safety around schools after these 500 meters can increase the safe walking distance and promote the children’s health factors.

Figure 6: Influence of crash distance from/to schools on school-aged vehicle-pedestrian crash severity.



To evaluate the safe distance, KDE analysis is used to identify vehicle-pedestrian crash risk areas for school-aged vehicle-pedestrian crashes. Figure 7 shows the results of KDE analysis for this crash type. Moreover, to identify crash black spots for severe crashes, data on fatal and serious crashes are extracted and used in another KDE analysis. Figure 8 shows the result of KDE analysis for severe school-aged vehicle-pedestrian crashes. In this figure, crash black spots are divided into four areas based on their KDE value. In this figure, areas with low and moderate crash risk are illustrated with white and green shades. While, areas with high density of vehicle-pedestrian crash risk are illustrated with orange and red shades.

Table 3 shows the percentage of crashes in 500 meters safe distances in different risk areas. Table 3 reveals that only 1.6% of crashes in 500 meters safe distance are in high and extreme crash risk area. Furthermore, Figure 8 and Table 3 show that only 26.6% of crashes in school-aged safe distance are in severe crash black spots. Also, these results show that 3.4% of vehicle-pedestrian crashes for this age group in 500 meters safe distance are in high and extreme severe crash risk area (see table 3).

These results indicate that the risk of fatal and severe crashes in 500 meters safe distance from schools is less than other areas. Therefore, it is possible to encourage people to walk more in this distance. Furthermore, these results can be used by traffic engineers and safety planners to find the target distances for safety projects and improve safety of pedestrian at roads in vicinity of schools. Improving safety of these areas could improve the walking and physical activity of school-aged pedestrians.

Figure 7: Crash blackspots and crashes in 500 meters safe distance around schools.

