An Analysis of Vehicle Occupants' Injury Severity in Crashes Occurred On Rural Freeways and Multilane Highways in Iran

Ali Tavakoli Kashani[1], Mohammad Mehdi Besharati[2]

Received: 25.05.2016 *Accepted: 24.08.2016*

Abstract:
Vehicle occupants comprise a considerable proportion of traffic crash victims in Iran. This paper has focused on vehicle occupants' injury severity and employed the Classification and Regression Tree (CART) technique in order to identify the most important variables affecting the injury severity of these road users in crashes occurred on rural freeways and multilane highways in Iran over a three year period (2006-2008). In the procedure adopted in this paper, the problem of three-class prediction was decomposed into four binary prediction models. Results revealed a high overall prediction accuracy of the models. Ten explanatory variables were considered in the current study in order to find the most important variables affecting the injury severity of occupants. In this regard, some "if-then" rules pertaining to the conditions that lead to more severe injuries are provided based on the decision tree analysis. Results confirm the already-known importance of seatbelt usage for preventing serious injuries in one hand, and imply the insufficiency of seatbelt usage for protecting the occupants from receiving serious injuries in some collision types, on the other hand. This underscores the need for more safety instruments (especially airbags for all occupants of the vehicle) to be installed in passenger cars in Iran.

Keywords: Vehicle occupant; seat belt; injury severity; rural freeway and multilane highway

Corresponding author E-mail: alitavakoli@iust.ac.ir
1 Assistant Professor, School of Civil Engineering, Iran University of Science and Technology, Tehran, Iran, and Road Safety Research Centre, Iran University of Science and Technology, Tehran, Iran
2 Ph.D. Student, School of Civil Engineering, Iran University of Science and Technology and Road Safety Research Centre, Iran University of Science and Technology, Tehran, Iran

1. Introduction

Completing a trip without injury or property damage has always been the primary concern of traffic engineers and traffic safety experts. The high rate of traffic crashes on Iran roads causes thousands of deaths, injuries, and economic loss every year. During 2006-2008, three people per hour died on average, due to traffic crashes in Iran (i.e. an average number of 26,000 fatalities annually) [F.M.O.I, 2012]. On the other hand, according to Iran Road Maintenance and Transportation Organization annual report in 2015, a considerable portion of road networks in Iran are comprised of rural freeways and multilane highways (19.7%), which constitute 16889 km of the total road network [R.M.T.O, 2015]. According to the statistics, about two-thirds of the above-mentioned fatalities occurred on rural roads and one-third on urban roads. This reveals the importance of paying considerable attention to rural freeway and multilane highways crashes in Iran.

Previous literature revealed that such factors as the location and the cause of the crash [Al-Ghamdi, 2002], car size [Wood and Simms, 2002], use of a seatbelt, drinking alcohol and the age and the gender of the drivers [Delen, Sharda, and Bessonov, 2006] might influence the injury severity car occupants.

Furthermore, in their paper, Anarkooli, and Hosseinlou, (2016) investigated lighting condition differences in the injury severity of crashes occurred on two-lane rural roads of the state of Washington. They highlighted the importance of deploying street lights at and near intersections (or access points) on two-lane rural roads to reduce the injury severity of crashes.

From the viewpoint of methodology, over the years, traffic safety experts and researchers have used a wide range of methods to assess the crash injuries, among which, Logistic regression, Multinomial logit and ordered probit models are the most popular models [Mirbaha, Saffarzadeh, and Noruzoliaee, 2012; Pour-Rouholamin and Zhou, 2016; Tavakoli Kashani, Rabieyan, and Besharati, 2015]. Savolainena, Mannering, Lord, and Quddus, (2011) have provided a thorough review of methodological alternatives used in this field.

On the other hand, when dealing with a large amount of data, Data Mining (DM) techniques are the most practical approach in which, unlike most regression models, no assumptions are made regarding the underlying distribution of predictors and dependent variables. During the last decade, several data mining techniques including Classification and Regression Trees(CART) method [Chang and Chen, 2005a; Tavakoli Kashani, Rabieyan, and Besharati, 2014; Tavakoli Kashani, Shariat-Mohaymany, and Ranjbari, 2012], Bayesian Networks [Mujalli and De ONa, 2011], Artificial Neural Networks [Delen, Sharda, and Bessonov, 2006] clustering analysis[Tavakoli Kashani and Besharati, 2016] and association rules discovery [Montella, 2011] have been used to conduct analysis on the crash data.

