Single and Multiple Vehicle Crashes due to Hydroplaning: A Case Study of Interpretable Machine Learning

Subasish Das, Ph.D.
(Corresponding author)
Associate Transportation Researcher, Texas A&M Transportation Institute
1111 RELLIS Parkway, Room 4414, Bryan, TX 77807
Email: s-das@tti.tamu.edu
ORCID: 0000-0002-1671-2753

Anandi Dutta, Ph.D.
Senior Lecturer, Computer Science and Engineering Dept., Ohio State University
2015 Neil Ave, Columbus, OH 43210
E-mail: dutta.34@osu.edu

Kakan Dey, Ph.D., P.E.
Assistant Professor, Civil and Environmental Engineering, West Virginia University
1374 Evansdale Drive, Morgantown, WV 26506
Email: kakan.dey@mail.wvu.edu

Mohammad Jalayer, Ph.D.
Assistant Professor, Department of Civil and Environmental Engineering
Rowan University, Glassboro, NJ 08028
Email: jalayer@rowan.edu

Abhisek Mudgal, Ph.D.
Assistant Professor, Department of Civil Engineering, Indian Institute of Technology (BHU)
Varanasi, Uttar Pradesh, India-221005
E-mail: abhisek.civ@iitbhu.ac.in

TOTAL WORDS: 6,427 words

5,927 words = text (including abstract and references)

500 = 2 tables

Submitted on August 1, 2019

ABSTRACT

Although hydroplaning is a major contributor to roadway crashes, it is not typically reported by police in crash databases. Hence, a framework to classify various crash attributes from police reports and identify hydroplane crashes is strongly needed. In this study, the research team applied text mining algorithms to seven years (2010-2016) of crash data from the Louisiana traffic crash database to identify hydroplaning related crashes. The two major objectives of this study are the following: 1) to develop a framework for applying machine learning models to unstructured textual content for classifying the number of vehicle involvements in a crash, and 2) to apply an interpretable machine learning framework that evaluates the effectiveness of keywords in determining the classification. The project team applied three machine learning algorithms. Of these algorithms, the XGBoost model was the most effective classifier. The proposed approach would be beneficial in improving road geometry and vehicles and in developing countermeasures that promote the overall safety of road users.

Keywords: hydroplaning crash, crash narrative, machine learning, precision, recall.

INTRODUCTION

Hydroplaning is a dangerous phenomenon that can cause roadway crashes and result in severe injuries or fatalities. The term hydroplaning is often used by roadway users to describe a wide variety of wet-roadway driving hazards. However, the more accurate definition of hydroplaning describes a specific condition of tire skidding on wet pavement in which the automobile tires float on a layer of water at a high speed, causing the driver to lose control of the vehicle. This occurs when the water underneath the tire cannot escape and hydrodynamic pressure builds up to sustain the floating tire over the area. Some common contributing factors of hydroplane crashes include vehicle speed, tire pressure, and tread depth, pavement texture, roadway and environmental conditions *(1)*.

Previous research on hydroplaning crashes has followed conventional methods to establish relationships between crash frequency and key contributing factors, such as the ones mentioned above. Many recent studies have focused on determining which factors significantly affect different crash characteristics. Conventional data analysis methods perform crash frequency or injury severity analysis on police-reported crash data. However, these reports usually contain a textual description of the crash event that has not been explored. Information from these narratives can require significant effort to understand the event circumstances as they are not stored electronically and they are presented as unstructured or semi-structured free-text data format. Because of this, precision information can be lost through manual and time-consuming interpretation of these text reports.

Text mining has become an increasingly popular research area since it is effective in identifying valuable patterns and hidden insights from plain texts. This study used text mining and machine learning algorithms to investigate hydroplaning crashes and develop text mining-based models to determine key crash contributing factors such as road, vehicle, and environmental conditions. In this study, the research team aims to evaluate various text mining classification techniques by measuring their ability to classify crash narratives for seven years of crash data obtained from Louisiana traffic crash database. The study evaluated three machine learning algorithms: support vector machine (SVM), random forest (RF), and XGBoost.

The structure of this paper is as follows: the literature review provides a comprehensive overview of previous studies focused on hydroplane crashes. The paper then introduces the modeling tools used in the study. The next sections demonstrate the data preparation and application of text mining techniques. Finally, this paper describes the results of this evaluation, followed by the study conclusions and discussion.

LITERATURE REVIEW

A majority of the studies included in this review focus on theoretical concept development for hydroplaning occurrences. There are a limited number of studies that aim to determine the key contributing factors of hydroplaning related crashes.

Studies on Hydroplaning Crashes

In 1976, Enustun *(2)* conducted a study that aimed to avoid the recurrence of identified hydroplaning crashes. To achieve this, Enustun identified the locations of crashes associated with hydroplaning to examine the road, vehicle, and weather conditions related to hydroplaning. The study concluded that true hydroplaning crashes on highways were rare and made up less than five percent of wet-pavement crashes. Enustun suggested that a greater effort be made to educate the public about the degree of tire capability deterioration on wet pavements under rapid speeds.

Black et al. (*3*) conducted a study examining the phenomenon of hydrodynamic drag, which is virtually anonymous to highway and transportation engineers. Aycock *(4)* investigated a 2003 hydroplaning crash on an interstate highway in Georgia and concluded that the speed of the vehicle, the condition of the tires, and the depth of the water on the roadway all contributed to the hydroplaning crash. Gunaratne et al. *(5)* developed statistical models that estimated wet weather speed reduction as well as analytical and empirical methods to predict hydroplaning speeds of trailers and heavy trucks. Using crash data, geometrical data, pavement condition data, and other relevant information available for Florida roadways, the study conducted a wet weather crash analysis. The results of this study indicated that wider sections increased the likelihood of hydroplaning crashes, and dense-graded pavements were more likely to induce conditions fit for hydroplaning than open-graded ones. In a follow-up study, Jayasooriya and Gunaratne (*6*) developed statistical models to estimate the thickness of water film that develops on roadways and classified the corresponding threshold hydroplaning speed into empirical, analytical, and numerical categories. Zhou et al. (*7*) conducted a study to measure the impacts of automated vehicles (AV) on roadway hydroplaning and pavement life in comparison to human-driven vehicles. The results showed that a uniformly distributed lateral wandering pattern for AVs prolonged pavement fatigue life, decreased pavement rut depths, and reduced hydroplaning potential.

Text Mining Studies on Crash Narratives

Text mining, an extremely useful data analysis technique, has been used to extract valuable information from large text-based datasets. Text mining methods can identify patterns and anomalies in data over time, determine key contributing factors, and develop predictive models to solve real-world problems (*8*, *9*). In-depth text mining has primarily been used in crash and injury analysis to gain insights from occupational crash reports (*10*-*17*), health care reports (*18*-*20*), automobile crash reports (*11*-*20*), and others (*8*-*9*).

