IATSS Research 44 (2020) 85–93

Contents lists available at ScienceDirect

IATSS Research

Research article

Severity analysis of tree and utility pole crashes: Applying fast and frugal heuristics

Subasish Das [a,*], Beverly Storey [a], Tahmida Hossain Shimu [a], Sudeshna Mitra [b], Magdalena Theel [a], Bita Maraghehpour [a]

[a] Texas A&M Transportation Institute, Texas A&M University System, 3135 TAMU, College Station, TX-77843, United States of America
[b] Civil Engineering Department, Indian Institute of Technology (IIT) Kharagpur, West Bengal 721302, India

ARTICLE INFO

Article history:
Received 27 October 2018
Received in revised form 4 May 2019
Accepted 19 August 2019
Available online 26 August 2019

Keywords:
Tree and utility pole/other pole crashes
Data mining
Market basket analysis
Rules mining

ABSTRACT

A roadway departure (RwD) crash is defined as a crash that occurs after a vehicle crosses an edge line or a center line, or otherwise leaves the designated travel path. RwD crashes account for approximately 50% of all traffic fatalities in the U.S. Additionally, crashes related to roadside fixed objects such as trees, utility poles, or other poles (TUOP) make up 12–15% of all fatal RwD crashes in the U.S. Data spanning over seven years (2010–2016) shows that TUOP crashes account for approximately 22% of all fatal crashes in Louisiana, which is significantly higher than the national statistic. This study aims to determine the effect of crash, geometric, environmental, and vehicle characteristics on TUOP crashes by applying the fast and frugal tree (FFT) heuristics algorithm to Louisiana crash data. FFT identifies five major cues or variable threshold attributes that contribute significantly to predicting TUOP crashes. These cues include posted speed limit, primary contributing factor, highway type, weather, and locality type. The balanced accuracy is around 56% for both training and test data. The current model shows higher accuracies compared to machine learning models (e.g., support vector machine, CART). The present findings emphasize the importance of a comprehensive understanding of factors that influence TUOP crashes. The insights from this study can help data-driven decision making at both planning and operation levels.

1. Introduction

According to the Federal Highway Administration (FHWA), the new definition of a roadway departure (RwD) crash is a "crash in which a vehicle crosses an edge line, a center line, or leaves the traveled way" [1]. From 2014 to 2016, an average of 18,779 fatalities resulted from RwD crashes, making up 53% of all traffic fatalities in the U.S. The majority of RwD crashes in the Fatality Analysis Reporting System (FARS) are crashes in which the first event for any vehicle involved in the crash is one of the following: running off road – right, running off road – left, crossing the median, or crossing the centerline [2]. In addition, there are several fixed object codes (e.g. trees, utility poles, other poles) that are included based on the idea that a vehicle must have left the roadway in order to come in contact with those objects. Tree and utility pole/other pole related (TUOP) crashes account for 12–15% of all RwD fatal crashes in the U.S. This statistic is significantly higher for the state of Louisiana, where TUOP crashes represent around 22% of all fatal crashes. Furthermore, all TUOP crashes increased by 5% from 2015 to 2016 [3]. These statistics prompted in-depth study of TUOP crashes in Louisiana.

In spite of recent advances in transportation safety research, there have been very few research efforts on TUOP crashes. This study applied the fast and frugal tree (FFT) heuristics algorithm to a seven-year (2010–2016) TUOP crash dataset of Louisiana in order to identify significant factors regarding crash, geometric, environmental and vehicle characteristics, and extract the decision rules for TUOP crashes. This study is based on the premise that TUOP crashes are caused in an environment resulting from roadway geometry and surrounding traffic conditions. Our emphasis of the current study is limited to identify the key contributing factors that influence the injury outcomes of TUOP crashes. The results of this study support the idea that the application lens of fast-and-frugal heuristics is well suited to describe and improve applied decision making to increase the safety of TUOP collisions. The findings of the current research will be helpful in making data-driven decisions for the reduction of TUOP crashes.

1.1. Literature review

The most common approach to examine crash frequencies and their injury severity is to start with a crash-frequency model and then

* Corresponding author.
E-mail addresses: s-das@tti.tamu.edu (S. Das), b-storey@tti.tamu.edu (B. Storey), tahmida@tamu.edu (T.H. Shimu), sudeshna@civil.iitkgp.ernet.in (S. Mitra), m-theel@tti.tamu.edu (M. Theel), b-maraghehpour@tti.tamu.edu (B. Maraghehpour).
Peer review under responsibility of International Association of Traffic and Safety Sciences.

https://doi.org/10.1016/j.iatssr.2019.08.001

consider the injury severity of the crash conditional on the crash having occurred. Lord and Mannering provided a systematic review of widely used crash frequency studies and their limitations [4], while Savolainen et al. conducted a similar study on injury severity associated studies in 2011 [5]. In 2014, Mannering and Bhat summarized different methodologies used in these two areas with the inclusion of directions for future studies [6]. The objective of most of the safety related studies is to identify how different variables affect crash occurrence or crash severity. Both statistical and machine learning methods are widely utilized in traffic safety analysis. Some of these methods are logistic regression [7,8], decision trees [9–12], support vector machines [13,14], rough sets [15], multiple correspondence analysis [16,17], association rules mining [18], and deep learning [19,20].

Roadway departure related crashes have been examined by many safety researchers. Based on a report by National Highway Traffic Safety Administration (NHTSA) [21], run-off-road (ROR) crashes (a variant of RwD crashes) accounted for nearly 65% of all crashes involving single vehicles in the U.S. between 1991 and 2007. The most frequent types of ROR crashes likely to result in a severe or fatal injury involve hitting fixed roadside objects, such as TUOP [22]. According to this study, these types of roadside hazards are the least treated ones, and more extensive research on countermeasure design planning and management would lower the severity of such crashes.

In 1980 Jones and Baum [23] conducted one of the earliest studies in the U.S. that examined the factors that increase the likelihood of utility pole related crashes using regression analysis. The factors identified by this research include number of roadside poles, lateral offset of the poles, pavement grade, roadway alignment, and posted speed limit. Factors that contribute to the severity of the pole related crashes include stiffness of the pole and impact speed. Another early study from South Australia revealed that from 1985 to 1996, 40% of all crashes resulted in at least one fatality where the immediate cause of death was identified as a roadside fixed object [24]. For more detailed knowledge of these early studies, readers should refer to a report developed by Motor Accident Commission of National Health and Medical Research Council [24].