The Classification and Regression Tree (CART) method, as one of the data mining techniques, is a useful tool to find the most appropriate predictors to classify the variables. The CART method can also handle numerical data that are highly skewed (such as traffic crashes) or multimodal, as well as categorical predictors with an ordinal or non-ordinal structure [Breiman, Friedman, Stone, and Olshen, 1984].

A review of past studies indicates that the CART method has successfully been employed in traffic safety researches. For instance, a study by Chang and Chen [Chang and Chen, 2005b] conducted a comparison between the negative binomial regression model and the CART method to analyse traffic road accidents on national freeways in Taiwan and concluded that the CART method is more accurate than the regression model. Chang and Wang (2006) employed the CART method in Taiwan to explore the significant relationship between the degree of injury severity and its associated explanatory variables. They found that the pedestrians, bicyclists, motorcyclists, and passengers were at more risk than drivers for severe injuries. In another study conducted in Iran, the CART and the logistic regression methods were compared to find the relationship between human factors and the injury severity of road crashes [Pakgohar, Sigari Tabrizi, Khalili, and Esmaeili, 2011].

Additionally, some previous studies focused on the issue of analysing crashes on freeways and multilane highways [Das and Abdel-Aty, 2010; Milton, Shankar, and Mannering, 2008; Savolainena, Mannering, Lord, and Quddus, 2011]. For instance, Das and Abdel-Aty, (2010) reported the vision obstruction and percentage of trucks as the variables that might lead to more severe crashes. Moreover, they reported that dry surface, wider shoulder and sidewalk widths decrease the crash severity.

Based on these studies, traffic safety researchers have gained good knowledge about crash severity. However, a review of the previous literature showed that very few previous studies have focused on the risk factors that might influence the injury severity of vehicle occupants (excluding the driver) especially in crashes occurred on rural freeways and multilane highways. Therefore, the main purpose of the present research is to explore the most significant factors affecting the injury severity of this group of road users in crashes occurred on rural freeway and multilane highways over three successive years (2006-2008) in Iran, using the CART technique. It is noteworthy that the current study was conducted over all rural freeway and multilane highways in Iran. The geographical extent of this study might help to uncover interesting patterns in these type of crashes.

2. Methodology

2.1 Data

For this study, Iran crash data maintained by the Iran Traffic Police from 2006 to 2008 has been used. These data are obtained from the Traffic Accident Record Form, KAM 114. This form covers different characteristics of traffic crash including environmental, human and vehicle attributes. In this database, the injury severity of the occupants involved in crash is recorded in terms of three levels: light injury, serious injury, and fatality. Evaluation of Injury severity is based on assessment by law enforcement officers at the crash scene. In this regard, “Serious injury” is any injury other than fatal injury, which prevents the injured person from walking, driving, or normally continuing the activities he/she was capable of performing before the injury occurred and forces the patient to be admitted to hospital as an “in-patient”. In addition, “Light injury” is possible or evident injury that is non-incapacitating and is treated on an “outpatient” basis. It also should be noted that in the current study, by the term “vehicle occupants”, the authors mean those passengers of the vehicle (except the driver), that have experienced a light or serious injury or died in the crash.

As the scope of the current study was to explore the influencing factors affecting the injury severity of vehicle occupants (excluding drivers) in crashes occurred on rural freeway and multilane highways; thus, the data pertaining to crashes occurred on the two-way two lane highways as well as minor roads were excluded. Finally, 6798 records were identified, each of which representing the crash attributes of a single occupant.

As shown in Table 1, the injury severity of the passenger was set as the target variable. Furthermore, ten independent variables were recommended to enter the CART model. In order to develop a CART model, the dataset might be randomly divided into 2 subsets of training and testing. Further explanations on this issue is provided in the subsequent sections. In this study, the model was trained by 70% of the data, and the remaining 30% was used in the testing process.

2.2 The Classification and Regression Tree (CART) Method

Figure 1 presents the principle of the CART method in developing the classification tree. In this regard, the data are first concentrated at a node located at the top of the tree (called root node or parent node). Then, the observations in the root node are divided into two subsets (called child nodes) based on an independent variable (called splitter) that creates the best homogeneity. In other words,¬ the data in each child node are more homogenous than those in the upper parent node. Next, one or both of these child nodes are further split into new subsets based on another splitter and this process is continued repeatedly until all the data in each child node have the greatest possible homogeneity or some stopping rules are triggered. Such a node that has no further branches is called a terminal node or leaf node (Figure 1). The fundamental idea is to select each split of a node so that the observations in each of the descendant nodes are purer or more homogeneous than those in the parent node. Thus, the principle behind tree growing is to recursively partition the target variable to minimize “impurity” in the terminal nodes.