Concept Chain Queries, a special case of text mining, is used to identify essential evidence trails across documents to explain relationships between two topics of interest (*9*). Researchers have developed a text search technique to explore the utility of crash narrative text analysis for generating code to determine injury mechanisms (*17*). Haddon’s matrix provided a conceptual framework for researchers to code text from work-related injury reports to identify contributing factors of these injuries (*15*). In another study, Haddon’s matrix separated the fatal incident into three event phases (pre-event, event, post-event) and provided a coded data set as well as coding rules (*16*).

Although the crash data are small and fairly homogenous, it is challenging to recognize patterns within the narrative text. A combined Naïve-Fuzzy Bayesian approach can be used to provide a more accurate narrative classification and to identify the data that require manual review, thus reducing the burden on human coders (*10*, *19*, *26*). Researchers have also utilized DUALIST to easily classify data; DUALIST is an online interactive program that allows novice users to quickly classify thousands of narratives after approximately sixty minutes of training (*8*). Studies have also evaluated the effectiveness of the Bayesian-based model in comparison to other machine learning algorithms. The Bayesian-based model was compared to the neural network (*8*), logistic regression model (*28*, *29*), and support vector machine model (*29*), all of which provided higher accuracy in classifying the emerging causes of occupational injury.

Another study compared a Semi-Supervised Set Covering Machine (S3CM) learning algorithm to the Transudative Vector Support Machine (TVSM), the original fully supervised Set

Covering Machine (SCM), and 'Freetext Matching Algorithm' natural language processor. The S3CM algorithm was developed to detect the presence of coronary angiograms and ovarian cancer diagnoses from electronic health record narratives. The model does not rely on linguistic rules; it worked effectively after being trained with pre-classified test data sets. The study found that S3CM results were better than TVSM and fully supervised SCM (*30*).

Text mining has become an increasingly popular crash analysis method in the area of transportation engineering. A study demonstrated a connectionist-based model to classify free-text crash descriptions (*21*); singular value decomposition was used for feature extraction and network training. This model was evaluated in comparison to a fuzzy Bayes model and a keyword-based model. All three models were applied to human classified data, and the study found that the connectionist and fuzzy Bayes model outperformed the keyword model. Another study performed exploratory text mining and empirical Bayes (EB) data mining to understand the associations between vehicle condition and automotive safety (*22*). These methods were used to identify key crash causing factors in terms of vehicular manufacturing defects (e.g., airbags, brake system, seat belts, and speed control).

Researchers developed a three-level hierarchical Bayesian model, named Latent Dirichlet Allocation (LDA), to identify major recurring crash topics from textual data compiled from the Federal Railroad Administration (FRA) reports. The researchers then applied a text clustering method to the text using Jigsaw text visualization software and found that both methods had equal effects. In another study, LDA, Random Forest, and Partial Least Squares techniques were combined and applied to identify the contributing factors and to accurately predict the cost of railroad crashes (*28*). Researchers also used logistic regression (*25, 26, 30*), clustering (*30*), and other computational tools (e.g., Statistical Analysis System (SAS) and Leximancer) (*27, 29*) to identify vehicle crash factors. In summary, computerized approaches and predictive models can be used to standardize crash narrative text analysis and reduce human error in crash and injury surveillance.

The literature review indicates that there is a need for further research and an in-depth investigation on hydroplaning crashes. As hydroplaning crashes can often involve multiple vehicles, this study aims to identify the trends from crash narratives by performing machine learning and classifying the nature of single and multiple vehicle involvement in hydroplaning crashes.

METHODOLOGY

Defining Hydroplaning Crashes

Hydroplaning refers to uncontrolled sliding of a vehicle on the wet road surface. It is triggered when a tire encounters more water than it can dissipate between the vehicle's tire tread and the roadway surface, resulting in a failure of traction and preventing the car from reacting to control inputs (see Figure 1). The pressure of water developed in the wheel front forces the water wedge under the tire's leading edge and causes the tire to get lifted from the road. Then the tire skates on the water sheet with little or no contact on the road surface, which makes the driver lose control of vehicle. The result is a hydroplaning crash *(44, 45)*. Yeager et al. *(46)* examined tire hydroplaning and its effect on wet traction and gathered wide range of data.

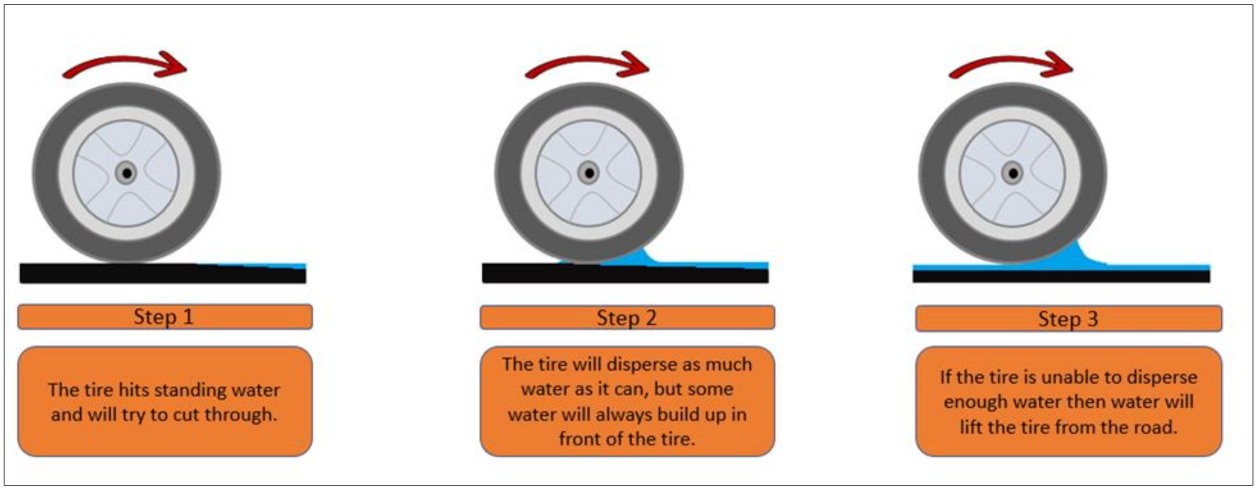


Figure 1 The Stepwise phenomenon of hydroplaning (Source: *44-46*).

