

# INFLUENCING FACTORS ON VEHICLE-PEDESTRIAN CRASH SEVERITY OF SCHOOL-AGED PEDESTRIANS

ALIREZA TORAN POUR<sup>1</sup>, SARA MORIDPOUR<sup>1</sup>, RICHARD TAY<sup>2</sup> & ABBAS RAJABIFARD<sup>3</sup>

<sup>1</sup>School of Engineering, RMIT University, Australia

<sup>2</sup>School of Business, IT and Logistics, RMIT University, Australia

<sup>3</sup>Department of Infrastructure Engineering, the University of Melbourne, Australia

## ABSTRACT

Every year, about 19 percent of vehicle-pedestrian crashes in Melbourne metropolitan area, Australia, involves pedestrians with less than 18 years of age or school-aged pedestrians. This paper aims to identify contributing factors on vehicle-pedestrian crash severity of this age group. Reasonable walking distance to schools is applied in Geographic Information Systems (GIS) to identify vehicle-pedestrian crashes around schools. Then Boosted Decision Tree (BDT) and Cross-Validation (CV) technique are applied to explore significant factors. Results show that the distance of pedestrians from school is a significant factor on vehicle-pedestrian crash severity for this age group. This result could assist in identifying safe distance and safe zone around schools. Furthermore, Public health indicators such as income and commuting type from or to school found as other contributing factors to this crash type.

*Keywords: vehicle-pedestrian crash, school-aged, BDT, Melbourne, GIS.*

## 1 INTRODUCTION

On average, every year about 34 pedestrians are killed in traffic crashes in Melbourne metropolitan area, representing 24% of the total traffic fatalities. Also, more than 9% of the pedestrians killed in those crashes were under the age of 18 (i.e. the age of school children) whereas 19% of the total pedestrians seriously injured were under the age of 18 [1]. To reduce the involvement of school children in such crashes, we need to identify contributing factors on vehicle-pedestrian crashes for this age group and improve safety in the school zones.

In recent years, many research conducted to identify contributing factors on vehicle-pedestrian crash frequency and severity [2]–[11]. For instance, Tay et al. [3], 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. [2] indicated that neighbourhood social characteristics were as important as traffic and infrastructure variables in severity of pedestrian crashes.

However, there is relatively fewer research focused on school-aged vehicle-pedestrian crashes. For instance, In addition, Graham and Glaister [12] found that the probability of child pedestrian casualties is higher in more deprived areas. In another study, Noland and Quddus [13] found that more severe pedestrian injuries are associated with the areas with lower income, higher percent of local roads, higher per capita expenditure on alcohol, and larger numbers of people. Abdel-Aty et al. developed a GIS base crash analysis and Log-linear model for pedestrians under 19 years of age in Florida [14]. In this research, they showed that majority of school-aged children crashes occurred in the areas near schools. Furthermore, in this study drivers' and pedestrians' age, 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 more strongly associated with built environment features [15].

Koopmans et al. [16], 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 on crash injury severity of pedestrians. Lee et al. [17], applied standard negative binomial and zero-inflated negative binomial models to identify the influencing environmental attributes of intersections on crashes involving children aged 10 to 12 years of age in Korea near elementary schools. 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-way intersections were significant and positively associated with perceived crash risk among school-aged children.

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

The objectives of this study are: (a) to identify contributing factors on vehicle-pedestrian crash severity for the school-aged pedestrians. For this reason, the Boosted Decision Tree (BDT) is developed and influencing factors on this crash types identified. b) to identify crash severity risk distance from schools using GIS analysis and BDT model development. Using Network analysis in GIS and Partial Dependence Plots in BDT model are developed to identify the in which distance from schools the severity of crashes in lower.

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 DATA AND METHODS

### 2.1 Dataset

Road Crash Information System (RCIS) is used to extract pedestrian crashes. 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 to at least one of the road users involved in the accident 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 the hospital [1]. 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. Of the total of 11,548 vehicle-pedestrian crashes, 2,161 are related to pedestrians less than 19 years of age. According to VicRoads severity classification, of the 2,161 school-aged vehicle-pedestrian crashes included in the study, 1.2% were fatal crashes, 43.5% were serious injury crashes, and 55.3% 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 [18]. 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 used in GIS. Tables 1 and 2 show a summary of the categorical and continuous variables used in this study, respectively.