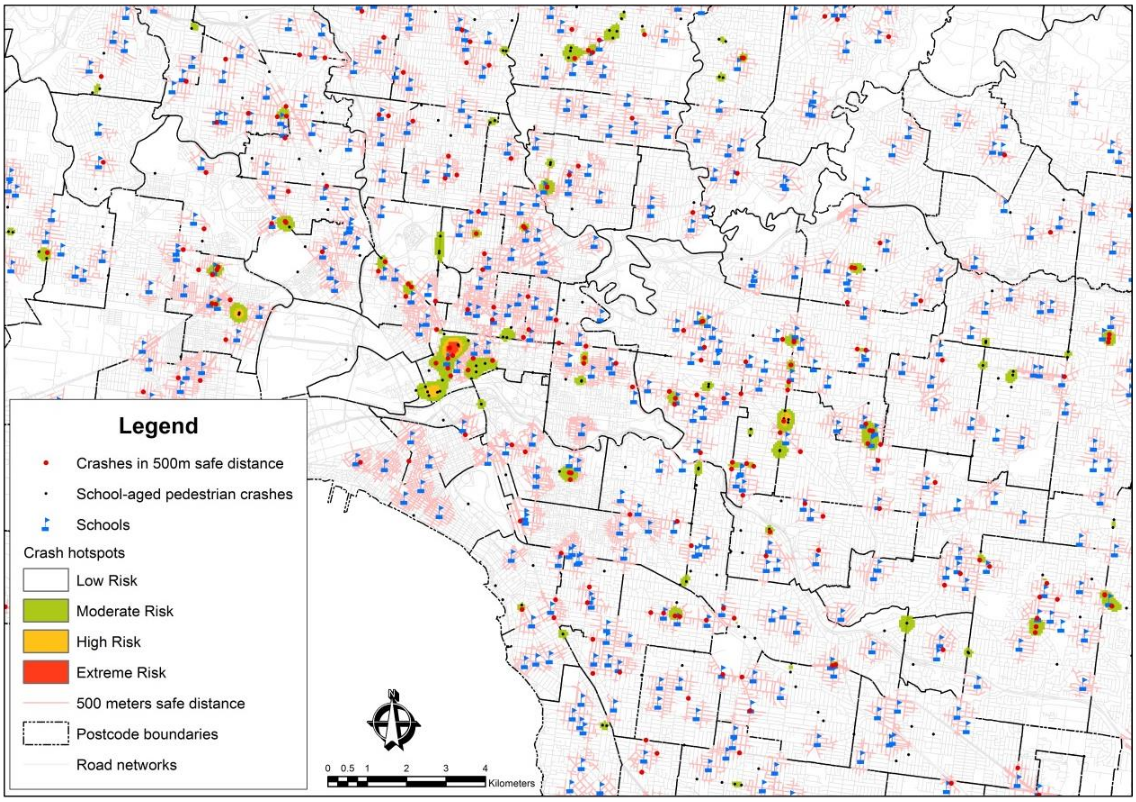


Figure 8: Severe crash blackspots and crashes in 500 meters safe distance around schools.

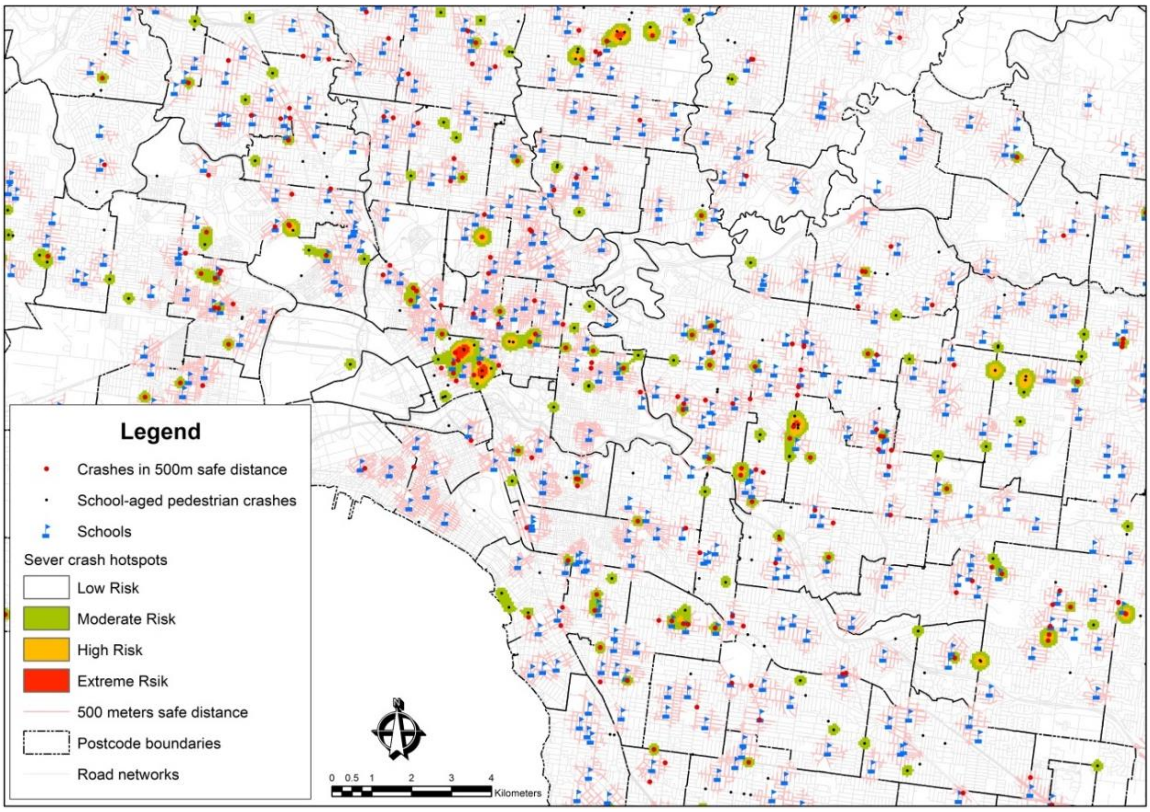


Table 3: Crash percentage in different risk areas.