Machine Learning Models

Machine learning models are gaining popularity among the researchers due to the high prediction power and flexibility. A machine learning model is created by a machine learning algorithm or a set of rules that helps in learning how to predict a specific outcome based on past events. Machine learning can identify patterns in the data and make predictions based on a training process. Once a model is developed, it can be used to predict measures or classify types based on the research question. In comparison to machine learning, conventional statistical modeling uses mathematical equations to identify associations between variables. It is also significantly beneficial because it has a higher interpretability than machine learning models. One major disadvantage of statistical modeling is its dependence on pre-determined assumptions. The major disadvantage of machine learning models is that it is more challenging to comprehend or describe the outcomes *(31, 36).*

Natural Language Processing (NLP) and Text Mining (TM), in conjunction with Machine Learning (ML), are an efficient data-centric tool to collect, analyze, and extract interesting findings or patterns from big data such as the collection of crash narrative reports. This study performed multiple longitudinal studies to illustrate various interesting patterns and anomalies in the data using text mining pipelines. The classification experiment designed in this study included three machine learning classifiers: random forest (RF), support vector machine (SVM), and XGBoost. Interested readers can consult Bishop's book for further information and detailed concepts *(37).*

Random Forest (RF)

The framework of random forest (RF) algorithm is based on the bagging principle *(39)* and random subspace method *(40).* The concept is based on constructing a collection of decision trees with random predictors. The two important byproducts in this process are 1) out-of-bag error rate (OOB) and 2) variable importance measures (VIM). The OOB is known as the misclassification rate and it decreases as the number of trees increase. The number of trees does not affect overfitting the data; therefore, a significant number of trees can be used. To decrease the bias and the correlation, the trees are grown to the maximum depth. VIM is measured by the classification accuracy and Gini impurity. This importance measure shows how much the mean squared error or the *impurity* increases when the specified variable is randomly permuted. If prediction error does not change by permuting the variable, the importance measures will not be altered significantly which in turn will change the mean squared error (MSE) of the variable only slightly (low values). This implies that the specified variable is not important. On the contrary, if the MSE significantly decreases during the permutation of the variable then the variable is deemed as important.

Support Vector Machine (SVM)

Vapnik and Lerner introduced the concept of Generalized Portrait algorithm in 1963 *(41).* It has been considered as the core algorithm for developing Support Vector Machine (SVM). SVMs are the statistical learning theory algorithms executing the structural risk minimization inductive principal to attain good generalization on a limited number of learning patterns. Later, Vapnik started the field of statistical learning theory in 1974 *(41).* Vapnik et al. introduced the current form of the SVM on a basis of a separable bipartition problem at the AT & T Bell Laboratories in 1992 *(42).* The basic idea of SVM is to map the data x into a high-dimensional feature space F via a nonlinear mapping and to perform linear regression in this space.

XGBoost

Tree boosting is a highly effective machine learning tool. XGBoost implements machine learning algorithms under the Gradient Boosting framework *(31).* It provides a parallel tree boosting algorithm that reaches to the optimized point in a fast and accurate way. The success of XGBoost relies on its scalability in different scenarios. This is due to several algorithmic optimizations. Some of the recent algorithmic improvements include a novel tree learning algorithm to handle sparse data and a theoretically justified weighted quantile sketch procedure enables handling instance weights in approximate tree learning.

Interpretability

Interpretability is a measure that is used to help the user to understand the cause of crux of the solution. A model with high interpretability is useful for the end-user to comprehend how the estimations were determined *(35).* The supervised machine learning algorithms are trained to predict the outcome variable based on the training cases. In many cases, the precision of prediction is more important than interpretability. For example, the classification tasks associated with a self-driving car to identify a pedestrian is an extremely high-risk task, and an incorrect prediction could end in a fatality. On the other hand, sometimes interpretation is more important than precision. A model with high interpretability means that the model can explain its decision, and the end-users can understand how to improve upon its decision-making function. Interpretability also works as a debugging tool to identify the biases. Better interpretability enables easier modification of the model or make it more robust. In summary, an interpretable model can ensure fairness, data protection or privacy, reliability, causality, and trust (*33, 36).*

DATA DESCRIPTION

Data Preparation

The dataset of the current study is comprised of crash narrative reports in text format that were collected from police-reported crashes in Louisiana from 2010 to 2016. The current dataset does not provide filtering options for hydroplaning related crashes. The researchers used a text searching algorithm to identify crashes associated with the keywords: ‘hydroplane’ and ‘hydroplaning.’ Initially, the search algorithm identified 703 crash reports that might be related to hydroplaning. Undergraduate students inspected those crash reports and removed reports that were out of scope of hydroplaning. This resulted in a total of 652 remaining crash reports.

It is important to know the locations of the hydroplaning crashes. Figure 2 illustrates the hydroplaning crashes for different parishes. Out of 64 Parishes, 51 Parishes experience at least one hydroplaning crash during 2010-2016. The illustration also shows that East Baton Rouge has a high number of hydroplaning crashes.

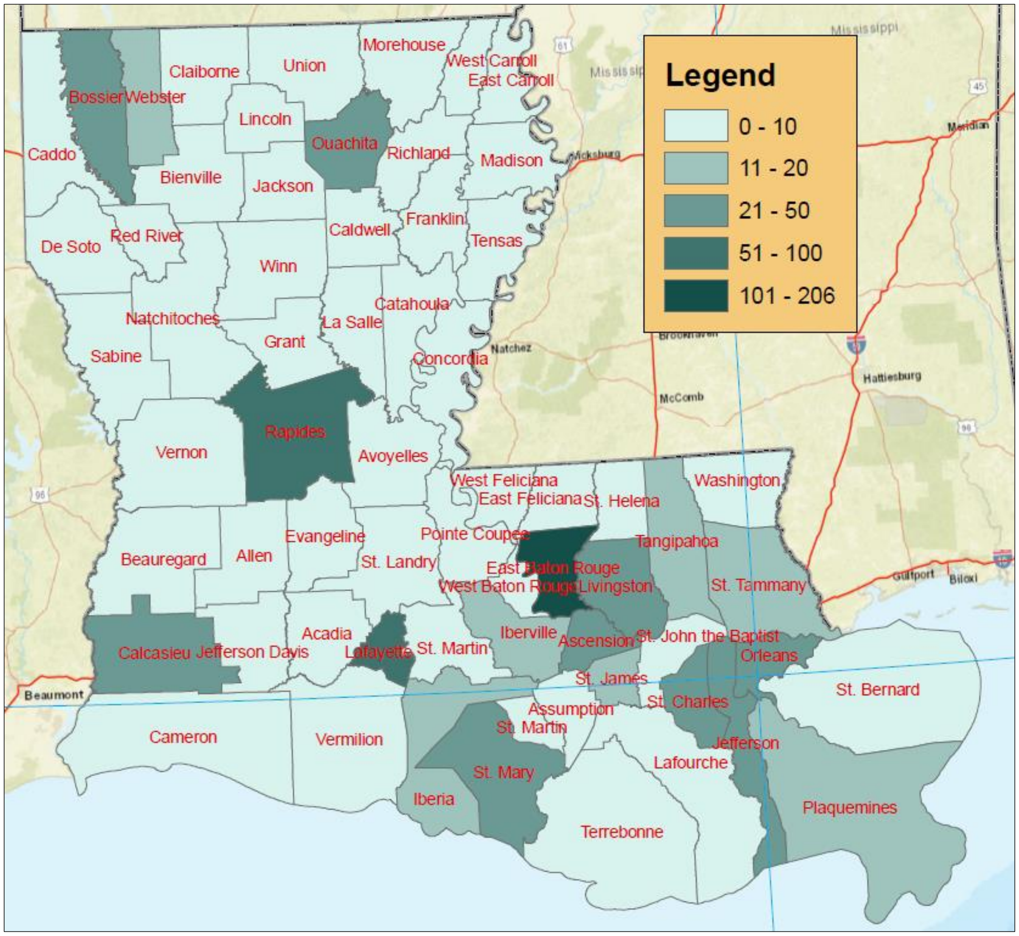


Figure 2 Hydroplaning crashes by Parishes.