## 2.2 Methods

In this research network analysis in ArcMap 10.3 is used to identify distance between crash locations and schools. Then this distance with other factors are applied in Boosted Decision Tree Model (BDT) to identify contributing factors on school-aged vehicle-pedestrian crash severity. Furthermore, Partial Dependence Plot (PDP) is applied to find crash severity risk distance from schools and this result is used in GIS to identify roads with higher probability of crash severity around schools.

Non-parametric techniques such as Decision Tree (DT) and support vector machine have been used in different traffic crash severity analysis [19]–[24]. 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 [20]. In DT models, the root node on top of tree

Table 1: Categorical explanatory variables applied in BDT model.

	Variables	Percent	Variables	Percent	
Severity	Fatal Crash	0.6	Light Condition	Day	98.1
	Serious Injury Crash	37.5		Dusk/Dawn	1.5
	Minor Injury Crash	61.9		Other	0.4
Day of Crash (school days)	Monday	18.3	Node Type	Intersection	49.5
	Tuesday	23.1		Mid-Blocks	49.8
	Wednesday	18.4		Other	0.7
	Thursday	19.7		Surface Condition	Dry
	Friday	20.5	Wet		9.1
Time of Crash	Firs Peak (7:00–9:00 am)	32.7	Atmosphere Condition	Other	3.3
	Off-Peak (10:00 am–3:00 pm)	67.3		Clear	90.4
Month of Crash	January	4.8	Pedestrian Gender	Rainy	5.8
	February	8.8		Other	3.8
	March	10.4		Male	54.2
	April	7.9	Driver Gender	Female	44.7
	May	11.4		Male	47.5
	June	7.7	Speed Limit	Female	43.8
	July	7.9		Under 50 km/h	49.6
	August	8.6		60–70 km/h	41.8
	September	9.8		>70 km/h	5.5
	October	8.2		Other	3.1
	November	7.2			
	December	7.4			

Table 2: Descriptive statistics for continuous variables applied in BDT model.

Variable	Unit	Mean	Std. Deviation
Traffic Volume	Vehicle per day	11694.3	8973.8
Distance to school	Meters	595.2	466.3
Distance to Public Transport Stops	Meters	133.1	169.0
Median income	AUS Dollars	599	155.6
Use Public Transport	Percent	10.4	6.5
Use other type of transport	Percent	3.3	2.3
Use Private	Percent	62.5	14.6
Use Walk	Percent	4.1	7.8
Use Multimodal	Percent	5.1	1.5

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 [25]. 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 [2], [26]–[30]. Since BDT has better performance in vehicle-pedestrian crash severity analysis than DT and BgDT models [2], 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 tree models and improve the quality of prediction [31]. This procedure introduces a stochastic gradient boosting procedure that can improve model performance and reduce the risk of over-fitting [32]. In stochastic gradient boosting procedure and 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.

In this approach, the residuals will be calculated after the first tree is fitted. In BDT, the difference between the target function 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 eqn (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 [33]. This procedure is repeated  $n$  times and on  $m$  models to adjust the estimated probabilities.

In this research, fitted BDT models are obtained using the “gbm” library [34], in the R software [35], in the caret package [36]. 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. In 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 ( $s \times k$ ) estimates of performance are then averaged to obtain the overall re-sampled estimate.

To illustrate the influence of factors on crash severity levels the PDP is applied. Partial Dependence Plots (PDP) depict the relationship between the response and one predictor variable, while accounting for the average effects of all other predictors [32], [37].

In this research, a 10-fold cross-validation with 5 iterations [2], is applied for each model and the performances of the models were estimated. In addition to the error rate, interaction depth and shrinkage are other parameters that are used to evaluate the BDT models. The shrinkage or learning rate is used to determine the contribution of each tree to the growing model. This parameter is used to decrease the contribution of each tree in the model. Tree complexity or interaction depth represented the depth of a tree and showed the interaction among predictor variables. Based on a previous research conducted by Toran Pour et al. [2], the interaction depth and shrinkage parameters are assumed to be 15 and 0.1, respectively and the model repeated 2,000 times (boosting iterations) to find the final model.

