ARTICLE IN PRESS
IATSSR-00288; No of Pages 7
IATSS Research xxx (xxxx) xxx
Contents lists available at ScienceDirect
IATSS Research

Research Article

Applying interpretable machine learning to classify tree and utility pole related crash injury types

Subasish Das [a,*], Songjukta Datta [a], Hamsa Abbas Zubaidi [b], Ihsan Ali Obaid [b]

[a] Texas A&M Transportation Institute, 1111 RELLIS Parkway, Bryan, TX 77807, United States of America
[b] Oregon State University, 233 Owen Hall, Corvallis, OR 97331-3212, United States of America

ARTICLE INFO

Article history:
Received 3 September 2020
Received in revised form 14 January 2021
Accepted 21 January 2021
Available online xxxx

Keywords:
TUOP crashes
Crash narratives
Text mining
Interpretable machine learning
Misclassification
Explanation

ABSTRACT

In spite of enormous improvements in vehicle safety, roadway design, and operations, there is still an excessive amount of traffic crashes resulting in injuries and major productivity losses. Despite the many studies on factors of crash frequency and injury severity, there is still further research to be conducted. Tree and utility pole/other pole related (TUOP) crashes present approximately 12 to 15% of all roadway departure (RwD) fatal crashes in the U.S. The count of TUOP crashes comprise nearly 22% of all fatal crashes in Louisiana. From 2010 to 2016, there were 55,857 TUOP crashes reported in Louisiana. Individually examining each of these crash reports is not a realistic option to investigate crash factors. Therefore, this study employed text mining and interpretable machine learning (IML) techniques to analyze all TUOP crashes (with available crash narratives) that occurred in Louisiana from 2010 to 2016. This study has two major goals: 1) to develop a framework for applying machine learning models to classify injury levels from unstructured textual content, and 2) to apply an IML framework that provides probability measures of keywords and their association with the injury classification. The present study employed three machine learning algorithms in the classification of injury levels based on the crash narrative data. Of the used modeling techniques, the eXtreme gradient boosting (XGBoost) model shows better performance, with accuracy ranging from 0.70 to 24% for the training data and from 0.30% to 16% for the test data.

1. Introduction

Although there have been major improvements in vehicle safety, roadway design, and operations over time, the cost of traffic crashes in terms of injury and productivity loss is still unreasonably high. There has been extensive research conducted on crash frequency, injury severity, and the influencing factors of these measurements; nevertheless, there is still further work to be done. In the years 2014 to 2016, roadway departure (RwD) crashes (a crash in which a vehicle leaves the designated traveled way) resulted in an average of 18,779 fatalities per year, which make up approximately 53% of all traffic fatalities in the U.S. [1]. Tree and utility pole/other pole related (TUOP) crashes represent approximately 12 to 15% of all RwD fatal crashes in the U.S. The TUOP crashes in Louisiana represent 22% of all fatal crashes. Additionally, from 2015 to 2016, TUOP crashes increased overall by 5%. In many cases however, TUOP is not the first crash event. Therefore, further investigation of crash reports is needed to determine the primary contributing factor of the crash occurrence. From 2010 to 2016, 55,857 TUOP crashes were reported in Louisiana. Individually examining each crash report is not a realistic option. Innovative text mining tools are the best alternative tool for this research.

The traditional approach to studying crash factors is to identify the association between crash count and the road conditions, traffic characteristics, and driver behavior. Recently, research focus has been directed towards identifying the factors that majorly affect driver injury severity in traffic crashes. Multiple methodologies have been used previously to study injury severity. The traditional approach with data analysis procedures uses police-reported structured crash data to conduct crash data analysis. Police crash reports generally contain a written description of the crash occurrence, but in many cases, these crash narratives are not stored electronically. Furthermore, the narratives are in free-text data format, which requires expansive manual efforts in extracting data from them. Due to the difficulties associated with investigating these crash reports, there is a high likelihood of losing essential information from these narratives. TUOP crashes typically involve a series of crash events. Comprehending injury severity data and information is possible with text mining algorithms. This study has two main objectives: 1) to develop a context for applying machine learning models to classify injury levels from unstructured textual content, and 2) to apply an interpretable machine learning (IML) framework that can provide probability values of the keywords to classify injury types. This study

* Corresponding author.
E-mail addresses: s-das@tti.tamu.edu (S. Das), s-datta@tti.tamu.edu (S. Datta), zubaidih@oregonstate.edu (H.A. Zubaidi), obaidi@oregonstate.edu (I.A. Obaid).
Peer review under responsibility of International Association of Traffic and Safety Sciences.

https://doi.org/10.1016/j.iatssr.2021.01.001
0386-1112/

employed machine learning algorithms such as random forest (RF), support vector machine (SVM), and eXtreme Gradient Boosting (XGBoost) to conduct an analysis on a TUOP crash dataset spanning seven years (2010–2016) in Louisiana and classify each crash narrative according to crash injury levels. IML techniques were then applied to the narratives to explain the terms and their associations with the injury level classification.

2. Literature review

Conventional traffic safety analysis can be divided into two broad groups: crash count analysis and crash injury analysis. Interested researchers can consult three major survey papers to understand the state-of-the-art methods and future research scopes [2–4]. The key methodology of most studies in these areas is identifying numerous contributing factors and their relationships with crash occurrence or crash severity. This section divides the literature review into two major sections: 1) TUOP studies, and 2) crash narrative investigations.

