Flooding related Traffic Crashes: Findings from Association Rules

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

Xiaoduan Sun, Ph.D., P.E.
Department of Civil Engineering, University of Louisiana at Lafayette
131 Rex Street, Lafayette, LA-70504
E-mail: xsun@louisiana.edu

Sarvesh Goel, MS. Candidate
Texas A&M Transportation Institute, Texas A&M University System
3135 TAMU, College Station, TX-77843
E-mail: s-goel@tti.tamu.edu

Ming Sun, Ph.D. candidate
Department of Civil Engineering, University of Louisiana at Lafayette
131 Rex Street, Lafayette, LA-70504
E-mail: C00115250@louisiana.edu

M. Ashifur Rahman, Ph.D. candidate
University of Louisiana at Lafayette
104 East University Avenue, Lafayette, LA-70504
E-mail: ashifur@louisiana.edu

Isha Narsaria, MS. Candidate
Zachry Department of Civil Engineering, Texas A&M University System
3136 TAMU, College Station, TX 77843-3136
E-mail: isha.narsaria@tamu.edu

TOTAL WORDS: 6,528 words

5,278 words = text (including abstract and references)

1,250= 5 tables

ABSTRACT

In the United States, flooding related crashes have exposed the vulnerability of transport networks. Without realizing its depth, some might attempt to cross a flooded roadway. At night during heavy storms, determining if a road is flooded is especially difficult. To sweep away a truck, only 18-24 inches of moving water are needed and only 6 inches to carry away a small car. Louisiana's geographic location places the state in a unique position to receive both frontal tropical hurricanes and large air masses, which may result in high air moistures in almost any direction and drop rain with heavy intensity. In 2016, 12% of flooding related crashes occurred in Louisiana. During 2010-2016, flooding resulted in a total of 449 crashes in Louisiana with a total of 22 fatalities. This study collected seven years (2010-2016) of flooding related crash data to identify the key contributing factors. The findings show that two-lane roadways with no separation, driver violations, and driver injury are frequently seen in the generated rules. This study provides a comprehensive picture of flooding related to traffic crashes; thus targeting suitable countermeasures can be devised to reduce the number of crashes.

Keywords: flood-related crashes, association rules, rules mining, countermeasures.

INTRODUCTION

On August 11, 2016, an extreme and slow-moving low-pressure weather system moved across the Gulf Coast, with the heaviest concentration of rainfall in southern Louisiana. Areas around Baton Rouge received rainfall in excess of 24 inches over multiple days. Many roadways, including both interstates in the Baton Rouge area, I-10 and I-12, were closed at points with water over the road. In 2016, there were 15 fatalities involved with flooding with a total of 194 traffic crashes *(1)*. During 2010-2017, a total of 449 crashes occurred in Louisiana due to flooding with a total of 22 fatalities *(2)*. Flooding related crash events in Louisiana have exposed the vulnerability of transport networks. People might attempt to cross a flooded roadway not realizing its depth or, especially at night during heavy storms, when it might be difficult to see that a road is flooded. The geography of Louisiana makes it a unique position to receive both frontal tropical hurricanes and large air masses that can bring in air moistures in almost any direction and drop rain with heavy intensity. A recent report measured that the monetary loss due to the 2016 flooding is $8.733 million *(3)*. This suggests a need to examine the key contributors of flooding-related crashes from historical crash data to exercise caution to avoid flooded areas when driving in those locations.

The conventional crash databases do not contain a filtering option in segregating flooding-related crashes. This study collected seven years (2010-2016) of crash data from Louisiana to determine the patterns of the key contributing factors. Regression models have been extensively used in many studies to estimate traffic crashes. It is important to note that regression models usually examine the mean effect of the contributing variables and ignore any cluster effect. As a result, interventions are often geared towards the mean effect, without consideration of any subgroup effect. On the other hand, machine learning models provide better prediction by considering the subgroup effect. However, these models are not useful for the practitioners due to lack of interpretability. The current study demonstrates the applicability of rule-based modeling techniques that can address the clustering effects with heterogeneous profiles without considering any prior assumptions by using flooding-related crash data in Louisiana. The findings of this study can help reduce flooding related crashes and provide guidelines for appropriate countermeasure development.

LITERATURE REVIEW

There are a limited number of studies on flooding-related crashes. Majority of the studies focused on low water crossings (LWCs) and associated countermeasures.

Human Factors

Road traffic crashes are responsible for a substantial fraction of mortalities. The key behavioral factors of road traffic crashes can be distinguished as 1.) Decreasing capability on a long-term basis (inexperience, aging, disease, disability, alcoholism, drug abuse); 2.) Decreasing capability on a short-term basis (drowsiness, fatigue, acute alcohol intoxication, short term drug effects, binge eating, acute psychological stress, temporary distraction); 3.) Risk-taking behavior with long-term impact (overestimation of capabilities, macho attitude, habitual speeding) and; 4.) Risk-taking behavior with short-term impact (moderate ethanol intake, psychotropic drugs, motor vehicle crime, suicidal behavior, compulsive acts).

An age evaluation of vehicle-related flash flood mishaps indicates that every age group is in danger of dying from floods, including a substantial number of sufferers in ages below 20 and above 60 years. Considerably, more males than females die in vehicle crashes during a flood. This finding agrees with research that addressed similarities in vehicle-related fatalities. Studies found that a female passenger considerably reduces the probability of a crash, and Chen et al. (*4*) stated that the presence of a male passenger increases the per capita death rate almost by double, regardless of who is driving. Rappaport (*5*), who amassed a database of tropical cyclone fatality rate using the NWS readings and newspaper reports, also found a high percentage of male deaths.

Most drivers are often vain in recognizing the dangers associated with flooded watercourses. Research examining driving through flooded watercourses led to several mutual findings: 1.) Most of the cars will start to float in as little as 12 inches of water; 2.) 6 inches of water will reach the bottommost point of most passenger vehicles which can cause loss of control and even delaying; 3.) Nearly all vehicles including 4-wheel drives will float in 24 inches of water.

Studies conclude that the consumption of alcohol while driving can lead to major crashes. Drobot et al. (*6*) found that people were more likely to drive onto flooded highways if they consumed alcohol. Many fatalities related to driving through floodwaters can be avoided, but further preventive action is necessary. Theory-based operations are more effective at endorsing health-protective conduct compared to a theoretical one, and by using theoretically developed approaches given the clearly quantifiable constructs, the evaluation of advertising countermeasures is more cost-effective and efficient. Peoples' verdicts to drive into or turn back from floodwater is determined because of both their combined impact and risk perception of all other factors (e.g. social, individual, environmental, etc.) that interdependently contribute to shaping decision-making. Theory of Planned Behavior (TPB) is an individual theory that has been utilized within the literature to recognize driver's preparedness to take risks.