There are several indexes for evaluating node impurity and growing the tree. The Gini index is one of the most common measures and is defined as:

$$P(j \mid m) = \frac{p(j,m)}{p(m)}, \quad P(j,m) = \frac{\pi(j) N_j(m)}{N_j}, \quad P(m) = \sum_{j=1}^{j} P(j,m) \tag{1}$$

$$Gini\ (m) = 1 - \sum_{j=1}^{j} P^2(j \mid m) \tag{2}$$

Table 1. Variable description

Variable	Description
Injury severity	Target variable: 1. Light injury 2. Serious injury 3. Fatality
Gender	1.Male, 2.Female
Seat belt	1. Used 2. Not used 3. Unknown
Lighting condition	1.Daylight, 2.Dark, 3.Dusk/dawn
Occurrence Location	1.On roadway, 2.On Shoulder, 3.In median, 4.On roadside, 5.Outside traffic way, 6.Other
Location Type	1. Segment 2.Intersection 3. Bridge 4. Tunnel 5. Roundabout 6. Other
Road surface condition	1. Dry 2.Wet 3. Icy 4. Gravel/Sand 5. Slush/Mud 6. Standing oil 7. Other
Weather condition	1.Clear 2.Foggy 3.Rainy 4.Snowy 5.Stormy 6.Cloudy 7.Dusty
Shoulder Type	1. None 2. Stabilized gravel 3. Paved
Collision type	1. Collision with motorcycle/bicycle 2. Two vehicle collision 3. Multi vehicle collision 4. Collision with pedestrian 5. Collision with animal 6.Fixed object 7. Overturn 8.Fire/Explosion 9. Motorcycle collision with ped/bicycle 10. Two motorcycle collision 11.Other
Primary cause of crash	1. Following too closely 2. Ignoring proper lateral distance 3. Ignoring right of way 4. Inattention 5. Inability to drive 6. Failure to control vehicle 7. Speeding 8. Improper overtaking 9. Straying to the right 10.Improper turning 11. Crossing prohibited place 12. Driving on the wrong side of the road 13.Improper backing 14.Vehicle defect 15. Swerving 16.Pedestrian violation 17.Improper packing 18.Improper towing 19.Red light running 20.Turning in no-turn zone 21.Other

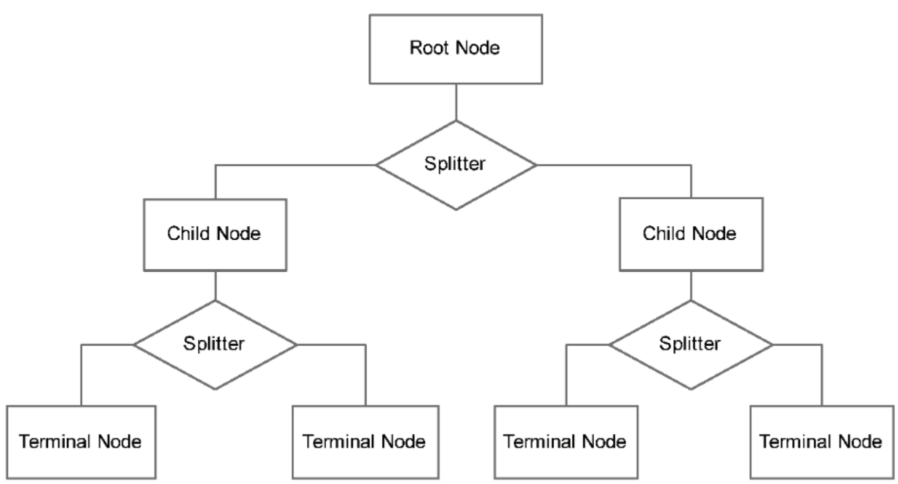

Figure 1. General structure of a decision tree

Gini (m) represents the Gini index which is an indication of how impure or heterogeneous the node m is. In Eq. 1, j denotes the number of target variables or classes, P(j|m) is the conditional probability of a record in class j provided that it is in node m, π(j) is the prior probability considered for class j, $N_j(m)$ is the number of records in class j of node m, and N_j indicates the number of records of class j.

If all observations for one node belong to a specific class, then the would be zero and this indicates the greatest homogeneity and purity in that node. The index for node m would reach its maximum value if there is the same ratio of observations in that node. The prior probability is an indicator for the percentage of observations in every class. Tree growth, based on the Gini index, starts from the root node, which contains all of the observations.

