Accident Analysis and Prevention 49 (2012) 58–72

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery

Alfonso Montella[a,*], Massimo Aria[b], Antonio D'Ambrosio[b], Filomena Mauriello[a]

[a] University of Naples Federico II, Department of Transportation Engineering "Luigi Tocchetti", Via Claudio 21, 80125 Naples, Italy
[b] University of Naples Federico II, Department of Mathematics and Statistics, Via Cinthia 26, 80126 Naples, Italy

ARTICLE INFO

Article history:
Received 31 July 2010
Received in revised form 8 April 2011
Accepted 25 April 2011

Keywords:
Powered two-wheelers
Data mining
Classification trees
Rules discovery
Hypothesis testing
Crash characteristics
Crash types
Crash severity

ABSTRACT

Aim of the study was the analysis of powered two-wheeler (PTW) crashes in Italy in order to detect interdependence as well as dissimilarities among crash characteristics and provide insights for the development of safety improvement strategies focused on PTWs. At this aim, data mining techniques were used to analyze the data relative to the 254,575 crashes involving PTWs occurred in Italy in the period 2006–2008.

Classification trees analysis and rules discovery were performed. Tree-based methods are non-linear and non-parametric data mining tools for supervised classification and regression problems. They do not require a priori probabilistic knowledge about the phenomena under studying and consider conditional interactions among input data. Rules discovery is the identification of sets of items (i.e., crash patterns) that occur together in a given event (i.e., a crash in our study) more often than they would if they were independent of each other. Thus, the method can detect interdependence among crash characteristics. Due to the large number of patterns considered, both methods suffer from an extreme risk of finding patterns that appear due to chance alone. To overcome this problem, in our study we randomly split the sample data in two data sets and used well-established statistical practices to evaluate the statistical significance of the results.

Both the classification trees and the rules discovery were effective in providing meaningful insights about PTW crash characteristics and their interdependencies. Even though in several cases different crash characteristics were highlighted, the results of the two the analysis methods were never contradictory. Furthermore, most of the findings of this study were consistent with the results of previous studies which used different analytical techniques, such as probabilistic models of crash injury severity. Basing on the analysis results, engineering countermeasures and policy initiatives to reduce PTW injuries and fatalities were singled out. The simultaneous use of classification trees and association discovery must not, however, be seen as an attempt to supplant other techniques, but as a complementary method which can be integrated into other safety analyses.

1. Introduction

Powered two-wheelers (PTWs) form a rapidly growing substantial and integrated part of the transportation system and provide the opportunity to make better use of the existing road system. They are flexible, small, faster than cars in congested traffic and easy to park. Furthermore, PTWs are gradually getting cleaner. Although their relative environmental performance has lagged behind the dramatic improvements achieved for passenger cars in recent years, the overall trend is heading for the right direction. Their CO_2 emission is less than 50% of that produced by cars covering the same distance (ACEM, 2006). Despite these positive characteristics, PTWs safety is a major concern. Crash data from Europe and U.S. indicate that powered two-wheelers are a particularly vulnerable group of road users because they run a relatively high risk in crashes compared with passenger car drivers and occupants.

In the EU27, more than 2.4 million new PTWs were registered in 2007, a number which is increasing every year. There are currently an estimated 33 million PTWs in circulation in the EU27 countries, from small mopeds to powerful motorcycles, which represent about 14% of the entire European private vehicle fleet (cars and PTWs only). In 2007, total road passenger transport activities in the EU27 are estimated to have amounted to 5381 billion pkm. Power two wheelers accounted for 2.8% of this total, but they accounted for more than 15% of the total traffic fatalities (European Commission, 2009).

* Corresponding author. Tel.: +39 0817683941; fax: +39 0817683946.
E-mail addresses: alfonso.montella@unina.it (A. Montella), aria@unina.it (M. Aria), antdambr@unina.it (A. D'Ambrosio), filomena.mauriello@unina.it (F. Mauriello).

0001-4575/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2011.04.025

Italy is the EU country with the greater number of PTWs in use and PTW fatalities. In Italy, PTWs in use were 5,538,653 in 1990 (15% of the total vehicles) and 9,609,094 in 2008 (19% of the total vehicles) (Italian Ministry of Infrastructures and Transports, 2010). A shift from mopeds towards motorcycles was observed. Indeed, in 1990 mopeds accounted for 55% of the total PTWs and motorcycles constituted the remaining 45%, whereas in 2008 mopeds accounted for only 39% of the total PTWs and motorcycles accounted for 61% of the total. PTW fatalities remained almost constant in the last 20 years (1452 in 1991 and 1484 in 2008), but their relative contribution to the global burden of road fatalities increased from 19% in 1991 to a dramatic 31% in 2008.

In the EU27, the increase in PTW fatalities was a significant obstacle towards the achievement of the ambitious target to halve the number of road deaths in the period 2001–2010 set out in the white paper European transport policy for 2010: time to decide (European Commission, 2001). In its communication "Towards a European road safety area: policy orientations on road safety 2011–2020", the European Commission (2010) proposes to continue with the target of halving the overall number of road deaths in the European Union by 2020 starting from 2010 and specifically refers to PTWs stating: "This ever-growing group of users is the one where it is the most difficult to attain a significant reduction in accidents and fatalities".

In the U.S., PTWs are emerging as a major safety issue. As a consequence, goal 11 in the AASHTO Strategic Highway Safety Plan (2005) is to improve motorcycle safety and increase motorcycle awareness; that is, the awareness by highway agencies of the unique characteristics of motorcycles and their needs on the roadway. PTW fatalities accounted for almost 14% of all traffic fatalities in 2008, annual PTW deaths have more than doubled, from 2077 in 1997 to 5091 in 2008, and PTW registrations have increased by about two-thirds, from 5,174,326 in 2000 to 9,850,301 in 2008 (Baer et al., 2010; Teoh, 2010).

Since PTW riders are more vulnerable of other vehicle drivers because of the lack of protection in the event of a crash, most studies on PTWs safety have investigated the factors affecting crash severity developing probabilistic models of crash injury severity (e.g., ACEM, 2009b; de Lapparent, 2006; Quddus et al., 2002; Savolainen and Mannering, 2007; Rifaat et al., 2011; Yannis et al., 2005). Only a few studies have conducted in-depth analyses focused on PTW crashes, such as the MAIDS study which investigated a total of 921 PTW crashes and their contributory factors in five EU countries (ACEM, 2009a). Previous studies produced mixed results, also because of the differences in driver behaviour, vehicle composition, traffic, and road environment in the different context analyzed (Rifaat et al., 2011). Thus, it is crucial to carry out studies on PTW safety related to specific contexts, especially in countries with high number of PTW fatalities, such as Italy.

To fill this research gap, this study was carried out through the analysis of a large data-base of PTW crashes in Italy. Specifically, the aim of the study was to: (1) detect interdependence as well as dissimilarities among crash characteristics; (2) find out non-trivial and unsuspected relations in the data; and (3) provide insights for the development of safety improvement strategies focused on PTWs. The analysis of the crash data was performed using data mining techniques, which are aimed at extracting knowledge from large amounts of data previously unknown and indistinguishable and do not need any assumptions and a priori probabilistic knowledge about the phenomena under studying. The reason for using data mining technique was related also to the nature of the crash phenomenon. A crash can defined as a rare, random, multi-factor event always preceded by a situation in which one or more road users fail to cope with the road environment (ROSPA, 2002). Each crash is the result of a chain of events which is, in its entirety, unique, but some factors are common to several crash circumstances and the identification of these factors and their interdependences by means of data mining techniques can provide insights useful for the development of effective countermeasures.

In detail, two complementary techniques were used: (1) classification trees and (2) rules discovery. Classification and regression tree, a non-parametric model without any pre-defined underlying relationship between the target (dependent) variable and the predictors (independent variables), has been employed in different contexts of application such as marketing (Kim et al., 2001), data editing (Petrakos et al., 2004), missing data imputation and data fusion (Conversano and Siciliano, 2009; D'Ambrosio et al., 2007), and road safety analyses (Chang and Chen, 2005; Chang and Wang, 2006; Harba et al., 2009; Montella et al., in press). When the value of the target variable is discrete, a classification tree is developed, whereas a regression tree is developed for the continuous target variable. Because this study was aimed at exploring categorical variables, classification trees were developed. The association discovery in data mining has been successfully used to uncover obscured patterns or rules in a variety of fields, including market basket analysis, product recommendation, and medical record analysis. Recently, Montella (2011) and Pande and Abdel-Aty (2009) used association rules to detect interdependence among crash characteristics. Similarly, Cheng et al. (2010) used association rules to explore relationships in occupational accidents in the construction industry. Association discovery is the identification of sets of items (i.e., crash patterns) that occur together in a given event (i.e., a crash in our study) more often than they would if they were independent of each other. In terms of understanding the results, association rules provide specific and easy to describe relationships between crash attributes. Due to the large number of patterns considered, classification trees and association rules suffer from an extreme risk of type I error, that is, of finding patterns that appear due to chance alone (Webb, 2007). To overcome this problem, in our study the sample data were randomly split in two data sets and well-established statistical practices were used to evaluate the statistical significance of the results.

The remainder of the paper is organized as follows: Section 2 provides details of the statistical methodologies used in the analysis; Section 3 describes the study data; Section 4 explains the results of the classification trees and of the association analysis methodology; and finally, a discussion places the results in the context of the practice of highway engineering, and conclusions are drawn.

2. Methodology

2.1. Classification trees

Tree-based methods are non-linear and non-parametric data mining tools for supervised classification and regression problems. A tree is an oriented graph formed by a finite number of nodes departing from the root node. In binary trees, each parent node is linked to only two children nodes, namely the left node and the right node. A branch of the tree is a sub-tree obtained pruning the tree at a given internal node. By definition, the terminal nodes present a low degree of impurity compared to the root node. In the tree growing, predictors generate candidate partitions (or splits) at each internal node of the tree, so that a suitable criterion needs to be defined to choose the best partition (or the best split) of the objects. In the tree structure, it is possible to read the conditional interactions among the predictors to explain the behaviour of the response variable.

Any segmentation methodology is characterized by the definition of the following steps: (1) the partitioning criterion to define the optimality function when choosing the best partition (or split) of the objects into homogeneous subgroups; (2) the stopping rule to arrest the growing procedure to build up the tree; and (3) the assignment rule to identify either a class or a value as label of each terminal node. In the following we will focus on the framework of classification trees according to the classification and regression trees methodology introduced by Breiman et al. (1984).

Let (Y, X) be a multivariate random variable where X is the vector of M predictors $(X_1, \ldots, X_m, \ldots, X_M)$ and Y is the criterion variable taking values in the set of prior classes $C = 1, \ldots, j, \ldots, J$. On the basis of a sample of N observations $C = \{(y_n, x_n);\ n = 1, \ldots, N\}$ taken from the distribution of (Y, X) a simple goal of exploratory trees is to uncover the predictive structure of the problem, and understanding which predictors and which interactions of predictors are the most significant to explain the response variable. Tree methods consist of a recursive partitioning procedure of observations into K disjoint classes such that observations are internally homogeneous within the classes and externally heterogeneous among the classes with respect to the response variable Y. The heterogeneity at any node t is evaluated in terms of an impurity measure $i_Y(t)$.

In our study, we expressed the impurity by the Gini heterogeneity index, which is calculated as follows:

$$i_Y(t) = 1 - \sum_j p(j|t)^2 \tag{1}$$

where $p(j|t)$ is the proportion of observations in the node t that belong to the class j.

The total impurity of any tree T is defined as follows:

$$i_Y(T) = \sum_{t \in \tilde{T}} i_Y(t) p(t) \tag{2}$$

where $i_Y(t)$ is the impurity of the node t $p(t) = N(t)/N$ is the weight of the node t, $N(t)$ is the number of observations falling in node t, and $\tilde{T}$ is the set of terminal nodes of the tree T.