2.1. TUOP studies

The National Highway Traffic Safety Administration published a report [5] stating that run-off-road (ROR) crashes (a subset of RwD crashes) represented approximately 65% of the total number of single-vehicle crashes in the U.S. from 1991 to 2007. Furthermore, the Federal Highway Administration published a study [6] on TUOP crashes, reporting that TUOP roadside hazards are rarely treated. The study also asserted that further extensive research is required to plan and manage countermeasure design to decrease the severity of TUOP crashes. In 1980 in the U.S., Jones and Baum examined the factors of utility pole crashes [7]. Another early study took place in South Australia and found that 40% of all crashes recorded at least one fatality in which the key harmful event is the collision with a roadside fixed object [8]. In another study, Nilsson et al. [9] explored countermeasure design in ROR crashes by conducting hierarchical agglomerative cluster analysis. Al-Bdairi et al. [10] applied ordered random parameter probit model to estimate the effects of a variety of variables in large-truck ROR crashes. In a follow-up study, Al-Bdairi et al. [11] applied mixed logit models to examine the effect of lighting conditions on large truck related ROR crashes. By applying logistic regression, Dissanayake and Roy [12] evaluated the impacts of key contributing crash factors in ROR crashes. In a later study, Dissanayake [13] researched the variables that affect injury severity of ROR single-vehicle crashes, focusing on crashes involving young drivers. In this study, the author found that neither severe weather conditions nor physical impairment (i.e., fatigue or illness) had a significant effect on the severity of single-vehicle ROR crashes. This discovery contradicted findings from previous studies [14,15]. Alruwaished [16] studied the factors of ROR crashes where the vehicle departed from the dedicated travel lane at a non-intersection area of the roadway and collided with either a fixed object or another vehicle. Another study utilized classification and regression tree (CART) approach to investigate variables that affect injury severity. The study found that human error was the most crucial factor in ROR and RwD crashes [17]. Watson et al. [18] explored the impacts of roadside fixed objects on traffic safety by analyzing crash indicator values for roadway segments where roadside fixed objects are present and segments where there are no roadside fixed objects. The Maine Department of Transportation conducted a study [19] and reported that the most significant variables that contribute to utility pole crashes include dark lighting condition, poor roadway condition, roadway curvature, rural environment, driver inattention, speeding, and the location and offset of utility poles. Dumbaugh [20] conducted an analysis of TUOP crash site locations and identified the crash factors in an urban environment. Many studies have also devoted their efforts to studying the safety effects of various countermeasures.

2.2. Incident narrative investigations

Text mining has demonstrated its usefulness in detecting valuable information from a large text based dataset. It is particularly used to find patterns and peculiarities in data, identify contributing factors, and develop predictive models used for guidance in real-world scenarios [21,22]. Previous studies employed different applications of natural language processing to gain insights from occupational incident reports [23–30], health care reports [31–33], automobile crash reports [24–33,44], and others [21,22].

In recent years, researchers started applying text mining for vehicle crash data analysis in the transportation research area. Chatterjee applied a connectionist-based model to classify free-text incident narratives [34]. Das et al. [35] applied both exploratory text mining and empirical Bayes (EB) data mining were applied in gaining associations between vehicle condition and automotive safety. Least Squares techniques was applied to predict railroad crash cost accurately and determine the contributing factors [41]. Previous researchers also used extensive interview data analysis [36], logistic regression [26,27,38,39], Haddon matrix [28,29,42], clustering [23,42], naïve Bayes [26,27,32], and latent Dirichlet allocation [43].

In summary, text mining and advanced models offer an extra potential to develop standardize incident narrative text analysis and reduce human error in crash and injury investigation. Table 1 provides a summary of previous studies on incident narrative reports.

3. Methodology

3.1. Data

The current study collected seven years, from 2010 to 2016, of traffic crash data from Louisiana. Louisiana crash data provides five consequent harmful events. The following filter is applied to identify the TUOP crashes:

Most harmful event = Tree (JJ) or utility pole (KK) or other poles (NN) OR First harmful event = Tree (JJ) or utility pole (KK) or other poles (NN) OR Second harmful event = Tree (JJ) or utility pole (KK) or other poles (NN) OR Third harmful event = Tree (JJ) or utility pole (KK) or other poles (NN) OR Fourth harmful event = Tree (JJ) or utility pole (KK) or other poles (NN)

After the filter was applied, the dataset contained 55,857 crash level data. The data summary shows that TUOP represents approximately 80% of the "most harmful" event scenarios. Furthermore, trees account for to 75% of TUOP crashes in the most harmful event scenarios. Table 2 displays yearly TUOP crashes by severity types (K = fatal, A = incapacitating injury, A = non-incapacitating injury, B = possible injury, O = no injury). Over that time period, the number of total crashes has increased by 6%. However, the number of TUOP fatal crashes has actually decreased by 8% from 2010 to 2016. Overall, in that time period, TUOP fatal crashes represent approximately 20% of all fatal crashes in Louisiana.