KDE Risk level	KDE for Total Crashes	KDE for Severe Crashes
Low	31.2	5.8
Moderate	14.3	17.4
High	1.2	3.1
Extreme	0.4	0.3
Total	47.2	26.6

In summary, results of BDT model identified that crash distance from schools is the most important influencing factor on the severity of school-aged vehicle-pedestrian crashes. Also, PDP showed that increasing the crash distance from 500 meters from schools could increases the risk of vehicle-pedestrian fatal crashes for this age group. Results of KDE for all vehicle-pedestrian crash severity levels and severe crashes for school-aged pedestrian identified that the probability of severe crashes within the 500 meters distance from schools is less than other areas. In term of crash severity, this distance could be introduced as safe walking distance for school-aged pedestrians. Using distances less than 500 meters for walking distance might have no significant influence on vehicle-pedestrian crash severity (See figure 6). However, considering walking distances of more than 500 meters around schools can increase the risk of vehicle-pedestrian fatal crashes for school-aged pedestrians. To encourage children to walk to school and parents to let their children walk to school, it is important to improve the road safety in school vicinity, especially after 500 meters from schools.

4. Conclusion

Walking to school can improve the children’s physical health, educational development and their social activities. Road safety is known as a barrier in improving children’s walking to schools. Also, it is identified that perceived level of children's safety, particularly by parents, is a key issue influencing a child's opportunity to be more active as part of a daily routine. Therefore, identifying the contributing factors on school-aged vehicle-pedestrian crash severity could assist transportation engineers, road safety professionals and policy makers in developing and implementing effective countermeasures around schools to reduce the number of deaths and injuries of these vulnerable road users and improving safety in vicinity of schools.

In this research, the BDT model was applied to identify the contributing factors on the school-aged vehicle-pedestrian crash severity. This study found that distance of crash to schools is the most important variable in determining vehicle-pedestrian crash severity around schools. Moreover, this research revealed that public wellbeing indicators such as median income and using public transport have an influence on severity of crashes in this age group.

PDP was applied in this study to identify the influence of contributing factors on safety levels. Appling PDP for crash distance from schools showed that the probability of vehicle-pedestrian crashes to be fatal increased after 500 meters from/to schools. Therefore, distance of less than 500 meters was defined as safe walking distance for school-aged pedestrians. Crash black spot analysis for fatal and serious crashes identified that less than 30% of school-aged crashes in safe walking distance were in black spots and only 3.4% of these crashes were in high severe crash risk areas.

Results from this research would provide valuable information to assist road safety professionals in targeting the right schools and distances to implement different safety measures and improve the safety school-aged pedestrians. Improving the road safety in vicinity of schools and increasing the safe walking distance increase the tendency of children to use active transport to commute to school.

Reference

ABDEL-ATY, M., CHUNDI, S. S. & LEE, C. 2007. Geo-spatial and log-linear analysis of pedestrian and bicyclist crashes involving school-aged children. *Journal of Safety Research,* 38**,** 571-579.

ABELLÁN, J., LÓPEZ, G. & DE OÑA, J. 2013. Analysis of traffic accident severity using Decision Rules via Decision Trees. *Expert Systems with Applications,* 40**,** 6047-6054.

ABS. 2013. *Australia's population by country of birth* [Online]. Canberra, ACT: Australian Bureau of Statistics 2016].

ASSAILLY, J.-P. 1997. Characterization and prevention of child pedestrian accidents: An overview. *Journal of applied developmental psychology,* 18**,** 257-262.

BEJLERI, I., STEINER, R. L., FISCHMAN, A. & SCHMUCKER, J. M. 2011. Using GIS to analyze the role of barriers and facilitators to walking in children's travel to school. *URBAN DESIGN International,* 16**,** 51-62.

BIDDLE, S. J., GORELY, T. & STENSEL, D. J. 2004. Health-enhancing physical activity and sedentary behaviour in children and adolescents. *Journal of sports sciences,* 22**,** 679-701.

BÍL, M., ANDRÁŠIK, R. & JANOŠKA, Z. 2013. Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation. *Accident Analysis & Prevention,* 55**,** 265-273.

CHAINEY, S., REID, S. & STUART, N. 2002. *When is a hotspot a hotspot? A procedure for creating statistically robust hotspot maps of crime*, Taylor & Francis, London, England.