The total impurity of any binary tree is reduced by finding at each node of the tree the best partition s^* of the observations into two disjoint classes such that it induces the highest decrease in the impurity of the response variable Y when passing from the node t to the children nodes t_l and t_r:

$$\max_{s \in S} \Delta i_Y(t, s) = \max_s \{i_Y(t) - p_l i_Y(t_l) - p_r i_Y(t_r))\} \tag{3}$$

where $s \in S$ includes the set of splits generated by all predictors, t_l and t_r are the children nodes of the node t, respectively the left node and the right node, and p_l and p_r are the proportions of observations in node t falling into the left and right node.

Tree growing was arrested basing on two criteria: (1) minimum decrease in the impurity equal to 0.001; and (2) maximum size of the tree, choosing the maximum number of levels of the tree equal to four.

Since the analysis was oriented towards a social problem, our objective was to identify specific features which explain the change in the response variables distribution. Thus, we introduced a posterior classification ratio (PCR) to assign response class to each node of the tree, instead of the mode. The PCR was calculated as follows:

$$PCR(j\,|t\,) = \frac{p(j\,|t\,)}{p(j\,|t_{\mathrm{root}}\,)} \tag{4}$$

where t_{root} is the root node of the tree.

The assignment of the class to each node was performed selecting the class j^* with the greater value of PCR:

$$j * |t : \max_j PCR(j|t) \tag{5}$$

Analyses were performed by use of the software SPSS.

2.2. Rules discovery

2.2.1. Association rules

Association discovery is the identification of sets of items (i.e., the crash patterns in our study) that occur together in a given event (i.e., a crash in our study). Association rules are based on the relative frequency of the number of times the sets of items occur alone and in combination in a database. The rules have the form "A → B", where A is the antecedent on the left hand side and B is the consequent on the right hand side. The rules can have multiple items as antecedent and consequent. It is worth mentioning that these rules should not be interpreted as a direct causation, but as associations between sets of items.

Associations discovery was performed using the a priori algorithm according to the methodology introduced by Agrawal et al. (1993). The a priori algorithm uses simple and step-by-step ways to repetitively examine candidate item-sets to find frequent item-sets. Then, it uses the new candidate item-sets produced using frequent item-sets to find new frequent item-sets until no more new item-sets can be produced.

In our study, the rules were filtered by support, confidence, and lift, where support is the percentage of the entire data set covered by the rule, confidence is the proportion of the number of examples which fit the right side among those that fit the left side, and lift is a measure of the statistical dependence of the rule.

Supports are calculated as follows:

$$\mathrm{Support(A \rightarrow B)} = \frac{\#(\mathrm{A} \cap \mathrm{B})}{N}; \quad \mathrm{Support(A)} = \frac{\#(\mathrm{A})}{N}; \quad \mathrm{Support(B)} = \frac{\#(\mathrm{B})}{N} \tag{6}$$

where support(A → B) is the support of the rule, support(A) is the support of the antecedent, support(B) is the support of the consequent, #(A ∩ B) is the number of crashes where both the condition A (antecedent) and the condition B (consequent) are verified, #(A) is the number of crashes where the antecedent is verified, #(B) is the number of crashes where the consequent is verified, and N is the total number of crashes.

Confidence is calculated as follows:

$$\mathrm{Confidence} = \frac{\mathrm{Support(A \rightarrow B)}}{\mathrm{Support(A)}} \tag{7}$$

Lift is calculated as follows:

$$\mathrm{Lift} = \frac{\mathrm{Support(A \rightarrow B)}}{\mathrm{Support(A) \times Support(B)}} \tag{8}$$

The lift of the rule relates the frequency of co-occurrence of the antecedent and the consequent to the expected frequency of co-occurrence under the assumption of conditional independence. A lift value equal to 1 indicates independence, and a value greater than 1 indicates positive interdependence (i.e., the number of times the sets of items occur together is greater than they would if they were independent of each other). The higher the lift, the greater is the strength of the association rule.

It is desirable for the rules to have a high level of support, a large confidence, and a lift value considerably greater than one. Since we have interest also in rare crash characteristics (such as fatal injuries), the support for some rules of interest could be much lower than the support typically used in other applications, such as the market basket analysis. Indeed, the lift value is more important for determining the strength of an association rule than the support and the confidence. Furthermore, to ensure that the patterns identified by the rules are observed with reasonable frequency and that the rules are sufficiently accurate, minimum thresholds also for support and confidence are needed. Given the nature of the data and the significant interest in fatal crashes, minimum values for support and confidence smaller than those reported in literature (Pande and Abdel-Aty, 2009) were selected. Specifically, the threshold values for support (S), confidence (C), and lift (L) were set as follows: $S \geq 0.1\%$, $C \geq 1\%$, and $L \geq 1.5$.

Similarly to the classification trees, where the tree growing was arrested if a minimum decrease in the impurity was not reached, each rule with n items was supplemented by adding one item in the antecedent only if the ratio between the lift values of the $n+1$-item and n-item rules was greater than 1.05.

Analyses were performed by use of the software Statsoft Statistica Data Miner V8.0.

2.2.2. Classification trees

To integrate the results of the classification trees and the association discovery, the results of the classification trees were converted into rules. Indeed, each terminal node of the trees represents a rule, with all the splits of the parent nodes being the antecedent and the class of the terminal node being the consequent.

For each terminal node t, support, confidence and lift were calculated: (a) the support is the ratio between the number of crashes belonging to the class j^* of the node t (i.e., the class with the greater value of the posterior classification ratio, see Eq. (5)) and the total number of crashes; (b) the confidence is the proportion of crashes in the node t that belong to the class j^*; and (c) the lift is the posterior classification ratio (see Eq. (4)).

2.3. Statistical evaluation

Due to the large number of patterns considered, classification trees and association rules suffer from an extreme risk of type I error, that is, of finding patterns that appear due to chance alone (Webb, 2007). To reduce the risk of type I error, we randomly split the dataset in two parts: (1) an exploratory sample (70% of the total dataset) and (2) an holdout sample (30% of the total dataset). The exploratory sample was used to generate the classification trees and all the potential rules satisfying the pre-defined threshold values for support (S), confidence (C), and lift (L). The holdout sample was used to validate the classification trees (see Fig. 1) and to evaluate the statistical significance of the rules through hypothesis testing (see Fig. 2).

The tree growing process was applied to the data of the exploratory sample using different response variables to obtain different trees (see Fig. 1). As a result, for each response variable, the tree structure and the learned tree were obtained. The application of the tree structure to the holdout sample produced the holdout tree that was used for validation. To reduce the risk that results were over-fitted to the sample, at each node of the holdout tree, the assignment of the class was compared with the assignment performed in the learned tree. As a result, only nodes with the same class in both the learned and the holdout trees were validated and the validated tree was obtained.

Similarly, the holdout sample was used to test the statistical significance of the rules discovered with the exploratory data by hypothesis testing (see Fig. 2). We used the binomial test to verify the statistical significance of deviations of the rule support measure from the theoretically expected value when antecedent and consequent items are independent, using the rules obtained with the exploratory data and applying the test to the holdout data. In the hypothesis testing, the p-value is compared to some predefined threshold α. If $p \leq \alpha$, the null hypothesis is rejected and the discovery is called significant at level α. Parameter α defines the probability of committing type I error, i.e., accepting a spurious rule. The problem of decide suitable thresholds of the level of significance of the tests is not trivial in the data mining where numerous patterns are tested. For example, if we use the threshold $\alpha = 0.10$, there is a 10% chance that a spurious rule passes the significance

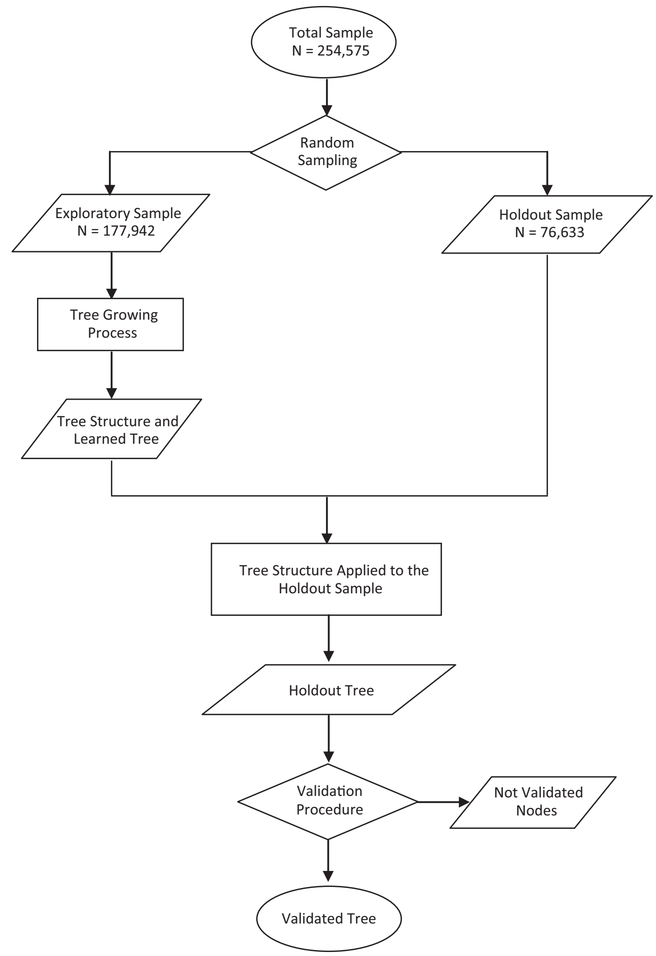


Fig. 1. Classification trees: learning and validation process.

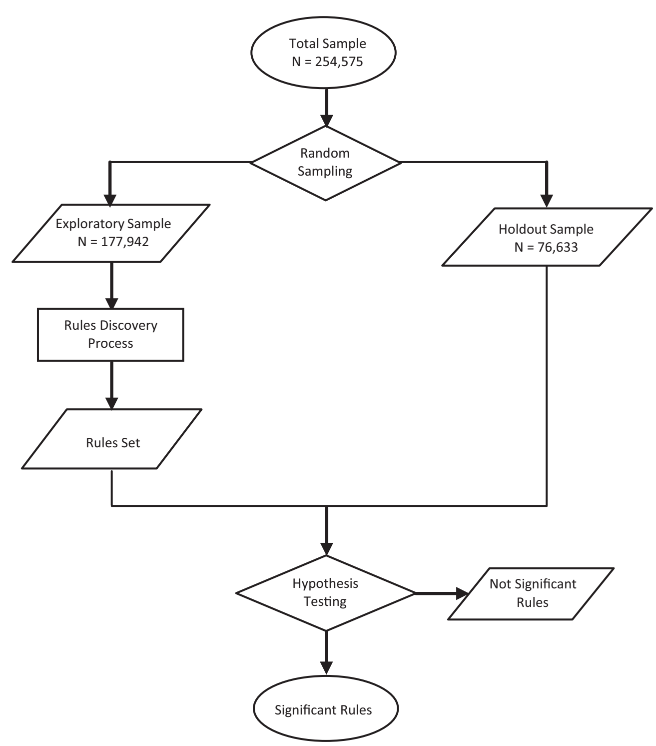


Fig. 2. Rules discovery process.

Table 1
Search space dimension for association rules discovery.

Consequent items	M_y	M_x	K	s
Severity items	2	71	4	2.06×10^6
Crash type items	12	59	4	5.82×10^6
Involved vehicles items	7	52	4	2.06×10^6
Alignment items	4	48	4	8.52×10^5

test. If we test 10,000 rules, it is likely that we will find 1000 spurious rules. This so-called multiple testing problem is inherent in the knowledge discovery, where we perform an exhaustive search over all possible patterns. As a solution, the more patterns we test, the stricter bounds for the significance we should use. Thus, we corrected the type I error level by the Bonferroni adjustment (Abdi, 2007; Dunn, 1961; Shaffer, 1995), where the desired significance level α is divided by the number of tests s, considering the dimension of the search-space in the pattern discovery process. The search space dimension s is defined as the number of all possible combinations of binary items under constraints of an upper level for rule complexity:

$$s = M_y \cdot \sum_{i=1}^{K} \binom{M_x}{i} \tag{9}$$

where M_y is the number of binary items candidate as consequent part of a rule, M_x is the number of binary items candidate as antecedent part of a rule and K is the rule complexity upper bound (it means the maximum number of antecedent items). The s value can be interpreted as the number of possible rules (and possible tests) which can be generated by the discovery pattern process (see Table 1). The significance level of the binomial test, adjusted by s, is equal to $0.1/s$.