3.2. Machine learning algorithms

Machine learning is a technique of training machines to learn patterns and associations from data and to make forecasts based on the knowledge learned throughout the training process. There are two major types of machine learning: supervised learning and unsupervised learning. A machine learning algorithm is essentially a set of rules that the machines must follow in order to learn and achieve a specific goal. A machine learning model can predict, classify or can fulfill other intentions based on the problem type. In contrast to machine learning, standard statistical modeling uses statistical equations to discover links between variables. The most significant advantage of conventional statistical modeling is its interpretability. However, one shortcoming of this method is its pre-determined assumptions must be made

Table 1
Studies on Incident Narrative Reports.

Area of study	Dataset	Approach	Key findings	Ref
Research collaboration and funding	14,000 project information from Research in Progress database by U.S. Department of Transportation (DOT) and State DOTs	Neural network	Agency's interest in various subject areas varies every year.	[21]
Document collection	9/11 attack report and aviation accident reports from 2001 to 2003 by the U.S. National Transportation Safety Board (NTSB)	Concept chain queries	Find the most crucial evidence trails across documents to describe connections between two topics of interest.	[22]
Occupational accidents	143 serious occupational accident with movement disturbance (OAMD) scenarios in construction and metallurgy sectors	Bayesian network model; Clustering	Extracted eight scenarios and 30 accident generic factors from 143 serious occupational accident narrative texts.	[23]
Occupational accidents	17,000 injury narratives between 2002 and 2004 extracted from workers' insurance claim	Fuzzy and naïve Bayesian models	Classification accuracies by the models compared to manual efforts	[24]
Occupational accidents	535,605 injury narratives regarding work-related ladder fall fractures.	Descriptive statistics	Provide relevant additional information on case identification, mechanisms, causes and outcomes for a severe injury.	[25]
Occupational accidents	7200 accident narratives extracted from compensation claims from 2001 to 2009 from Ohio Bureau of Workers' Compensation (OHBWC) database	Naïve Bayes; Logistic regression model	The logistic model performed better than the naïve Bayes model. Inclusion of two-word sequences as opposed to single keywords marginally improved the overall accuracy	[26]
Occupational accidents	30,000 injury narratives extracted from workers' compensation (WC) claim	Naïve Bayes; Support Vector Machine (SVM); Logistic regression model	SVM performs better than other models for large dataset. Classify the large set of claim data into Bureau of Labor Statistics (BLS) OIICS event codes.	[27]
Occupational accidents	4000 injury reports during the construction of Denver International Airport (DIA)	Haddon matrix	Developed coded dataset and a set of coding rules from injury reports to help reviewers interpret the narrative text.	[28]
Occupational accidents	69 Kentucky Fatality Assessment and Control Evaluation (FACE) agricultural tractor fatality reports from 1994 to 2004	Haddon matrix; Univariate and multivariate logistic regressions model	Identify influential factors to tractor fatalities	[29]
Occupational accidents	National datasets of occupational fatalities Australia, United States and New Zealand	Text search technique	Narrative coding was more useful for some types of injury than others.	[30]
Health care	Narrative text of traumatic brain injuries (TBIs) in the National Electronic Injury Surveillance System (NEISS)	DUALIST- an interactive machine learning program	DUALIST reduces time frame from a few days to minutes after nearly sixty minutes of training.	[31]
Health care	30,000 injury narratives extracted from workers' compensation (WC) claim	Naïve Bayesian model	Overall accuracy is 87%; positive predictive values across all two-digit BLS event categories.	[32]
Health care	Electronic health records from the UK General Practice Research Database (GPRD)	Semi-supervised Set Covering Machine (S3CM) model; Transudative Vector Support Machine (TVSM).	S3CM works better than TVSM and fully supervised SCM for coronary angiogram detection.	[33]
Vehicle crashes	3680 accident reports	Connectionist based model; Fuzzy Bayes model, Keyword model	Connectionist and fuzzy Bayes model better performed than keyword model	[34]
Vehicle defects	National Highway Traffic Safety Administration's (NHTSA) vehicle complaint and Fatality Analysis Reporting System (FARS)	Exploratory text mining; Empirical Bayes (EB)	Several key association-patterns	[35]
Vehicle crashes (inattention and distraction)	856 crash report from Australian National Crash In-depth Study during the period of 2000 to 2011	In-depth analysis using interviews and external verifications	Classify five types: restricted attention, incorrectly prioritized attention, neglected attention, cursory attention, and diverted attention.	[36]
Rail crashes	Rail accidents report from 2001 to 2012 maintained by Federal Railroad Administration (FRA)	Probabilistic Topic Modeling; Latent Dirichlet Allocation (LDA); Random Forest (RF); Partial Least Squares	Text mining can improve understanding of the contributing factors.	[37]
Vehicle crashes (speeding)	Crash narrative texts from 2012 to 2014 by the state of Massachusetts DOT	Logistic regression model	Prediction accuracy is 53%.	[38]
Vehicle crashes (bicycle)	Bicycling and other sports injury narratives by U.S. National Electronic Injury Surveillance System (NEISS), 2005–2011	Text algorithm, Logistic regression model	Demonstrated the possible use of simple text-search algorithms to detect supplementary variables in unstructured data.	[39]
Vehicle crashes	Crash report data in Queensland from year 2004 and year 2005	Clustering; Leximancer- a tool based on the Bayesian theory	Higher likelihood of the second vehicle being involved in a crash; a right-turn crash as opposed to a left-turn crash; person's inability to control speed or stop resulting in a rear-end crash; and multiple vehicles being involved in an intersection crash.	[40]
Vehicle crashes	Accident narrative texts from Liberty Mutual Insurance Company in 1991	Text mining; Haddon injury epidemiology model	About 26% of the crashes involved a stopped or slowing vehicle in the work zone. With 31% of the crashes, rear-ends were the most common crashes.	[42]
Rail crashes	Railroad equipment accident 2005–2015	LDA	Both LDA and clustering produced equivalent results	[43]

in the first place, which could be questionable. On the other hand, majority of the machine learning model lacks substantial amount of interpretability.