Crash and Geometric Characteristics

Roadway-related crashes are among the five leading causes of death worldwide. Adverse weather conditions are well documented as an important environmental factor leading to a higher risk of motor vehicle crash occurrences (*7-9*). Several studies have investigated the safety effects of different weather variables, including temperature (*10*), fog and smog (*11*), precipitation (*8, 12, 13*). Studies confirmed the significant impact of precipitation on crash risk and frequency. However, this impact has various compounding factors such as visibility impairment and road friction, which were considered across several studies (*8, 13, 14*).

From 1969–1981, French et al. (*15*) found that the majority of the flash floods occurred during the warmer season spanning between July to September, with September demonstrating the highest mortality rate. In the U.S., flood-related mortalities have been an important topic of several studies. Vehicle-associated deaths were the leading cause in both studies conducted by Mooney (*16*) and French et al. (*15*), constituting 49 percent (*16*) and 42 percent (*15*) of all logged fatalities. Floods, the most extensive all-natural calamity, are the leading cause of drowning related deaths worldwide (*17, 18*). Sinking risks increase with floods, specifically in middle and low-income countries, and it is projected that between 1980 and 2009, there were more than 500,000 deaths from floods worldwide. Research suggests that driving through floodwaters is a common type of flood practice; however, little is known about the danger issues of motor vehicle-related drowning (*19*).

Geomorphology

Geomorphologic and hydraulic data collection and analysis of river and stream channels in the Edwards Plateau enhanced the understanding of processes contributing to bedload transport and deposition at LWC (*20*). Surveys also compared pre-flood and post-flood channel conditions (*21*).

This study documented the state-of-the-art practices for flood-related crashes and indicated a need for an in-depth study of flooding related crashes. Due to the data limitations faced by past studies and the recent opportunities presented by vastly improved methodologies to investigate this research gap in a innovative manner, this study aims to achieve a better understanding of flooding related crashes by using association rules mining.

METHODOLOGY

With the intent of identifying valid and understandable patterns in a large data set, data mining aims to extract refined usable knowledge. Data mining involves machine learning, statistical learning, modeling concepts, and database management skillset and focuses on the search and finding of patterns in data rather than the confirmation of hypotheses. For this reason, the data mining method is not only concerned with algorithmic capabilities but also provides tools to analyze work without any prior assumptions.

Overview of Association Rules Mining

There are plenty of previous research on developing algorithms to solve the itemset problem. In one study, Agrawal and Srikant developed a level-wise, breadth-first algorithm called *a priori* to count transactions and mine frequent itemsets, maximal itemsets and closed frequent itemsets. *A priori* can also be used to develop association rules. The concept of *a priori* assumes that if an itemset is frequent, then its subsets must also be frequent. The algorithm works by mining itemsets or subsequences occurring frequently from a comprehensively large dataset. This process allows users to identify associations between different items. In recent years, several studies applied association rules mining techniques to determine the latent trends in the complex crash databases *(22-26)*. Many studies included the general concepts of association rules mining. The authors consulted a Das et al. *(25)* study to provide a brief overview on the association rules mining.

By considering I= $\{i_1, i_2, \ldots, i_m\}$ as a set of items (e.g., a set of crash categories for a particular crash record) and $C = \{c_1, c_2, \ldots, c_n\}$ as a set of database crash information (i.e., transaction) where each crash record, c_i, contains a subset of items chosen from I, an itemset with k items can be defined as a k-itemset.

An association rule can be represented as A → B, where A and B are disjoint itemsets. Here, A is the antecedent, and B is the consequent. The association rule’s significance is determined by the values of support, confidence, and lift. The equations of support are listed below.

$$S(A) = \frac{\sigma(A)}{N} \tag{1}$$

$$S(B) = \frac{\sigma(B)}{N} \tag{2}$$

$$S(A \to B) = \frac{\sigma(A \cap B)}{N} \tag{3}$$

where,

$\sigma(A)$= Number of incidents with X antecedent

$\sigma(B)$ =Number of incidents with Y consequent

$\sigma(A \cap B)$ =Number of incidents with both X antecedent and Y consequent
N= Total number of incidents
S(A)= Support of antecedent
S(B)= Support of consequent
S(A → B)= Support of the association rule (A → B)

The equations of confidence and lift are listed in equations 4 and 5. The measure of reliability for the inference of a generated rule is called confidence. High confidence for a given A → B means that B has a strong presence in transactions with A antecedent. The lift represents the relation with the frequency and expected the frequency of co-occurrence of the antecedent and the consequent.

$$C(A \to B) = \frac{S(A \to B)}{S(A)} \tag{4}$$

$$L(A \to B) = \frac{S(A \to B)}{S(A).S(B)} \tag{5}$$

where,
C(A → B) = Confidence of the association rule (A → B)
L(A → B) = Lift of the association rule (A → B)

A lift value greater than one indicates positive interdependence between the antecedent and the consequent, while a value smaller than one indicates negative interdependence. A value of one signifies independence. Depending on the number of antecedents and consequents, rules are defined as either a 2-product rule or a 3-product rule. If there is more than one of either, an antecedent or consequent, then it is a 3-product rule. It is important to note that generated rules only imply association and not causation.

Data Integration

The dataset of the current study is police-reported crashes in Louisiana from 2010 to 2016. Louisiana crash datasets also include crash narrative in a database format. There is no filter in the crash database that can identify flood-related crashes from the crash database. To identify the flooding-related crash database, the research team applied a text mining algorithm to distinguish the flooding related crashes. Three terms were used to identify the flooding-related crashes: flood, flooding, and flooded. After extracting these crashes, a manual approach of reading the crash reports was conducted to remove the redundant crashes. In seven years, 449 crashes in Louisiana occurred due to the flooding. From 2015 to 2016, the state of Louisiana experienced a 234% increase of flooding-related crashes. Prolonged rainfall in August 2016 resulted in catastrophic flooding for the state of Louisiana in which thousands of houses and businesses were submerged. Many rivers and waterways reached record levels, and rainfall exceeded 20 inches in multiple parishes. The flowchart of data integration and analysis is shown in Figure 1.