Contexts of Single and Multiple vehicle Hydroplaning Crashes

One of the most important aspects of understanding hydroplaning crashes is to understand the mechanism of crash events. It is also important to know the patterns of single and multiple vehicles crashes associated with hydroplaning. Louisiana crash data provides sequences of harmful events for each crash records. These sequences are:

- First harmful event
- Second harmful event
- Third harmful event
- Fourth harmful event

Figure 3 illustrates the network plot of the association between the sequences of the harmful events for the hydroplaning crashes. Red and green (high and moderate respectively) indicates the node density associated with the harmful event types. The nodes at the right (MV in transport or multiple vehicle crash, ran off-road, and ditch) show higher density. Most of the harmful events are associated with single-vehicle crashes or ran off-road crashes; however, multiple vehicle crashes are also fairly common. The final dataset shows that 32 percent of hydroplaning crashes are multiple vehicle crashes. A study done by Khattak et al. also showed similar statistics (*48*).

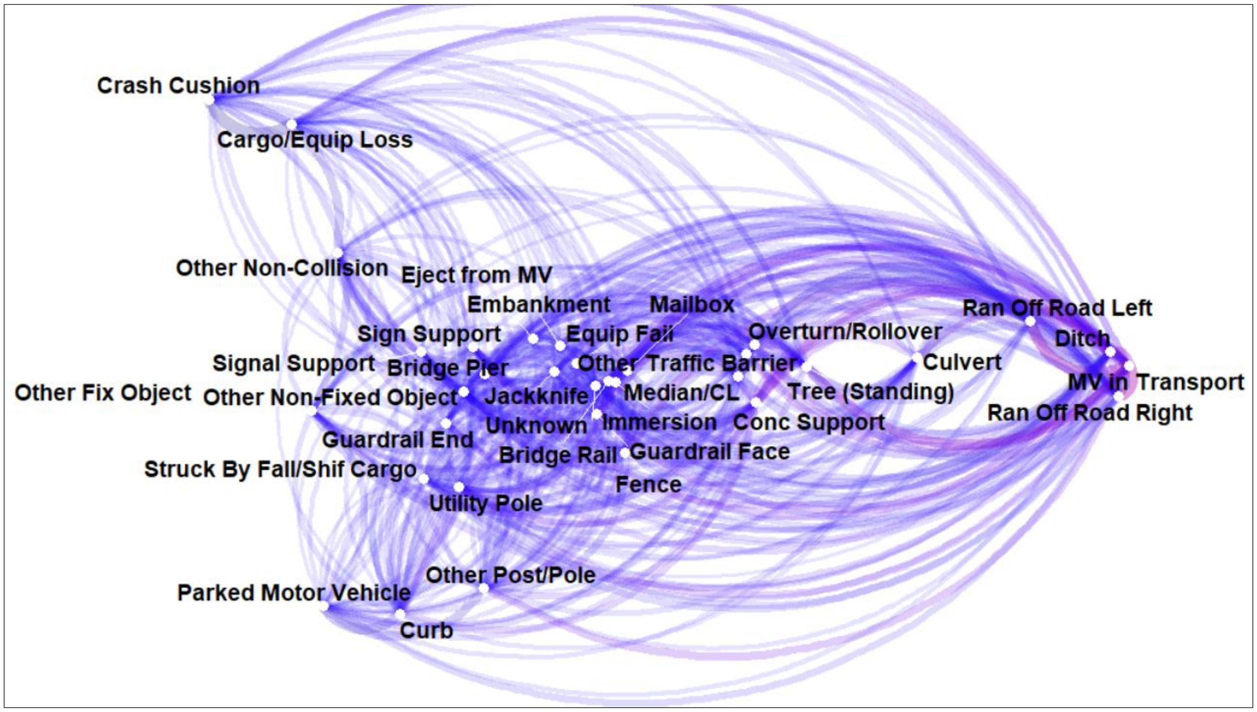


Figure 3 Association between sequences of harmful events in hydroplaning crashes.

Table 1 lists the chi-square test values (a convenient test to determine the difference between the dataset attributes) and descriptive statistics of the key variables (variables listed here have p-value < 0.05). The p-values from the chi-square tests indicate that some of the major contributing factors are significantly different for two types of crashes (single vehicle vs. multiple vehicle).

Table 1 Chi Squared Tests and Descriptive Statistics for Key Variables

Attributes	**Multiple**	**Single**	**p-value**
	N=575	**N=566**	
Access Control			<0.001
No Control (Unlimited Access to Roadway)	70.70%	58.70%	
Full Control (Only Ramp Entrance & Exit)	20.20%	33.20%	
Partial Control (Limited Access to Roadway)	9.06%	7.60%	
Other	0.00%	0.35%	
Unknown	0.00%	0.18%	
Highway Type			<0.001
Interstate	28.90%	38.70%	
State Hwy	24.00%	22.00%	
U.S. Hwy	16.40%	19.30%	
Parish Road	11.10%	12.80%	
City Street	19.50%	7.27%	

Road Condition			<0.001
No Abnormalities	78.60%	60.20%	
Water on Roadway	20.70%	36.70%	
Bumps	0.35%	0.71%	
Deep Ruts	0.00%	0.35%	
Other	0.02%	0.54%	
Weather			<0.001
Rain	78.90%	90.80%	
Cloudy	12.90%	6.54%	
Clear	7.84%	1.41%	
Fog/Smoke	0.35%	0.35%	
Other	0.00%	0.53%	
Sleet/Hail	0.00%	0.35%	
Day of the Week			0.023
FSS	43.00%	49.80%	
MTWT	57.00%	50.20%	
Driver Condition			<0.001
Normal	71.30%	61.20%	
Inattentive	23.80%	32.40%	
Apparently Asleep/Blackout	0.28%	0.46%	
Unknown	2.98%	3.19%	
Driver Distraction			0.001
Not Distracted	82.00%	75.40%	
Cell Phone	0.35%	0.18%	
Other Electronic Device (Pager, Palm Pilot, Navigation Device, Etc.)	0.01%	0.18%	
Other Inside the Vehicle	0.01%	0.71%	
Other Outside the Vehicle	2.29%	1.06%	
Unknown	15.30%	22.50%	
Driver Gender			0.001
F	41.00%	36.70%	
M	55.80%	62.70%	
U	3.13%	0.53%	

Framework for Crash Narrative Analysis

The current study developed a framework, as shown in Figure 4, for the application of interpretable machine learning in solving the classification problem using crash narrative data. The steps are the following:

- **Step 1: Data collection.** The first step is to collect the electronic data of the crash narratives. In many cases, the crash reports are hand-written and are not electronically recorded. However, many states have begun the process of transferring all crash reports to electronic versions; this includes Louisiana, which currently maintains an electronic database on crash reports. The current dataset has crash narratives for at least 50 percent of the crash records starting from 2010.
- **Step 2: Data cleaning.** Data cleaning can be performed by using text mining algorithms. Available lexicons can be used to remove redundant words, but there is still a need for domain-specific lexicons. For example, numbers are sometimes important while investigating crash reports; in a two-vehicle crash, vehicle 1 and vehicle 2 usually indicate at-

fault and not-at-fault vehicles, respectively. Therefore, removing all numbers from the text data would not be a practical strategy for crash narrative analysis.