3. Data

The study data were the micro data provided by the Italian National Institute of Statistics (ISTAT) for the Italian road network, relative to the three-year period 2006–2008. These data include, for each crash, in ASCII format, 159 fields containing all the information of the ISTAT CTT/INC report. The original data base consisted of 687,958 crashes. Only the 254,575 crashes where at least one PTW was involved (37% of the total crashes) were extracted from the data base and were used for the subsequent analyses.

The ISTAT data were relative only to the injury and fatal crashes. Furthermore, only one injury level was considered. As shown in Table 2, the data base was rearranged and 19 categorical variables were selected: (1) area, (2) road type, (3) lighting, (4) weather, (5) pavement, (6) driver PTW gender, (7) driver PTW age, (8) driver PTW outcome, (9) Vehicle B driver gender, (10) Vehicle B driver age, (11) Vehicle B driver outcome, (12) pedestrian gender, (13) pedestrian age, (14) pedestrian outcome, (15) alignment, (16) involved vehicles, (17) PTW type, (18) crash type, and (19) severity.

The road type classification was an administrative classification based on the Italian Highway Code requirements. Noteworthy is that the sections of national and provincial highways crossing urban communities with population less than 10,000 inhabitants were classified as urban national and urban provincial. PTW types were: (1) light weight PTWs (mopeds and scooters) with a cylinder capacity less than or equal to 50 cm^3; (2) scooters and light weight motorcycles with a cylinder capacity greater than 50 cm^3 and less than or equal to 250 cm^3; and (3) heavy scooters and motorcycles with a cylinder capacity greater than 250 cm^3.

The data used in the analyses have some limitations. First, these data were based on police reports and the main scope of the police officers was to investigate the drivers' faults, whereas more focus on the interaction between the road users and the road environment would be helpful for a better analysis of the crash patterns. In addition, some relevant information were not reported in the database, such as the availability of street lighting during nighttime at the crash location, the location of the pedestrian crashes with respect to the pedestrian crosswalks, the geometric features of the crash scene, and the drivers' physical and psychological conditions.

4. Results

In the classification trees, the following response variables were assessed: (1) severity (Fig. 3), which is defined as the level of injury sustained by the most severely injured person involved in the crash and measures the crash cost; (2) crash type (Fig. 4), which explains the manner of collision; (3) involved vehicles (Fig. 5), which identifies the vehicles involved in the crash; and (4) alignment (Fig. 6), which identifies the crash location along the road network. The classification trees were developed in sequential order, removing the variable severity in the second tree, the variables severity and crash type in the third tree, and the variables severity, crash type, and involved vehicle in the fourth tree.

Association rules with severity (Table 3), crash type (Tables 4 and 5), involved vehicle (Table 6), and alignment (Table 7) as consequent were extracted from the set of the rules identified with the a priori algorithm. The rules with the same consequent were then ordered according the decreasing value of the lift of the two-items rules. For each two-items rule, the three item rules having the same two items of the parent rule were ordered again according the decreasing value of the lift, and so on. Furthermore, rules identified with the a priori algorithm and the classification trees (Table 8) were compared.

Tree growing produced with the exploratory sample produced 54 nodes. Of these, 2 nodes were not validated and the validated trees consisted of 52 nodes. The application of the a priori algorithm to the exploratory sample discovered 172 rules. The statistical tests identified 54 not significant rules and 118 significant rules. In addition, the classification trees identified 28 validated rules.

4.1. Severity

Severity was classified in two categories really overbalanced: fatal crashes (1.8%) and injury crashes (98.2%). The classification tree produced 12 validated terminal nodes (Fig. 3) and the a priori algorithm identified 20 significant rules (Table 3). The predictors of the classification tree were the variables road type, crash type, PTW type, lighting, and involved vehicles. The same variables were also identified as antecedents in the association rules. Furthermore, the a priori algorithm identified as significant the variables alignment, pavement, and PTW driver gender.

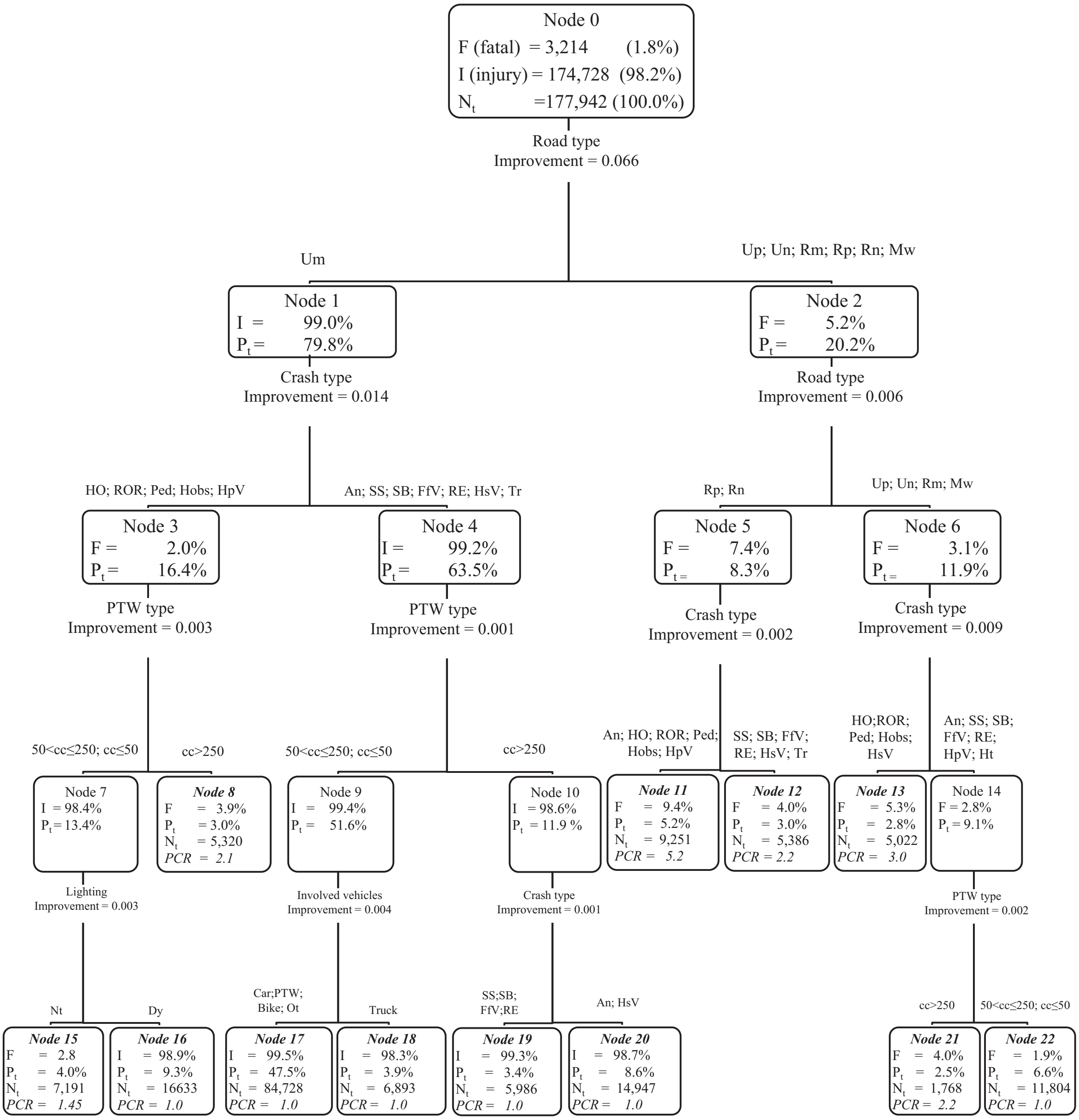
The primary split of the classification tree was the road type. Interestingly, the classification tree keep the urban municipal roads (node 1) away from all the other road types (node 2). Fatal crashes proportion in the node 2 was substantially greater than in the node 1 (5.2% vs. 1.0%). Node 1 is constituted by the urban municipal roads and represents 80% of the total crashes, whereas node 2 is constituted by the urban national and provincial roads and all the rural roads. A further split of the node 2 was based again on the road type, and the road types with the greater severity were the rural provincial and national roads. Consistently with these results, the a priori algorithm identified 2-item rules with fatal severity as consequent and the following road types as antecedent: rural provincial roads (rule 1, L = 4.39), rural national roads (rule 5, L = 3.80), and rural municipal roads (rule 18, L = 2.14). In addition, also rural area (which includes the previous road types and the motorways) was associated with fatal severity (rule 7, L = 3.46).

Table 2
Descriptive statistics of crash data.

Variable	Code	Count	Percent	Variable	Code	Count	Percent
Area				Vehicle B driver outcome			
Urban	U	224,576	88.2	Uninjured	Unj	184,746	88.6
Rural	R	29,999	11.8	Injured	I	23,646	11.3
Road type				Dead	Dd	190	0.1
Urban municipal	Um	203,260	79.8	Pedestrian gender			
Urban provincial	Up	11,585	4.6	Male	Ma	4985	44.9
Urban national	Un	9731	3.8	Female	Fe	6112	55.1
Rural municipal	Rm	4801	1.9	Missing		1	0.0
Rural provincial	Rp	11,532	4.5	Pedestrian age			
Rural national	Rn	9438	3.7	0–18	0–18	554	5.0
Motorway	Mw	4228	1.7	19–25	19–25	342	3.1
Lighting				26–45	26–45	1162	10.5
Day	Dy	195,710	76.9	46–65	46–65	1244	11.2
Night	Nt	58,865	23.1	>65	>65	1274	11.5
Weather				Missing	Missing	6522	58.8
Clear	Cl	223,509	87.8	Pedestrian outcome			
Rainy	Rn	15,246	6.0	Uninjured	Unj	1	0.0
Foggy	Fg	926	0.4	Injured	I	10,855	97.8
Snow	Sw	131	0.1	Dead	Dd	242	2.2
Other	Ot	14,763	5.8	Alignment			
Pavement				Tangent	Tan	103,775	40.8
Dry	D	224,636	88.2	Curve	Cu	18,231	7.2
Wet	W	24,819	9.7	Intersection	Int	129,860	51.0
Slippery	Sl	4723	1.9	Other	Ot	2709	1.1
Frozen	Fr	297	0.1	Involved vehicles			
Snowy	Swy	100	0.0	PTW-car	Car	176,251	69.2
PTW driver gender				PTW single vehicle	SV	33,130	13.0
Male	Ma	216,290	85.0	PTW-pedestrian	Ped	11,098	4.4
Female	Fe	38,285	15.0	PTW-truck	Truck	16,566	6.5
PTW driver age				PTW-ptw	PTW	11,277	4.4
0–18	0–18	24,880	9.8	PTW-bicycle	Bike	4095	1.6
19–25	19–25	38,678	15.2	PTW-other vehicles	Ot	2158	0.8
26–45	26–45	106,481	41.8	PTW type			
46–65	46–65	45,312	17.8	Light weight PTW	$cc \leq 50$	43,707	17.2
>65	>65	13,623	5.4	Scooter or light weight motorcycle	$50 < cc \leq 250$	63,190	24.8
Missing	Missing	25,601	10.1	Heavy scooter or motorcycle	$cc > 250$	53,355	21.0
PTW driver outcome				Missing	Missing	94,323	37.1
Uninjured	Unj	21,935	8.6	Crash type			
Injured	I	228,678	89.8	Angle	An	107,140	42.1
Dead	Dd	3,962	1.6	Falling from the vehicle	FfV	8547	3.4
Vehicle B driver gender				Head-on	HO	18,171	7.1
Male	Ma	151,822	72.8	Hit obstacle in carriageway	Hobs	6112	2.4
Female	Fe	55,691	26.7	Hit parked vehicle	HpV	1760	0.7
Missing	Missing	1069	0.5	Hit pedestrian	Ped	11,098	4.4
Vehicle B driver age				Hit stopped vehicle	HsV	8122	3.2
0–18	0–18	7433	3.6	Hit train	Tr	5	0.0
19–25	19–25	28,804	13.8	Rear-end	RE	27,843	10.9
26–45	26–45	86,992	41.7	Run-off-the-road	ROR	17,214	6.8
46–65	46–65	47,603	22.8	Sideswipe	SS	47,306	18.6
>65	>65	16,159	7.7	Sudden braking	SB	1257	0.5
Missing	Missing	21,598	10.4	Severity			
				Fatal	F	4626	1.8
				Injury	I	249,949	98.2

Rural provincial and national roads were split in relation to the crash type, with head-on, run-off-the-road, hit pedestrian, hit obstacle in carriageway, and hit parked vehicle crash types belonging to the node with the greater severity (node 3, fatal crashes = 9.1%). Similarly, the other road types were split in relation to the crash type with the difference that angle and hit parked vehicle crashes belonged to the nodes with the smaller severity (node 4, fatal crashes = 0.8%; node 14, fatal crashes = 2.8%), while hit stopped vehicle crashes belonged to the nodes with the greater severity (node 5, fatal crashes = 2.0%; node 13, fatal crashes = 5.3%). The association between head-on crashes and fatal severity was also a significant rule (rule 19, L = 2.21). Furthermore, the combination of curve alignment and run-off-the-road crash type was strongly associated with fatal severity (rule 14, L = 3.42).