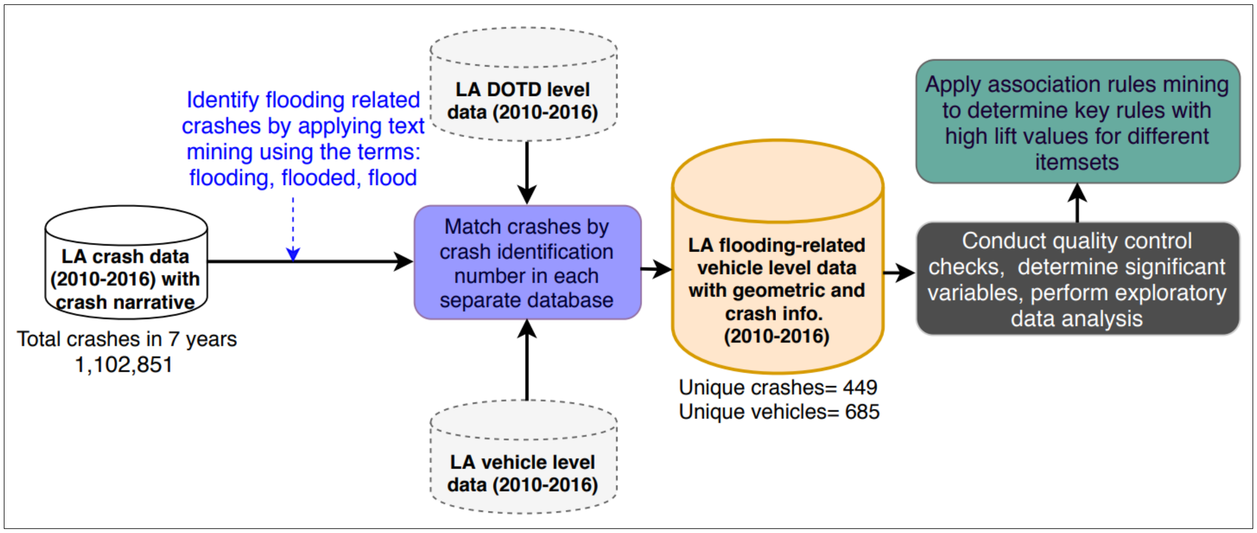

Figure 1 Flowchart of data integration.

A majority of the Parishes of District 3, 61, and 62 were designated as federal disaster areas by the Federal Emergency Management Agency (FEMA) in the aftermath of the floods. Figure 2 illustrates the heatmap of flood-related crashes by the Louisiana Department of Transportation and Development (DOTD) Districts and year. Both East and West Baton Rouge were in District 61. District 61 has been ranked one due to a high number of flooding-related crashes. The number of crashes by year is illustrated in the top section of Figure 2. The line chart clearly indicated the sudden rise of flooding-related crashes in Louisiana in 2016.

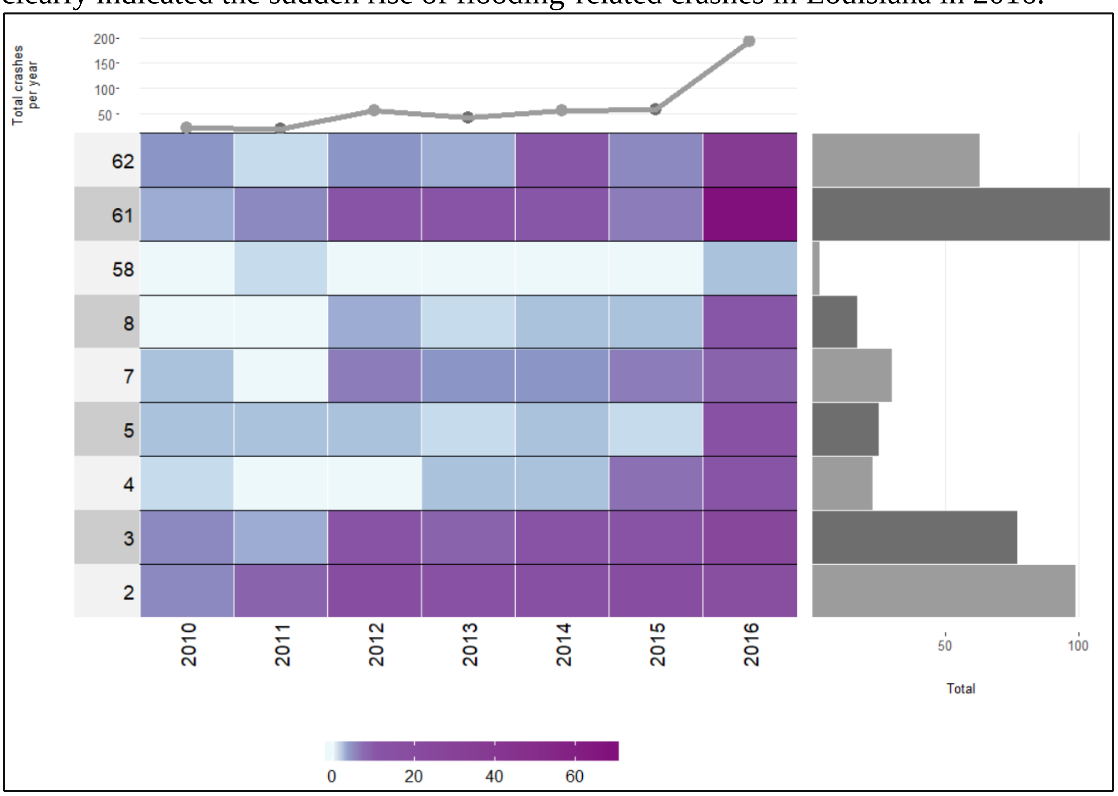

Figure 2 Flood-related crashes by DOTD districts.

Table 1 exhibits the attributes leading to road traffic crashes and the frequency distribution among each characteristic. Attributes such as age, vehicle type, season, and more are listed below. Approximately 70 percent of the flooding crashes are segment related. Areas with posted speed limits (PSL) from 25-50 mph are more likely to be involved in a road traffic crash than areas with PSL of 20 mph or less and 55 mph or more. Summer months experience more crashes than any other season. Warmer months are more susceptible to crashes in comparison to months in the winter season. Rear-end crashes have the highest frequency in flooding related crashes. Two-lane roadways with no physical separation display high frequencies in flooding related crashes.