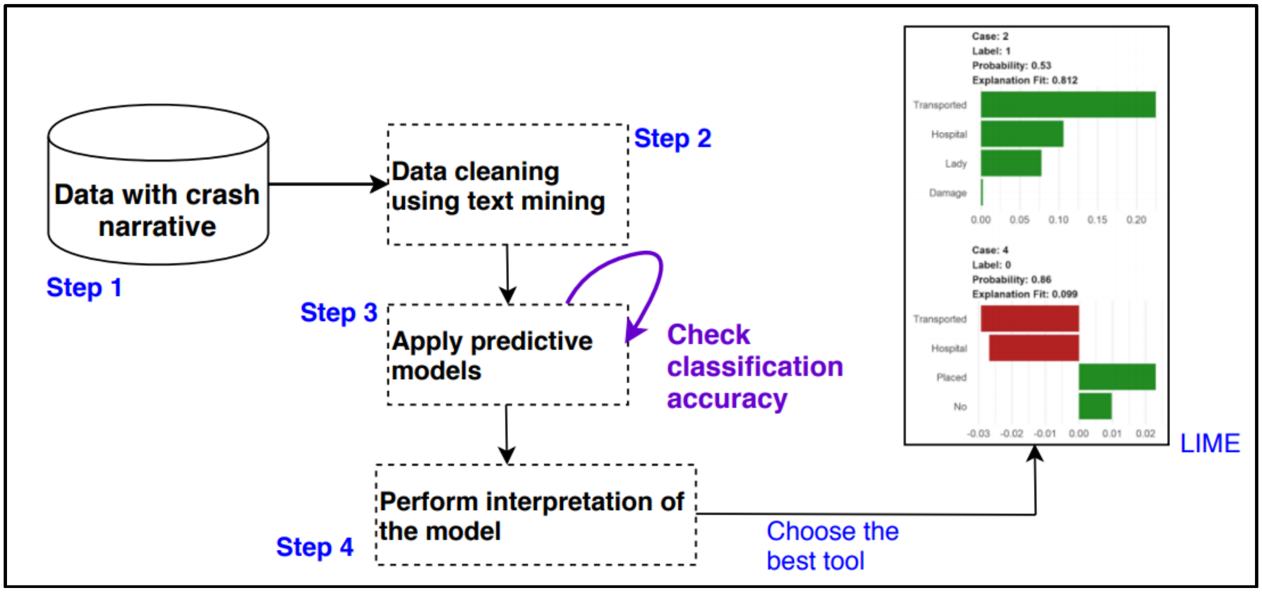


Figure 4 Framework for crash narrative analysis.

- **Step 3: Application of predictive modeling.** In recent years, innovative machine learning tools have been developed to solve the classification problem. Several machine learning models can be evaluated by selecting the best model with the lowest misclassification rate. This study applied three different machine learning models to determine the most suitable model.
- **Step 4: Application of interpretable machine learning model.** In recent years, several interpretable machine learning techniques have been introduced. The techniques include partial dependence plot (PDP), feature interaction, feature importance, Individual conditional expectation (ICE), Local interpretable model-agnostic explanations (LIME), global surrogate models, and Shapley value explanations *(31-35, 38).* This study applied LIME to explain some randomly selected crash narratives.

Before applying the machine learning algorithms, the text mining tools were applied to make the dataset less noisy. The most common problem in text data analysis is the excessive number of terms and redundant features in a single narrative. Additionally, words or parts of the words with similar meaning are compressed into the same term to make the classification more robust. This study performed a basic redundant word removal to prepare the final dataset. Future work may include more robust data cleaning that improves model precision based on domain-specific lexicons and stop word list. Figure 5 shows a sample example of the original versus the modified crash narrative.

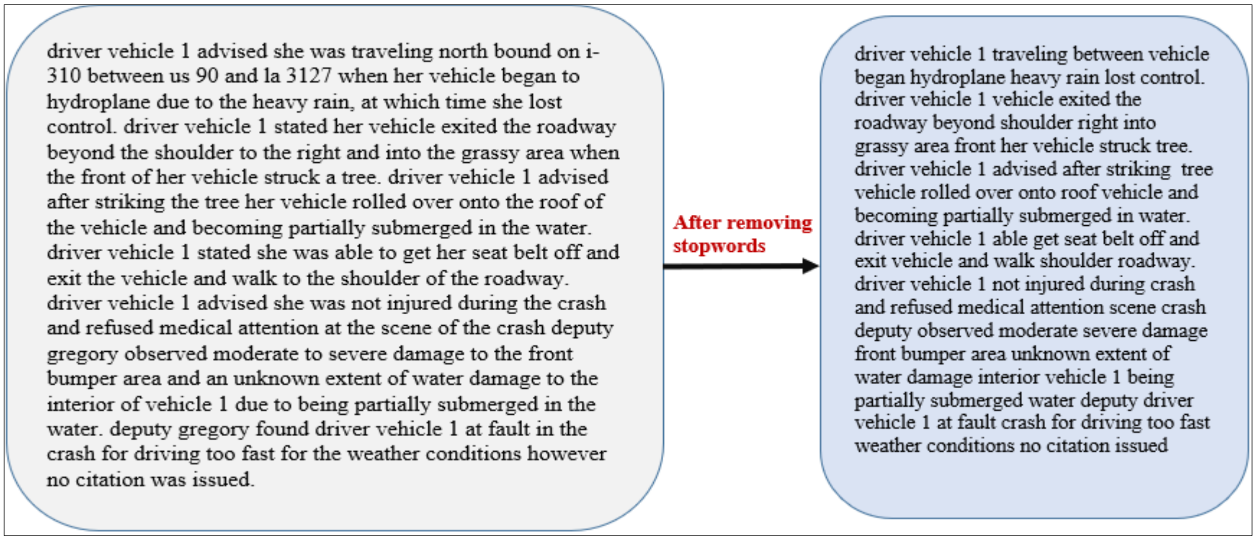


Figure 5 Sample example of original and cleaned text narrative.