Various statistical techniques have been utilized to analyze crash data and identify associated variables influencing ROR type crashes in the past. Nilsson et al. [25] conducted hierarchical agglomerative cluster analysis on a set of crash data to identify similarities among crash variables for countermeasure design in ROR crashes. The authors identified driver-related factors (e.g., drifting off, speeding), and environmental factors (e.g., snowy and wet roadways, sharp curvatures) that were statistically significant crash attributes for ROR type crashes. Al-Bdairi et al. [26] developed an ordered random parameter probit model to estimate the effects of various factors in ROR crashes for large-trucks. In a later study, Al-Bdairi et al. [27] employed mixed logit models to understand the effect of lighting conditions in ROR crashes involving large-trucks. The findings of these studies suggest that collision with a fixed object significantly increases the likelihood of high injury severity in ROR crashes, and darkness increases the probability of crash outcomes due to the decreased visibility as opposed to well-lit conditions.

Information gained from these studies provides valuable information for designing test-scenarios and countermeasures that can be used to study ROR crashes and reduce incidences of these types of crashes resulting in injuries or fatalities. Dissanayake and Roy [28] used a binary logit model to investigate crash variables in ROR crashes and found that driver age, impairment, license status, speeding, road surface condition, dark lighting condition, type of vehicle, and loss of control are the most significant parameters in predicting crash severity for ROR crashes. In another study, Dissanayake [29] investigated the factors influencing injury severity of single-vehicle ROR crashes involving young drivers. According to this study, factors such as driving under influence, rural environment, roadway curvatures, impact speed and location were found to be most significant. The author also argued that severe weather condition and physical impairment due to fatigue or illness do not significantly affect the severity of single-vehicle ROR crashes, which is contradictory to other studies [30,31].

In 2009, the NHTSA conducted a national study on ROR crashes using FARS data on fatal crash data on passenger cars from 1991 to 2007 [30]. They found that roadway curvature, speeding, rural environment, posted high speed limit, adverse weather conditions, and driver impairment in terms of fatigue, sleep, and alcohol are the significant factors affecting the severity of ROR crashes. Additionally, the results suggest that under adverse weather conditions, young drivers and vehicle speeding increase the likelihood of fatal single-vehicle ROR crashes. In another research effort, the NHTSA [21] studied the pre-crash events, the critical crash causation during crash events, and the factors in the pre-crash phase of single-vehicle ROR crashes using National Motor Vehicle Crash Causation Survey (NMVCCS) data from 2005 to 2007. Results from a logistic regression model revealed that the most significant factors contributing to single-vehicle ROR type crashes are driver fatigue, inattention, and drivers being in a hurry.

Crashes involving roadside fixed objects in general can draw better association to crashes involving TUOP, which is the primary focus of this paper. In a comprehensive study, Alruwaished [32] investigated the factors influencing ROR crashes in which vehicles collided with either another vehicle or a fixed object after leaving its designated travel lane at a non-intersection area of the roadway. The ROR model successfully identified nine variables that influence crash severity. These variables include roadway surface condition, type of collision, driving under influence, speed limit, type of vehicle, and driver age and gender. Another study used similar classification and regression tree modeling to analyze factors influencing injury severity, and it found human errors to be the most significant variables in ROR and RwD crashes [33]. Watson et al. [34] examined the effects of roadside fixed objects by considering crash indicator values for segments with and without roadside fixed objects. The findings revealed that there are consistently lower crash indicators for roadways with no fixed roadside objects as opposed to those that have fixed objects on the roadside, meaning that roadways without fixed roadside objects have a much lower risk of crashes occurring.

A study by Maine Department of Transportation revealed that [35], the most prevalent factors contributing to utility pole crashes are rural environment, roadway curvature, dark lighting condition, driver inattention, speeding, poor roadway condition, and location and offset of utility poles. Dumbaugh [36] analyzed tree and utility pole crash site locations to identify the reasons behind these types of crashes in an urban environment. Many studies invested research efforts in analyzing safety effects of various countermeasures. For more detailed knowledge on these countermeasures, readers can refer to these studies [37–44]. Table 1 represents the key variables used by other studies found in the existing literature.

According to the literature review under the scope of this paper, a majority of studies identified roadway curvature, vertical grade, posted speed limit, roadway surface condition, speeding, loss of control and physical impairment of drivers as the variables that most significantly affected RwD crashes. It is important to note that very few studies focused only on TUOP related crashes. Additionally, the past studies have not considered algorithm-based modeling techniques to identify the effect of the contributing factors. This study focuses only on TUOP crashes to identify the key associated factors by using an algorithm-based technique FFT. This research focuses on studying the association of geometric, crash, environmental and vehicle-related variables with the level of injury caused by the crash. This study demonstrated the applicability of FFT heuristics to better understand the association between the key contributing factors.

Table 1
Findings from existing literature.

Variable		Studies	Key association patterns
Driver/Occupant	Speeding	Nilsson et al. [25], Dissanayake and Roy [28], Dissanayake [29], NHTSA [30], Alruwaished [32], Rovšek et al. [33], MDOT [35]	High speed is associated with higher number of RwD crashes. NHTSA [31] indicates an odds ratio of 1.30 for fatal ROR crashes.
	Departure angle	Nilsson et al. [25]	Determination of the simulator-based departure angles of the wheels for RwD crashes.
	Loss of control	NHTSA [21], Nilsson et al. [25], Al-Bdairi et al. [26], Al-Bdairi et al. [27], Dissanayake and Roy [28], Rovšek et al. [33]	Loss of control is a significant crash contributing factor.
	License status	Al-Bdairi et al. [26], Dissanayake and Roy [28]	Young drivers, irrespective of their license permit status, are disproportionately involved in RwD crashes.
	Seatbelt use	Al-Bdairi et al. [26], Al-Bdairi et al. [11], Dissanayake [29]	No seat-belt usage is associated with more severe crashes.
	Driving under influence (DUI)	NHTSA [21], Al-Bdairi et al. [27], Dissanayake and Roy [28], Dissanayake [29], NHTSA [30], Alruwaished [32],	DUI is positively associated with RwD crash occurrences. NHTSA [31] indicates an odds ratio of 1.94 for fatal ROR crashes.
	Sleeping, fatigue, inattention, disabilities and other physical impairment	NHTSA [21], Al-Bdairi et al. [26], NHTSA [30], Rovšek et al. [33], Dissanayake and Roy [28], MDOT [35]	NHTSA [31] indicates an odds ratio of 3.21 for fatal ROR crashes.
	Gender and age of the driver	Dissanayake and Roy [28], Dissanayake [29], Alruwaished [32]	Females are involved in less severe RwD crashes. Young drivers are disproportionately involved in RwD crashes.
Vehicle	Number of vehicles involved	Al-Bdairi et al. [26], Al-Bdairi et al. [27]	Majority of these crashes are single vehicle crashes.
	Type of vehicle	Dissanayake and Roy [28], Alruwaished [32]	Majority of the involved vehicles are passenger cars.
Environmental, Roadway and Crash	Type of road	Nilsson et al. [25]	Rural two-lane roadways are dominant.
	Lighting condition	Nilsson et al. [25], Dissanayake and Roy [28], MDOT [35]	No lighting is associated with more severe RwD crashes.
	Road surface condition	NHTSA [21], Nilsson et al. [25], Al-Bdairi et al. [26], Dissanayake and Roy [28], Alruwaished [32], MDOT [35]	Contradictory results are found.
	Median type	Al-Bdairi et al. [26]	'No median' is potential concern for 'centerline crossing' RwD crashes.
	Crash type	Al-Bdairi et al. [26], Al-Bdairi et al. [27], Alruwaished [32]	Majority of the crashes are single vehicle crashes.
	Roadway curvature (horizontal or vertical)	NHTSA [21], Jones and Baum [23], Nilsson et al. [25], Al-Bdairi et al. [26], Al-Bdairi et al. [27], NHTSA [30], Dissanayake [29], MDOT [35]	Horizontal curvature is associated with higher number of RwD crashes.
	Rural environment	Dissanayake [29], NHTSA [30], MDOT [35]	Rural roadways are dominant.
	Posted Speed Limit	Jones and Baum [23], Dissanayake and Roy [28], Dissanayake [29], NHTSA [30], Alruwaished [33], MDOT [35]	NHTSA [31] indicates an odds ratio of 1.30 for fatal ROR crashes.
	Adverse weather condition	Nilsson et al. [25], NHTSA [30]	NHTSA [31] indicates an odds ratio of 1.24 for fatal ROR crashes.
	Location of poles or trees	Jones and Baum [23], MDOT [35]	Poles nearer to the travel paths have negative safety effect.