In urban municipal roads, the third split was related to the PTW type and the class heavy scooter or motorcycle ($cc > 250$) belonged to the nodes with the greater proportion of fatal crashes (nodes 8 and 10). In urban provincial and national roads, rural municipal roads, and motorways, the fourth split was related to the PTW type. As in the other road types, the class heavy scooter or motorcycle belonged to the node with the greater severity (node 21). The association between heavy scooter or motorcycle and fatal severity was confirmed by the significant rule 20 (L = 1.99). In addition, also the combination of curve alignment and PTW type heavy scooter or motorcycle was associated with fatal severity (rule 13, L = 4.05). In urban municipal roads, the fourth split was related to lighting and involved vehicles. Truck involvement and nighttime belonged to the nodes with the greater proportion of fatal crashes (nodes 15 and 18). The combination truck and rural area was significantly associated with fatal severity (rule 8, L = 5.56). Furthermore, fatal severity was significantly associated with the combinations (a) rural provincial roads and nighttime (rule 2, L = 5.73), (b) rural national roads and nighttime (rule 6, L = 4.63), and (c) rural area and nighttime (rule 10, L = 4.38).



An = Angle; cc>250 = Heavy scooter or motorcycle; 50< cc≤250 = Scooter or light weight motorcycle; cc≤50 = Light weight PTW; Dy = Day; F = Fatal; FfV = Falling from the vehicle; HO = Head-on; Hobs = Hit obstacle in carriageway; HpV = Hit parked vehicle; HsV = Hit stopped vehicle; I = Injury; RE = Rear-end; Rm = Rural municipal; Rn = Rural national; ROR = Run -off-the-road; Rp = Rural provincial; SB = Sudden braking; Tr = Hit train; Um = Urban municipal; Un = Urban national; Up = Urban provincial .

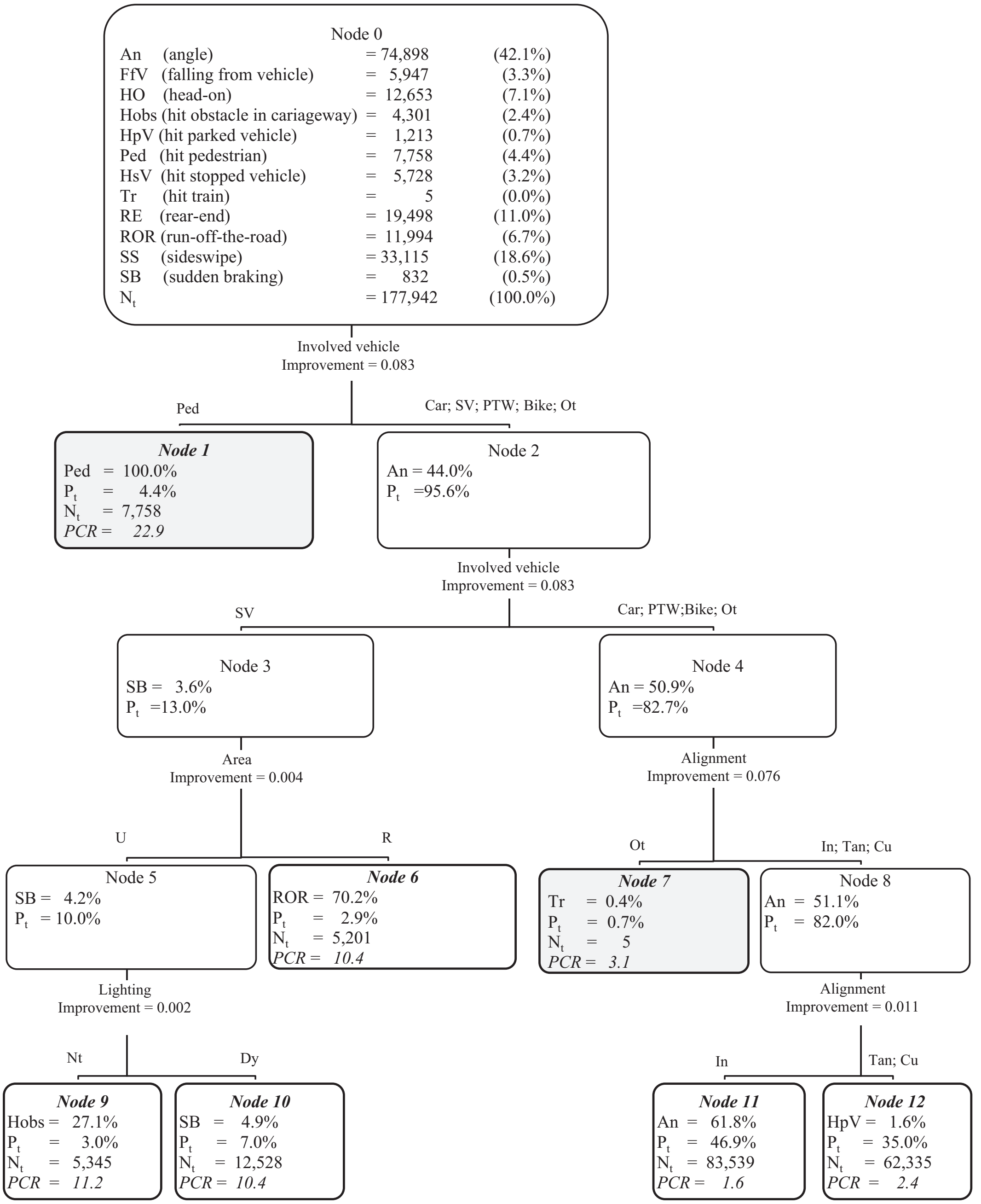
Fig. 3. Classification tree with severity as response variable.

As far as the alignment is concerned, the curve alignment was significantly associated with fatal severity both as single crash characteristic (rule 12, L = 2.53) and in combination with other patterns such as rural provincial roads (rule 3, L = 5.31), rural area (rule 9, L = 4.53), PTW type heavy scooter or motorcycle (rule 13, L = 4.05), crash type run-off-the-road (rule 14, L = 3.42), single vehicle involvement (rule 15, L = 3.04), dry pavement (rule 16, L = 2.74), and PTW driver gender male (rule 16, L = 2.74).

4.2. Crash type

Crash type was classified in twelve categories, with the largest being angle (42.1%), sideswipe (18.5%), rear-end (10.9%), head-on (7.1%), run-off-the-road (6.8%), and hit pedestrian (4.4%). The classification tree produced 5 validated terminal nodes (Fig. 4) and the a priori algorithm identified 76 significant rules (Tables 4 and 5).

Single vehicle crashes in rural area showed greater propensity towards run-off-the-road (node 6 and rule 79), whereas in urban area they showed propensity towards hit obstacle in nighttime (node 9) and towards sudden braking in daytime (node 10). In multi vehicle crashes, the crash type distribution was significantly affected by the road alignment. At intersections, a greater increase in the proportion of angle crashes was observed (node 11). In tangents and curves, the greater change was observed in hit parked vehicle crashes (node 12), which did not represent a significant proportion of the total crashes (1.6%).



An = Angle; Bike = PTW -bicycle; Car = PTW -car; Cu = Curve; Dy = Day; FfV = Falling from the vehicle; HO = Head -on; Hobs = Hit obstacle in carriageway; HpV = Hit parked vehicle; HsV = Hit stopped vehicle ; Int = Intersection; Mw = Motorway; Ot = Other; Ped = Hit pedestrian; Ped = PTW - pedestrian; R = Rural; RE = Rear -end; ROR = Run-off-the-road; SB = Sudden braking; SS = Sideswipe; SV = PTW sv; Sw = Snow; Swy = Snowy; Tan = Tangent; Tr = Hit train; Truck = PTW-truck; U = Urban.

Fig. 4. Classification tree with crash type as response variable.

Furthermore, the a priori algorithm identified several rules showing a significant effect on the crash type of different crash characteristics and their combination. The most critical variables were involved vehicles, road type, area, alignment, and lighting. Run-off-the-road and hit pedestrian were the crash types giving rise to more significant association rules. Run-off-the-road crashes were associated with numerous combinations of curve alignment and other characteristics, such as road type (rural national, provincial, and municipal), rural area, slippery pavement, nighttime, PTW type (heavy scooter or motorcycle), and PTW driver age (older than 65 and between 26 and 45). Hit pedestrian crashes were associated with nighttime, rainy weather, tangent alignment, and their different combinations with other characteristics, such as wet pavement, urban municipal road type, and scooter or light weight motorcycle PTW type. Noteworthy is the association between crashes in curves with slippery pavement and falling from the vehicle crash type (rule 22, $L = 8.67$).

4.3. Involved vehicles

Involved vehicles were classified in seven categories, with the largest being PTW-car (69.2%) and PTW single vehicle (13.0%). The classification tree produced 7 validated terminal nodes (Fig. 5) and the a priori algorithm identified 12 significant rules (Table 6).