MODELING RESULTS

Data were split into a training set that included 434 crashes (67%), and a testing set that included 218 crash records (33%) using stratified resampling. The models that were built using the training data were applied to the testing data to evaluate their performance. Table 2 includes the results of models in classifying single and multiple vehicle hydroplaning crashes. The values show that XGBoost algorithm performed better than both SVM and RF. The accuracy ranged from 55 % to 71 % and 42 % to 65 % for the training and testing dataset, respectively. These accuracies can be further improved by developing a robust crash narrative lexicon that includes stop words and trigger words, or words that are highly associated with crash outcomes.

Table 2 Confusion Matrix of the Predicted Classes

Model	Class (Observed)	Train (438 crashes)		Test (218 crashes)	
		Single (Predicted)	Multiple (Predicted)	Single (Predicted)	Multiple (Predicted)
SVM	Single	200	103	87	53
	Multiple	59	72	38	40
RF	Single	195	108	81	59
	Multiple	56	75	33	45
XGBoost	Single	215	88	92	48
	Multiple	57	74	26	52

True positive (TP) and false positive (FP) are instances of correct and incorrect classifications per actual class, respectively. True negative (TN) and false negative (FN) are instances of correct and incorrect rejections per actual class, respectively *(47).* Some of the common measures of model performance and accuracy are showed in Table 3.

Table 3 Performance Measures

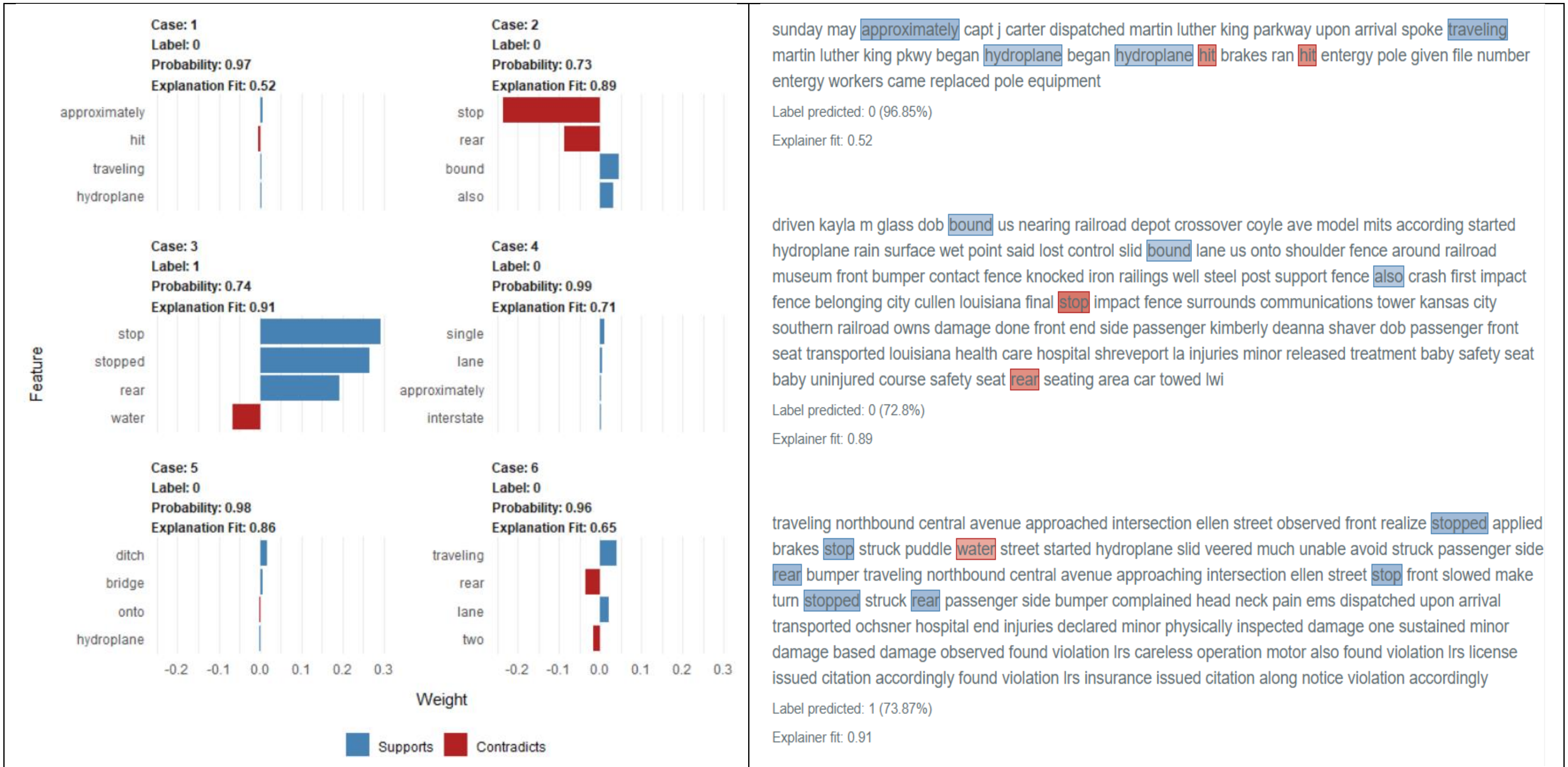
Measure	Definition	What is measured
$Recall\ or\ sensitivity$	$\frac{TP}{TP+FN}$	Effectiveness of positive level identifications
$Specificity$	$\frac{TN}{TN+FP}$	Effectiveness negative level identifications
$Precision$	$\frac{TP+TN}{TP+TN+FP+FN}$	Class agreement of the data labels with the positive labels
$Accuracy$	$\frac{TP+TN}{TP+TN+FP+FN}$	Overall accuracy
$Balanced\ Accuracy$	$\frac{TP}{TP+FN} \times 0.5 + \frac{TN}{TN+FP} \times 0.5$	Average of the proportion corrects of each class individually
$F-score$	$\frac{2 \times Precision \times Recall}{Precision + Recall}$	Weighted average of the recall and precision

Table 4 lists the performance measures of the machine learning models for both training and testing data. Based on the performance measures described above, XGBoost shows better performances than the other two models.