Table 1 Descriptive Statistics

Attribute	**Percent**	**Attribute**	**Percent**
Intersection (Int)		**Season (Season)**	
No	68.76	Fall	24.82
Yes	31.24	Spring	27.59
Posted Speed Limit (PSL)		Summer	35.18
20 or less	11.68	Winter	12.41
25-35	44.23	**Driver Injury (DrInj)**	
40-50	30.51	KABC[1]	26.28
55-65	11.82	PDO[2]	73.72
70 or above	1.75	**Driver Condition (Cond)**	
Collision Type (CollTyp)		Normal	52.41
Rear End	31.68	Inattentive	26.72
Single	23.94	Impaired	2.63
Sideswipe	14.74	Others	18.25
Right Angle	7.01	**Driver Age (Age)**	
Head-On	3.07	15-24	20.88
Turning	4.82	25-34	24.38
Others	14.74	35-44	20.58
Major Contributing Factor (Cont)		45-54	14.01
Violations	61.02	55-64	13.43
Movement Prior To Crash	17.08	65-74	5.69
Roadway Condition	10.07	> 74	1.02
Condition of Driver	4.82	**Vehicle Type (VehTyp)**	
Weather	7.01	Car	46.42
Roadway Type (RdTyp)		Lt. Truck	24.53
2-way no Sep.	68.18	Suv	15.47
2-way with Sep	20.73	Truck/Bus	6.57
2-way with Barrier	3.21	Van	2.63
Others	7.88	Others	4.38

Note: [1]KABC indicates Fatal and Injury Crashes (K= Fatal, A= Incapacitating Injury, B= Non-Incapacitating Injury, C= Minor Injury), [2]PDO (Property Damage Only) indicates No injury.

RESULTS AND DISCUSSIONS

Association rules mining can handle complex datasets with a large number of variables. Unlike the parametric models, there is no need to predetermine the assumptions and functional forms in association rules mining. Additionally, the generated rules have potential in exhibiting latent patterns in data. The current study used 'a priori' algorithm to perform the analysis. The rules are generated by using 'arules' package of the open-source R Software *(27).* To get meaningful and significant results, there is a need to determine the minimum support and confidence. The current study used trial and error processes to set the minimum values of support and confidence, as recommended by a previous study *(23).* A similar approach has been adopted by other studies *(24, 25).* This study used a priori algorithm that includes two separate steps: 1) minimum support is used to find the frequent itemsets in the database, 2) these frequent itemsets and the minimum confidence constraint are used to form rules. For real-world implications, researchers can perform more tests to determine the suitable values of support and confidence based on different decision criteria. Moreover, there is a need to investigate the rules with more items based on specified conditions. The higher the lift, the higher the associations between the variables. For this study, the minimum support is considered as 0.05, and the minimum confidence was determined as 0.30 after several trial and errors.

Table 2 depicts the variables considered as antecedent-consequent rules for n-itemsets that contribute to road traffic crashes. The rules and their corresponding antecedent-consequent were ordered according to decreasing lift value. Each rule has a lift value greater than 1, indicating positive interdependence between the antecedent and the consequent. In addition to the lift value, each antecedent lists their consequent, support, confidence, and count. Table 2 has 2-itemset rules to 5-itemset rules. Two attributes *DrInj=KABC* and *Cond=Others* are present in all rules. The other attributes that are presented in multiple times in the rules are: *RdTyp=2-way no Sep.* (in 9 rules), *Int=No* (in 8 rules), *PSL=25-35* (in 6 rules), and *Cont=Violations* (in 6 rules). The rule with highest lift value is ***Rule_All01:*** *RdTyp=2-way no Sep., PSL=25-35, DrInj=KABC* → *Cond=Others* (Support=0.063, Confidence=0.662, Lift=3.625). The explanation of the first rule is: 6.5% of flooding related KABC crashes occurred on two-way roadways with no physical separation with posted speed limit (PSL) 25 to 35 miles/hour; out of all flooding related crashes with the antecedent (*RdTyp=2-way no Sep., PSL=25-35, DrInj=KABC*), 66.2% crashes occur with driver condition as others; the proportion of KABC flooding related crashes (with driver condition as others) on rural two-lane roadways with no physical separation and PSL 25-35 mph was 3.625 times the proportion of flooding related crashes with driver as others.

Table 2 Top 20-Rules from N-Itemsets

Rule No.	Antecedent	Consequent	Support	Confidence	Lift	Count
Rule_All01	RdTyp=2-way no Sep., PSL=25-35, DrInj=KABC	Cond=Others	0.063	0.662	3.625	43
Rule_All02	PSL=25-35, DrInj=KABC	Cond=Others	0.076	0.605	3.313	52
Rule_All03	Cont=Violations, Int=No, Cond=Others	DrInj=KABC	0.074	0.864	3.290	51
Rule_All04	Cont=Violations, RdTyp=2-way no Sep., Int=No, Cond=Others	DrInj=KABC	0.055	0.864	3.287	38
Rule_All05	Int=No,PSL=25-35, DrInj=KABC	Cond=Others	0.057	0.591	3.238	39
Rule_All06	Int=No,Cond=Others, PSL=25-35	DrInj=KABC	0.057	0.830	3.158	39
Rule_All07	Cont=Violations, Cond=Others	DrInj=KABC	0.099	0.829	3.156	68
Rule_All08	RdTyp=2-way no Sep., Cond=Others, PSL=25-35	DrInj=KABC	0.063	0.827	3.147	43
Rule_All09	Cont=Violations, RdTyp=2-way no Sep., Cond=Others	DrInj=KABC	0.076	0.825	3.141	52

Rule_All10	RdTyp=2-way no Sep., Int=No,Cond=Others	DrInj=KABC	0.083	0.803	3.055	57
Rule_All11	Cond=Others, PSL=25-35	DrInj=KABC	0.076	0.800	3.044	52
Rule_All12	RdTyp=2-way no Sep., Int=No,DrInj=KABC	Cond=Others	0.083	0.553	3.033	57
Rule_All13	Int=No, Cond=Others	DrInj=KABC	0.104	0.789	3.002	71
Rule_All14	RdTyp=2-way no Sep., DrInj=KABC	Cond=Others	0.107	0.545	2.985	73
Rule_All15	RdTyp=2-way no Sep., Cond=Others	DrInj=KABC	0.107	0.760	2.894	73
Rule_All16	Int=No, DrInj=KABC	Cond=Others	0.104	0.522	2.861	71
Rule_All17	Cont=Violations, RdTyp=2-way no Sep., DrInj=KABC	Cond=Others	0.076	0.520	2.850	52
Rule_All18	Cond=Others	DrInj=KABC	0.134	0.736	2.801	92
Rule_All19	DrInj=KABC	Cond=Others	0.134	0.511	2.801	92
Rule_All20	Cont=Violations, DrInj=KABC	Cond=Others	0.099	0.507	2.781	68

Table 3 depicts the variables considered as antecedents that contribute to road traffic crashes. A rule number labels each antecedent, and a rule is defined as a 2-item rule if there are a single antecedent and single consequent. The rules and their corresponding antecedent were ordered according to decreasing lift value. The lift value of each rule is greater than one implying a positive interdependence between the antecedent and the consequent. In addition to the lift value, each antecedent lists their consequent, support, confidence, and count. The rules generated for 2-itemset rules include additional attributes in the top 20 rules. The other attributes that are presented in multiple times in the rules are: *Cont=Violations* (in 5 rules), *RdTyp=2-way no Sep.* (in 5 rules), *CollTyp=Rear End* (in 5 rules), and *DrInj=KABC* (in 4 rules). Majority of the 2-itemsets rules with driver injury are associated with fatal or injury crashes (exception is ***Rule_2I13:*** *Cont=Roadway Condition → DrInj=PDO*).