3.2.1. RF

Random forest, a supervised classification algorithm, generates a comprehensive set of decision trees by subsetting the training dataset and accumulates the prediction from each tree using voting. A single decision tree works well on the training dataset. However, it cannot always provide an accurate prediction. Random Forest is built on several decision trees with randomly selected samples and variables [45,46]. To create a random forest, there are several steps:

(1) Create a bootstrapped dataset by randomly selecting several samples in the original dataset. Each sample can be selected more than once.
(2) Create a decision tree using the bootstrapped dataset, but, at each node, only select a random subset of variables (columns)

Table 2
Crash Injury Counts by Severity Levels.

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

to determine which one is the best one to separate the samples.

(3) Repeat step (1) to (2) many times. This will result in several decision trees that form a random forest.
(4) For the test dataset, run each sample through all decision trees in the random forest and summarize the results. The final prediction of a test sample is determined by the result with the most vote from RF.

The parameters in RF are the total number of trees, the number of randomly selected variables at a node split, and the maximum tree depth. The reason for randomly selecting a subset of the variables is to avoid the correlation of the trees: if one or a few features are very strong predictors for the target, these features will be selected to split examples into many trees. This will result in many correlated trees in the random forest. One of the advantages of RF is that by building multiply samples of the original dataset, it prevents overfitting problems. The algorithm is effective in handling missing values.

3.2.2. SVM

Support Vector Machine is based on finding an optimum hyperplane in a high dimensional space. In this case, a hyperplane is a classifier to classify data into two groups. The goal is to select an optimal hyperplane with the maximum possible margin between two groups or classes. Margin is the closest distance between the nearest data points from two sides. The maximum possible margin provides chances of higher prediction accuracies [47]. The equation of the hyperplane can be written as:

$$wx - b = 0$$

For x_i in one class, it should satisfy:

$$wx_i - b \leq -1$$

For x_i in the other class, it should satisfy:

$$wx_i - b \geq 1$$

w is a real-valued vector of the same dimensionality as the input feature vector x. b is a real number.

Margin can be calculated as:

$$\frac{2}{\|w\|}$$

To get the maximum margin, $\|w\|$ should be minimized. However, in the case where many outliers exist in the opposite class, if the hyperplane is still drawn based on the nearest data points from two sides, many future data will be classified into the wrong class. Here we need a soft margin that can tolerate some outliers be classified into the opposite class. Then, for further data, they will be less likely to be classified into the wrong class. The hyperplane is drawn at the center of this soft margin. There are many methods, including minimizing cost function and cross validation, to determine the optimum soft margin. With respect to multiclass problem, SVM uses different kernel functions to transform data into a higher dimension that can be classified into two groups where a hyperplane can be found. The advantages of SVM classifier is that it can handle outlier and overlapping problems. Moreover, when the original dataset does not have a clear boundary which can divide the data into two classes. SVM provides a good solution by transforming the dataset into a higher dimension.

3.2.3. XGBoost

XGBoost is one of the most popular Gradient Boosting method. Gradient boosting is a regression model to make predictions based on several decision trees. The first step of gradient boost is to select an initial prediction f_0. Assuming there is a dataset $\{(x_i, y_i)\}_{i=1}^{n}$, the initial prediction is the average of y_i, Then a new label will be added into the training dataset which is called pseudo residual $\widehat{y_i}$ and is calculated as: $\widehat{y_i} = y_i - f_0$.

Then, use the training dataset with pseudo residuals to build a new decision tree model, f_1. Now the new prediction for each sample becomes to $f = f_0 + af_1$. Where a is the learning rate.

Then, repeat the above process and get the second decision trees f_2 based on the pseudo residuals. Then the new prediction becomes to $f = f_0 + af_1 + af_2$. Repeat this process until the maximum number of decision trees [48,49].

For XGBoost algorithm, different from gradient boosting, it creates a XGBoost tree at each step. At each root of a XGBoost tree, a similarity score is calculated as:

$$\frac{(Sum\ of\ residual)^2}{number\ of\ residual + \lambda}$$

λ is regularization parameter to prevent overfitting. The default value is 0.

At each root, the algorithm picks the optimum threshold by choosing the one which results in the maximum Gain. Gain is calculated as: *Gain = sum of the similarity scores at all leaves − similarity score at the root.*

After reaching the maximum number of leaves, the next step is to prune the XGBoost trees. A parameter Gamma (user defined tree complexity parameter) is introduced. For the roots where Gain is smaller than gamma, this root will be deleted. Parameter Gamma is used to prevent overcomplexity. Finally, for all remaining leaves in a XGBoost tree, the output value is calculated as

$$\frac{(Sum\ of\ residual)^2}{number\ of\ residual + \lambda}$$

New prediction is calculated as:

$$new\ prediction = previous\ prediction + a{*}output$$

Then based on the new residual, create a new XGBoost tree. Repeat this process until reaching the maximum number of trees or new residual is very small.