In the CART method, tree growth will continue until there are only similar observations in each terminal node. In such a case, the maximal tree that over fits the training data is created. As the maximal tree over fits the training dataset, the tree classifier might have very small classification errors on the training dataset, but could perform poorly on a new testing dataset. Therefore, in the final step, the tree identified by the CART method is needed to be verified using an independent dataset which is called testing dataset. To lessen the complexity of the maximal tree, the CART method prunes the tree to an optimal size by the cost-complexity pruning method. In this regard, the "misclassification error rate" or "misclassification cost" for each tree is calculated as

misclassification error rate =

$$\sum_{m=1}^{M} p(m)[1-\sum_{j=1}^{j} p^2(j|m)] \quad (3)$$

where p(m) is the proportion of existing observations in the terminal node or leaf m (from all observations) and M is the number of terminal nodes. The above equation shows the data that are misclassified in unrelated classes (e.g., injury observations that are placed in a fatality terminal node).

As shown in Figure 2, the misclassification error rate for training dataset continuously decreases by increasing the complexity of terminal nodes. In contrast, for the testing dataset the misclassification error rate reaches a minimum and then increases as the complexity increases (Figure 2). This reflects the fact that an overfitted and too complex tree will not perform well on a new set of data. After pruning a branch, if the increase in the misclassification cost is sufficiently lower than the decrease in the complexity cost, that branch will be pruned, and a new tree is created. The branches are pruned repeatedly and according to Figure 2, the optimal tree is the one that has the lowest misclassification cost for the test data. For more detailed description of CART method, the readers are referred to [Han and Kamber, 2006].

2.3 Variable Importance Measure (VIM)

The importance score for a particular variable is the sum of the improvement of impurity measures across all nodes in the tree when it acts as a primary or substitute splitter:

$$VIM\ (x_j) = \sum_{t=1}^{T} \frac{\eta t}{N} \Delta Gini\ (S(x_j, t)) \quad (4)$$

Where VIM (x_j) is the importance score for jth independent variable (xj), $\Delta Gini\ (S(x_j, t))$ denotes the reduction of Gini index at node t which is obtained by splitting variable xj, and $\frac{\eta t}{N}$ is the proportion of the observations that belongs to node t. T indicates the total number of nodes and N is the representative for the total number of observations. Such a measure is calculated for all independent variables and is scaled so that the summation for all variables is one. Consequently, the variable with the largest value is the most important compared to the other variables [Pham, 2006].

3. Results and Discussion

3.1 Reduction of The Multi-Class Problem to a Series of Two-Class (Binary) Problems

The target variable in this research (injury severity) was categorized into three imbalanced classes (light injury, serious injury, and fatality). The number of serious injuries (3311) was almost four times greater than that of fatalities (886) which leads to a decrease in the prediction accuracy of the model or each class of the model. It can be inferred from Table 2 that after running the model with the three-class target variable (injury severity) only 31.87% of the data on serious injury crashes were classified correctly. To circumvent this problem, it has been proposed

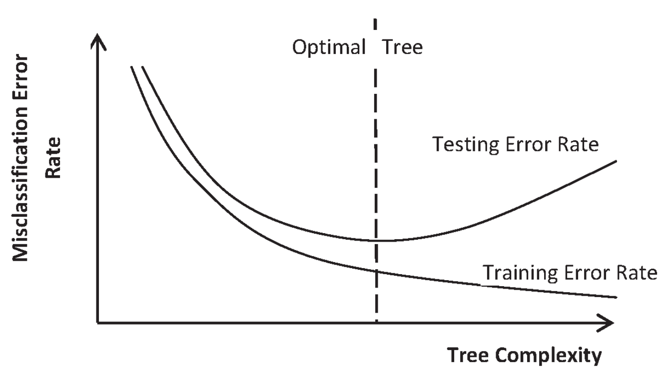

Figure 2. misclassification error rates for training and testing data as a function of tree complexity

to convert it into separate problems with two-class target variable (binary) through ones-vs-all (OVA) and all-vs-all (AVA) methods. However, the OVA and AVA cannot be employed in the present study since the target variable has three levels and combining the light injuries and fatalities into one class is rather misleading. To handle this problem, the previous literature proposed a method to consider the lower injury severity levels versus the higher injury severity levels (i.e. serious injuries combined with fatalities) [Delen et al., 2006].

The combination of the target variables for the four models is presented in Table 3. As shown in model 1.1, the data on light injury and serious injury crashes were combined into one class and labeled as 0 and fatalities were categorized into the other class with the label of 1. Therefore, this model might be labelled as "fatality vs. non-fatality" model. In model 2.2, the light injuries were not imported to the CART method and only the fatal and serious injuries were compared.