### 3 RESULTS AND DISCUSSION

Fig. 1 shows the top 10 most important predictor variables for school-aged vehicle-pedestrian crash severity. As shown in this figure. ‘Crash Distance to School’ is the most important contributing factor on the severity of school-aged vehicle-pedestrian crashes, showing that this variable is a significant influencing variable on vehicle-pedestrian crash severity at this age group. The results from this study show that this factor needs to be considered in vehicle-pedestrian crash studies around schools and for school-aged pedestrians.

Median income of crash location neighbourhoods is the second 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 [13], [38]. This result will assist transportation engineers, planners and policy makers in identifying the target customer segments for improving child pedestrian safety. Knowing the right target audience is critical for the success of safety education and communications programs, such as publishing child pedestrian safety bulletins, safety programs at schools, and using warning messages on billboards.

According to Fig. 1, Traffic volume and Distance of crash to public transport stops are two next contributing factors on the vehicle-pedestrian crash severity of this age group. These results are consistent with results from other studies that found traffic volume and public transport stops could influence vehicle-pedestrian crashes [2], [39]–[41]. Furthermore, Fig. 1 indicates that type of commuting to/from school could be a contributing factor on school-aged vehicle-pedestrian crash severity. School-home commuting type could influence walking distance and exposure of school-aged pedestrians to vehicular traffic. Finally, this figure shows that month of crash (March) is another factor that is identified as influencing factor on this type of crashes. In Australia, March is the last month of school term 1 and it is one of the busiest months for schools and school-aged pedestrians.

Fig. 2 shows the partial dependence plots for the first 10 top factors for different levels of vehicle-pedestrian crash severity. In this figure, it is possible to identify the influence of different variables on vehicle-pedestrian crash severity levels. Fig. 2(a) illustrates that increasing the distance of crash location from/to schools from about 450 meters will increase the probability of fatal crashes. Therefore, it is possible to define this distance as safe walking distance toward schools for school-aged pedestrians. Furthermore, this result assists in identifying target roads to apply road safety strategies and plans for school-aged pedestrians including speed calming, warning systems or child pedestrian crossing supervision.

Fig. 2(b) presents the influence of median income on crash severity of school-aged pedestrians around the schools. This figure shows that the probability of a vehicle-pedestrian crash to be fatal is much higher in suburbs around schools with less weekly income. According to this figure, the risk of fatal crashes that occurs in suburbs with median weekly income between \$370.00 and \$450.00 per week (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 [6], [13], [42]. Our results, as highlighted in Fig. 2(b), suggest 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 child living in these targeted suburbs.

Moreover, Fig. 2(c) shows 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 crash. 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 increasing traffic volume can increase pedestrian crash frequency and the probability of pedestrian crash severity [2], [40]–[44]. These results suggest that transportation engineers and planners may want to target roads with more than 15,000 vehicles per day to improve the safety of these vulnerable road users around schools. More pedestrian crossings, pedestrian signals and flashing lights on these roads may assist in improving the safety of school-aged pedestrians.

Fig. 2(d) shows that the severity of vehicle-pedestrian crashes for this age group and around schools is decreased by increase in the distance of crash to/from public transport stops (from 0 to 50 meters) and then remained stable (from 50 to 110 meters). According to this figure the probability of fatal crashes then increases with the increase in the distance of pedestrian crash locations to public transport stops from 160 to 450m. Using different warning signals and signs around public transport stops may increase the attention given by drivers to pedestrians. More research is required to analyse vehicle-pedestrian crashes in the vicinity of public transport stops and further the 450m distance to identify appropriate pedestrian safety programs [2].