2. Methodology

2.1. Data preparation

The dataset used in this study is police-reported crashes in Louisiana Department of Transportation and Development (LADOTD) from 2010 to 2016. The Louisiana crash database contains three major data tables: 1) crash table, 2) roadway inventory table, known as DOTD table, and 3) vehicle table. Vehicle table contains five variables related to harmful events (first harmful event, second harmful event, third harmful event, fourth harmful event, and most harmful event). Once the filter is applied, the dataset contains 56,426 vehicle level information. A summary statistic of the dataset reveals that TUOPs make up 95% of the most harmful or the first harmful event scenarios. Out of these, trees contribute to 82% of TUOP crashes. Later, this table was merged with crash and DOTD table. Fig. 1 illustrates the data preparation task in a flowchart.

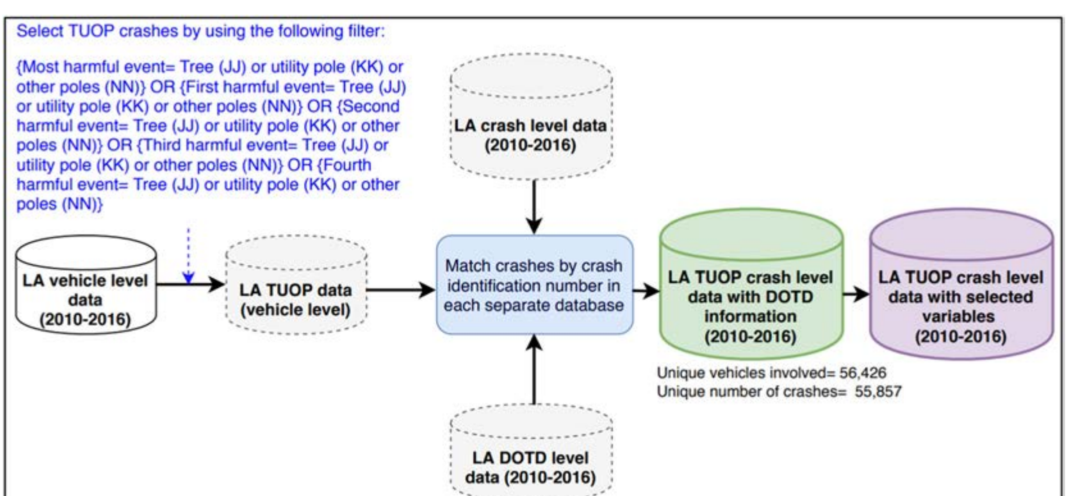


Fig. 1. Flowchart of data preparation.

2.2. Descriptive statistics

Based on the literature review, twenty preliminary variables were selected for analysis. Some variables were removed due to redundancy. Additionally, variables with over-representation of an attribute were omitted. For example, normal weather (in weather variable) and dry pavement condition (in pavement condition variable) represent above 95%. Other variables were removed due to their high correlations with other variables. For example, crash hour shows high correlation with lighting condition. The injury classification system (known as KABCO) divides crash severity into five major groups: 1) fatal injury (K), 2) incapacitating injury (A), 3) non-incapacitating injury (B), 4) minor injury (C), and 5) non-injury or property damage only (PDO) or O. Table 2 provides a summary of TUOP crashes from 2010 to 2016. The primary dataset contains 56,426 vehicle level crash information with 20 variables. The research team performed variable importance (by using random forest algorithm) to determine the significant factors. After removing the variables with less variable importance measures, the final dataset contains 55,857 crash level data with 12 variables (11 predictor variables and one response variable).

Fig. 2 illustrates the total number of TUOP crashes by Louisiana parishes. The colors used in the map indicate the frequency ranges (light yellow for low frequencies to dark red for high frequencies). East Baton Rouge, Orleans, and Jefferson are the parishes with high number of TUOP crashes. East Baton Rouge observes highest number of TUOP crashes (average 612 crashes per year). Tensas shows the lowest number of TUOP crashes (yearly average 3 crashes) among all parishes. The top 20 parishes with high number of TUOP crashes contain approximately 72% of all TUOP crashes. The remaining 44 parishes contain the other 28%. Although it is very difficult to obtain inventory on specific tree and utility pole locations, locations with high crash proneness from TUOP can be identified from historical crash data. Crash proneness determination using a smaller spatial unit (for example, U.S. Census tract, block group, or block) works better as it will identify the specific zones with higher TUOP crashes. Future studies can focus on this issue.