Table 3 Top 20-Rules from 2-Itemsets

Rule No.	**Antecedent**	**Consequent**	**Support**	**Confidence**	**Lift**	**Count**
Rule_2I01	Cond=Others	DrInj=KABC	0.134	0.736	2.801	92
Rule_2I02	Season=Winter	CollTyp=Single	0.051	0.412	1.720	35
Rule_2I03	RdTyp=2-way with Sep	CollTyp=Rear End	0.105	0.507	1.601	72
Rule_2I04	CollTyp=Rear End	RdTyp=2-way with Sep	0.105	0.332	1.601	72
Rule_2I05	Age=55-64	DrInj=KABC	0.053	0.391	1.489	36
Rule_2I06	Age=15-24	CollTyp=Single	0.069	0.329	1.373	47
Rule_2I07	RdTyp=2-way with Sep	PSL=40-50	0.085	0.408	1.339	58
Rule_2I08	CollTyp=Single	Cond=Inattentive	0.085	0.354	1.324	58
Rule_2I09	Cont=Movement Prior To Crash	PSL=25-35	0.098	0.573	1.295	67
Rule_2I10	PSL=40-50	CollTyp=Rear End	0.123	0.402	1.269	84
Rule_2I11	DrInj=KABC	Cont=Violations	0.196	0.744	1.220	134

Rule_2I12	Cont=Violations	DrInj=KABC	0.196	0.321	1.220	134
Rule_2I13	Cont=Roadway Condition	DrInj=PDO	0.091	0.899	1.219	62
Rule_2I14	Age=25-34	Cond=Inattentive	0.079	0.323	1.210	54
Rule_2I15	RdTyp=2-way with Sep	Cond=Normal	0.131	0.634	1.209	90
Rule_2I16	CollTyp=Other	PSL=25-35	0.079	0.535	1.209	54
Rule_2I17	CollTyp=Right Angle	Cont=Violations	0.051	0.729	1.195	35
Rule_2I18	RdTyp=2-way with Sep	Int=Yes	0.077	0.373	1.195	53
Rule_2I19	CollTyp=Rear End	Cont=Violations	0.231	0.728	1.193	158
Rule_2I20	Cont=Violations	CollTyp=Rear End	0.231	0.378	1.193	158

Table 4 depicts the variables considered as antecedents that contribute to road traffic crashes. A rule number labels each antecedent, and a 3-item rule indicates that there are two antecedents and one consequent or one antecedent and two consequents. The rules and their corresponding antecedent were ordered according to decreasing lift value. The lift value of each rule is greater than one implying a positive interdependence between the antecedent and the consequent. In addition to the lift value, each antecedent lists their consequent, support, confidence, and count. ***Rule 3I20:*** *Cont=Violations, Cond=Normal → CollTyp=Rear End* occurred the most often with a count of 100, and ***Rule 3I16:*** *RdTyp=2-way with Sep, VehTyp=Car → CollTyp=Rear End* occurred the least often with a count of 35. The attributes that are present in multiples rules are: *CollTyp=Rear End* (in 10 rules), *DrInj=KABC* (in 7 rules), *Cond=Others* (in 7 rules), and *Cont=Violations* (in 6 rules). The findings are in line with other studies (*10*, *11*). Moreover, Tamerius et al. (*8*) showed similar findings as they found the significant effects of precipitation on the severity and number of crashes.

Table 4 Top 20-Rules from 3-Itemsets

Rule No.	**Antecedent**	**Consequent**	**Support**	**Confidence**	**Lift**	**Count**
Rule_3I01	PSL=25-35, DrInj=KABC	Cond=Others	0.076	0.605	3.313	52
Rule_3I02	Cont=Violations, Cond=Others	DrInj=KABC	0.099	0.829	3.156	68
Rule_3I03	Cond=Others, PSL=25-35	DrInj=KABC	0.076	0.800	3.044	52
Rule_3I04	Int=No, Cond=Others	DrInj=KABC	0.104	0.789	3.002	71
Rule_3I05	RdTyp=2-way no Sep., DrInj=KABC	Cond=Others	0.107	0.545	2.985	73
Rule_3I06	RdTyp=2-way no Sep., Cond=Others	DrInj=KABC	0.107	0.760	2.894	73
Rule_3I07	Int=No, DrInj=KABC	Cond=Others	0.104	0.522	2.861	71
Rule_3I08	CollTyp=Rear End, Season=Summer	RdTyp=2-way with Sep	0.055	0.418	2.014	38
Rule_3I09	RdTyp=2-way with Sep, Season=Summer	CollTyp=Rear End	0.055	0.623	1.966	38
Rule_3I10	Cont=Violations, RdTyp=2-way with Sep	CollTyp=Rear End	0.082	0.609	1.921	56
Rule_3I11	CollTyp=Single, Cont=Violations	Cond=Inattentive	0.053	0.500	1.872	36

Rule_3I12	RdTyp=2-way with Sep, Int=No	CollTyp=Rear End	0.073	0.562	1.773	50
Rule_3I13	RdTyp=2-way with Sep, Cond=Normal	CollTyp=Rear End	0.073	0.556	1.754	50
Rule_3I14	CollTyp=Rear End, Cont=Violations	RdTyp=2-way with Sep	0.082	0.354	1.710	56
Rule_3I15	CollTyp=Rear End, Cond=Normal	RdTyp=2-way with Sep	0.073	0.347	1.675	50
Rule_3I16	RdTyp=2-way with Sep, VehTyp=Car	CollTyp=Rear End	0.051	0.530	1.674	35
Rule_3I17	CollTyp=Rear End, DrInj=PDO	RdTyp=2-way with Sep	0.088	0.347	1.673	60
Rule_3I18	Cont=Weather, DrInj=PDO	Cond=Normal	0.054	0.841	1.605	37
Rule_3I19	Cont=Violations, RdTyp=2-way with Sep	PSL=40-50	0.066	0.489	1.603	45
Rule_3I20	Cont=Violations, Cond=Normal	CollTyp=Rear End	0.146	0.508	1.602	100

Table 5 depicts the variables considered as antecedents that contribute to road traffic crashes. The rules and their corresponding antecedent were ordered according to decreasing lift value. Each rule has a lift value greater than one indicating positive interdependence between the antecedent and the consequent. In addition to the lift value, each antecedent lists their consequent, support, confidence, and count. The frequency of each rule was evenly distributed in the 4-itemset. The attributes that are present in multiples rules are: *Cont=Violations* (in 11 rules), *CollTyp=Rear End* (in 10 rules), *Cond=Others* (in 10 rules), *DrInj=KABC* (in 10 rules), and *Int=No* (in 9 rules). This finding is in line with one study by Sorock et al. (*28*).