3.3. Interpretable machine learning (IML)

The supervised machine learning (IML) algorithms learn from the training data. They are trained to predict based on the training cases. In a low-risk scenario such as a movie recommendation system, a wrong prediction is acceptable. On the other hand, the classification task for a self-driving car to identify a pedestrian is an example of a high-risk task. In many cases, a wrong prediction for high-risk tasks is not affordable. Therefore, there is a need to know how the model made its decision. Interpretability provides justification of the developed models. Moreover, it is important to know about the biases in the model obtained from the training data. Interpretability works as a debugging tool to identify the biases. Better interpretability opens options to fix the model or to make it more robust [50–52].

3.4. Crash narrative framework

The current study developed an IML framework in solving the research problem. The steps are the following:

- Step 1: Data compilation. This study collected electronic format of the crash narrative data in a structured dataset in spreadsheet. In the recent years, some U.S. states have started the procedure of transferring the paper crash reports into electronic versions. For example, Louisiana now maintains an electronic database of crash reports.
- Step 2: Data cleaning. Text mining algorithms are used to perform data cleaning. Redundant words can be removed with available lexicons. However, availability of transportation safety specific dictionaries would be helpful in reducing noise from textual contents associated with transportation safety.
- Step 3: Apply predictive modeling. Many recent studies have applied different machine learning tools for solving the classification problem. The best model can be selected by evaluating various machine learning models according to their misclassification rates. This study evaluated three different machine learning models to select the best option.
- Step 4: Apply IML model. Several IML techniques have been introduced in recent years such as partial dependence plot (PDP), and local interpretable model-agnostic explanations (LIME) [53,54]. This study employed LIME to explain some randomly selected crash narratives.

4. Results and discussions

The final dataset used for this analysis contains 23,416 TUOP crashes (around 50% of all TUOP crashes) and their crash narratives (see Table 3). However, the database has many missing values in the crash narrative columns. Additionally, some entries contain redundancies in the crash narrative columns (for example, 'see supplement narrative', 'see attached narrative', 'see narrative supplements'), which were subsequently excluded in the final dataset preparation.

Standard text mining steps such as removal of stop words and punctuations, removal of numbers, and lemmatization, before applying the machine learning algorithms to reduce noise in the dataset. The most common issue in text mining is the abundance of redundant information in the form of words. Domain specific lexicon is needed to reduce the number of redundancies and the associated noise. Furthermore, words or word fragments with similar meanings should be condensed into the same term (known as lemmatization) to reduce misclassifications. The final dataset was prepared with the performance of a basic redundant word removal. Future studies can perform further robust data cleaning to improve model precision.

As mentioned earlier, this study used three different machine learning algorithms to perform the analysis. The researchers provide a brief explanation of these algorithms. Table 4 lists the misclassification or accuracy rates for each algorithm. The training and test data are used from the dataset described in Table 3. The values in Table 4 show that the prediction accuracies are higher in XGBoost algorithm. The accuracy of the XGBoost algorithm ranges from 0.70 to 24% for the training data and 0.30 to 16% for the test dataset. The prediction accuracies in the test data are usually smaller than the accuracies of the training dataset as the test data was not included while developing the models. This study can be considered as a starting point of IML implementation in crash narrative analysis. The prediction accuracies can be improved by using additional structured information such as collision type, number of vehicles, final location of the vehicles and other relevant information. The current analysis is limited to use only crash narrative information to determine the injury type. Additionally, imbalanced data issue was not handled in this study to provide the accuracy results from the original number of crash reports based on different injury levels.

Table 3
Dataset with crash narratives.

Dataset	Count	K	A	B	C	O
Train data	15,610	169	411	2224	3994	8812
Test data	7806	87	210	1069	2013	4427
Total	23,416	256	621	3293	6007	13,239

Table 4
Accuracies by different algorithms.

Dataset	Severity type	XGBoost	SVM	RF
Train data	K	0.007	0.002	0.000
	A	0.027	0.013	0.023
	B	0.135	0.130	0.091
	C	0.240	0.221	0.203
	O	0.206	0.203	0.201
Test data	K	0.003	0.000	0.000
	A	0.013	0.005	0.000
	B	0.080	0.031	0.023
	C	0.163	0.110	0.117
	O	0.159	0.109	0.101

It is important to note that confusion matrix provides understanding of the model performance in the form of misclassification measures. On the other hand, model explanation provides justification of the models in the form of interpretability. This paper employed a recent IML algorithm, known as LIME, to provide contexts of the prediction results [53]. Open source R package 'lime' was utilized in this study [54]. IML provides interpretation of randomly selected cases with some explanation parameters such as probability and explanation fit. These measures are different from accuracies used in Table 4. Table 4 indicates overall performance measures of the used algorithms. The explanation parameters in Fig. 1 indicates the performance of the algorithm for a randomly selected case. For instance, six cases of non-debilitating crash narratives were selected from the test data (see Fig. 1). The interpretation of each of six random cases in Fig. 1 is below:

- Case: indicates randomly selected case number.