Table 4 Performance Measures if the Machine Learning Models

Model	Data	Sensitivity	Specificity	Accuracy	Balanced Accuracy	Precision	F-score
SVM	Train	0.6601	0.5496	0.6267	0.6048	0.7722	0.7117
	Test	0.6214	0.5128	0.5826	0.5671	0.6960	0.6566
RF	Train	0.6436	0.5725	0.6221	0.6080	0.7769	0.7040
	Test	0.5786	0.5769	0.5780	0.5777	0.7105	0.6378
XGBoost	Train	0.7096	0.5649	0.6659	0.6372	0.7904	0.7478
	Test	0.6571	0.6667	0.6606	0.6619	0.7797	0.7132

As previously mentioned, “trusting a prediction” is important to “trust the model.” However, the confusion matrix may not always be suitable to evaluate the model; hence, model explanation is crucial. This study applied a recently developed interpretable machine learning algorithm, Local Interpretable Model-Agnostic Explanations (LIME), to explain the predictions of any classifiers in an interpretable manner. This study used the R package ‘lime’ to interpret the model outcomes *(38)*. For example, considering the target ‘multiple-vehicle crash,’ six crash reports are randomly selected to visualize the association of the keywords and the output (Figure 6a). Label one will indicate that the XGBoost model predicts the crash as a multiple-vehicle hydroplaning crash. Similarly, label zero will indicate the crash as a single-vehicle crash. The explanation models considered the 10 words with the highest associations in this study (blue indicates the word supports in true classification and red means the word contradicts in the classification task). For explanation plots, only the top four highly associated words were displayed. Figure 6a shows that all cases (except Case 3) are single-vehicle crashes; the bars



(a) Randomly selected six reports (Level 1: multiple vehicles, level 0: single vehicle)	(b) Text explanation of first 3 reports in the left-hand side (Level 1: multiple vehicles, level 0: single vehicle)

Figure 6 Interpretation by keywords from randomly selected crash reports from the test dataset.

indicate the association with the classification. For example, Case 1 shows that the probability is 99 percent for this case to be considered as single-vehicle crash. The words in the crash narrative identify the class as single-vehicle are approximately, traveling, and hydroplane. Case 3 shows that the probability is 74 percent for this case to be considered as multiple-vehicle. High associations are seen for the words stop, stopped, and rear. The word 'water' shows a negative association. The plots show that the word 'rear' is in conflict with single-vehicle crashes, which is obvious as 'rear-end' crashes are usually multiple vehicle crashes. The other cases can also be explained based on the associated values of each word and the overall probability and explanation fit value, a value which represents how much the explanation can be done by using the top 10 words.

Figure 6b is another excellent display of the mechanisms of the interpretable machine learning algorithm. It shows the condensed form of the crash reports and red and blue identifiers for the classification task for first three cases in the left-hand side of Figure 6.

CONCLUSIONS

This study is unique in that it is the first to apply three machine learning methods to classify the number of vehicle involvements in hydroplaning crashes. The crash characteristics that were identified in this study can help researchers better understand the contributing factors of these crashes. The research team assessed the predictive performance of text mining to detect single and multiple vehicle crashes from hydroplaning crash-related studies. They also evaluated the efficacy of text mining using free text from crash narratives. The data used in this study included the crash narratives in electronic form for all hydroplaning crashes in Louisiana from 2010 to 2016.

The research team achieved two major objectives in this study: 1) it developed a framework for applying machine learning models to unstructured textual content in order to classify the number of vehicle involvements, and 2) it applied an interpretable machine learning framework which can identify the effectiveness of keywords in determining the classification mechanism. The study also included a comprehensive literature review of previous studies that have employed text mining to a traffic crash and occupational injury data. The framework developed by the research team can also be used to conduct other crash-related classifications (e.g., collision type) from the crash narratives. The XGBoost classifiers demonstrated a high prediction power; the three machine learning models predicted the severity of injury outcomes for the data with an accuracy of 66%. This proves that underlying trends can be revealed and analysed through machine learning models. Furthermore, this indicates that quantitative modelling techniques can be used to address safety concerns.

The study has several limitations. Ideally, the classification accuracies should be higher. Furthermore, the findings from this study are dependent on the accuracy of the information provided in the text of police reports. Future research should aim to develop a comprehensive lexicon of trigger words, or words that are highly associated with crash injury, in crash narratives to help reduce misclassification.

ACKNOWLEDGMENT

The authors appreciate the assistance provided by the students on this project: Magdalena Theel, Bita Maraghehpour, Ly-Na Tran, and Alence Poudel.

DISCLAIMER

The crash narrative data were collected from Dr. Xiaoduan Sun. 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).

AUTHOR CONTRIBUTION STATEMENT

The authors confirm the contribution to the paper as follows: study conception and design: Subasish Das; data collection: Subasish Das; analysis and interpretation of results: Subasish Das; draft manuscript preparation: Subasish Das, Anandi Dutta, Kakan Dey, Mohammad Jalayer, and Abhisek Mudgal. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. Federal Highway Administration. Safety and Pavement Friction. https://safety.fhwa.dot.gov/roadway_dept/pavement_friction/. Accessed: Jun. 9, 2019.
2. Enustun, N. Study to Identify Hydroplaning Accidents. 1976, p. 48 p.
3. Black, G. W., L. E. Jackson, and Institute of Transportation Engineers (ITE). Pavement Surface Water Phenomena and Traffic Safety. ITE Journal, Vol. 70, No. 2, 2000, pp. 32–37.
4. Aycock, E. Hydroplaning: The Effect of Water on the Roadway. Collision. The International Compendium for Crash Research, Vol. 3, No. 2, 2008, p. pp 78-85.
5. Gunaratne, M., Q. Lu, J. Yang, J. Metz, W. Jayasooriya, M. Yassin, S. Amarasiri. University of South Florida, Florida Department of Transportation, and Federal Highway Administration. Hydroplaning on Multi Lane Facilities, 2012, p. 130p.
6. Jayasooriya, W., and M. Gunaratne. Evaluation of Widely Used Hydroplaning Risk Prediction Methods Using Florida's Past Crash Data. Transportation Research Record: Journal of the Transportation Research Board, No. 2457, 2014, p. pp 140–150.
7. Zhou, F., S. Hu, S. T. Chrysler, Y. Kim, I. Damnjanovic, A. Talebpour, and A. Espejo. Optimization of Lateral Wandering of Automated Vehicles to Reduce Hydroplaning Potential and to Improve Pavement Life. Transportation Research Record, 2019, p. 0361198119853560.
8. Gopalakrishnan, K., and S. K. Khaitan. Text Mining Transportation Research Grant Big Data: Knowledge Extraction and Predictive Modeling Using Fast Neural Nets. International Journal for Traffic & Transport Engineering, 2017, 7: 354–367.
9. Jin, W., R. K. Srihari, H. Hay Ho, and X. Wu. Improving Knowledge Discovery in Document Collections through Combining Text Retrieval and Link Analysis Techniques, 2007.
10. Abdat, F., S. Leclercq, X. Cuny, and C. Tissot. Extracting Recurrent Scenarios from Narrative Texts Using a Bayesian Network: Application to Serious Occupational Accidents with Movement Disturbance. Accident Analysis & Prevention, 2014, 70: 155–166.
11. Marucci-Wellman, H., M. Lehto, and H. Corns. A Combined Fuzzy and Naïve Bayesian Strategy Can Be Used to Assign Event Codes to Injury Narratives. Injury Prevention, 2011, 17: 407–414.
12. Smith, G. S., R. A. Timmons, D. A. Lombardi, D. K. Mamidi, S. Matz, T. K. Courtney, and M. J. Perry. Work-Related Ladder Fall Fractures: Identification and Diagnosis Validation Using Narrative Text. Accident Analysis & Prevention, 2006, 38: 973–980.