Table 5 Top 20-Rules from 4-Itemsets

Rule No.	**Antecedent**	**Consequent**	**Support**	**Confidence**	**Lift**	**Count**
Rule_4I01	RdTyp=2-way no Sep., PSL=25-35, DrInj=KABC	Cond=Others	0.063	0.662	3.625	43
Rule_4I02	Cont=Violations, Int=No, Cond=Others	DrInj=KABC	0.074	0.864	3.290	51
Rule_4I03	Int=No,PSL=25-35, DrInj=KABC	Cond=Others	0.057	0.591	3.238	39
Rule_4I04	Int=No, Cond=Others, PSL=25-35	DrInj=KABC	0.057	0.830	3.158	39
Rule_4I05	RdTyp=2-way no Sep., Cond=Others, PSL=25-35	DrInj=KABC	0.063	0.827	3.147	43
Rule_4I06	Cont=Violations, RdTyp=2-way no Sep., Cond=Others	DrInj=KABC	0.076	0.825	3.141	52
Rule_4I07	RdTyp=2-way no Sep., Int=No, Cond=Others	DrInj=KABC	0.083	0.803	3.055	57
Rule_4I08	RdTyp=2-way no Sep., Int=No, DrInj=KABC	Cond=Others	0.083	0.553	3.033	57
Rule_4I09	Cont=Violations, RdTyp=2-way no Sep., DrInj=KABC	Cond=Others	0.076	0.520	2.850	52
Rule_4I10	Cont=Violations, Int=No, DrInj=KABC	Cond=Others	0.074	0.495	2.713	51

Rule_4I11	Cont=Violations, RdTyp=2-way with Sep, Int=No	CollTyp=Rear End	0.060	0.732	2.311	41
Rule_4I12	Cont=Violations, RdTyp=2-way with Sep, Cond=Normal	CollTyp=Rear End	0.055	0.704	2.221	38
Rule_4I13	Cont=Violations, RdTyp=2-way with Sep, DrInj=PDO	CollTyp=Rear End	0.066	0.625	1.973	45
Rule_4I14	CollTyp=Rear End, Cont=Violations, Cond=Normal	RdTyp=2-way with Sep	0.055	0.380	1.833	38
Rule_4I15	CollTyp=Rear End, Cont=Violations, DrInj=PDO	RdTyp=2-way with Sep	0.066	0.375	1.809	45
Rule_4I16	Cont=Violations, Int=No, Cond=Normal	CollTyp=Rear End	0.107	0.570	1.800	73
Rule_4I17	Cont=Violations, Cond=Normal, PSL=40-50	CollTyp=Rear End	0.064	0.564	1.781	44
Rule_4I18	CollTyp=Rear End, Cond=Normal, DrInj=PDO	RdTyp=2-way with Sep	0.063	0.361	1.743	43
Rule_4I19	RdTyp=2-way with Sep, Cond=Normal, DrInj=PDO	CollTyp=Rear End	0.063	0.551	1.740	43
Rule_4I20	RdTyp=2-way with Sep, Int=No, DrInj=PDO	CollTyp=Rear End	0.057	0.542	1.710	39

Visualization of association rules is challenging because of the high number of rules and the generated texts for the rules. Figure 3 shows the scatter plots of the top twenty rules based on support, confidence, and lift for different itemset groups. This plot provides an overall trend of the generated top rules and the clustering patterns of the rules based on the parameter. The size and color of the circles indicate the value of the lift. The key insights from this figure are the following:

- A limited number of rules have high support values. For example, for 4-itemsets, one rule (***Rule_4I14***: *CollTyp=Rear End, Cont=Violations, Cond=Normal → RdTyp=2-way with Sep*) has high support value. This rule has an uncommon trait (*Cond=Normal*) in flooding involved association rules.
- Two clusters in 3-itemsets and 4-itemsets group have moderate values of support, confidence, and lift. For 3- itemsets, this cluster has 3 rules: ***Rule_3I02****: Cont=Violations, Cond=Others→ DrInj=KABC;* ***Rule_3I04****: Int=No, Cond=Others→DrInj=KABC;* ***Rule_3I06****: RdTyp=2-way no Sep., Cond=Others→DrInj=KABC*. For 4- itemsets, this cluster has 3 rules: ***Rule_4I02****: Cont=Violations, Int=No, Cond=Others → DrInj=KABC;* ***Rule_4I06****: Cont=Violations, RdTyp=2-way no Sep., Cond=Others → DrInj=KABC;* ***Rule_4I07****: RdTyp=2-way no Sep., Int=No, Cond=Others → DrInj=KABC.* Two key attributes are common in all of these rules: DrInj=KABC, and Cond=Others.
- Majority of the rules with high lift values have high confidence scores.