Table 3 represents the proportional distribution of the 11 key predictor variables used in the final dataset. Here, 'access control' accounts for the type of regulation imposed on a highway to limit the access of traffic flow from other roadways. A fully access-controlled highway operates without any traffic lights, intersections, or property access to provide an unhindered high-speed movement of traffic. According to the statistics of the Louisiana dataset, 'no control' roadways account for over 85% of all TUOP crashes, which is in line with the findings of past studies [29, 30, 35]. The report developed by NHTSA [30] provides evidence of association between rural highways with high speed limits and susceptibility to ROR crashes. Dissanayake [29] argued that rural high-speed highways increase the likelihood of ROR crashes with a possible explanation that young drivers have tendency of speed on such roadways. According to the research by Maine Department of Transportation (MDOT) [35], over 87% of all TUOP crashes in the state of Maine happened on rural areas with speed limit over 45 mph. 'Alignment' was found to be one of the most significant factors in ROR and RwD crashes by majority of the studies. However, roadways with no alignment constitute over 68% of TUOP crashes in Louisiana. Curve related crashes represent 25% of the crash data, which is disproportionately high when compared with all crashes. According to previous studies [21, 23, 25–27, 29, 30, 35], roadway curvature and vertical grades impose high probability of ROR and RwD crashes. These are aggravated by human errors such as speeding, inattention, and other related issues.

Approximately 72% of all TUOP crashes in the dataset occurred on two-way undivided roadways, which supports the findings by Dumbaugh [36]. This study found that medians can help reduce roadside and midblock crashes, including TUOP crashes. However, residential and open county represent >60% of the TUOP crashes. Furthermore, TUOP crashes are more dominant on lower functional class roadways (making up approximately 80% of all TUOP crashes).

Nearly two-third (67.21%) of TUOP crashes occurred on roadways with posted speed in between 35 and 60 mph, which is a typical range on principal arterial, collector, or local roadways. Posted speed limit was identified in previous studies as another significant factor in TUOP crashes [23,28–30,33,35], in which this variable was found to be significantly affect ROR and RwD crashes through various statistical analyses. Evidence of lighting condition (predominant variable being daylight with 50.10% of crash occurrence) and weather effect (predominant variable being clear weather with 64.73% of crash occurrence) contributing to ROR crashes are found in some studies [25,28,30,35]. Some studies [25,28,35] found association of adverse weather with a higher rate of RwD crashes. However, other studies [29] aligned more closely with the data from Louisiana and found evidence to the contrary. Passenger cars have the highest percentage distribution (45.88%) in the vehicle category, which is also evident in the FARS data [30]. Of the primary contributing factors, violations of the drivers and condition of the drivers demonstrate the most significant effect on frequency of TUOP crashes.

Table 2
TUOP crash severity distribution.

Year	K	A	B	C	O	All
2010	154	159	1098	2307	4022	7740
2011	153	184	1124	2258	4055	7774
2012	137	168	1101	2368	4031	7805
2013	157	149	1065	2256	4338	7965
2014	163	162	1068	2360	4200	7953
2015	158	180	1104	2446	4487	8375
2016	141	171	1109	2401	4423	8245
Grand total	1063	1173	7669	16,396	29,556	55,857

3. Fast and frugal tree: theory

Fast and frugal tree (FFT) is a heuristic algorithm for binary decision making. FFT is defined as a decision tree that contains $n + 1$ exits, with one exit for each of the first $n - 1$ cues, and two exits for the last cue or variable threshold. The FFT algorithm makes predictions about what cues or variable categories will influence decisions as well as how decision makers might utilize these cues. Initially, an FFT starts by checking the thresholds on the first cue to explore the exit condition then it continues to the other cues one after another until the final exit criteria are met [45].

Like all other relevant heuristic approaches, FFTs considers selected binary variables in order to simplify decision-making issues. FFT is generally composed of three major building blocks:

- Explore rules: Explore predictor variables and associated attributes (variable categories) in the order of their importance.
- Stopping rule: Stop search as soon as one predictor allows it.
- Decision rule: Classify according to this predictor variable.

Laura Martignon first introduced FFT in 2003 [45]. The three commonly used FFT algorithms are max, zigzag, and fan. There are four tasks to construct FFT algorithms: select cues or variables; determine a decision threshold for each cue; provide cue rankings; and determine cue exits [45].

The first step of constructing FFT is to determine the decision threshold for each cue. Thresholds are single values for numerical cues and sets of factor values for nominal or categorical cues. Applying each cue into training dataset while ignoring others, the single value or factor maximizing the cue's accuracy is selected [45].

A confusion matrix table generates a simple representation of the model's accuracy and the types of associated errors in the model. The measures *tp* and *tn* infer the correct prediction (true positive and true negative) respectively, whereas *fp* and *fn* measures refer to errors (false positive and false negative). The algorithm of FFT is designed to maximize frequencies in *tp* and *tn* while minimizing those in *fp* and *fn*. This study considered five important measures out of several

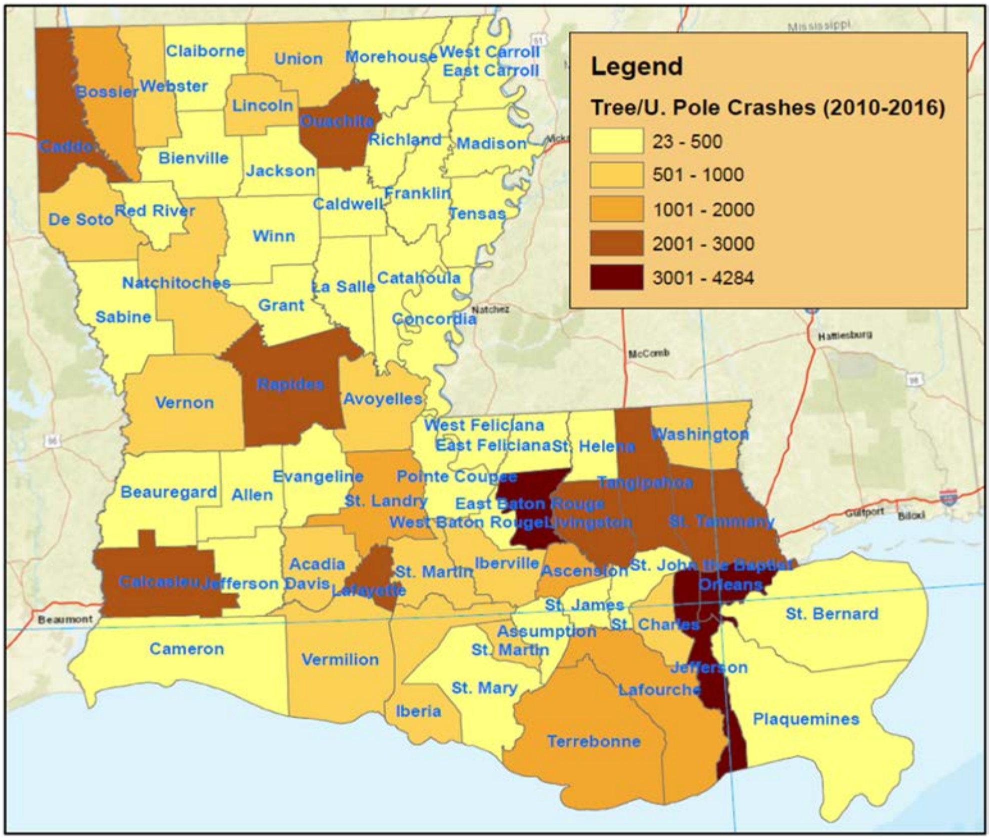


Fig. 2. TUOP crashes in Louisiana parishes.