These four binary models resulted from combining the target variables were analysed by employing the CART method. For growing the tree, the Gini index was applied and the prior probabilities, $\pi(j)$s, were considered to be equal for all models. The prior probability indicates the proportion of each class in the population. However, when the proportion of one class is much more than that of another class, and the prior probabilities for both classes are adjusted based on the proportion of each class in the training data, then, all the data in the dominant class will be predicted by the resulting model. Therefore, this would lead to an increase in the overall model accuracy. When examining the injury severity, since the percentage of fatal crashes is commonly less than that of injury crashes, the prediction accuracy decreases. A previous study [Steinberg and Golovnya, 2007] suggested considering the equal prior probabilities and taking the variables with low proportions into consideration when the target variables have imbalanced proportions but have nearly the same prediction accuracy importance. Although the overall model accuracy decreases by adopting this approach, the accuracy of predicting the data with the least proportion increases. Such improvements are essential and important to decision makers. Table 4 demonstrates the accuracy of the four developed models as well as the overall model for the training and testing data. The accuracy of the overall model (Table 4) was improved between 17.3% and 22.81% compared to that of the models with three classes (Table 2).

3.2 Variable Importance

Table 5 presents the results of the relative importance of the variables for all binary models. As shown, in all models, the "cause of the crash" was identified as one of the

Table 2. Prediction accuracy of three-class model

	Training (%)	Testing (%)
Light injury	47.14	44.59
Serious injury	33.62	31.87
Fatality	49.65	45.91
Overall	40.88	38.54

Table 3. Graphical illustration of binary configuration

Model No	Model outcome	Light injury 38.26%	Serious injury 48.70%	Fatality 13.03%
1.1	fatality vs. non-fatality	At most serious injury		Fatality
1.2	light injury vs. severe injury	Light injury	Serious injury	
2.1	light injury vs. severe injury or fatality	Light injury	At least serious injury	
2.2	fatality vs. non-fatality		Serious injury	Fatality

Class Label 0 ☐ Class Label 1 ☐

two most important variables. This is in line with findings of a previous study by Al-Ghadimi [Al-Ghadimi, 2002] in which the cause of the crash was reported to be a significant factor that increase the crash severity. Moreover, the seatbelt use in models 1.2 and 2.1 is the most important variable and in models 1.1 is among the three most important variables. It can be inferred from the corresponding decision trees that when the occupants do not fasten seatbelts the probability of receiving serious injuries increases. This is consistent with the findings of some previous studies [Bédard, Guyatt, Stones, and Hirdes, 2002; Delen et al., 2006; Sohn and Shin, 2001; Valent et al., 2002]. It is noteworthy that, since the drivers were excluded from the dataset in this study, results suggest the importance of seat belt use for other car occupants rather than the driver. In models 1.1 and 2.2 in which, fatality was considered as separate class, the collision type is also among the three most important variables. This imply that the type of collision is a significant factor that affects the fatality risk of occupants.

In general, based on Table 5, when fatality vs. non-fatality among the car occupants was modelled as the outcome (model 1.1 and 2.2), "Cause of crash ", "Collision type" and "Seat belt" were among the most important variables. Moreover, When light injury vs. severe injury or fatality among the car occupants was modelled (model 1.2 and 2.1), "Seat belt", "Cause of crash" and "Occurrence location' were among the most important variables.

3.3 The Decision Tree

A substantial advantage of the decision tree over the other methods for decision makers is that the decision tree answers "if-then" questions very clearly. As follows, the decision tree for model 2.1 is illustrated in Figure 3. As previously stated, the binary outcome in this model is "light injury" vs. "at least serious injury". As shown, node 0 as the parent or root node is split into one child node (node 1) and one terminal node (node 2) based on the seatbelt use. This confirms the importance of the seatbelt use as a significant variable to predict the injury severity of car occupants. Node 2, as a terminal node presents the data related to no seatbelt use or the unknown conditions of seatbelt use. As displayed in Figure 3, in such conditions the occupants are predicted to receive at least a serious injury (Class Label 1). On the left branch of the tree, there are four terminal nodes (Nodes 3, 5, 7, and 8). Node 1 on the left side pointing out to the cases in which the occupant used a seat belt, is then split into two other nodes based on the variable of crash occurrence location. As shown, for those occupants that used seatbelt, if the crash occurs on the shoulder, in the median, or outside of traffic way (node 3), then the crash would be so severe that the occupant would at-least experience a serious injury (i.e. Class Label 1). It can be inferred that although the occupants used their seatbelts, the severity of these types of crashes is so high that seatbelt might not be enough to prevent severe injuries. This might indicate the need for more safety instruments such as airbags to be installed in passenger cars not only for the driver but also for occupants.