The first split of the classification tree was based on the alignment. At the intersections, the involved vehicles were related to

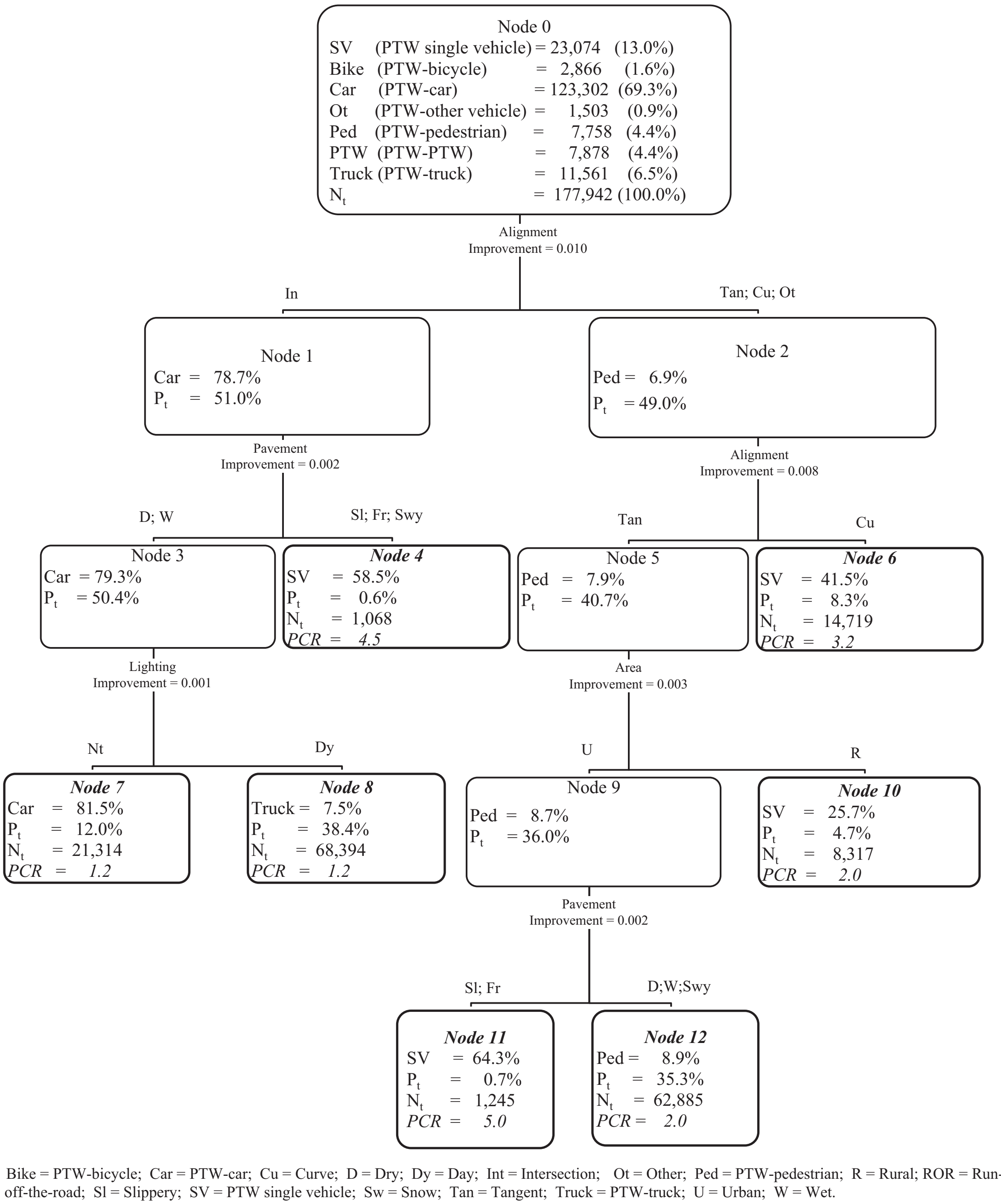


Fig. 5. Classification tree with involved vehicles as response variable.

the pavement and lighting conditions: (a) in dry or wet pavement, the most overrepresented involved vehicle was the car, in nighttime (node 7), or the truck, in daytime (node 8); (b) in slippery, frozen, or snowy pavement, a greater propensity towards single vehicle was observed (node 4). In the tangents, the involved vehicles were related to the area and to the pavement conditions: (a) in urban area, single vehicle was the most frequent scenario in slippery or frozen pavement (node 11), whereas pedestrian was over involved in dry, wet, or snowy pavement (node 12); (b) in rural area, a greater propensity towards single vehicle was observed (node 10 and rule 105). In the curves, a greater propensity towards single vehicle was observed (node 6 and rule 99) and this propensity further increased in rural area (rule 104, L = 4.14). In addition, the a priori algorithm showed other significant associations such as between slippery pavement and single vehicle (rule 97, L = 5.19) and between rainy weather and pedestrian involvement (rule 102, L = 5.19).

4.4. Alignment

Alignment was classified in four categories: intersection (51.0%), tangent (40.8%), curve (7.2%), and other (1.1%). The classification tree produced 5 validated terminal nodes (Fig. 6) and the a priori algorithm identified 8 significant rules (Table 7).

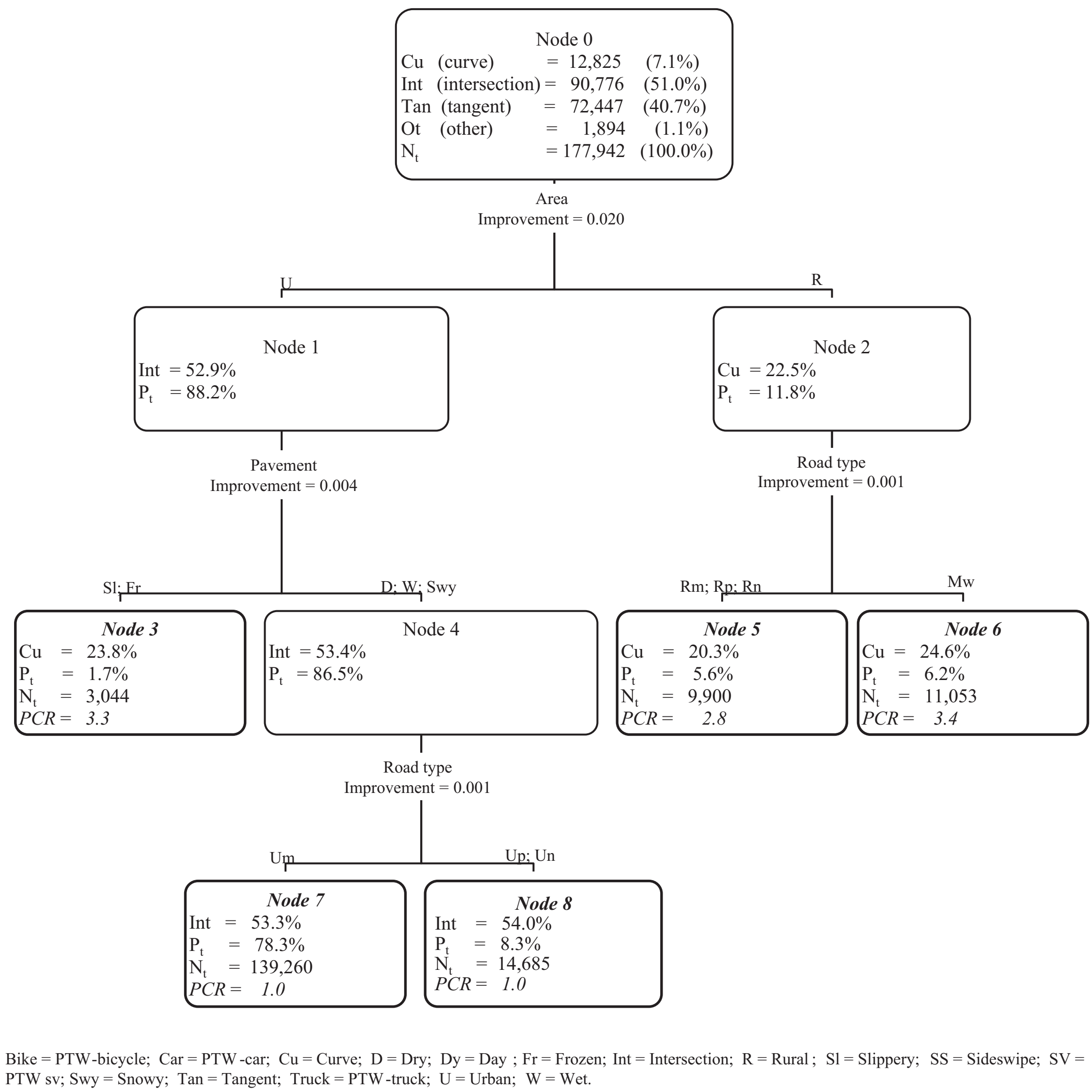


Fig. 6. Classification tree with alignment as response variable.

Table 3
Rules with fatal severity as consequent.

Rule ID	Association rule – Antecedent	Association rule – Consequent	*S*%	*C*%	*L*	*p*-Value
1	Road type = Rp	Severity = F	0.4	7.9	4.39	$<1\times10^{-10}$
2	Road type = Rp & lighting = Nt	Severity = F	0.1	10.4	5.73	$<1\times10^{-10}$
3	Road type = Rp & alignment = Cu	Severity = F	0.1	9.6	5.31	$<1\times10^{-10}$
4	Road type = Rp & alignment = Tan	Severity = F	0.1	8.7	4.82	$<1\times10^{-10}$
5	Road type = Rn	Severity = F	0.3	6.9	3.80	$<1\times10^{-10}$
6	Road type = Rn & lighting = Nt	Severity = F	0.1	8.4	4.63	$<1\times10^{-10}$
7	Area = R	Severity = F	0.7	6.3	3.46	$<1\times10^{-10}$
8	Area = R & involved vehicles = Truck	Severity = F	0.1	10.2	5.66	$<1\times10^{-10}$
9	Area = R & alignment = Cu	Severity = F	0.2	8.2	4.53	$<1\times10^{-10}$
10	Area = R & lighting = Nt	Severity = F	0.2	7.9	4.38	$<1\times10^{-10}$
11	Area = R & involved vehicles = SV	Severity = F	0.2	6.7	3.73	$<1\times10^{-10}$
12	Alignment = Cu	Severity = F	0.3	4.6	2.53	$<1\times10^{-10}$
13	Alignment = Cu & PTW type = *cc* > 250	Severity = F	0.2	7.3	4.05	$<1\times10^{-10}$
14	Alignment = Cu & crash type = ROR	Severity = F	0.1	6.2	3.42	$<1\times10^{-10}$
15	Alignment = Cu & involved vehicles = SV	Severity = F	0.2	5.5	3.04	$<1\times10^{-10}$
16	Alignment = Cu & pavement = D	Severity = F	0.3	5.0	2.74	$<1\times10^{-10}$
17	Alignment = Cu & PTW driver gender = Ma	Severity = F	0.3	4.9	2.71	$<1\times10^{-10}$
18	Road type = Rm	Severity = F	0.1	3.9	2.14	$<1\times10^{-10}$
19	Crash type = HO	Severity = F	0.3	4.0	2.21	8.87×10^{-9}
20	PTW type = *cc* > 250	Severity = F	0.8	3.6	1.99	$<1\times10^{-10}$

$\alpha_{crit} = 0.10$; $\alpha_{crit}/s = 4.85\times10^{-8}$. *cc* > 250 = heavy scooter or motorcycle; Cu = curve; D = dry; HO = head-on; Ma = male; Nt = night; R = rural; Rm = rural municipal; Rp = rural provincial; SV = PTW single vehicle; Tan = tangent; Truck = PTW-truck.

Table 4
Rules with crash type as consequent (part A).

