Crash Contributing Factors and Patterns Associated with Fatal Truck-involved Crashes in Bangladesh: Findings from Text Mining Approach

Conference Paper · January 2023

DOI: 10.6084/m9.figshare.21937229.v1

4 authors:

Ahmed Hossain
University of Louisiana at Lafayette

Xiaoduan Sun
University of Louisiana at Lafayette

Shah Alam
Engineering Alliance

Subasish Das
Texas State University

Crash Contributing Factors and Patterns Associated with Fatal Truck-involved Crashes in Bangladesh: Findings from Text Mining Approach

Ahmed Hossain
(Corresponding Author)
Ph.D. Student
Department of Civil Engineering
University of Louisiana at Lafayette, Lafayette, Louisiana, 70504
Email: ahmed.hossain1@louisiana.edu
ORCID ID: 0000-0003-1566-3993

Xiaoduan Sun, Ph.D., P.E.
Professor
Department of Civil Engineering
University of Louisiana at Lafayette, Lafayette, Louisiana, 70504
Email: xiaoduan.sun@louisiana.edu
ORCID ID: 0000-0001-7282-1340

Shah Alam
Lecturer
Department of Civil Engineering
Rajshahi Science & Technology University, Natore, Bangladesh, 6400
Email: shahalamruet4@gmail.com
ORCID ID: 0000-0001-8766-1503

Subasish Das, Ph.D.
Assistant Professor
College of Science of Engineering, Texas State University
601 University Drive, San Marcos, Texas 78666-4684
E-mail: subasish@txstate.edu
ORCID ID: 0000-0002-1671-2753

Word Count: 7033 words + 5 tables (250 words per table) + 10 figures = 8,283 words

First Submission: July 30, 2022
Updated Manuscript Submission: January 1, 2023

This conference paper has been accepted for publication by the Transportation Research Board (TRB) 102nd Annual Meeting

ABSTRACT

Despite extensive research on traffic injury severities, relatively little is known about the factors contributing to truck-involved crashes in developing countries, especially in the context of Bangladesh. Due to the unavailability of authentic crash data sources, this study collected data from alternative sources such as online English news media reports. The current study prepared a database of 144 truck-involved fatal crash reports during the period of twelve months (January 2021 to December 2021). The crash reports contain a bag of 15,300 words. Several state-of-the-art text mining tools were utilized to identify crash patterns, including word cloud analysis, word frequency analysis, word co-occurrence network analysis, rapid automatic keyword extraction, and topic modeling. The analysis revealed several important crash contributing factors such as the type of vehicle involved (auto-rickshaw, bus, van, motorcycle), manner of collision (head-on), time of the day (morning, night), driver behavior (speeding, overtaking, wrong-way driving), and environmental factors (dense fog). In addition, 'coming from opposite direction' and 'head-on collision' are two important sequences of events in truck-involved crashes. Truck drivers are also involved in crashes with trains at the rail crossing. The findings of this research can assist policymakers in identifying crash avoidance strategies to lower truck-related crashes in Bangladesh.

Keywords: Fatal, head-on, speeding, overtaking, wrong-way driving, rail crossing

INTRODUCTION

In low and middle-income countries (LMIC) like Bangladesh, fatalities, and serious injuries due to traffic crashes have been a major health concern. According to the 2018 World Health Organization (WHO) Global Status Report (*1*), the WHO estimate of 2016 traffic fatalities in Bangladesh was 24,954 (95% confidence intervals: 20,730 – 29,177). The report also provides some other valuable crash statistics: 67% of roadway crash fatalities and injuries in the economically productive age groups (15-64 years old); 15.3 fatalities per 100K population; estimated serious injuries 374, 310; and cost of fatalities and series injuries of $11,630 (5.3% of country GDP). Due to the underdeveloped crash reporting system in Bangladesh (*2*), it is challenging to collect comprehensive data on these collisions and their underlying causes.

Bangladesh has experienced a sharp rise in crashes involving heavy vehicles, particularly trucks, in recent years (*3*). Due to the mass of large trucks, the count and severity of injuries are higher in number in truck-involved crashes. The number of fatalities in truck-involved crashes in Bangladesh is usually extremely high as these crashes are associated with vulnerable road users such as human haulers, cycle or motorcycle riders, rickshaws, and auto-rickshaws. As a result, the morbidity and mortality rates of truck-involved crash injuries have increased, making it one of the most serious safety concerns. For example, **TABLE 1** lists some of the major truck-involved crashes, which were reported in different online news media. The fatal truck-involved crash that occurred on July 11, 2011, in Mirsarai, Chattogram resulted in 44 fatalities, which is an extremely high number for a single road traffic crash.