CHANG, L.-Y. & CHIEN, J.-T. 2013. Analysis of driver injury severity in truck-involved accidents using a non-parametric classification tree model. *Safety Science,* 51**,** 17-22.

CHANG, L.-Y. & WANG, H.-W. 2006. Analysis of traffic injury severity: An application of non-parametric classification tree techniques. *Accident Analysis & Prevention,* 38**,** 1019-1027.

CHUNG, Y.-S. 2013. Factor complexity of crash occurrence: An empirical demonstration using boosted regression trees. *Accident Analysis & Prevention,* 61**,** 107-118.

COLLINS, D. C. A. & KEARNS, R. A. 2001. The safe journeys of an enterprising school: negotiating landscapes of opportunity and risk. *Health & Place,* 7**,** 293-306.

COOPER, A. R., PAGE, A. S., FOSTER, L. J. & QAHWAJI, D. 2003. Commuting to school: are children who walk more physically active? *American journal of preventive medicine,* 25**,** 273-276.

D'HAESE, S., DE MEESTER, F., DE BOURDEAUDHUIJ, I., DEFORCHE, B. & CARDON, G. 2011. Criterion distances and environmental correlates of active commuting to school in children. *International Journal of Behavioral Nutrition and Physical Activity,* 8**,** 88.

DAVISON, K. K. & LAWSON, C. T. 2006. Do attributes in the physical environment influence children's physical activity? A review of the literature. *International journal of behavioral nutrition and physical activity,* 3**,** 19.

DELLINGER, A. M., CONTROL, C. F. D. & PREVENTION, U. 2002. Barriers to children walking and biking to school--United States, 1999. *MMWR: Morbidity and mortality weekly report,* 51**,** 701-704.

ELITH, J., LEATHWICK, J. R. & HASTIE, T. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology,* 77**,** 802-813.

FRIEDMAN, J. H. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics,* 29**,** 1189-1232.

FRIEDMAN, J. H. & MEULMAN, J. J. 2003. Multiple additive regression trees with application in epidemiology. *Statistics in medicine,* 22**,** 1365-1381.

GRAHAM, D. J. & GLAISTER, S. 2003. Spatial variation in road pedestrian casualties: the role of urban scale, density and land-use mix. *Urban Studies,* 40**,** 1591-1607.

JIANG, X., ABDEL-ATY, M., HU, J. & LEE, J. 2016. Investigating macro-level hotzone identification and variable importance using big data: A random forest models approach. *Neurocomputing,* 181**,** 53-63.

JUNG, S., QIN, X. & OH, C. 2016. Improving strategic policies for pedestrian safety enhancement using classification tree modeling. *Transportation Research Part A: Policy and Practice,* 85**,** 53-64.

KASHANI, A. T. & MOHAYMANY, A. S. 2011. Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models. *Safety Science,* 49**,** 1314-1320.

KOOPMANS, J. M., FRIEDMAN, L., KWON, S. & SHEEHAN, K. 2015. Urban crash-related child pedestrian injury incidence and characteristics associated with injury severity. *Accident Analysis & Prevention,* 77**,** 127-136.

KUHN, M. 2008. Caret package. *Journal of Statistical Software,* 28.

KUTER, S., USUL, N. & KUTER, N. 2011. Bandwidth determination for kernel density analysis of wildfire events at forest sub-district scale. *Ecological Modelling,* 222**,** 3033-3040.

LASCALA, E. A., GRUENEWALD, P. J. & JOHNSON, F. W. 2004. An ecological study of the locations of schools and child pedestrian injury collisions. *Accident Analysis & Prevention,* 36**,** 569-576.

LEE, C. & LI, X. 2015. Predicting Driver Injury Severity in Single-Vehicle and Two-Vehicle Crashes with Boosted Regression Trees. *Transportation Research Record: Journal of the Transportation Research Board***,** 138-148.

LEE, G., PARK, Y., KIM, J. & CHO, G.-H. 2016. Association between intersection characteristics and perceived crash risk among school-aged children. *Accident Analysis & Prevention,* 97**,** 111-121.