performance measures: sensitivity (sens), specificity (spec), overall accuracy (acc), weighted accuracy, and balanced accuracy (bacc). These terms are described below:

$$sens = \frac{tp}{tp + fn} \quad (1)$$

$$spec = \frac{tn}{tn + fp} \quad (2)$$

$$acc = \frac{tp + tn}{tp + fp + fn + tn} \quad (3)$$

$$\mathrm{wacc} = \frac{tp}{tp + fn} \times w + \frac{tn}{tn + fp} \times (1 - \mathrm{w}) = sens \times w + spec \times (1 - w) \quad (4)$$

where,

$sens$ = Sensitivity,
$spec$ = Specificity,
acc = Accuracy,
$wacc$ = Weighted Accuracy,
tp = true positive,
tn = true negative,
fp = false positive,
fn = false negative,
w = weighting factor

The notation *'sens'* indicates the percentage of cases with positive criterion values that are correctly predicted by the FFT algorithm. On the other hand, '*spec*' infers the percentage of cases with negative criterion values correctly predicted by the FFT algorithm. The next three measures ('acc,' 'wacc,' and 'bacc') define accuracy across all cases. The notation *'acc'* is defined as the overall percentage of correct decisions by ignoring the difference between true positive and true negative. To make a balance between sensitivity and specificity, one can use the measure named as '*wacc*'. This measure depends on a weighting factor, which ranges between 0 and 1. In cases where sensitivity is more relevant compared to specificity, the user can consider the weighting factor above 0.50. For an ideal balanced situation, the user can use 0.50, which is narrated as *'bacc'* [45].

3.1. Results and findings

The final dataset used for FFT analysis consists of 55,857 TUOP crashes with twelve variables. The response variable of this study is the injury type (KABC or PDO) of TUOP crashes. To evaluate the performance of the model, both train (70% of the randomly selected data), and test (the rest 30%) datasets were used in the final analysis. The research team extensively used R package 'FFTrees' to perform the analysis as well as for data visualization [45]. Fig. 3 illustrates the outputs of FFTs by using train and test data. The output visualization has three basic parts: top, middle, and bottom part.

The top part of Fig. 3 shows information about the dataset, including the counts and percentage distributions of PDO and KABC crashes. For example, top part of Fig. 3 indicates that the sample size of the dataset is 39,100. The count of KABC TUOP crashes is 18,588 which represents 48% of the total TUOP crashes in the training dataset.

The middle part contains the deceision tree and icon arrays showing the frequency and confusion table matrices at each node. The results showed that FFT developed for each case (70% data as train data, and 30% data as test data) is the same. It indicates that three major criteria can identify the severity type of TUOP crashes: posted speed limit, primary contributing factor, and highway type. The interpretations from the FFTs are following:

- Training data: 1) If the posted speed limit is not 35–60 mph, and the primary contributing factor is condition of the driver, there is a high likelihood of the occurrence of a KABC TUOP crash; 2) If the posted speed limit is not 35–60 mph, the primary contributing factor is not

Table 3
Descriptive statistics of the key variables.

Code	Description	Freq.	Perc.
Access control			
A	No Control	47,586	85.19
C	Full Control	5595	10.02
B	Partial Control	2300	4.12
Z	Others	376	0.67
Alignment			
A	Straight-Level	38,395	68.74
C	Curve-Level	12,257	21.94
F	On Grade-Curve	1494	2.67
Z	Others	3711	6.65
Road type			
B	Two Way-No Physical Sep	40,134	71.85
C	Two Way-Physical Sep	10,663	19.09
A	One Way	2998	5.37
D	Two Way-Physical Barr	1688	3.02
Z	Others	374	0.67
Locality			
E	Res Scattered	13,763	24.64
G	Open Country	12,621	22.60
D	Residential	11,205	20.06
C	Mixed Bus/Res	10,250	18.35
B	Bus Continuous	6258	11.20
Z	Others	1760	3.15
Highway type			
C	State Hwy	21,441	38.39
D	Parish Road	11,962	21.42
E	City Street	11,214	20.08
A	Interstate	5710	10.22
B	US Highway	5088	9.11
Z	Others	442	0.79
Traffic control condition			
A	Control Functioning	43,053	77.08
E	Lane Marking Unclear or Defective	11,599	20.77
Z	Others	1205	2.16
Posted speed limit			
0–30	0–30 mph	10,896	19.51
35–45	35–45 mph	20,889	37.40
50–60	50–60 mph	16,650	29.81
> 60	>60 mph	5182	9.28
	Unknown	2240	4.00
Lighting			
A	Daylight	27,983	50.10
B	Dark-No Street Light	13,938	24.95
C	Dark-Continuous Street Lts	9422	16.87
D	Dark-Str Lits-Intersect Only	2259	4.04
Z	Others	2255	4.04
Weather			
A	Clear	36,154	64.73
C	Rain	9187	16.45
B	Cloudy	9059	16.22
Z	Others	1457	2.61
Vehicle type			
A	Passenger Car	25,628	45.88
B	Lt Trk/Pickup	16,160	28.93
S	SUV	8919	15.97
C	Van	1320	2.36
Z	Others	3830	6.85
Primary contributing factor			
A	Violations	35,211	63.04
D	Condition of Driver	8941	16.01
B	Movement Prior to Crash	7416	13.28
E	Vehicle Conditions	1308	2.34
G	Roadway Condition	1157	2.07
Z	Others	1824	3.27

the condition of the driver, the highway type is either state or U.S. highway, and weather is either clear or cloudy, there is a high likelihood of the occurrence of a KABC TUOP crash;

- Test data: 1) If the posted speed limit is not 35–60 mph, and the highway type is either state or U.S. highway, there is a high likelihood of the occurrence of a KABC TUOP crash; 2) If the posted speed limit is not 35–60 mph, the highway type is neither state nor U.S. highway, and the primary contributing factor is condition of the driver, there is a high likelihood of the occurrence of a KABC TUOP crash; 3) If the posted speed limit is not 35–60 mph, the highway type is neither state nor U.S. highway, the primary contributing factor is not condition of the driver, and the weather is either clear or cloudy, there is a high likelihood of the occurrence of a KABC TUOP crash;