- Label: 1 indicates that the XGBoost model calculates that the crash is non-incapacitating injury and 0 implies that the crash is not non-incapacitating injury.
- Probability: indicates probability of a case to be a particular label (either 1 or 0) by XGBoost
- Explanation Fit: indicates the best fit measures for that particular case. The higher is the value, the better is the fitness.

The explanation models considered the top twenty words with prediction probability measures. The explanation plots only show top four keywords and their probability measures in the form of bar plots. Fig. 1 shows that only Case 2 is a non-incapacitating crash. For example, Case 1 illustrates that the probability is 60% as non-incapacitating. The words in the crash narrative identify the injury level are 'hospital,' 'bleeding,' 'head,' and 'highway.' Compared to the term 'highway,' the other three words perform better in classifying the injury level. Case 2 illustrates that the probability is 53 as non-incapacitating injury crash. The words with high probability measures are 'transported,' 'hospital,' 'lady,' and 'damage.' The general explanation comes from some of the key words and their associations with a non-incapacitating crash. The terms 'damage,' 'notified,' and 'advised' are associated with the outcomes of non-incapacitating crash narratives. Conversely, some terms like transported and hospital are making balance in determining the

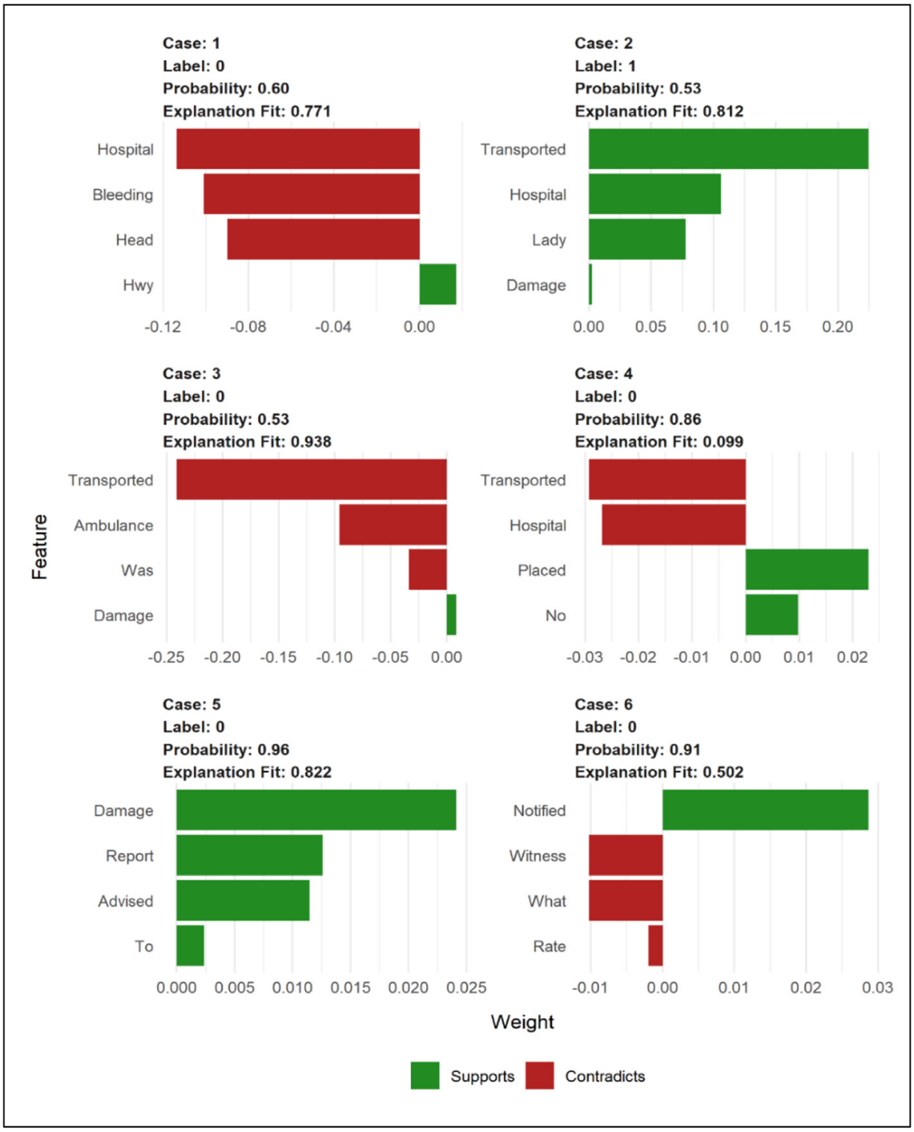

Fig. 1. Explanations of six randomly chose crash narratives with non-incapacitating severity (label = 1).

weightage towards the classification determination. The findings of this study are in line with a recent study on TUOP crashes [55].