Node 5 implies that in those crashes occurred on the road way, on the roadside, or other conditions (i.e. cases 1, 4, or 6), in which, the occupant used a seatbelt and the primary cause of the crash was either following too closely, ignoring proper lateral distance, ignoring right of way, inability to drive, speeding, or crossing prohibited place (i.e. cases 1, 2, 3, 5, 7, or 11), the occupant is expected not to receive a serious injury.

Additionally, node 7 indicates that for those passengers that used seatbelt, if the crash occurs because of the col-

Table 4. Prediction accuracy of models with binary class labels

	Model 1.1		Model 1.2		Model 2.1		Model 2.2	
	Training (%)	Testing (%)	Training (%)	Testing (%)	Training (%)	Testing (%)	Training (%)	Testing (%)
Class Label 0	64.7	63.3	55.6	54.2	56.6	55.6	63.2	62.6
Class Label 1	56.1	49.8	60.4	57.1	60.0	57.9	53.3	52.5
Overall	63.6	61.4	58.2	55.8	58.7	57.1	61.2	60.4

Table 5. Relative importance of variables

Model 1.1		Model 1.2		Model 2.1		Model 2.2	
"fatality" vs. "non-fatality"		"light injury" vs. "serious injury"		"light injury" vs. "at least serious injury"		"fatality" vs. "non-fatality"	
Cause of crash	0.66	Seat belt	0.68	Seat belt	0.53	Cause of crash	0.39
Seat belt	0.17	Cause of crash	0.12	Cause of crash	0.25	Collision type	0.35
Collision type	0.17	Occurrence location	0.09	Occurrence location	0.05	Seat belt	0.04
Occurrence location	0	Collision Type	0.04	Collision Type	0.05	Occurrence location	0.04
Shoulder type	0	Road surface condition	0.04	Weather condition	0.05	Location type	0.04
Location type	0	Weather condition	0.04	Location Type	0.03	Lighting condition	0.04
Weather condition	0	Shoulder type	0	Road surface condition	0.03	Weather condition	0.04
Road surface condition	0	Gender	0	Shoulder type	0	Road surface condition	0.03
Gender	0	Location Type	0	Lighting condition	0	Gender	0.03
Lighting condition	0	Lighting condition	0	Gender	0	Shoulder Type	0

lision with animals or with fixed objects (i.e. cases 5 and 6), the occupants are predicted to receive severe injuries (Class Label 1). In contrast, if the crash takes place due to collision with motorcycle/bicycle, with another vehicle, with more than on vehicle, overturn, or motorcycle collision with pedestrian/bicycle (i.e. cases 1, 2, 3, 7, or 9), the occupants would receive light injuries (Class Label 0). Finally, in terms of prioritizing traffic safety measures, one should focus on those factors that significantly affect serious injury. As presented in Figure 3, node 2 is of a great importance confirming the already-known significance of the seatbelt usage. Moreover, nodes 3 and 7 imply the insufficiency of seatbelt usage for preventing serious injuries in some crash types.

4. Conclusion

In this study, the CART algorithm was employed to explore the most significant factors affecting the injury severity of vehicle occupants (excluding the drivers) in crashes occurred on rural freeways and multilane highways in Iran. Results indicates that three variables of "cause of the crash", "seatbelt use" and "collision type" significantly increase the injury severity of vehicle occupants in crashes occurred on rural freeway and multilane highways. More specifically, the model predicted severe injuries for those occupants that did not used seat belt. This confirms the need for more law enforcement over seatbelt use for all vehicle occupants (not only drivers), as well as awareness campaigns about the importance of seat belt use for all vehicle occupants to reduce the injury severity of vehicle occupants.

Moreover, results indicated that despite using seatbelt, if 1) the crash occurs on the shoulder, in the median, or outside of traffic way; or 2) the vehicle collides with an animal or fixed objects; then the crash would be so severe that fastening seat belt could not be enough to protect the vehicle occupant from receiving severe injuries. This underscores the need for more safety instruments (especially airbag for vehicle occupants) to be installed in passenger cars in Iran.

5. Acknowledgements

The authors would like to acknowledge Mr. Sabbaq and Mr. Mishani in Traffic Administration for their relentless and genuine collaboration in providing the crash data.