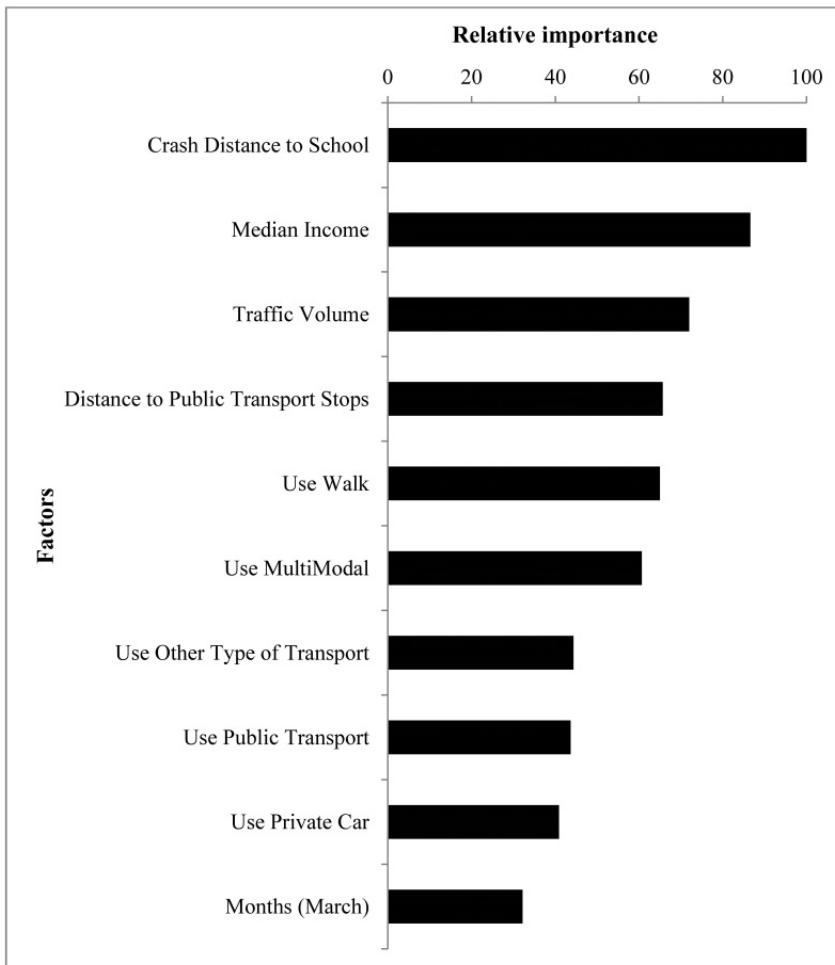


Figure 1: Top 10 relative importance of predictor variables for school-aged vehicle-pedestrian crashes in BDT model.

Figs 2(e) to (i) show the influence of using different transport type to commute between home and schools on school-aged vehicle-pedestrian crash severity. These figures show that different commute type could influence the vehicle-pedestrian crashes differently. For instance, increasing the percentage of people that walk or use public transport to commute between home and schools could slightly increase the risk of fatal vehicle-pedestrian crashes. However, using multimodal transport (e.g. Private-public transport) could decrease this risk.

Finally, Fig. 2(j) shows that the risk of school-aged vehicle-pedestrian crashes could vary at different months of year. This figure shows that the risk of fatal crash is increased in March. In Victoria, Australia, March is the last month of term 1 for schools and in this month students may be more active than other months. This result could assist in identifying the month with higher risk of accident to apply safety programs for students and parents to improve the safety of school-aged pedestrians.