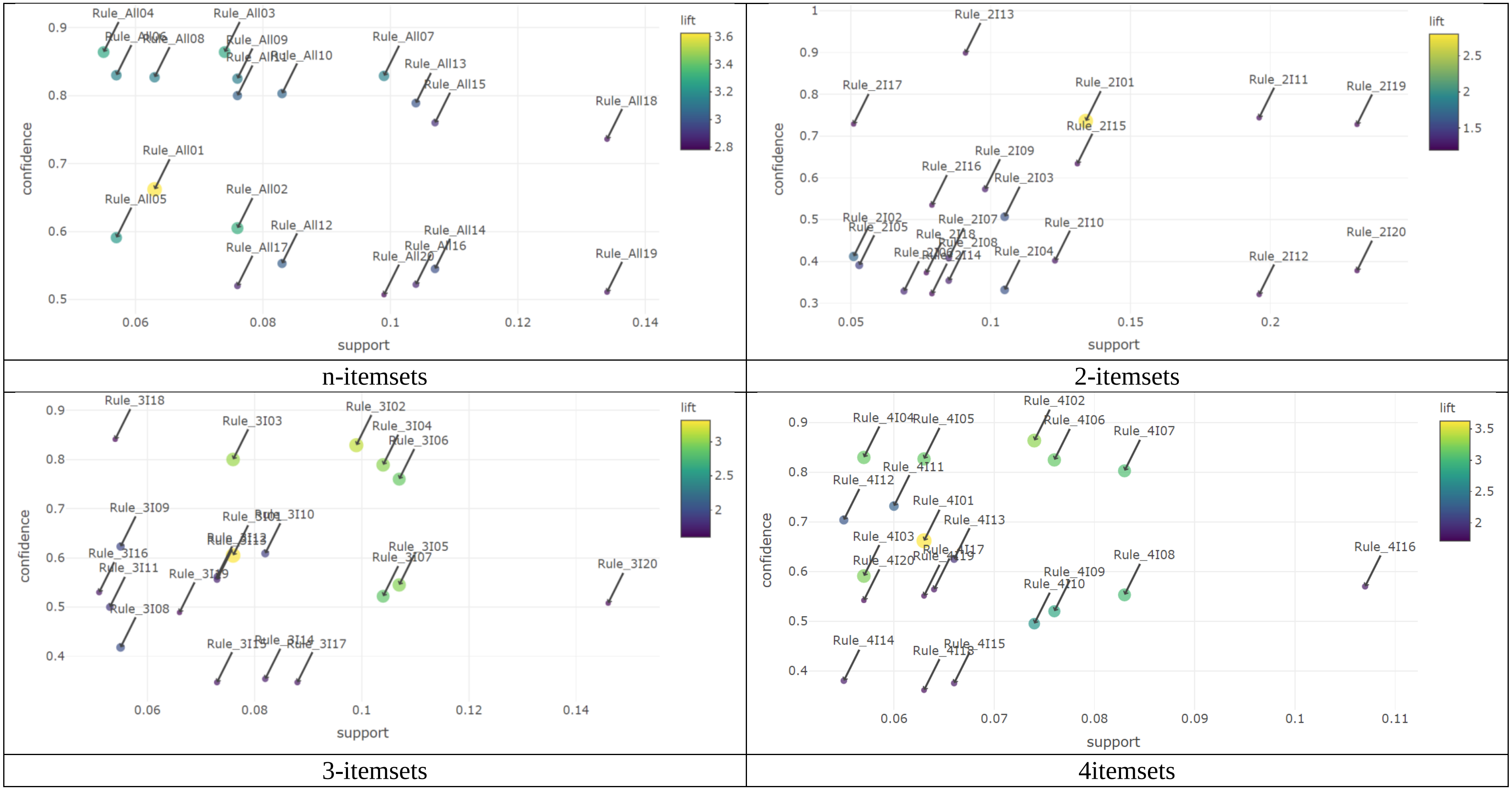

Figure 3 Scatter plots based on support, confidence, and lift values.

Countermeasures and Design Alternatives

To reduce flooding related crashes, several countermeasures can be proposed. Static warning signs to the approach of a low water crossing or frequently flooded locations, indicators along the verge of hardened surfaces, and water depth gauges are the conventional countermeasures. Many drivers have trouble arbitrating the depth and speed of water over the roadway and enter the flooded road. In the nonexistence of any helpful signs, drivers set their standards to determine whether the road is drivable, which may lead to loss of property and life. For the safety of the commuters in the flooded areas, it is suggested to use one regulatory two warning sign in advance of LWC. Barricades and signs can be vague because they are not usually present at all crossings and can remain in position when the location is dry. For example, Ashley and Ashley (*17*), recommended that future structural modifications of flood control designs might not be associated with fatality reduction. Solutions such as the installation of traffic barricades, alarm signs, and water depth gauges at hazardous locations are also insufficient.

It was evident that the trust of information (and trust in the source of information) is an important aspect in encouraging appropriate behaviors and should be consequently developed more broadly across flood education programs. Many fatalities related to floods can be evaded. In Queensland, Australia, strategy makers adopted a campaign with the slogan *"If it's Flooded, Forget it"* after January 2011 floods, but despite these campaigns, a reliable number of motor vehicle-related drowning losses are still being described each year in Australia (about 7 percent between 2012 and 2013). To efficiently reduce the rate of vehicle-related drowning, it is always necessary to first create empirical evidence on key factors that guide commons choice to drive through flooded watercourses (*20*). To better understand individual behavior, innovative analytical models will help in the growth of more effective interference programs *(29-31)* to battle risky driving behavior.

Meeting road management objectives while fulfilling site-specific biological and geomorphic goals requires a truly interdisciplinary approach in which a biologist and hydro geomorphologist work with the design engineer. A strong structure must integrate the engineering requirements with hydrologic and biological factors. Between the years of 2005-2011, at least four concrete-cable LWCs were installed across Camp Atterbury Joint Maneuver Training Center or CAJMTC (*32*). Most crossings were orientated using the existing trail path. With respect to the stream, the orientations are now under consideration as LWCs have been found as sources of sediment deposition, especially as vehicles traverse.

CONCLUSIONS

The framework presented in this study provided a thorough investigation on how association rules mining approach can be utilized to identify contributing factors that affect flooding involved crashes. Understanding the impact of flooding related crashes is crucial in identifying safety-related issues and countermeasures. One of the major contributions of this study is that more domain-specific patterns were developed in association with several factors, including geometric characteristics; driver demographics; as well as environmental and traffic characteristics associated with flooding related crashes. From the generated rules, several attributes have been identified as the key contributing traits in flooding related crashes. Some of the key findings are below:

- Violation of rules has been found in many of the rules. Other studies also show that many drivers underestimate the depth of the flooding water and speed of water over the roadway and enter the flooded road (*6, 17, 20*). Safety campaigning like *"If it's*

Flooded, Forget it" could help to educate drivers about the dangers of driving in the flooding water.
- The rules indicate that rules with high lift values are associated with fatal or injury crashes. Barriers and signs and other related countermeasures can be used to protect drivers to enter into a flooding zone to avoid crashes.
- Two-lane roadways with no separation are also present in several rules. Roadway marking and other retroreflective markers in the centerline and edge line can assist drivers in partially flooded areas.

This research also has some limitations with respect to flooding related crash identification, data size, data structure and attributes, and result presentation. Further research is desirable to enhance current model performance and better investigate a comprehensive list of rules with added variables of interests. Limitations of the current study offer directions for future research in this domain. Data reliability is an issue that is frequently encountered in the modeling of crashes, particularly when crash data is collected from the police reports.

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 & Development (LADOTD).

ACKNOWLEDGMENT

The authors appreciate the assistance provided by Dr. Chiara Silvestri Dobrovolny. The authors appreciate the assistance provided by the students on this project: Bita Maraghehpour, and Magdalena Theel.