6. References

- Al-Ghadimi, A. S. (2002)"Using logistic regression to estimate the influence of accident factors on accident severity". Accident Analysis and Prevention, Vol.34, No. 6, pp.729-741.
- Anarkooli, A. J. and Hosseinlou, M. H. (2016) "Analysis

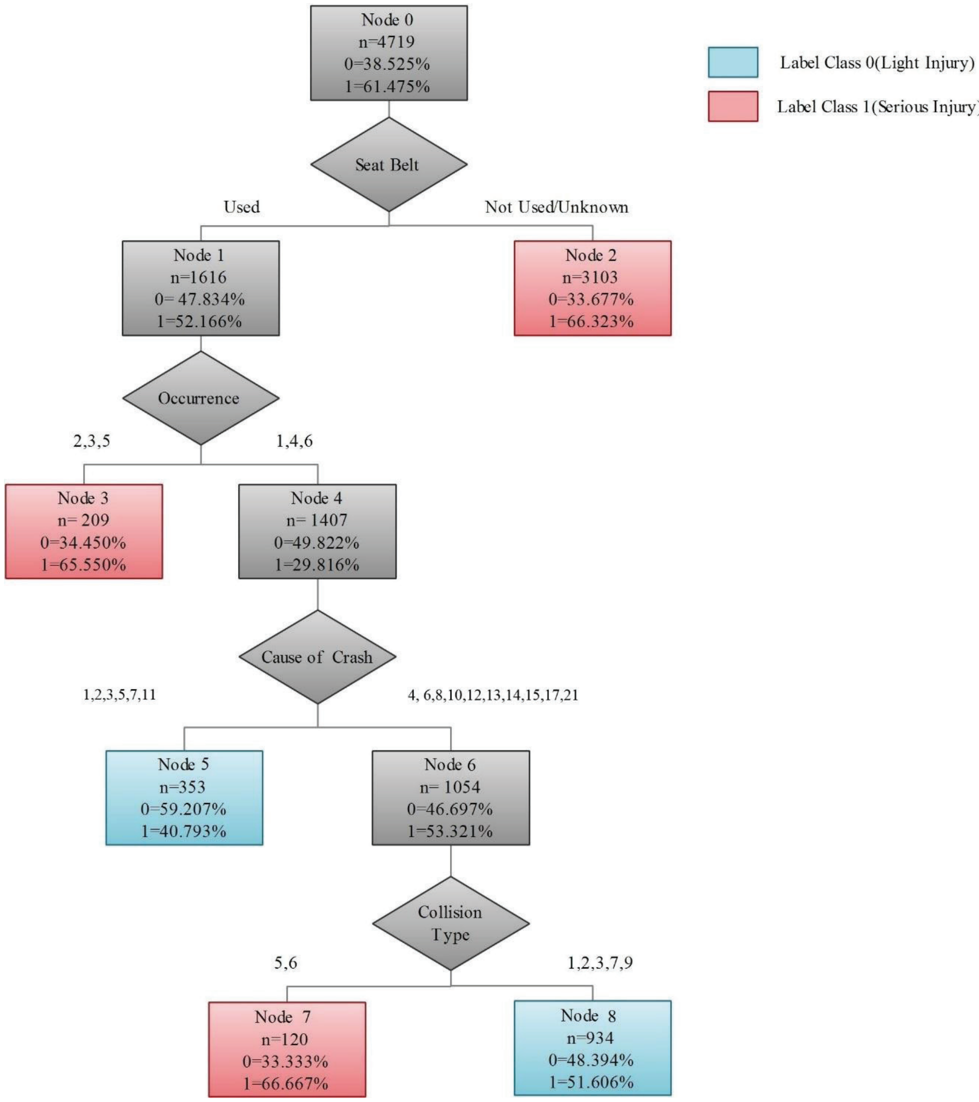

Figure 3. Decision tree for model 2.1

of the injury severity of crashes by considering different lighting conditions on two-lane rural roads", Journal of Safety Research, Vol. 56, pp.57-65.

- Bédard, M., Guyatt, G. H., Stones, M. J. and Hirdes, J. P. (2002) "The independent contribution of driver, crash, and vehicle characteristics to driver fatalities", Accident Analysis and Prevention, Vol. 34, No. 6, pp.717-727.

- Breiman, L., Friedman, J., Stone, C. J. and Olshen, R. A. (1984) "Classification and Regression Trees", USA: CRC press.

- Chang, L.-Y. and Chen, W.-C. (2005a) "Data mining of tree-based models to analyze freeway accident frequency", Journal of Safety Research, Vol. 36, pp. 365-375.