13. Bertke, S. J., A. R. Meyers, S. J. Wurzelbacher, A. Measure, M. P. Lampl, and D. Robins. Comparison of Methods for Auto-Coding Causation of Injury Narratives. Accident Analysis & Prevention, 2016, 88: 117–123.
14. Vallmuur, K., H. R. Marucci-Wellman, J. A. Taylor, M. Lehto, H. L. Corns, and G. S. Smith. Harnessing Information from Injury Narratives in the 'Big Data' Era: Understanding and Applying Machine Learning for Injury Surveillance. Injury Prevention, 2016, 22: 34–42.
15. 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. American Journal of Industrial Medicine. 2005, 48: 373–380.
16. Bunn, T. L., S. Slavova, and L. Hall. Narrative Text Analysis of Kentucky Tractor Fatality Reports. Accident Analysis & Prevention, 2008, 40: 419–425.
17. Williamson, A., A. M. Feyer, N. Stout, T. Driscoll, and 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 Prevention, 2001, 7: 15–20.
18. Chen, W., K. K. Wheeler, S. Lin, Y. Huang, and H. Xiang. Computerized "Learn-As-You-Go" Classification of Traumatic Brain Injuries Using NEISS Narrative Data. Accident Analysis & Prevention, 2016, 89: 111–117.
19. Marucci-Wellman, H. R., M. R. Lehto, and 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. Accident Analysis & Prevention, 2015, 84: 165–176.
20. Wang, Z., A. D. Shah, A. R. Tate, S. Denaxas, J. Shawe-Taylor, and H. Hemingway. Extracting Diagnoses and Investigation Results from Unstructured Text in Electronic Health Records by Semi-Supervised Machine Learning. PLOS ONE, 2012, 7: e30412.
21. Chatterjee, S. A Connectionist Approach for Classifying Accident Narratives. Theses and Dissertations Available from ProQuest, 1998, 1–345.
22. Das, S., A. Mudgal, A. Dutta, and S. R. Geedipally. Vehicle Consumer Complaint Reports Involving Severe Incidents: Mining Large Contingency Tables. Transportation Research Record: Journal of the Transportation Research Board, 2018.
23. Beanland, V., M. Fitzharris, K. L. Young, and M. G. Lenné. Driver Inattention and Driver Distraction in Serious Casualty Crashes: Data from the Australian National Crash In-Depth Study. Accident Analysis & Prevention, 2013, 54: 99–107.
24. Brown, D. E. Text Mining the Contributors to Rail Accidents. IEEE Transactions on Intelligent Transportation Systems, 2016, 17: 346–355.
25. Fitzpatrick, C. D., S. Rakasi, and M. A. Knodler. An Investigation of the Speeding-Related Crash Designation through Crash Narrative Reviews Sampled via Logistic Regression. Accident Analysis & Prevention, 2017, 98: 57–63.
26. Graves, J. M., J. M. Whitehill, B. E. Hagel, and F. P. Rivara. Making the Most of Injury Surveillance Data: Using Narrative Text to Identify Exposure Information in Case-Control Studies. Injury, 2015, 46: 891–897.
27. Nayak, R., N. Piyatrapoomi, and J. Weligamage. Application of Text Mining in Analysing Road Crashes for Road Asset Management. Engineering Asset Lifecycle Management, 2010, 49–58.
28. Pollack, K. M., N. Yee, M. Canham-Chervak, L. Rossen, K. E. Bachynski, and S. P. Baker. Narrative Text Analysis to Identify Technologies to Prevent Motor Vehicle Crashes: Examples from Military Vehicles. Journal of Safety Research, 2013, 44: 45–49.

29. Sorock, G. S., T. A. Ranney, and M. R. Lehto. Motor Vehicle Crashes in Roadway Construction Workzones: An Analysis Using Narrative Text from Insurance Claims. Accident Analysis & Prevention, 1996, 28: 131–138.
30. Williams T. P., and J. F. Betak. Identifying Themes in Railroad Equipment Accidents Using Text Mining and Text Visualization. International Conference on Transportation and Development, 2016.
31. Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine. Annals of Statistics, JSTOR, 2001, pp. 1189–1232.
32. Friedman, J. H., and B. Popescu. Predictive Learning via Rule Ensembles. The Annals of Applied Statistics, JSTOR, 2018, pp. 916–54.
33. Fisher, A, C. Rudin, and F. Dominici. Model Class Reliance: Variable Importance Measures for any Machine Learning Model Class, from the 'Rashomon' Perspective, 2018. http://arxiv.org/abs/1801.01489. Accessed: July 27, 2018.
34. Ribeiro, M. T., S. Singh, and C. Guestrin. Why should I trust you? Explaining the predictions of any classifier. In the proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2016.
35. Lundberg, S., and S. Lee. An unexpected unity among methods for interpreting model predictions, 2016. http://arxiv.org/abs/1611.07478. Accessed: July 27, 2018.
36. Molnar, C. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. 2018. https://christophm.github.io/interpretable-ml-book/. Accessed: July 27, 2018.
37. Bishop, C. Pattern Recognition and Machine Learning. Springer, 2011.
38. Pedersen, T., and M. Benesty. LIME: Local Interpretable Model-Agnostic Explanations. R package version 0.4.0. https://CRAN.R-project.org/package=lime. Accessed: July 27, 2018
39. Breiman, L. Random forests. Machine learning, Vol. 45(1), 2001, pp. 5-32.
40. Ho, T. K. The random subspace method for constructing decision forests. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 8, 1998, pp. 832–844.
41. History of SVM. http://www.svms.org/history.html. Last accessed: July 2016.
42. Smola, A. J., and B. Schölkopf. A tutorial on support vector regression. Statistics and Computing, Vol. 14, No. 3, 2004, pp. 199–222.
43. Cornejo-Bueno, L., J. N. Borge, E. Alexandre, K. Hessner, and S. Salcedo-Sanz. Accurate estimation of significant wave height with Support Vector Regression algorithms and marine radar images. Coastal Engineering, Vol. 114, 2016, pp. 233–243.
44. Traction Friction of Tires by Ron Kurtus - Physics Lessons: School for Champions. https://www.school-for-champions.com/science/friction_rolling_traction.htm#.XSTX4uhKiUk. Accessed Jul. 9 2019.
45. Roadway Hydroplaning - The Trouble with Highway Cross Slope. http://www.crashforensics.com/papers.cfm?PaperID=8. Accessed Jul. 9, 2019.
46. Yeager, R. W. Tire Hydroplaning: Testing, Analysis, and Design. The Physics of Tire Traction: Theory and Experiment (D. F. Hays and A. L. Browne, eds.), Springer US, Boston, MA, 1974, pp. 25–63.
47. Labatut, V., and H. Cherifi. Accuracy measures for the comparison of classifiers. In the Proceedings of the fifth international conference on information technology, 2011, p. 3790.

48. Khattak, A. J., P. Kantor, and F. M. Council. Role of adverse weather in key crash types on limited-access: roadways implications for advanced weather systems. *Transportation research record*, *1621*(1), 1998, 10-19.