AUTHOR CONTRIBUTION STATEMENT

The authors confirm the contribution to the paper as follows: study conception and design: Subasish Das; data collection: Subasish Das and Sarvesh Goel; analysis and interpretation of results: Subasish Das; draft manuscript preparation: Subasish Das, Xiaoduan Sun, Sarvesh Goel, Ming Sun, M. Ashifur Rahman, & Isha Narsaria. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. National Weather Service (NWS). NWS Preliminary US Flood Fatality Statistics. https://www.weather.gov/arx/usflood Accessed: July 2019.
2. Highway Safety Research Group. Louisiana Crash Data Reports. http://datareports.lsu.edu/Reports/SpecialReports/2017_Factbook.pdf Accessed: July 2019.
3. Terrell, D. The Economic Impact of the August 2016 Floods on the State of Louisiana. L. T. and Associates, LLC, 2017.
4. Chen, L., S. P. Baker, E. R. Braver, and G. Li. Carrying passengers as a risk factor for crashes fatal to 16 and 17-year-old drivers. Journal of the American Medical Association, 2000, 283, 1578–1582.
5. Rappaport, E. N. Loss of Life in the United States Associated with Recent Atlantic Tropical Cyclones. Bulletin of the American Meteorological Society, 2000, 81, 2065–2074.
6. Drobot, S. D., C. Benight, and E. C. Gruntfest. Risk factors for driving into flooded roads. Environ. Hazards, Advances and challenges in flash flood warnings, 2007, 7, 227–234.

7. Mannering, F. L., and C. R. Bhat. Analytic methods in accident research: Methodological frontier and future directions. Analytic Methods in Accident Research, 2014, 1, 1–22.
8. Tamerius, J. D., X. Zhou, R. Mantilla, and T. Greenfield-Huitt. Precipitation Effects on Motor Vehicle Crashes Vary by Space, Time, and Environmental Conditions. Weather, Climate, and Society, 2016, 8, 399–407.
9. Peng, Y., M. Abdel-Aty, Q. Shi, and R. Yu. Assessing the impact of reduced visibility on traffic crash risk using microscopic data and surrogate safety measures. Transportation Research Part C: Emerging Technologies, 2017, 74, 295–305.
10. Yu, R., M. Abdel-Aty, and M. Ahmed. Bayesian random effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors. Accident Analysis & Prevention, 2013, 50, 371–376.
11. Abdel-Aty, M., A. A. Ekram, H. Huang, and K. Choi. A study on crashes related to visibility obstruction due to fog and smoke. Accident Analysis & Prevention, 2011, 43, 1730–1737.
12. Jung, S., X. Qin, and D. A. Noyce. Rainfall effect on single-vehicle crash severities using polychotomous response models. Accident Analysis & Prevention, 2010, 42, 213–224.
13. Black, A. W., G. Villarini, and T. L. Mote. Effects of rainfall on vehicle crashes in Six U.S. States. Weather Clim. Soc., 2016, 9 (1), 53-70.
14. Kouchaki, S., H. Roshani, J. A. Prozzi, and J. H. Bernardo. Evaluation of aggregates surface micro-texture using spectral analysis. Construction and Building Materials, 2017, 156, 944–955.
15. French, J. G., R. Ing, S. V. Allmen, and R. Wood. Mortality from flash floods: A review of the National Weather Service reports, 1969–81. Public Health Rep., 1983, 98, 584–588.
16. Mooney, L. E. Applications and implications of fatality statistics to the flash flood problem. Fifth Conference on Hydrometeorology, American Meteorological Society, 1983, 127–129.
17. Ashley, S. T., and W. S. Ashley. Flood Fatalities in the United States. Journal of Applied Meteorology and Climatology, 2008, 47, 805–818.
18. Berz, G., W. Kron, T. Loster, E. Rauch, J. Schimetschek, and A. Schmieder. World map of natural hazards – a global view of the distribution and intensity of significant exposures. Nat. Hazards, 2001, 23, 443–465.
19. Yale, J. D., T. B. Cole, H. G. Garrison, C. W. Runyan, and J. K. R. Ruback. Motor vehicle-related drowning deaths associated with inland flooding after hurricane Floyd: a field investigation. Traffic Injury Prevention, 2003, 4, 279–284.
20. Clarkin, K., G. Keller, T. Warhol, and S. Hixson. Low-water crossings: geomorphic, biological, and engineering design considerations. Washington, D.C, 2006.
21. Heitmuller, F. T., W. H. Asquith, X. Fang, D. B. Thompson, and K. Wang. Guidance for Design in Areas of Extreme Bed-Load Mobility, Edward Plateau, Texas, Texas Department of Transportation, 2005.
22. Das, S., A. Dutta, M. Jalayer, A. Bibeka, and L. Wu. Factors Influencing the Patterns of Wrong-Way Driving Crashes on Freeway Exit Ramps and Median Crossovers: Exploration Using 'Eclat' Association Rules to Promote Safety. International Journal of Transportation Science and Technology, Vol. 7, No. 2, 2018, pp. 114–123. https://doi.org/10.1016/j.ijtst.2018.02.001.
23. Das, S., A. Dutta, R. Avelar, K. Dixon, X. Sun, and M. Jalayer. Supervised Association Rules Mining on Pedestrian Crashes in Urban Areas: Identifying Patterns for Appropriate Countermeasures. International Journal of Urban Sciences, 2018, pp. 1–19.

24. Das, S., L. Minjares-Kyle, and R. Avelar. Improper Passing-Related Crashes on Rural Roadways: Using Association Rules Negative Binomial Miner, 2017.
25. Das, S., X. Kong, and I. Tsapakis. Hit and Run Crash Analysis Using Association Rules Mining. Journal of Transportation Safety & Security, Vol. 0, No. 0, 2019, pp. 1–20. https://doi.org/10.1080/19439962.2019. 1611682.
26. Weng, J., J. Z. Zhu, X. Yan, and Z. Liu. Investigation of Work Zone Crash Casualty Patterns Using Association Rules. Accident Analysis & Prevention, Vol. 92, 2016, pp. 43–52.
27. Hahsler, M., B. Grün, K. Hornik. Introduction to arules – a computational environment for mining association rules and frequent item sets. J. Stat. Softw. 14 (15), 2015, 1–25.
28. Abraham, M. J., J. Price, F. A. Whitlock, and G. Williams. The Brisbane floods, January 1974, their impact on health. Medical Journal of Australia-2, 1976, 936–939.
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*, Vol. 28, No. 1, 1996, pp. 131–138. https://doi.org/10.1016/0001-4575(95)00055-0.
30. Ajzen, I. The theory of planned behaviour. Organ. Behav. Hum. Decis. Process, 1991, 50, 179–211.
31. Michie, S., and M. M. Johnston. Theories and techniques of behavior change: Developing a cumulative science of behavior change. Health Psychology Review, 2012, 6, 1, 1-6.
32. Clarkin, K., G. Keller, T. Warhol, and S. Hixson. Low-water crossings: geomorphic, biological, and engineering design considerations. Washington, D.C, 2006.