The above interpretations from the FFT algorithm associating driver condition, posted speed limit, and highway type is supported by the findings from some previous studies, in which similar statistical association between these factors were found. Dissanayake's [29] binary regression models illustrated positive coefficients for posted speed limit (odd's ratio > 1.0) and intoxication (odd's ratio > 3.0), indicating a high probability of more severe crash outcomes for those variables. However, freeway as crash locations, bad weather, and other physical condition of the driver (for example, fatigue, illness) were found to the statistically non-significant in the models of Dissanayake's study, which contradicts with the findings of the FFT. Alruwaished [32] also found positive correlation between alcohol consumption and posted speed limit affecting injury severity in ROR crashes. A study by NHTSA [31] showed a higher probability of fatal single-vehicle ROR crashes associated with factors such as sleeping, alcohol impairment, high posted speed limit, and adverse weather effect. The odd's ratio of these factors in the logistic regression models were found to be 3.21, 1.94, 1.30 and 1.24 respectively indicating a higher likelihood of ROR crashes occurring in these conditions.

The rules that are generated for the training and the test data are not widely varied. However, the training data shows higher specificity, and the test data shows higher sensitivity. Whether a crash was coded as KABC or as PDO was largely affected by certain key factors including the posted speed limit (either lower than 30 mph or above 60 mph), if the highway type was either a state or U.S. highways, clear versus cloudy weather, and the condition of the driver as the primary contributing factor.

The bottom part of Fig. 3 shows the FFT's performance in receiver operating characteristics (ROC) curve, confusion matrix, and levels for a range of performance measures. The '*ifan*' algorithm explicitly selects and ranks cues' (threshold of variable category or category-groups) accuracies. Visualizing marginal cue helps in understanding the ranking of each cues in terms of *wacc*. The top five variable categories are colored and described in the legend. Fig. 4 shows the resulting plot for the training data. The graph reveals that the three cues (posted speed limit, primary contributing factor, and highway type) used in FFT (shown as #1 in Fig. 3) have the highest individual balanced accuracies. Fig. 4 also shows that the two next best cues are weather and locality. The extracted knowledge can be useful in guiding a top-down process of future FFT development using similar datasets.

One of the most convenient ways to present the characteristics of a diagnostic test is the ROC curve. An ROC curve visually displays the trade-off between '*sens*' and '*spec*' in classification algorithms. As '*sens*' increases, '*spec*' decreases (i.e., 1 − *spec* increases). The bottom right plot (Fig. 3) shows the performance of all FFTs in ROC space (green circles with numbers correspond to FFTs). FFT (shown as #1) has the highest weighted accuracy (colored in solid green). Additional points in this plot correspond to the performance of competing classification algorithms: standard decision trees (CART), logistic regression (LR), random forests (RF), and support vector machines (SVM). In this case, FFT has a higher '*sens*' than LR, CART, and SVM. The current analysis showed that a simplified tree algorithm can predict severity of TUOP crashes more precisely than sophisticated machine learning tools such as CART and SVM.