5. Conclusions

This study considered all TUOP crashes (with crash narratives in electronic format) in Louisiana over a seven-year period (2010–2016). The current research has two main contributions: 1) it developed a applied framework to utilize machine learning models to categorize crash injury types from unstructured crash reports, and 2) it developed an IML framework that can interpretations on the developed algorithms so that users understand the key factors that are associated with the injury levels. The framework created in this study has the potential to be utilized in other traffic crash related categorization (i.e., type of collision) based on crash narratives. Many of the crash narratives that were studied were previously unused. This study also demonstrated that, using the XGBoost model, these crash narratives can be used to identify injury severity with an accuracy rate ranging from 0.30 to 24%. The present study should be used as a jumping off point for future studies that may implement IML and use it as a key research tool for future crash narrative data analysis.

The current study design does have limitations, including at least two major ones that should be addressed. Firstly, the injury prediction accuracy rates are not high due to the standalone usage of crash narrative report contents. Future studies should focus on the development of a robust transportation safety lexicon on the stop words and dominant/redundant words to help reduce classification errors. Additionally, prediction accuracies can be improved with inclusion of additional metadata from the structured crash information. Secondly, the current paper lacks explanations and descriptions of all injury levels related to TUOP crashes. The current research team will continue to pursue research focusing on data extraction from crash narrative data of TUOP crashes.

Declaration of Competing Interest

The author "Subasish Das" has the affiliation with "Texas A&M Transportation Institute". All the other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

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 2014 (Accessed: July 27, 2018).

[2] 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.

[3] 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.

[4] F. Mannering, C. Bhat, Analytic methods in accident research: methodological frontier and future directions, Analytic Methods in Accident Research, 1, 2014, pp. 1–22.

[5] U.S. Department of Transportation, National Highway Traffic Safety Administration, Run-Off-Road Crashes: An On-Scene Perspective, Report no. DOT HS 811 500 2011.

[6] U.S. Department of Transportation, Federal Highway Administration, Noteworthy Practices: Roadside Tree and Utility Pole Management, Report no. FHWA-SA-16-043 2016.

[7] I.S. Jones, A.S. Baum, An analysis of the urban utility pole accident problem, Publication FHWA-RD-79-117, FHWA, U.S. Department of Transportation, 1980.

[8] 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.

[9] D. Nilssona, 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.

[10] 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.

[11] 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. Part A 144 (1) (2018).

[12] S. Dissanayake, U. Roy, Crash severity analysis of single vehicle run-off-road crashes, J. Transp. Technol. 4 (2014) 1–10.

[13] S. Dissanayake, Young drivers and run-off-the-road crashes, Proceedings of the 2003 Mid-Continent Transportation Research Symposium, Ames, Iowa, August 2003.

[14] U.S. Department of Transportation, National Highway Traffic Safety Administration, Factors Related to Fatal Single-Vehicle Run-Off-Road Crashes, Report no. DOT HS 811 232 2009.

[15] C. Liu, R. Subramanian, Factors related to fatal single-vehicle run-off-road crashes, Report DOT HS 811 232, 2009.

[16] 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.

[17] 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.

[18] D.C. Watson Jr., A. Al-Kaisy, N.D. Anderson, Examining the effect of speed, roadside features, and roadway geometry on crash experience along a rural corridor, J. Modern Transp. 22 (2014) 84–95.

[19] B. Marquis, Utility pole crash modeling, Report No. ME 00-8, Maine Department of Transportation, Transportation Research Division, 2001.

[20] E. Dumbaugh, Design of Safe Urban Roadsides: An Empirical Analysis, Transportation Research Record: Journal of the Transportation Research Board, 2006 74–82.

[21] K. Gopalakrishnan, S.K. Khaitan, Text mining transportation research grant big data: knowledge extraction and predictive modeling using fast neural nets, Int. J. Traffic Transp. Eng. 7 (2017) 354–367.

[22] W. Jin, R.K. Srihari, H. Hay Ho, X. Wu, Improving Knowledge Discovery in Document Collections through Combining Text Retrieval and Link Analysis Techniques, 2007.

[23] F. Abdat, S. Leclercq, X. Cuny, C. Tissot, Extracting recurrent scenarios from narrative texts using a Bayesian network: application to serious occupational accidents with movement disturbance, Accid. Anal. Prev. 70 (2014) 155–166.

[24] H. Marucci-Wellman, M. Lehto, H. Corns, A combined fuzzy and Naïve Bayesian strategy can be used to assign event codes to injury narratives, Injury Prev. 17 (2011) 407–414.

[25] G.S. Smith, R.A. Timmons, D.A. Lombardi, D.K. Mamidi, S. Matz, T.K. Courtney, M.J. Perry, Work-related ladder fall fractures: identification and diagnosis validation using narrative text, Accid. Anal. Prev. 38 (2006) 973–980.

[26] S.J. Bertke, A.R. Meyers, S.J. Wurzelbacher, A. Measure, M.P. Lampl, D. Robins, Comparison of methods for auto-coding causation of injury narratives, Accid. Anal. Prev. 88 (2016) 117–123.

[27] K. Vallmuur, H.R. Marucci-Wellman, J.A. Taylor, M. Lehto, H.L. Corns, G.S. Smith, Harnessing information from injury narratives in the 'big data' era: understanding and applying machine learning for injury surveillance, Injury Prev. 22 (2016) 34–42.

[28] Bondy, J., H. Lipscomb, K. Guarini, and J. E. Glazner. Methods for using narrative text from injury reports to identify factors contributing to construction injury. Am. J. Ind. Med. 48: 373–380.