LEVINE, N. & CRIMESTAT, I. 2002. A spatial statistics program for the analysis of crime incident locations. *Ned Levine and Associates, Houston, TX, and the National Institute of Justice, Washington, DC*.

LI, D., RANJITKAR, P., ZHAO, Y., YI, H. & RASHIDI, S. 2016. Analyzing pedestrian crash injury severity under different weather conditions. *Traffic Injury Prevention***,** 00-00.

LI, X., LORD, D., ZHANG, Y. & XIE, Y. 2008. Predicting motor vehicle crashes using support vector machine models. *Accident Analysis & Prevention,* 40**,** 1611-1618.

LOO, B. P. Y., SHENJUN, Y. & JIANPING, W. Spatial point analysis of road crashes in Shanghai: A GIS-based network kernel density method. Geoinformatics, 2011 19th International Conference on, 24-26 June 2011 2011. 1-6.

LORD, D., VAN SCHALKWYK, I., CHRYSLER, S. & STAPLIN, L. 2007. A strategy to reduce older driver injuries at intersections using more accommodating roundabout design practices. *Accident Analysis & Prevention,* 39**,** 427-432.

MATIGNON, R. 2007. *Data Mining Using SAS Enterprise Miner,* South Sanfrancisco, CA, John Wiley & Sons.

MCDONALD, N. C. 2008. Children's mode choice for the school trip: the role of distance and school location in walking to school. *Transportation,* 35**,** 23-35.

MIRANDA-MORENO, L. F., MORENCY, P. & EL-GENEIDY, A. M. 2011. The link between built environment, pedestrian activity and pedestrian–vehicle collision occurrence at signalized intersections. *Accident Analysis & Prevention,* 43**,** 1624-1634.

MORENCY, P., GAUVIN, L., PLANTE, C., FOURNIER, M. & MORENCY, C. 2012. Neighborhood Social Inequalities in Road Traffic Injuries: The Influence of Traffic Volume and Road Design. *American Journal of Public Health,* 102**,** 1112-1119.

NOLAND, R. & QUDDUS, M. 2004. Analysis of pedestrian and bicycle casualties with regional panel data. *Transportation Research Record: Journal of the Transportation Research Board***,** 28-33.

O'SULLIVAN, D. & UNWIN, D. J. 2014. *Geographic information analysis,* Hoboken, New Jersey, John Wiley & Sons.

PHAM, M.-H., BHASKAR, A., CHUNG, E. & DUMONT, A.-G. Random forest models for identifying motorway rear-end crash risks using disaggregate data. Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on, 2010. IEEE, 468-473.

PLUG, C., XIA, J. C. & CAULFIELD, C. 2011. Spatial and temporal visualisation techniques for crash analysis. *Accident Analysis & Prevention,* 43**,** 1937-1946.

PULUGURTHA, S. S., KRISHNAKUMAR, V. K. & NAMBISAN, S. S. 2007. New methods to identify and rank high pedestrian crash zones: An illustration. *Accident Analysis & Prevention,* 39**,** 800-811.

PULUGURTHA, S. S. & SAMBHARA, V. R. 2011. Pedestrian crash estimation models for signalized intersections. *Accident Analysis & Prevention,* 43**,** 439-446.

RIDGEWAY, G. 2007. Generalized Boosted Models: A guide to the gbm package. *Update,* 1**,** 2007.

RIFAAT, S. M., TAY, R., RAIHAN, S. M., FAHIM, A. & TOUHIDDUZZAMAN, S. M. 2017. Vehicle-Pedestrian crashes at Intersections in Dhaka city. *The Open Transportation Journal,* 11.

ROSSEN, L. M., POLLACK, K. M., CURRIERO, F. C., SHIELDS, T. M., SMART, M. J., FURR-HOLDEN, C. D. M. & COOLEY-STRICKLAND, M. 2011. Neighborhood Incivilities, Perceived Neighborhood Safety, and Walking to School Among Urban-Dwelling Children. *Journal of Physical Activity and Health,* 8**,** 262-271.