Rule ID	Association rule – Antecedent	Consequent	S%	C%	L	p-Value%
21	Pavement = Sl	Crash type = FfV	0.5	25.1	7.50	$<1 \times 10^{-10}$
22	Pavement = Sl & alignment = Cu	Crash type = FfV	0.1	29.0	8.67	$<1 \times 10^{-10}$
24	Alignment = Cu	Crash type = FfV	0.7	9.4	2.82	$<1 \times 10^{-10}$
25	Alignment = Cu & PTW driver gender = Fe	Crash type = FfV	0.1	14.6	4.36	$<1 \times 10^{-10}$
26	Alignment = Cu & road type = Um	Crash type = FfV	0.4	11.0	3.30	$<1 \times 10^{-10}$
27	Alignment = Cu & area = U	Crash type = FfV	0.5	10.6	3.17	$<1 \times 10^{-10}$
28	Alignment = Cu & involved vehicles = Car	Crash type = HO	0.7	22.3	3.13	$<1 \times 10^{-10}$
29	Alignment = Cu & involved vehicles = truck	Crash type = HO	0.1	21.0	2.95	$<1 \times 10^{-10}$
30	Alignment = Cu & PTW driver age = 0–18	Crash type = HO	0.0	14.9	2.10	4.18×10^{-09}
31	Pavement = Sl	Crash type = Hobs	0.3	15.1	6.24	$<1 \times 10^{-10}$
32	Alignment = Cu	Crash type = Hobs	0.5	7.3	3.03	$<1 \times 10^{-10}$
33	Alignment = Cu & area = U	Crash type = Hobs	0.4	8.1	3.34	$<1 \times 10^{-10}$
34	Alignment = Tan & lighting = Nt	Crash type = Hobs	0.5	5.6	2.32	$<1 \times 10^{-10}$
35	Alignment = Tan & lighting = Nt	Crash type = HpV	0.2	2.2	3.28	$<1 \times 10^{-10}$
36	Alignment = Tan & involved vehicles = Car	Crash type = HpV	0.5	1.9	2.78	$<1 \times 10^{-10}$
37	Alignment = Tan & road type = Um	Crash type = HpV	0.5	1.4	2.03	$<1 \times 10^{-10}$
38	Lighting = Nt	Crash type = Ped	1.3	5.5	1.26	2.41×10^{-09}
39	Lighting = Nt & weather = Rn	Crash type = Ped	0.3	15.9	3.65	$<1 \times 10^{-10}$
40	Lighting = Nt & weather = Rn & alignment = Tan	Crash type = Ped	0.2	27.0	6.20	$<1 \times 10^{-10}$
41	Lighting = Nt & weather = Rn & alignment = Tan & road type = Um	Crash type = Ped	0.2	28.8	6.61	$<1 \times 10^{-10}$
42	Lighting = Nt & weather = Rn & PTW type = $50 < cc \le 250$	Crash type = Ped	0.1	25.8	5.92	$<1 \times 10^{-10}$
43	Lighting = Nt & weather = Rn & PTW type = $50 < cc \le 250$ & alignment = Tan	Crash type = Ped	0.1	39.9	9.16	$<1 \times 10^{-10}$
44	Lighting = Nt & weather = Rn & Road type = Um	Crash type = Ped	0.3	17.3	3.96	$<1 \times 10^{-10}$
45	Lighting = Nt & pavement = W	Crash type = Ped	0.4	12.5	2.88	$<1 \times 10^{-10}$
46	Lighting = Nt & pavement = W & alignment = Tan	Crash type = Ped	0.3	21.7	4.97	$<1 \times 10^{-10}$
47	Lighting = Nt & weather = Rn & PTW type = $50 < cc \le 250$	Crash type = Ped	0.1	25.8	5.92	$<1 \times 10^{-10}$
48	Lighting = Nt & pavement = W & PTW type = $50 < cc \le 250$ & alignment = Tan	Crash type = Ped	0.1	30.5	7.01	$<1 \times 10^{-10}$
49	Lighting = Nt & pavement = W & road type = Um & alignment = Tan	Crash type = Ped	0.3	22.9	5.25	$<1 \times 10^{-10}$
50	Lighting = Nt & alignment = Tan	Crash type = Ped	1.0	10.4	2.38	$<1 \times 10^{-10}$
51	Weather = Rn	Crash type = Ped	0.6	9.6	2.21	$<1 \times 10^{-10}$
52	Weather = Rn & alignment = Tan	Crash type = Ped	0.4	16.1	3.69	$<1 \times 10^{-10}$
53	Alignment = Tan	Crash type = Ped	3.2	7.9	1.80	$<1 \times 10^{-10}$
54	Alignment = Tan & pavement = W	Crash type = Ped	0.6	13.1	2.99	$<1 \times 10^{-10}$
55	Alignment = Tan & road type = Um	Crash type = Ped	3.0	9.0	2.07	$<1 \times 10^{-10}$
56	Alignment = Tan & area = U	Crash type = Ped	3.2	8.7	2.00	$<1 \times 10^{-10}$
57	Involved vehicles = truck & alignment = Tan	Crash type = HsV	0.3	10.8	3.34	$<1 \times 10^{-10}$
58	Involved vehicles = truck & alignment = Tan & road type = Um	Crash type = HsV	0.2	12.1	3.75	$<1 \times 10^{-10}$
59	Involved vehicles = truck & alignment = Tan & road type = Um & PTW type = $50 < cc \le 250$	Crash type = HsV	0.1	13.4	4.17	$<1 \times 10^{-10}$
60	Alignment = Tan & involved vehicles = car	Crash type = HsV	1.8	7.0	2.16	$<1 \times 10^{-10}$
61	Alignment = Tan & involved vehicles = car & road type = Um	Crash type = HsV	1.6	7.6	2.37	$<1 \times 10^{-10}$
62	Alignment = Tan & road type = Um	Crash type = HsV	1.9	5.8	1.81	$<1 \times 10^{-10}$
63	Road type = Mw	Crash type = ROR	0.5	29.7	4.41	$<1 \times 10^{-10}$
64	Road type = Mw & alignment = Cu	Crash type = ROR	0.2	44.7	6.63	$<1 \times 10^{-10}$

$\alpha_{crit} = 0.10$; $\alpha_{crit}/s = 1.70 \times 10^{-8}$. Car = PTW-car; $50 < cc \le 250$ = scooter or light weight motorcycle; $cc > 250$ = heavy scooter or motorcycle; $cc \le 50$ = light weight PTW; Cu = curve; Fe = female; FfV = falling from the vehicle; HO = head-on; Hobs = hit obstacle in carriageway; HpV = hit parked vehicle; HsV = hit stopped vehicle; Mw = motorway; Nt = night; Ped = hit pedestrian; PTW = PTW-PTW; R = rural; RE = rear-end; Rm = rural municipal; Rn = rainy; Rn = rural national; ROR = run-off-the-road; Rp = rural provincial; SB = sudden braking; Sl = slippery; SS = sideswipe; SV = PTW single vehicle; Sw = snow; Swy = snowy; Tan = tangent; Tr = hit train; Truck = PTW-truck; Um = urban municipal; Un = urban national; Up = urban provincial; W = wet.

The significant variables were area, road type, pavement, and weather. In urban area, more propensity for crashes at intersections was observed (node 1). With slippery or frozen pavement, crash at curves were the most overrepresented (node 3). In the other pavement conditions, crash at intersections were the most overrepresented (node 4). In rural area, there was more propensity for crashes at curves (node 2 and rule 112). The greater proportion of crashes at curves was observed in the motorways (node 6 and rule 109). Interestingly, the combination rural area, slippery pavement, and clear weather exhibited the greater association with crashes at curves (rule 114, $L = 7.20$).

5. Discussion

Analysis results showed that PTW crashes are strongly sensitive to several combinations of road, environment, and drivers attributes.

As far as the road is concerned, it is stressed that the pavement conditions play a very significant role. Contrary to cars and other four-wheeled vehicles, a PTW has only two points of contact with the surface and, therefore, the consistency of grip of the tyres on the surface is critical for the stability of the PTW (ACEM, 2006). To negotiate a curve in the road motorcyclists lean over at an angle whose acuteness is related to speed and to radius of the curve, and any change in grip of the tyres on the surface can destabilize the machine. Any deviation from a consistent surface can seriously impair the grip of the motorcycle on the road. A sudden change in surface level rapidly charges and discharges the shock absorbers, thus reducing the grip of the front wheel on the road surface. Our results are consistent with these theoretical considerations. Indeed, slippery pavement were significantly associated with crashes at curves, that is, the most unfavourable pavement condition was associated with the alignment characteristic most demanding in terms of negotiation. This result was found by both the classification

Table 5
Rules with crash type as consequent (part B).

Rule ID	Association rule: Antecedent	Consequent	*S*%	*C*%	*L*	*p*-Value%
65	Alignment = Cu	Crash type = ROR	1.9	25.6	3.80	$<1\times10^{-10}$
66	Alignment = Cu & road type = Rn	Crash type = ROR	0.3	41.4	6.15	$<1\times10^{-10}$
67	Alignment = Cu & road type = Rn & lighting = Nt	Crash type = ROR	0.1	49.0	7.26	$<1\times10^{-10}$
68	Alignment = Cu & area = R	Crash type = ROR	1.1	39.8	5.91	$<1\times10^{-10}$
69	Alignment = Cu & area = R & pavement = Sl	Crash type = ROR	0.1	49.8	7.38	$<1\times10^{-10}$
70	Alignment = Cu & area = R & lighting = Nt	Crash type = ROR	0.2	48.3	7.17	$<1\times10^{-10}$
71	Alignment = Cu & road type = Rp	Crash type = ROR	0.4	38.2	5.67	$<1\times10^{-10}$
72	Alignment = Cu & road type = Rp & lighting = Nt	Crash type = ROR	0.1	49.8	7.39	$<1\times10^{-10}$
73	Alignment = Cu & road type = Rm	Crash type = ROR	0.1	34.7	5.14	$<1\times10^{-10}$
74	Alignment = Cu & pavement = Sl	Crash type = ROR	0.2	32.2	4.78	$<1\times10^{-10}$
75	Alignment = Cu & PTW type = *cc* > 250	Crash type = ROR	0.7	32.4	4.80	$<1\times10^{-10}$
76	Alignment = Cu & PTW driver age >65	Crash type = ROR	0.1	28.9	4.29	$<1\times10^{-10}$
77	Alignment = Cu & lighting = Nt	Crash type = ROR	0.5	30.5	4.52	$<1\times10^{-10}$
78	Alignment = Cu & PTW driver age = 26–45	Crash type = ROR	0.9	28.2	4.18	$<1\times10^{-10}$
79	Area = R	Crash type = ROR	2.1	17.4	2.59	$<1\times10^{-10}$
80	Area = R & lighting = Nt	Crash type = ROR	0.5	20.8	3.09	$<1\times10^{-10}$
81	Road type = Rn	Crash type = ROR	0.6	15.7	2.34	$<1\times10^{-10}$
82	Road type = Rn & lighting = Nt	Crash type = ROR	0.1	17.4	2.59	$<1\times10^{-10}$
83	Road type = Rn & alignment = Tan	Crash type = ROR	0.2	16.2	2.41	$<1\times10^{-10}$
84	Road type = Rp	Crash type = ROR	0.7	15.4	2.28	$<1\times10^{-10}$
85	Road type = Rp & lighting = Nt	Crash type = ROR	0.2	20.1	2.97	$<1\times10^{-10}$
86	Pavement = Sl	Crash type = ROR	0.5	25.6	3.79	$<1\times10^{-10}$
87	Road type = Rm	Crash type = ROR	0.3	14.8	2.20	1.21×10^{-08}
88	Road type = Rm & alignment = Tan	Crash type = ROR	0.1	15.5	2.29	1.50×10^{-08}
89	Road type = Mw	Crash type = RE	0.5	30.9	2.82	$<1\times10^{-10}$
90	Road type = Mw & alignment = Tan	Crash type = RE	0.4	37.7	3.44	$<1\times10^{-10}$
91	Road type = Mw & alignment = Tan & lighting = Nt	Crash type = RE	0.1	41.0	3.74	$<1\times10^{-10}$
92	Road type = Mw & lighting = Nt	Crash type = RE	0.1	32.8	3.00	$<1\times10^{-10}$
93	Involved vehicles = PTW-PTW & alignment = Tan	Crash type = RE	0.4	23.5	2.14	$<1\times10^{-10}$
94	Area = R & alignment = Tan	Crash type = RE	1.1	24.4	2.22	$<1\times10^{-10}$
95	Alignment = Tan & involved vehicles = Truck	Crash type = RE	0.5	21.1	1.93	2.31×10^{-09}
96	Alignment = Tan & involved vehicles = Car	Crash type = RE	5.0	19.4	1.77	$<1\times10^{-10}$

$\alpha_{crit}=0.10$; $\alpha_{crit}/s=1.70\times10^{-8}$. Car = PTW-car; $50<cc\leq250$ = scooter or light weight motorcycle; $cc>250$ = heavy scooter or motorcycle; $cc\leq50$ = light weight PTW; Cu = curve; Fe = female; FfV = falling from the vehicle; HO = head-on; Hobs = hit obstacle in carriageway; HpV = hit parked vehicle; HsV = hit stopped vehicle; Mw = motorway; Nt = night; Ped = hit pedestrian; PTW = PTW-PTW; R = rural; RE = rear-end; Rm = rural municipal; Rn = rainy; Rn = rural national; ROR = run-off-the-road; Rp = rural provincial; SB = sudden braking; Sl = slippery; SS = sideswipe; SV = PTW single vehicle; Sw = snow; Swy = snowy; Tan = tangent; Tr = hit train; Truck = PTW-truck; Um = urban municipal; Un = urban national; Up = urban provincial; W = wet.

tree analysis and the rules discovery. In rural area, where generally the greater inconsistencies between drivers' speed behaviour and road alignment are observed (Montella et al., 2008), the number of crashes at curves with slippery pavements resulted 6.87 times the expected number of crashes if rural area and slippery pavements were independent of the curve alignment (rule 113). Clear weather, in rural area with slippery pavement, further increased the association with curve alignment (rule 114, $L=7.20$), probably because of the higher driving speed. Slippery pavements were also significantly associated with PTW single vehicle crashes, run-off-the-road crashes, and with fall from the vehicle crash type. At intersections, where multi vehicle crashes are generally prevalent, in slippery, frozen, or snowy pavement conditions, single vehicle crashes were 61% of the total crashes (node 4 in Fig. 5), clearly highlighting the association between pavement conditions and the manner of collision. In light of these results, engineering measures focused on the improvement of pavement friction and evenness, especially in rural area, have the potential to significantly reduce PTW crashes, other than reduce crashes involving other vehicle types (e.g., Montella, 2005; Lyon and Persaud, 2008).