- Chang, L.-Y. and Chen, W.-C. (2005b) "Data mining of tree-based models to analyze freeway accident frequency", Safety Research, Vol. 36, No. 4, pp. 365-375.

- Das, A. and Abdel-Aty, M. (2010). "A genetic programming approach to explore the crash severity on multi-lane roads". Accident Analysis and Prevention, Vol. 42, No. 2, pp. 548-557.

- Delen, D., Sharda, R. and Bessonov, M. (2006) "Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks", Accident Analysis and Prevention, Vol. 38, No. 3, pp. 434-444.

- F.M.O.I. Forensic Medicine Organization of Iran (2012) "FMOI statistical data", Accidents (in Farsi), Retrieved September 3, 2012, from http://www.lmo.ir/?siteid=1andpageid=1347

- Han, J. and Kamber, M. (2006). "Data Mining: Concepts and Techniques (2nd ed.)". Elsevier.

- Milton, J. C., Shankar, V. N. and Mannering, F. L. (2008) "Highway accident severities and the mixed logit model: An exploratory empirical analysis", Accident Analysis and Prevention, Vol. 40, No.1, pp. 260-266.

- Mirbaha, B., Saffarzadeh, M. and Noruzoliaee, M. H. (2012)" A model for predicting schoolchildren accidents in the vicinity of rural roads based on geometric design and traffic conditions", International Journal of Transportation Engineering, Vol. 1, No. 1, pp. 25-33.

- Montella, A. (2011) "Identifying crash contributory factors at urban roundabouts and using association rules to explore their relationships to different crash types", Accident Analysis and Prevention, Vol. 43, No.4, pp. 1451-1463.

- Mujalli, R. O. and De ONa, J. (2011) "A method for simplifying the analysis of traffic accidents injury severity on two-lane highways using Bayesian networks", Journal of Safety Research, Vol. 42, No. 5, pp. 317-326.

- Pakgohar, A., Sigari Tabrizi, R., Khalili, M. and Esmaeili, A. (2011) "The role of human factor in incidence and severity of road crashes based on the CART and LR regression: a data mining approach", Prosedia Computer Science, Vol. 3, pp. 764-769.

- Pham, H. (2006) "Handbook of engineering statistics", Germany: Springer.

- Pour-Rouholamin, M. and Zhou, H. (2016) "Investigating the risk factors associated with pedestrian injury severity in Illinois", Journal of Safety Research, Vol. 57, pp. 9-17.

- Iran Road Maintenance and Transportation Organization - R.M.T.O (2015) "I.R. of Annual report (in Farsi) http://www.rmto.ir/Pages/SalnameAmari.aspx

- Savolainena, P. T., Mannering, F. L., Lord, D. and Quddus, M. A. (2011) "The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives", Accident Analysis and Prevention, Vol. 43, No. 5, pp. 1666-1676.

- Sohn, S. Y. and Shin, H. (2001) "Pattern recognition for road traffic accident severity in Korea", Ergonomics, Vol. 44, No. 1, pp. 107-117.

- Steinberg, D. and Golovnya, M. (2007) "CART 6.0 User's Guide", San Diego, California: Salford Systems.

- Tavakoli Kashani, A. and Besharati, M. M. (2016) "Fatality rate of pedestrians and fatal crash involvement rate of drivers in pedestrian crashes: a case study of Iran", International Journal of Injury Control and Safety Promotion, pp. 1-10.

- Tavakoli Kashani, A., Rabieyan, R. and Besharati, M. M. (2015) "Modeling the effect of operator and passenger characteristics on the fatality risk of motorcycle crashes", Journal of Injury And Violence Research, Vol. 8, No. 1.

- Tavakoli Kashani, A., Rabieyan, R. and Besharati, M. M. (2014) "A data mining approach to investigate the factors influencing the crash severity of motorcycle pillion passengers", Journal of Safety Research, Vol. 51, pp. 93-98.

- Tavakoli Kashani, A., Shariat-Mohaymany, A. and Ranjbari, A. (2012) "Analysis of factors associated with traffic injury severity on rural roads in Iran", Journal of Injury And Violence Research, Vol. 4, NO. 1, pp. 36.

- Valent, F., Schiava, F., Savonitto, C., Gallo, T., Brusaferro, S. and Barbone, F. (2002) "Risk factors for fatal road traffic accidents in Udine, Italy", Accident Analysis and Prevention, Vol. 34, No. 1, pp. 71-84.

- Wood, D. P. and Simms, C. K. (2002) "Car size and injury risk: a model for injury risk in frontal collisions", Accident Analysis and Prevention, 34, No.1, pp.93-99.