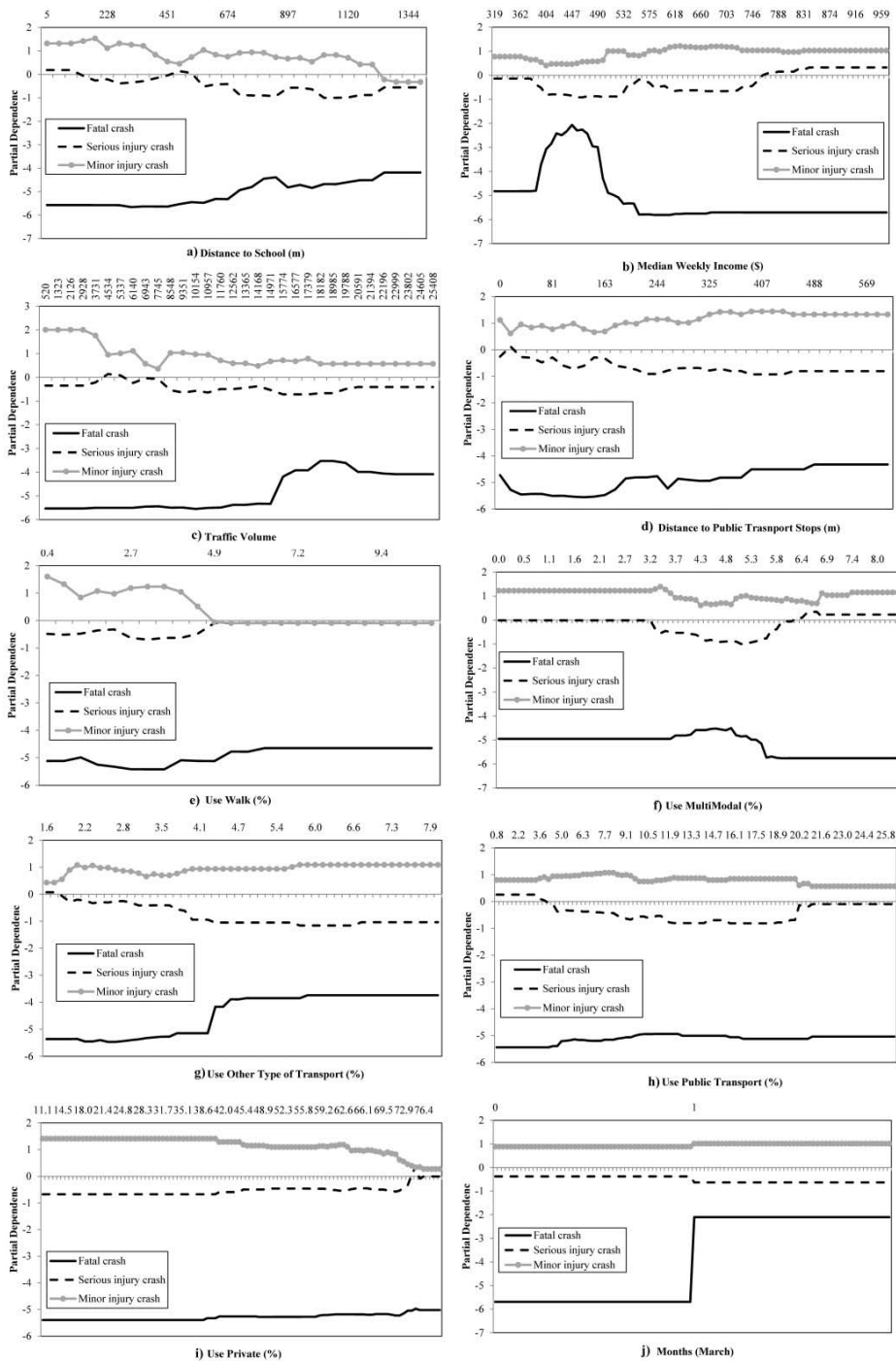


Figure 2: Partial Dependence Plot for first 10 top contribution factors to the school-aged vehicle-pedestrian crash severity.





#### 4 CONCLUSION

Identifying 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 pedestrian deaths and injuries of these vulnerable road users. In this research, the BDT model was applied to identify the contributing factors on the school-aged vehicle-pedestrian crashes. Also, GIS is applied to identify the distance of crash location to/from schools, and also extract socio-economic factors such as income and commute type related to location of crashes. Results from this research would provide valuable information to assist road safety professional in targeting the right neighbourhoods to implement different safety measures related to pedestrians and drivers, as well as targeting site specific safety measures to reduce vehicle-pedestrian crashes for school-aged pedestrians.

This study found that distance of crash to schools is the most important variable in vehicle-pedestrian crash severity around schools. Moreover, this research revealed that public wellbeing indicators such as median income and using public transport have influence on severity of crashes in this age group.

This research found that traffic volume and distance of crashes to/from public transport stop are two other contributing factors on school-aged vehicle-pedestrian crashes in Melbourne metropolitan area. Furthermore, this research showed that month of the year could be an important factor for vehicle-pedestrian crash severity in this age group. These results could assist transport and safety planners in identifying target suburbs/roads and introducing appropriate countermeasures such as pedestrian safety educational programs to improve the safety of school-aged pedestrians.