The road alignment significantly affected the crash severity, the involved vehicles, and the crash type. Consistently with U.S. data

Table 6
Rules with involved vehicles as consequent.

Rule ID	Association rule: Antecedent	Consequent	*S*%	*C*%	*L*	*p*-Value%
97	Pavement = Sl	Involved vehicles = SV	1.2	67.3	5.19	$<1\times10^{-10}$
98	Vehicle B driver age = 0–18	Involved vehicles = PTW	0.4	15.1	3.41	$<1\times10^{-10}$
99	Alignment = Cu	Involved vehicles = SV	3.1	43.2	3.33	$<1\times10^{-10}$
100	Road type = Mw	Involved vehicles = SV	0.7	39.9	3.07	$<1\times10^{-10}$
101	Vehicle B driver age = 0–18	Involved vehicles = Bike	0.2	5.2	3.22	$<1\times10^{-10}$
102	Weather = Rn	Involved vehicles = Ped	0.6	9.6	2.21	$<1\times10^{-10}$
103	Area = R	Involved vehicles = SV	2.9	24.8	1.91	$<1\times10^{-10}$
104	Area = R & Alignment = Cu	Involved vehicles = SV	1.4	53.7	4.14	$<1\times10^{-10}$
105	Area = R & Alignment = Tan	Involved vehicles = SV	1.2	25.7	1.98	$<1\times10^{-10}$
106	Area = U & Alignment = Tan	Involved vehicles = Ped	3.2	8.7	2.00	$<1\times10^{-10}$
107	Area = U & Alignment = Tan & Weather = Rn	Involved vehicles = Ped	0.4	17.1	3.93	$<1\times10^{-10}$
108	Vehicle B driver gender = Ma	Involved vehicles = Truck	6.2	10.4	1.60	3.74×10^{-08}

$\alpha_{crit}=0.10$; $\alpha_{crit}/s=4.86\times10^{-8}$. Cu = curve; Ma = male; Mw = motorway; R = rural; Rn = rainy; Rn = rural national; Sl = slippery; SV = PTW single vehicle; Tan = tangent; Truck = PTW-truck.

Table 7
Rules with alignment as consequent.

Rule ID	Association rule – Antecedent	Association rule – Consequent	S%	C%	L	p-Value%
109	Road type = Mw	Alignment = Cu	0.5	29.4	4.08	$<1 \times 10^{-10}$
110	Pavement = Sl	Alignment = Cu	0.5	27.1	3.76	$<1 \times 10^{-10}$
111	Road type = Rp	Alignment = Cu	1.0	22.8	3.17	$<1 \times 10^{-10}$
112	Area = R	Alignment = Cu	2.7	22.6	3.13	$<1 \times 10^{-10}$
113	Area = R & pavement = Sl	Alignment = Cu	0.1	49.5	6.87	$<1 \times 10^{-10}$
114	Area = R & pavement = Sl & Weather = Cl	Alignment = Cu	0.1	51.9	7.20	$<1 \times 10^{-10}$
115	Road type = Rm	Alignment = Cu	0.4	20.5	2.84	$<1 \times 10^{-10}$
116	Road type = Rn	Alignment = Cu	0.7	20.2	2.80	$<1 \times 10^{-10}$

$\alpha_{crit} = 0.10$; $\alpha_{crit}/s = = 1.17 \times 10^{-7}$. Cl = clear; Cu = curve; Mw = motorway; Ot = other; R = rural; Rm = rural municipal; Rn = rural national; Rp = rural provincial; Sl = slippery; W = wet.

(Potts et al., 2008), crash at curves were associated with fatal severity. As an example, in the curves of the rural provincial roads, the number of PTW fatal crashes resulted 5.72 times the expected number of fatal crashes if curve alignment and rural provincial road type were independent of the crash severity (rule 3). Furthermore, the curve alignment was significantly associated with run-off-the-road crash type and single vehicle involvement. As a result, consistent highway alignments, that are alignments which ensure that successive elements are coordinated in such a way as to produce harmonious and homogeneous driver performances along the road (Cafiso et al., 2007; Montella, 2009) and which ensure the reduction of crashes at curves, might contribute to considerably reduce PTW fatalities and injuries.

As expected, crashes in rural area exhibited the higher crash severity. Association discovery showed that, in decreasing order of lift, the rural provincial, national, and municipal road types were significantly associated with fatal severity. Furthermore, classification trees showed that, in urban area, the provincial and

Table 8
Rules identified with the classification tree.

Rule ID	Association rule – Antecedent	Association rule – Consequent	S%	C%	L
T1_8	Road type = Up/Un/Rm/Rp/Rn/Mw & PTW type = $cc > 250$	Severity = F	0.1	3.9	2.13
T1_11	Road type = Rp/Rn & crash type = An/HO/ROR/Ped/Hobs/HpV	Severity = F	0.5	9.4	5.22
T1_12	Road type = Rp/Rn & crash type = SS/SB/FfV/RE/HsV/Tr	Severity = F	0.1	4.0	2.24
T1_13	Road type = Up/Un/Rm/Mw & crash type = HO;ROR; Ped; Hobs; HsV	Severity = F	0.2	5.3	2.95
T1_15	Road type = Up/Un/Rm/Rp/Rn/Mw & PTW type = $50 < cc \leq 250/cc \leq 50$ & lighting = Nt	Severity = F	0.1	2.8	1.52
T1_16	Road type = Up/Un/Rm/Rp/Rn/Mw & PTW type = $50 < cc \leq 250/cc \leq 50$ & lighting = Dy	Severity = I	9.2	98.9	1.01
T1_17	Road type = Um & crash type = An/SS/SB/FfV/RE/HsV/Tr & PTW type = $50 < cc \leq 250/cc \leq 50$ & involved vehicles = car/SV/PTW/bike	Severity = I	47.4	99.5	1.01
T1_18	Road type = Um & crash type = An/SS/SB/FfV/RE/HsV/Tr & PTW type = $50 < cc \leq 250/cc \leq 50$ & involved vehicles = Truck/Ot	Severity = I	3.9	98.3	1.00
T1_19	Road type = Um & crash type = An/SS/SB/FfV/RE/HsV/Tr & PTW type = $cc > 250$ & crash type = SS/SB/FfV	Severity = I	3.4	99.3	1.01
T1_20	Road type = Um & crash type = An/SS/SB/FfV/RE/HsV/Tr & PTW type = $cc > 250$ & crash type = An/RE/HsV	Severity = I	8.4	98.4	1.00
T1_21	Road type = Up/Un/Rm/Mw & crash type = An/SS/SB/FfV/RE/HpV/Tr & PTW type = $cc > 250$	Severity = F	0.1	4.0	2.22
T1_22	Road type = Up/Un/Rm/Mw & crash type = An/SS/SB/FfV/RE/HpV/Tr & PTW type = $50 < cc \leq 250/cc \leq 50$	Severity = F	0.1	1.9	1.03
T2_1	Involved vehicles = Ped	Crash type = Ped	4.4	100.0	22.94
T2_6	Involved vehicles = SV & area = R	Crash type = ROR	2.1	70.2	10.42
T2_7	Involved vehicles = Ot	Crash type = Tr	0.0	2.1	3.09
T2_9	Involved vehicles = SV & area = U & lighting = Nt	Crash type = Hobs	0.8	27.1	11.23
T2_10	Involved vehicles = SV & area = U & lighting = Dy	Crash type = SB	0.3	4.9	10.40
T2_11	Involved vehicles = Car/Truck/PTW/Bike/Ot & alignment = Int	Crash type = An	29.0	61.8	1.47
T2_12	Involved vehicles = Car/Truck/PTW/Bike/Ot & alignment = Tan/Cu	Crash type = HpV	0.6	1.6	2.37
T3_4	Alignment = Int & pavement = Sl/Sny	Involved vehicle = SV	0.4	58.5	4.51
T3_6	Alignment = Cu/Ot	Involved vehicle = SV	3.4	41.5	3.20
T3_7	Alignment = Int & pavement = D/W/Fr & lighting = Nt	Involved vehicle = Car	9.8	81.5	1.18
T3_8	Alignment = Int & pavement = D/W/Fr & lighting = Dy	Involved vehicle = Truck	2.9	7.5	1.16
T3_10	Alignment = Tan & Area = R	Involved vehicle = SV	1.2	25.7	1.98
T3_11	Alignment = Tan & area = U & pavement = Sl/Fr	Involved vehicle = SV	0.4	64.3	4.96
T3_12	Alignment = Tan & area = U & pavement = D/W/Sny	Involved vehicle = Ped	3.1	8.9	2.04
T4_3	Area = U & pavement = Sl/Fr	Alignment = Cu	0.4	23.8	3.30
T4_5	Area = R & roadtype = Rm/Rn	Alignment = Cu	1.1	20.3	2.82
T4_6	Area = R & roadtype = Rp/Mw	Alignment = Cu	1.5	24.6	3.41
T4_7	Area = U & Pavement = D/W/Sny & roadtype = Um	Alignment = Int	41.7	53.3	1.05
T4_8	Area = U & pavement = D/W/Sny & roadtype = Up/Un	Alignment = Int	4.5	54.0	1.06

An = angle; Bike = PTW-bicycle; Car = PTW-car; $cc > 250$ = heavy scooter or motorcycle; $50 < cc \leq 250$ = scooter or light weight motorcycle; $cc \leq 50$ = light weight PTW; Cu = curve; D = dry; Dy = day; F = fatal; FfV = falling from the vehicle; HO = head-on; Hobs = hit obstacle in carriageway; HpV = hit parked vehicle; HsV = hit stopped vehicle; I = injury; Int = intersection; Mw = motorway; Ot = other; Ped = hit pedestrian; Ped = PTW-pedestrian; R = rural; RE = rear-end; Rm = rural municipal; Rn = rainy; Rn = rural national; ROR = run-off-the-road; Rp = rural provincial; SB = sudden braking; Sl = slippery; SS = sideswipe; SV = PTW single vehicle; Sw = snow; Swy = snowy; Tan = tangent; Tr = hit train; Truck = PTW-truck; U = urban; Um = urban municipal; Un = urban national; Up = urban provincial; W = wet.

national roads exhibited crash severity substantially greater than the municipal roads. Indeed, urban provincial and national roads are segments of rural roads which cross small urban centres and drivers generally maintain high operating speeds crossing the small urban areas. An effective countermeasure to mitigate this problem is the implementation of perceptual cues such as gateways and traffic calming devices, according to the principle of the self-explaining roads (Wegman and Aarts, 2006), aimed at influencing the drivers' behaviour to recognize the road type and to drive accordingly, in particular at the appropriate speed. Driving simulator studies showed that the combination of gateways at the urban area entrance and integrative traffic calming devices along the urban area have the potential to significantly reduce mean speeds and, consequently, to reduce the crash severity (Galante et al., 2010).