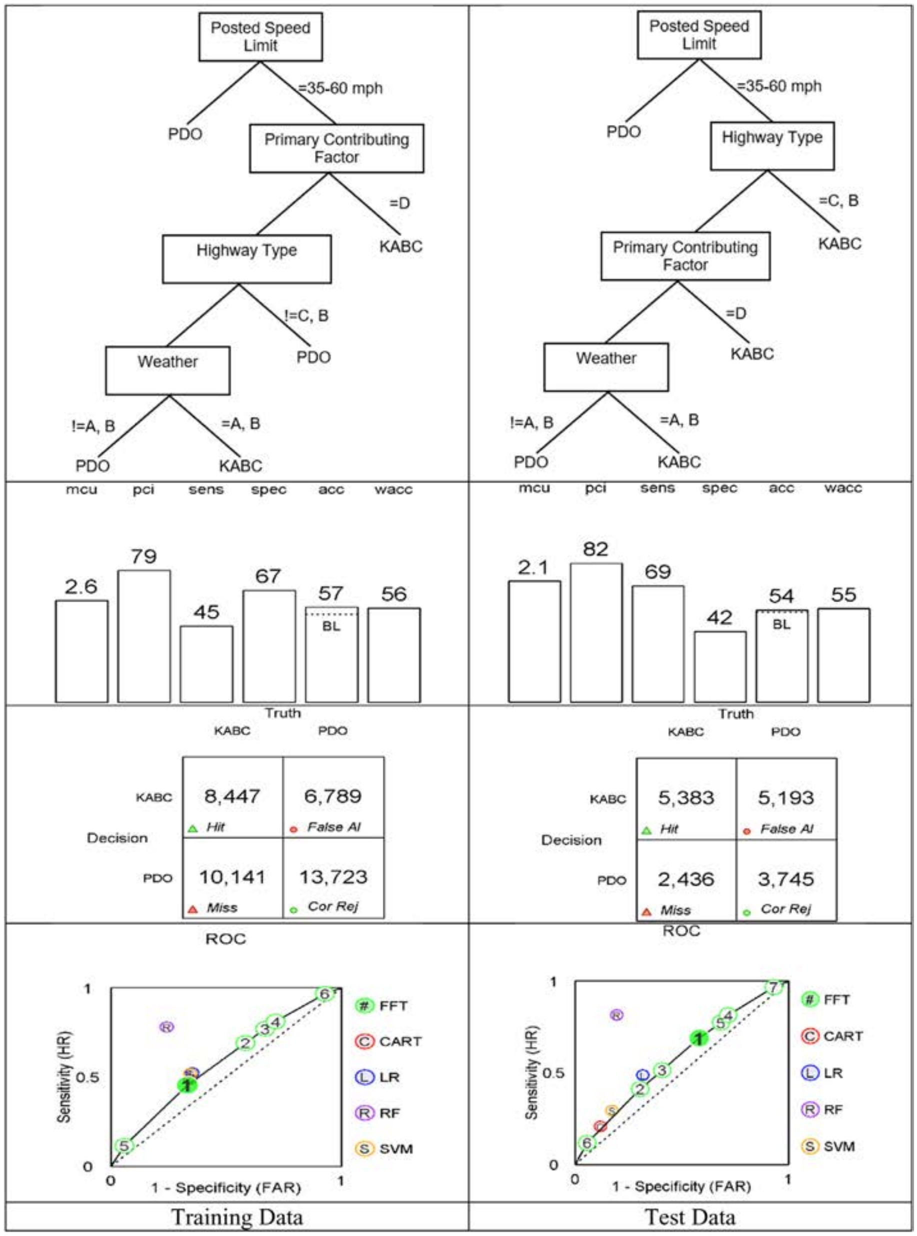


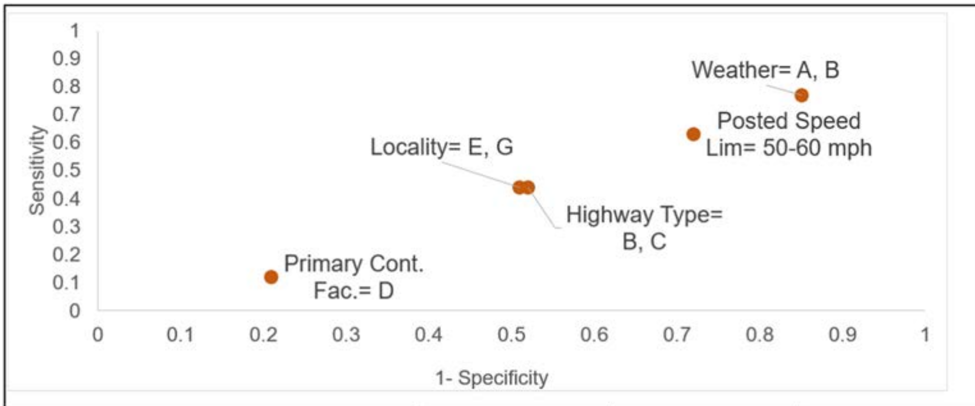
Fig. 3. Fast and frugal tree outputs for the training data.

4. Conclusions

This study analyzed all TUOP crashes in Louisiana that took place over the course of seven years (2010–2016) and identified key variables that contributed to these crashes. In this study, a variety of variables were considered, including access control type, alignment, road type, locality, highway type, traffic control condition, posted speed limit, lighting, weather, vehicle type, and primary contributing factor. Results from FFT heuristics revealed that posted speed limit, primary contributing factor, highway type, weather, and locality type were the key predictors of TUOP crashes. The key attributes that affected the injury severity (KABC and PDO crashes) are posted speed limit (either lower than 30 mph or above 60 mph), state and U.S. highways, clear versus cloudy weather, and condition of the driver as the primary contributing factor.

Reduction of TUOP crashes could significantly support the states' vision zero plans. However, designs of countermeasures to remove these roadside fixed objects are difficult to implement due to two reasons: 1) trees provide aesthetics and removal of trees would affect public sentiment, as well as other environmental considerations, and 2) the ownership of utility and other poles is controlled by various private companies rather than state agencies. Nonetheless, it is important to research the causes of TUOP crashes in order to make relevant policies and new countermeasures. Together with the visualizations of the rules, the current method provides interpretable results to the transportation safety practitioners.

This study provides valuable information regarding associative patterns of contributing variables in TUOP crashes. However, the current study is not without limitations. One of the limitations is that the current study is kept restricted to the identification of geometric, crash, environmental, and vehicle-related factors. Future studies can incorporate additional variables, like population and demographic characteristics, driver information, and real-time driver behavior. Additionally, additional insights from the the FFT algorithms can be achieved by extending the tree structures. Another limitation is that FFT is that does not provide quantitative impact of the contributing factors. It is important to note that ignoring information may be hard to justify in many cases. For complex problems, FFT can be used as the balance between high-cost and low-cost data driven decision making. Notwithstanding



Cue and Threshold	Sensitivity	Specificity	wacc
Posted Speed Lim= 50-60 mph	0.72	0.37	0.55
Primary Cont. Fac.= D	0.21	0.88	0.54
Highway Type= B, C	0.52	0.56	0.54
Weather= A, B	0.85	0.23	0.54
Locality= E, G	0.51	0.56	0.53

Fig. 4. Accuracies of the top five variable categories (training data).

the current limitations, this study provided a simplified decision mechanism (with high classification accuracies) to identify injury severity of TUOP crashes instead of incorporating uninterpretable machine learning algorithms.

Disclaimer

The contents of this paper reflect the views of the authors and not the official views or policies of the Louisiana Department of Transportation and Development (LADOTD).

Acknowledgments

The authors like to thank two anonymous reviewers for their valuable suggestions. The current version is much improved due to their well-thought suggestions and feedbacks.