ROTHMAN, L., MACARTHUR, C., TO, T., BULIUNG, R. & HOWARD, A. 2014. Motor Vehicle-Pedestrian Collisions and Walking to School: The Role of the Built Environment. *Pediatrics,* 133**,** 776.

SALMON, J. & SHILTON, T. 2004. Endorsement of physical activity recommendations for children and youth in Australia. *Journal of Science and Medicine in Sport,* 7**,** 405-406.

SCHABENBERGER, O. & GOTWAY, C. A. 2004. *Statistical methods for spatial data analysis,* Boca Raton, FL, CRC Press.

SIDDIQUI, C., ABDEL-ATY, M. & CHOI, K. 2012. Macroscopic spatial analysis of pedestrian and bicycle crashes. *Accident Analysis & Prevention,* 45**,** 382-391.

SPALLEK, M., TURNER, C., SPINKS, A., BAIN, C. & MCCLURE, R. 2006. Walking to school: distribution by age, sex and socio-economic status. *Health Promotion Journal of Australia,* 17**,** 134-138.

SZE, N. N. & WONG, S. C. 2007. Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes. *Accident Analysis & Prevention,* 39**,** 1267-1278.

TARAWNEH, M. S. 2001. Evaluation of pedestrian speed in Jordan with investigation of some contributing factors. *Journal of Safety Research,* 32**,** 229-236.

TAY, R., CHOI, J., KATTAN, L. & KHAN, A. 2011. A multinomial logit model of pedestrian–vehicle crash severity. *International journal of sustainable transportation,* 5**,** 233-249.

TEAM, R. D. C. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013. ISBN 3-900051-07-0.

TIMPERIO, A., CRAWFORD, D., TELFORD, A. & SALMON, J. 2004. Perceptions about the local neighborhood and walking and cycling among children. *Preventive Medicine,* 38**,** 39-47.

TORAN POUR, A., MORIDPOUR, S., TAY, R. & RAJABIFARD, A. A Partial Proportional Odds Model for Pedestrian Crashes at Mid-Blocks in Melbourne Metropolitan Area. MATEC Web of Conferences, 2016. EDP Sciences, 02020.

TORAN POUR, A., MORIDPOUR, S., TAY, R. & RAJABIFARD, A. 2017a. Influence of Pedestrians' Age and Gender types on Spatial and Temporal Distribution of Pedestrian Crashes. *Traffic Injury Prevention*.

TORAN POUR, A., MORIDPOUR, S., TAY, R. & RAJABIFARD, A. 2017b. Modelling pedestrian crash severity at mid-blocks. *Transportmetrica A: Transport Science,* 13**,** 273-297.

TULU, G. S., WASHINGTON, S., HAQUE, M. M. & KING, M. J. 2015. Investigation of pedestrian crashes on two-way two-lane rural roads in Ethiopia. *Accident Analysis & Prevention,* 78**,** 118-126.

VERZOSA, N. & MILES, R. 2016. Severity of road crashes involving pedestrians in Metro Manila, Philippines. *Accident Analysis & Prevention,* 94**,** 216-226.

VICROADS 2016. interactive crash statistics application CrashStats 2010-2016. 2016 ed. Victoria, Melbourne: Roads Corporation of Victoria.

XIE, Z. & YAN, J. 2008. Kernel Density Estimation of traffic accidents in a network space. *Computers, Environment and Urban Systems,* 32**,** 396-406.

XU, C., LIU, P., WANG, W. & LI, Z. 2014. Identification of freeway crash-prone traffic conditions for traffic flow at different levels of service. *Transportation Research Part A: Policy and Practice,* 69**,** 58-70.

ZEGEER, C., STEWART, J., HUANG, H. & LAGERWEY, P. 2001. Safety effects of marked versus unmarked crosswalks at uncontrolled locations: analysis of pedestrian crashes in 30 cities. *Transportation Research Record: Journal of the Transportation Research Board***,** 56-68.

ZHU, X. & LEE, C. 2008. Walkability and Safety Around Elementary Schools: Economic and Ethnic Disparities. *American Journal of Preventive Medicine,* 34**,** 282-290.