#### REFERENCES

- [1] Interactive crash statistics application CrashStats 2010–2016 [Internet]. Roads Corporation of Victoria. 2016 [cited 2016]. <https://www.vicroads.vic.gov.au/>.
- [2] Toran Pour, A., Moridpour, S., Tay, R. & Rajabifard, A., Modelling pedestrian crash severity at mid-blocks. *Transportmetrica A: Transport Science*, **13**(3), pp. 273–297, 2017.
- [3] Tay, R., Choi, J., Kattan, L. & Khan, A., A multinomial logit model of pedestrian–vehicle crash severity. *International journal of sustainable transportation*, **5**(4), pp. 233–249. 2011.
- [4] Tarawneh, M.S., Evaluation of pedestrian speed in Jordan with investigation of some contributing factors. *Journal of Safety Research*, **32**(2), pp. 229–236, 2001.
- [5] Sze, N.N. & Wong, S.C., Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes. *Accident Analysis & Prevention*, **39**(6), pp. 1267–1278, 2007.
- [6] Siddiqui, C., Abdel-Aty, M. & Choi, K., Macroscopic spatial analysis of pedestrian and bicycle crashes. *Accident Analysis & Prevention*, **45** pp. 382–391, 2012.
- [7] Rifaat, S.M., Tay, R., Raihan, S.M., Fahim, A. & Touhidduzzaman, S.M., Vehicle-Pedestrian crashes at Intersections in Dhaka city. *The Open Transportation Journal*, **11**(1), 2017.
- [8] Verzosa, N. & Miles, R., Severity of road crashes involving pedestrians in Metro Manila, Philippines. *Accident Analysis & Prevention*, **94**, pp. 216–226, 2016.
- [9] Li, D., Ranjitkar, P., Zhao, Y., Yi, H. & Rashidi, S., Analyzing pedestrian crash injury severity under different weather conditions. *Traffic Injury Prevention*, 2016.



- [10] Tulu, G.S., Washington, S., Haque, M.M. & King, M.J., Investigation of pedestrian crashes on two-way two-lane rural roads in Ethiopia. *Accident Analysis & Prevention*, **78** pp. 118–126, 2015.
- [11] Toran Pour, A., Moridpour, S., Tay, R. & Rajabifard, A, eds. *A Partial Proportional Odds Model for Pedestrian Crashes at Mid-Blocks in Melbourne Metropolitan Area*. MATEC Web of Conferences: EDP Sciences, 2016.
- [12] Graham, D. J. & Glaister, S., Spatial variation in road pedestrian casualties: the role of urban scale, density and land-use mix. *Urban Studies*, **40**(8) pp. 1591–1607, 2003.
- [13] Noland, R. & Quddus, M., Analysis of pedestrian and bicycle casualties with regional panel data. *Transportation Research Record: Journal of the Transportation Research Board*. (1897) pp. 28–33, 2004.
- [14] Abdel-Aty, M. & Chundi, S.S. & Lee, C., Geo-spatial and log-linear analysis of pedestrian and bicyclist crashes involving school-aged children. *Journal of Safety Research*, **38**(5) pp. 571–579, 2007.
- [15] Rothman, L., Macarthur, C., To, T., Buliung, R. & Howard, A., Motor Vehicle-Pedestrian Collisions and Walking to School: The Role of the Built Environment. *Pediatrics*, **133**(5) p. 776, 2014.
- [16] Koopmans JM, Friedman L, Kwon S, Sheehan K. Urban crash-related child pedestrian injury incidence and characteristics associated with injury severity. *Accident Analysis & Prevention*, **77**, pp. 127–36, 2015.
- [17] Lee, G., Park, Y., Kim, J. & Cho, G.H., Association between intersection characteristics and perceived crash risk among school-aged children. *Accident Analysis & Prevention*, **97**, pp. 111–21, 2016.
- [18] ABS. Australia's population by country of birth Canberra, ACT: Australian Bureau of Statistics, 2013.
- [19] Chang, L.Y. & Wang, H.W., Analysis of traffic injury severity: An application of non-parametric classification tree techniques, *Accident Analysis & Prevention*, **38**(5), pp. 1019–1027, 2006.
- [20] Kashani, A.T. & Mohaymany, A.S., Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models. *Safety Science*, **49**(10), pp. 1314–1320, 2011.
- [21] Li, X., Lord, D., Zhang, Y. & Xie, Y., Predicting motor vehicle crashes using support vector machine models. *Accident Analysis & Prevention*, **40**(4), pp. 1611–1618, 2008.
- [22] Abellán, J., López, G. & de Oña, J., Analysis of traffic accident severity using Decision Rules via Decision Trees. *Expert Systems with Applications*, **40**(15), pp. 6047–6054, 2013.
- [23] Chang, L.Y. & Chien, J.T., Analysis of driver injury severity in truck-involved accidents using a non-parametric classification tree model. *Safety Science*, **51**(1), pp. 17–22, 2013.
- [24] Jung, S., Qin, X. & Oh, C., Improving strategic policies for pedestrian safety enhancement using classification tree modeling. *Transportation Research Part A: Policy and Practice*, **85**, pp. 53–64, 2016.
- [25] Lord, D., van Schalkwyk, I., Chrysler, S. & Staplin, L., A strategy to reduce older driver injuries at intersections using more accommodating roundabout design practices. *Accident Analysis & Prevention*, **39**(3), pp. 427–432, 2007.
- [26] Pham, M.H., Bhaskar, A., Chung, E. & Dumont, A.G., eds, 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*.