References

[1] U.S. Department of Transportation, Revised Data Definitions for FHWA Safety Focus Areas, https://safety.fhwa.dot.gov/fas/docs/focus_area_data_mem.pdf 2018, Accessed date: 27 July 2019.
[2] Fatality Analysis Reporting System (FARS) Encyclopedia, https://www-fars.nhtsa.dot.gov/Main/index.aspx 2019, Accessed date: 4 May 2019.
[3] Louisiana Crash Data Reports, http://datareports.lsu.edu/ 2019, Accessed date: 4 May 2019.
[4] D. Lord, F. Mannering, The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives, Transp. Res. A Policy Pract. 44 (2010) 291–305.
[5] P. Savolainen, F. Mannering, D. Lord, M. Quddus, The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives, Accid. Anal. Prev. 43 (2011) 1666–1676.
[6] F. Mannering, C. Bhat, Analytic methods in accident research: methodological frontier and future directions, Anal. Methods. Acci. Res. 1 (2014) 1–22.
[7] S. Cafiso, G. Cava, G. Pappalardo, A logistic model for powered two-wheelers crash in Italy, Procedia Soc. Behav. Sci. 53 (2012) 880–889.
[8] M. Moomen, M. Rezapour, K. Ksaibati, An investigation of influential factors of downgrade truck crashes: a logistic regression approach, J. Traffic. Transp. Eng. (Engl. Ed.) 6 (2) (2019) 185–195.
[9] S. Das, A. Dutta, X. Kong, X. Sun, Hit and run crashes: knowledge extraction from bicycle involved crashes using first and frugal tree, Int. J. Transp. Sci. Technol. 8 (2) (2019) 146–160.
[10] Y. Chung, Factor complexity of crash occurrence: an empirical demonstration using boosted regression trees, Accid. Anal. Prev. 61 (2013) 107–118.
[11] da Cruz Figueira, Cira Souza Pitombo Aurenice, Ana Paula Camargo Larocca, Identification of rules induced through decision tree algorithm for detection of traffic accidents with victims: a study case from Brazil, Case Stud. Transp. Policy 5 (2) (2017) 200–207.
[12] G. Khan, A. Bill, D. Noyce, Exploring the feasibility of classification trees versus ordinal discrete choice models for analyzing crash severity, Transp. Res. C 50 (2015) 86–96.
[13] C. Cong, G. Zhang, Z. Qian, R. Tarefder, Z. Tian, Investigating driver injury severity patterns in rollover crashes using support vector machine models, Accid. Anal. Prev. 90 (2016) 128–139.
[14] X. Sun, S. Das, N. Broussard, Developing crash models with supporting vector machine for urban transportation planning, 17th International Conference Road Safety on Five Continents (RS5C 2016), Rio de Janeiro, Brazil, 2016.
[15] K. Karl, P. Pant, E. Yamashita, Hit-and-run crashes: use of rough set analysis with logistic regression to capture critical attributes and determinants, Transp. Res. Rec. 2083 (2018) 114–121.
[16] S. Das, X. Sun, Association knowledge for fatal run-off-road crashes by multiple correspondence analysis, IATSS Res. 39 (2) (2016) 146–155.
[17] S. Das, X. Sun, Factor association with multiple correspondence analysis in vehicle–pedestrian crashes, Transp. Res. Rec. 2519 (1) (2015) 95–103.
[18] S. Das, A. Dutta, R. Moran, K. Dixon, X. Sun, M. Jalayer, Supervised association rules mining on pedestrian crashes in urban areas: identifying patterns for appropriate countermeasures, Int. J. Urban Sci. (2018) 1–19.
[19] S. Das, A. Dutta, K. Dixon, L. Minjares-Kyle, G. Gillette, Using deep learning in severity analysis of at-fault motorcycle rider crashes, Transp. Res. Rec. 2672 (34) (2018) 122–134.
[20] S. Dabiri, K. Heaslip, Developing a twitter-based traffic event detection model using deep learning architectures, Expert Syst. Appl. 118 (2019) 425–439.
[21] C. Liu, T.J. Ye, Run-Off-Road Crashes: An On-Scene Perspective, Publication DOT HS 811 500 NHTSA, U.S. Department of Transportation, 2011.
[22] G.J. Jones, Noteworthy Practices: Roadside Tree and Utility Pole Management, Publication FHWA-SA-16-043 FHWA, U.S. Department of Transportation, 2016.
[23] I.S. Jones, A.S. Baum, An Analysis of the Urban Utility Pole Accident Problem, Publication FHWA-RD-79-117 FHWA, 1980.
[24] C.N. Kloeden, A.J. McLean, M.R.J. Baldock, A.J.T. Cockington, Severe and fatal car crashes due to roadside hazards, Report to Motor Accident Commission, NHMRC Road Accident Research Unit, the University of Adelaide, 1999.
[25] D. Nilsson, M. Lindmanb, T. Victora, M. Dozzaa, Definition of run-off-road crash clusters—for safety benefit estimation and driver assistance development, Accid. Anal. Prev. 113 (2018) 97–105.
[26] N.S.S. Al-Bdairi, S. Hernandez, An empirical analysis of run-off-road injury severity crashes involving large trucks, Accid. Anal. Prev. 102 (2017) 93–100.
[27] N.S.S. Al-Bdairi, S. Hernandez, J. Anderson, Contributing factors to run-off-road crashes involving large trucks under lighted and dark conditions, J. Transp. Eng. A. 144 (2018).
[28] S. Dissanayake, U. Roy, Crash severity analysis of single vehicle run-off-road crashes, J. Transp. Technol. 4 (2014) 1–10.
[29] S. Dissanayake, Young drivers and run-off-the-road crashes, Proceedings of the 2003 Mid-Continent Transportation Research Symposium, August 2003 , Ames, Iowa.
[30] U.S. Department of Transportation, National highway traffic safety administration, Factors Relateed to Fatal Single-Vehicle Run-off-Road Crashes, 2009 , Report No. DOT HS 811 232.
[31] C. Liu, R. Subramanian, Factors related to fatal single-vehicle run-off-road crashes, Rep. DOT HS 811 (2009) 232.
[32] A.F. Alruwaished, Characteristics of Drivers who Cause run-off-Road-Crashes on Ohio Roadways, Master's Thesis the School of Engineering of the University of Dayton, 2014.

[33] V. Rovšek, M. Batista, B. Bogunović, Identifying the key risk factors of traffic accident injury severity on Slovenian roads using a non-parametric classification tree, Transport 32 (2017) 272–281.
[34] Watson Jr., D.C.A. Al-Kaisy, N.D. Anderson, Examining the effect of speed, roadside features, and roadway geometry on crash experience along a rural corridor, J. Mod. Transp. 22 (2014) 84–95.
[35] B. Marquis, Utility Pole Crash Modeling, Report no. ME 00–8 Maine Department of Transportation, Transportation Research Division, 2001.
[36] E. Dumbaugh, Design of safe urban roadsides: an empirical analysis, Transp. Res. Rec. (2006) 74–82.
[37] L. Wu, S.R. Geedipally, A.M. Pike, Safety evaluation of alternative audible lane departure warning treatments in reducing traffic crashes: an empirical Bayes observational before–after study, Transp. Res. Rec. 2672 (21) (2018) 30–40.
[38] FHWA, Strategic Approach to Roadway Departure Safety, 2017.
[39] R.B. Patel, F.M. Council, M.S. Griffith, Estimating safety benefits of shoulder rumble strips on two-lane rural highways in Minnesota: empirical Bayes observational before-and-after study, Transp. Res. Rec. 2019 (2007) 205–211.
[40] T. Sayed, P. deLeur, J. Pump, Impact of rumble strips on collision reduction on highways in British Columbia, Canada comprehensive before-and-after safety study, Transp. Res. Rec. 2148 (2010) 9–15.
[41] J. Park, M. Abdel-Aty, Development of adjustment functions to assess combined safety effects of multiple treatments on rural two-lane roadways, Accid. Anal. Prev. 75 (2015) 310–319.
[42] M. Rys, D. Karkle, E. Russell, Study of KDOT Policy on Lane and Shoulder Minimum Width for Application of Centerline Rumble Strips, Report No. K-TRAN: KSU-10-7 Kansas Department of Transportation, Topeka, 2012.
[43] M. Khan, A. Abdel-Rahim, C.J. Williams, Potential crash reduction benefits of shoulder rumble strips in two-lane rural highways, Accid. Anal. Prev. 75 (2015) 35–42.
[44] H. Fu, C.G. Wilmot, Assessing Performance of Alternative Pavement Marking Materials, Report 08-4TA Louisiana Transportation Research Center, Baton Rouge, 2008.
[45] N. Phillips, H. Neth, J. Woike, W. Gaissmaer, FFTrees: Generate, Visualise, and Evaluate Fast-and-Frugal Decision Trees, R Package Version 1.3.5 2018.