Our results showed that PTW run-off-the-road crashes are a major safety issue, mainly because of their association with fatal severity, which was found also in previous studies (Daniello and Gabler, 2011; Montella and Pernetti, 2010; Perandones et al., 2008). It is stressed that there is a critical need to adopt improved safety barrier designs and new crash test procedures to protect PTWs. A positive example is the Spanish UNE 135900 standard, which defines procedures for the assessment of the performance of the motorcyclist protective devices' (MPDs). MPDs are devices installed on a steel safety barrier to protect motorcyclists from impacts against the barrier posts or to prevent motorcyclists from crossing the safety barrier through the space between the posts. Study results support both the use of MPDs on existing road restraint systems and the design of new devices more friendly for PTWs. The costs of fitting these devices can be reduced by selecting road sections where collisions by motorcycles is more frequent, i.e., in tight curves in rural areas. A significant driver factor related to run-off-the-road crash type was the PTW driver age. Specifically, drivers aged over 65 showed greater propensity for run-off-the-road crashes in curve. Legislative measures relative to the driving licence of the PTW drivers older than 65 might also be considered.

Environmental factors significantly associated with PTW crashes were nighttime and rainy weather. Consistently with literature findings (Quddus et al., 2002; Rifaat et al., 2011; Savolainen and Mannering, 2007), nighttime was associated with an increase in fatal severity. Moreover, several combinations of factors including nighttime were associated with hit pedestrian crashes. As an example, hit pedestrian crashes in nighttime, rainy weather, tangent alignment, and involving a scooter or a light weight motorcycle resulted 9.16 times the number of crashes expected if the above crash characteristics were independent. Both PTWs and pedestrians have low conspicuity and this phenomenon is accentuated in the night. Dark lighting severely reduces both PTW drivers' and pedestrians' sight and in turn their reaction times to avoid crashes increase. Consequently, these conditions increase braking distance of vehicles and lead to higher impact at the time of crashes. Further, higher driving speeds and longer response times from emergency crews are generally observed at night (de Lapparent, 2006). Several measures might contribute to reduce PTW crash severity and frequency at night. Motorcyclists can improve conspicuity, and thus their safety, by wearing retro-reflective material on their clothes and helmets. At the same time, to reduce PTW-pedestrian crashes at night measures relative to the road and the pedestrian are needed, such as improvement of the pedestrian crosswalks lighting, increase of the night time visibility of crosswalks markings and signs, pedestrian-activated overhead flashing beacons or high-intensity activated crosswalk devices, and use of reflective vests and other clothing for pedestrians.

As shown by both classification trees (Fig. 3, nodes 8, 10, and 21) and the rules discovery (rule 20), PTWs with greater cylinder capacity were significantly associated with fatal severity and this results is consistent with previous studies (de Lapparent, 2006; Quddus et al., 2002; SafetyNet, 2009; Yannis et al., 2005). Our results highlight the necessity to develop and maintain information campaigns about road safety concerning the exposition to risk of fatal injury of drivers of heavy scooters and motorcycles (with cylinder capacity greater than 250 cm^3), because of the central role of driving speeds to PTW safety. Also electronic speed limiters would give rise to relevant safety benefits for more powerful PTWs.

6. Conclusions

Both the classification trees and the a priori algorithm were effective in providing meaningful insights about PTW crash characteristics and their interdependencies. Even though in several cases different crash characteristics were highlighted, the results of the two the analysis methods were never contradictory. Furthermore, most of the findings of this study were consistent with the results of previous studies which used different analytical techniques, such as probabilistic models of crash injury severity. Basing on the analysis results, engineering countermeasures and policy initiatives to reduce PTW injuries and fatalities were singled out.

One possible criticism to the data mining techniques used in the study is that they might produce not significant results because, due to the large number of patterns considered, they suffer from an extreme risk of type-I error, that is, of finding patterns that appear due to chance alone to satisfy the constraints on the sample data. To overcome this problem, in our study we randomly split the sample data in two data sets and used well-established statistical practices to evaluate the statistical significance of the results. The framework defined in this study is of general validity and might me effectively repeated also in other cases.

As a result, our study showed that classification trees and association discovery are complementary and their joint use allows a better understanding of the crash characteristics. The use of classification trees and association discovery must not, however, be seen as an attempt to supplant other techniques, but as a complementary method which can be integrated into other safety analyses.

References

AASHTO – American Association of State Highway and Transportation Officials, 2005. AASHTO Strategic Highway Safety Plan. A Comprehensive Plan to Substantially Reduce Vehicle-Related Fatalities and Injuries on the Nation's Highways. AASHTO – American Association of State Highway and Transportation Officials, Washington, DC.

Abdi, H., 2007. Bonferroni and Šidák corrections for multiple comparisons. In: Salkind, N.J. (Ed.), Encyclopedia of Measurement and Statistics. Sage, Thousand Oaks, CA.

ACEM – Association des Constructers Européens de Motocycles, 2006. Guidelines for PTW-Safer Road Design in Europe. Brussels.

ACEM – Association des Constructers Européens de Motocycles, 2009a. MAIDS, In-depth Investigations of Accidents Involving Powered Two Wheelers, Final Report, Version 2.0. Brussels.

ACEM – Association des Constructers Européens de Motocycles, 2009b. MAIDS, In-depth Investigations of Accidents Involving Powered Two Wheelers: Multivariate Analysis of MAIDS Fatal Accidents. Brussels.

Agrawal, R., Imielinski, T., Swami, A., 1993. Mining association rules between sets of items in large databases. In: Proceedings of the ACM SIGMOD , Washington, DC, pp. 207–216.

Baer, J., Ayotte, K., Baldi, S., 2010. Evaluation of State Motorcycle Safety Programs. National Highway Traffic Safety Administration Report DOT HS 811 269, Washington, DC.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth International Group, Belmont, CA.

Cafiso, S., La Cava, G., Montella, A., 2007. Safety index for evaluation of two-lane rural highways. Transportation Research Record 2019, 136–145.

Chang, L., Chen, W., 2005. Data mining of tree-based models to analyze freeway accident frequency. Journal of Safety Research 36, 365–375.

Chang, L., Wang, H., 2006. Analysis of traffic injury severity: an application of non-parametric classification tree techniques. Accident Analysis and Prevention 38, 1019–1027.

Cheng, C., Lin, C., Leu, S., 2010. Use of association rules to explore cause–effect relationships in the Taiwan construction industry. Safety Science 48, 436–444.

Conversano, C., Siciliano, R., 2009. Incremental tree-based missing data imputation with lexicographic ordering. Journal of Classification 26, 361–379.
D'Ambrosio, A., Aria, M., Siciliano, R., 2007. Robust tree-based incremental imputation method for data fusion. Lecture notes in computer science 4723. In: Advances in Intelligent Data Analysis. Springer-Verlag, pp. 174–183.
Daniello, A., Gabler, H.C., 2011. Fatality risk in motorcycle collisions with roadside objects in the United States. Accident Analysis and Prevention 43, 1167–1170.
de Lapparent, M., 2006. Empirical Bayesian analysis of accident severity for motorcyclists in large French urban areas. Accident Analysis and Prevention 38, 260–268.
Dunn, O.J., 1961. Multiple comparisons among means. Journal of the American Statistical Association 56, 52–64.
European Commission, 2001. White Paper – European Transport Policy for 2010: Time to Decide. Brussels.
European Commission, 2009. EU Energy and Transport in Figures. Brussels.
European Commission, 2010. Towards a European Road Safety Area: Policy Orientations on Road Safety 2011–2020. COM(2010) 389, Brussels.
Galante, F., Mauriello, F., Montella, A., Pernetti, M., Aria, M., D'Ambrosio, A., 2010. Traffic calming along rural highways crossing small urban communities: driving simulator experiment. Accident Analysis and Prevention 42, 1585–1594.
Harba, R., Yanb, X., Radwanc, E., Sud, X., 2009. Exploring precrash maneuvers using classification trees and random forests. Accident Analysis and Prevention 41, 98–107.
Italian Ministry of Infrastructures and Transports, 2010. National Count of on Infrastructures and Transports: Years 2008–2009. Rome.
Kim, J.W., Lee, B.H., Shaw, M.J., Chang, H., Nelson, M., 2001. Application of decision-tree induction techniques to personalized advertisements on internet storefronts. International Journal of Electronic Commerce 5, 45–62.
Lyon, C., Persaud, B., 2008. Safety effects of targeted program to improve skid resistance. Transportation Research Record 2068, 135–140.
Montella, A., 2005. Safety reviews of existing roads: quantitative safety assessment methodology. Transportation Research Record 1922, 62–72.
Montella, A., 2009. Safety evaluation of curve delineation improvements: empirical Bayes observational before-and-after study. Transportation Research Record 2103, 69–79.
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 43, 1451–1463.
Montella, A., Colantuoni, L., Lamberti, R., 2008. Crash prediction models for rural motorways. Transportation Research Record 2083, 180–189.
Montella, A., Pernetti, M., 2010. In-depth investigation of run-off-the-road crashes on the Motorway Naples-Candela. In: Presented at the 4th International Symposium on Highway Geometric Design of the Transportation Research Board , Valencia.
Montella, A., Aria, M., D'Ambrosio, A., Mauriello, F. Data mining techniques for exploratory analysis of pedestrian crashes. Paper TRB-11-0283. In: Presented at Transportation Research Board Annual Meeting, Washington, DC, January 2011. Transportation Research Record, in press.
Pande, A., Abdel-Aty, M., 2009. Market basket analysis of crash data from large jurisdictions and its potential as a decision supporting tool. Safety Science 47, 145–154.
Perandones, J.M., Molinero, A., Martin, C., Mansilla, A., Pedrero, D., 2008. Recommendations for location of motorcyclist protection devices in Spanish regional road network of Castilla y Leon. In: Presented at the 87th Annual Meeting of the Transportation Research Board , Washington, DC.
Petrakos, G., Conversano, C., Farmakis, G., Mola, F., Siciliano, R., Stavropoulos, P., 2004. New ways to specify data edits. Journal of Royal Statistical Society Series A 167, 249–274.
Potts, I., Garets, S., Smith, T., Pferer, R., Neuman, T.R., Slack, K.L., Hardy, K.K., Nichols, J., 2008. NCHRP Report 500: Guidance for Implementation of the AASHTO Strategic Highway Safety Plan. Volume 22: A Guide for Addressing Collisions Involving Motorcycles. Transportation Research Board, Washington, DC.
Quddus, M.A., Noland, R.B., Chin, H.C., 2002. An analysis of motorcycle injury and vehicle damage severity using ordered probit models. Journal of Safety Research 33, 445–462.
Rifaat, S.M., Tay, R., de Barros, A., 2011. Severity of motorcycle crashes in Calgary. Accident Analysis and Prevention, doi:10.1016/j.aap.2011.02.025.
ROSPA – The Royal Society for Prevention of Accidents, 2002. Road Safety Engineering Manual. Birmingham.
SafetyNet, 2009. Powered Two Wheelers. Brussels.
Savolainen, P., Mannering, F., 2007. Probabilistic models of motorcyclists' injury severities in single- and multi-vehicle crashes. Accident Analysis and Prevention 39, 955–963.
Shaffer, J., 1995. Multiple hypothesis testing. Annual Review of Psychology 46, 561–584.
Teoh, E.R., 2010. Effectiveness of Antilock Braking Systems in Reducing Motorcycle Fatal Crash Rates. Insurance Institute for Highway Safety, Arlington, VA.
Webb, G.I., 2007. Discovering significant patterns. Machine Learning 68, 1–33.
Wegman, F., Aarts, L., 2006. Advancing Sustainable Safety: National Road Safety Outlook for 2005–2020. SWOV, Institute for Road Safety Research, Leidschendam.
Yannis, G., Golias, J., Papadimitriou, E., 2005. Driver age and vehicle engine size effects on fault and severity in young motorcyclists accidents. Accident Analysis and Prevention 37, 327–333.