- [27] Jiang, X., Abdel-Aty, M., Hu, J. & Lee, J., Investigating macro-level hotzone identification and variable importance using big data: A random forest models approach. *Neurocomputing*, **181**, pp. 53–63, 2016.
- [28] Chung, Y.S., Factor complexity of crash occurrence: An empirical demonstration using boosted regression trees. *Accident Analysis & Prevention*, **61**, pp. 107–118, 2013.
- [29] Lee, C. & Li, X., Predicting Driver Injury Severity in Single-Vehicle and Two-Vehicle Crashes with Boosted Regression Trees. *Transportation Research Record: Journal of the Transportation Research Board*, **2514**, pp. 138–48, 2015.
- [30] Xu, C., Liu, P., Wang, W. & Li, Z., Identification of freeway crash-prone traffic conditions for traffic flow at different levels of service. *Transportation Research Part A: Policy and Practice*, **69**, pp. 58–70, 2014.
- [31] Elith, J., Leathwick, J.R. & Hastie, T., A working guide to boosted regression trees. *Journal of Animal Ecology*, **77**(4), pp. 802–813, 2008.
- [32] Friedman, J.H., Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pp. 1189–1232, 2001.
- [33] Matignon, R., *Data Mining Using SAS Enterprise Miner*, South Sanfrancisco, CA: John Wiley & Sons, p. 584, 2007.
- [34] Ridgeway, G., *Generalized Boosted Models: A guide to the gbm package*, **1**(1), 2007.
- [35] Team RDC, *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria, 2014. ISBN 3-900051-07-0.
- [36] Kuhn, M., Caret package. *Journal of Statistical Software*, **28**(5), 2008.
- [37] Friedman, J.H. & Meulman, J.J., Multiple additive regression trees with application in epidemiology. *Statistics in Medicine*, **22**(9), pp. 1365–1381, 2003.
- [38] LaScala, E.A., Gruenewald, P.J. & Johnson, F.W., An ecological study of the locations of schools and child pedestrian injury collisions. *Accident Analysis & Prevention*, **36**(4) pp. 569–576, 2004.
- [39] Assailly, J.P., Characterization and prevention of child pedestrian accidents: An overview. *Journal of applied developmental psychology*, **18**(2), pp. 257–262, 1997.
- [40] Zegeer, C., Stewart, J., Huang, H. & Lagerwey, P., 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*, **1773**, pp. 56–68, 2001.
- [41] Morency, P., Gauvin, L., Plante, C., Fournier, M. & Morency, C., Neighborhood social inequalities in road traffic injuries: The influence of traffic volume and road design. *American Journal of Public Health*, **102**(6), pp. 1112–1119, 2012.
- [42] Zhu, X. & Lee, C., Walkability and safety around elementary schools: Economic and ethnic disparities. *American Journal of Preventive Medicine*, **34**(4), pp. 282–290, 2008.
- [43] Pulgurtha, S.S. & Sambhara, V.R., Pedestrian crash estimation models for signalized intersections. *Accident Analysis & Prevention*, **43**(1), pp. 439–446, 2011.
- [44] Miranda-Moreno, L.F., Morency, P. & El-Geneidy, A.M., The link between built environment, pedestrian activity and pedestrian–vehicle collision occurrence at signalized intersections. *Accident Analysis & Prevention*, **43**(5), pp. 1624–1634, 2011.