[29] T.L. Bunn, S. Slavova, L. Hall, Narrative text analysis of Kentucky tractor fatality reports, Accid. Anal. Prev. 40 (2008) 419–425.

[30] A. Williamson, A.-M. Feyer, N. Stout, T. Driscoll, H. Usher, Use of narrative analysis for comparisons of the causes of fatal accidents in three countries: New Zealand, Australia, and the United States, Injury Prev. 7 (2001) 15–20.

[31] W. Chen, K.K. Wheeler, S. Lin, Y. Huang, H. Xiang, Computerized "learn-as-you-go" classification of traumatic brain injuries using NEISS narrative data, Accid. Anal. Prev. 89 (2016) 111–117.

[32] H.R. Marucci-Wellman, M.R. Lehto, H.L. Corns, A practical tool for public health surveillance: semi-automated coding of short injury narratives from large administrative databases using Naïve Bayes algorithms, Accid. Anal. Prev. 84 (2015) 165–176.

[33] Z. Wang, A.D. Shah, A.R. Tate, S. Denaxas, J. Shawe-Taylor, H. Hemingway, Extracting diagnoses and investigation results from unstructured text in electronic health records by semi-supervised machine learning, PLoS One 7 (2012), e30412.

[34] S. Chatterjee, A Connectionist Approach for Classifying Accident Narratives, Theses and Dissertations Available from ProQuest 1998 1–345.

[35] S. Das, A. Mudgal, A. Dutta, S.R. Geedipally, Vehicle consumer complaint reports involving severe incidents: mining large contingency tables, Transportation Research Record: Journal of the Transportation Research Board 7 (2018).

[36] V. Beanland, M. Fitzharris, K.L. Young, M.G. Lenné, Driver inattention and driver distraction in serious casualty crashes: data from the Australian national crash in-depth study, Accid. Anal. Prev. 54 (2013) 99–107.

[37] D.E. Brown, Text mining the contributors to rail accidents, IEEE Trans. Intell. Transp. Syst. 17 (2016) 346–355.

[38] C.D. Fitzpatrick, S. Rakasi, M.A. Knodler, An investigation of the speeding-related crash designation through crash narrative reviews sampled via logistic regression, Accid. Anal. Prev. 98 (2017) 57–63.

[39] J.M. Graves, J.M. Whitehill, B.E. Hagel, F.P. Rivara, Making the most of injury surveillance data: using narrative text to identify exposure information in case-control studies, Injury 46 (2015) 891–897.

[40] R. Nayak, N. Piyatrapoomi, J. Weligamage, Application of text mining in analysing road crashes for road asset management, Eng. Asset. Lifecycle Manag. (2010) 49–58.

[41] K.M. Pollack, N. Yee, M. Canham-Chervak, L. Rossen, K.E. Bachynski, S.P. Baker, Narrative text analysis to identify technologies to prevent motor vehicle crashes: examples from military vehicles, J. Saf. Res. 44 (2013) 45–49.

[42] G.S. Sorock, T.A. Ranney, M.R. Lehto, Motor vehicle crashes in roadway construction Workzones: an analysis using narrative text from insurance claims, Accid. Anal. Prev. 28 (1996) 131–138.

[43] Trefor P. Williams, John F. Betak, Identifying Themes in Railroad Equipment Accidents Using Text Mining and Text Visualization. International Conference on Transportation and Development, 2016.

[44] S. Das, M. Le, B. Dai, Application of machine learning tools in classifying pedestrian crash types: a case study, Transp. Saf. Environ. 2 (2) (2020) 106–119.

[45] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.

[46] T.K. Ho, The random subspace method for constructing decision forests, IEEE Trans. Pattern Anal. Mach. Intell. 20 (1998) 832–844.

[47] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, New York, 2011.

[48] J.H. Friedman, Greedy Function Approximation: A Gradient Boosting Machine, JSTOR, Annals of Statistics, 2001 1189–1232.

[49] J.H. Friedman, B. Popescu, Predictive Learning Via Rule Ensembles. The Annals of Applied Statistics, JSTOR, 2018 916–954.

[50] A. Fisher, C. Rudin, F. Dominici, Model Class Reliance: Variable Importance Measures for any Machine Learning Model Class, From the 'Rashomon' Perspective, http://arxiv.org/abs/1801.01489 2018.

[51] C. Molnar, Interpretable Machine Learning: A Guide for Making Black Box Models Explainable, https://christophm.github.io/interpretable-ml-book/ 2018 Accessed July 27, 2018.

[52] M.T. Ribeiro, S. Singh, C. Guestrin, Why should I trust you? Explaining the predictions of any classifier, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2016.

[53] S. Lundberg, S. Lee, An Unexpected Unity Among Methods for Interpreting Model Predictions, 2016, 2016. http://arxiv.org/abs/1611.07478 Accessed Sept 2, 2020.

[54] T. Pedersen, M. Benesty, lime: Local Interpretable Model-Agnostic Explanations. R Package Version 0.4.0, https://CRAN.R-project.org/package=lime Accessed Sept 2, 2020.

[55] S. Das, B. Storey, T. Shimu, S. Mitra, M. Theel, B. Maraghepour, Severity Analysis of Tree and Utility Pole Crashes: Applying Fast and Frugal Heuristics, IATSS Research, 2020.