TABLE 1 Some major truck-involved crashes in Bangladesh

Source	Location	Date	Crash Scenario	Killed	Injured
(*4*)	Mirsarai, Chattogram	July 11, 2011	A pick-up truck veered of the road and toppled into a roadside pond.	44	0
(*5*)	Chunti area, Chittagong district	March 22, 2020	Truck rammed the passenger vehicle.	15	7
(*6*)	Chuadanga, Khulna district	March 26, 2017	A collision between a truck and a mini truck.	10	12
(*7*)	Sherpur, Bogura district	February 21, 2021	A bus crashed into the stone-laden truck.	6	15
(*8*)	Chapainawabganj	May 15, 2021	A truck hit the human hauler from the opposite direction.	3	6
(*9*)	Rangpur	April 26, 2020	Speeding truck hit the auto-rickshaw from behind after the truck driver lost control over the steering.	3	5
(*10*)	Tangail, Dhaka	March 28, 2020	A cement-laden truck overturned.	5	11
(*11*)	Ashulia, Dhaka	January 30, 2019	A brick-laden truck fell into a 40-foot-deep ditch.	4	3
(*12*)	Mirsarai, Chattogram	September 26, 2018	Truck ploughed into a CNG-run auto-rickshaw.	5	4
(*13*)	Natore	January 12, 2022	A collision between a bus, a truck, and a rickshaw van.	2	5

Due to the absence of authentic crash records, this study explored alternative sources to determine the contributing factors to truck-involved crashes in Bangladesh. The current study followed the approach developed by Das et al. (*14*). Das et al. collected fatal crash reports from online English daily news media sources by using Google News Alert. After collecting the data, this study identified critical factors by applying different natural language processing (NLP) tools. The current study focused on only fatal track-involved crashes. State-of-the-art text mining tools were used to identify insights and factors from reported text contents on fatal track-involved crashes. The findings of this study are expected to provide a better understanding of the crash contributing factors to truck-involved crashes in the context of Bangladesh.

LITERATURE REVIEW

The literature review focuses on three key areas: 1) major U.S. studies on truck-involved crashes, 2) major studies on crashes and truck-involved crashes in Bangladesh, and 3) studies on news media-related information extraction for transportation safety.

Major U.S. Studies

Truck-involved crash is a key safety concern in the U.S. As there is detailed information on crash data specifics, studies focused on many issues such as key contributing factors, patterns of risk factors, severity patterns and classification, and frequency analysis using advanced statistical and machine learning models. Dong et al. estimated large truck-related crash frequencies by developing a multivariate Poisson-lognormal (MVPLN) model (*15*). Comparison between car, car–truck, and truck crash counts with respect to different contributing factors were identified in the developed models. Cheung explored undercounting of large truck crash frequencies in federal and state crash databases (*16*). The results show that large truck involvement-related misclassification is significantly high. Large Truck Crash Causation Study (LTCCS) provides in-depth information on crash data, which is usually unavailable in state-maintained traffic crash data. Using LTCCS and naturalistic driving data, Hickman et al. conducted a synthetic risk ratio analysis (*17*). Zheng applied gradient boosting and data mining methods to identify the impact of key contributing factors on crash severities associated with commercial trucks (*18*). This study examined 25 variables and 22 of them were identified as significant variables contributing to injury severities. In Texas, truck-related crash frequencies have increased by 82% from 2009 to 2012. Zhao et al. identified hot spots and developed collision diagrams by collecting traffic crash data from Texas during 2011-2015 (*19*). Rahimi et al. applied a high-dimensional clustering approach to identify patterns of large truck crashes by using crash data from Florida between 2007 and 2016 (*20*). The results suggested that truck-related crash data can be explored in major clusters such as same-direction, opposing-direction, and single-vehicle crashes to better understand the patterns and develop relevant countermeasures. Das et al. applied an innovative dimension reduction method known as taxicab correspondence analysis (TCA) on fatal crash data involving large truck (*21*). Using 2010–2015 large truck fatal crash data from the Fatality Analysis Reporting System (FARS), this study identified five clusters with attributes such as two-lane undivided roadways, intersection types, posted speed limit, crash types, number of vehicles, driver impairment, and weather.

Major Traffic Safety Analysis in Bangladesh

Traffic safety-related studies in Bangladesh are limited. Most of these studies used aggregate data and performed either exploratory data analysis (EDA) or simple statistical analysis.

According to the authors' knowledge, there are two studies that focused solely on truck-related crashes in Bangladesh. A study conducted by Sufian et al. collected data on road traffic accidents in Bangladesh from 1998 to 2010 (*22*). The EDA shows that truck driver's activities contributed to 95 percent of truck crashes. Other factors were associated with pedestrians, vehicular properties, and roadway environment. Conducting a field observation on some bus terminals (Gabtoli truck terminal and Dholaikhal), this study also identified that vehicle-related defects are significantly high in trucks, which could be associated with traffic crashes. Another study utilized the 'Quasi-Static Rollover Model' to investigate the 'rollover criteria' of heavy trucks in Single Vehicle Run-off-Road (SVROR) crashes in Bangladesh (*23*). The study used a hypothetical combination of the geometric dimension of the truck and loading condition to find the critical condition in which rollover occurs. Some of the other research conducted in Bangladesh only briefly examined truck-related collisions. **TABLE 2** below highlights a list of information (year, data source, location, methods, and key findings on truck-involved crashes) from a few previous studies in the context of Bangladesh. It is important to mention here that, the findings of these studies are limited and difficult to assess how basic human, roadway, or environmental factors are associated with truck-involved crashes.

TABLE 2 Findings from the previous study

Study	Year; Data Source	Location	Methods	Key findings on truck-involved crash
(*24*)	Bangladesh police report from 2004-2015	Bangladesh	Ordered probit model	• There is a greater likelihood of a car-truck rear-end accident when the brakes are applied suddenly to a moving vehicle. • Serious injuries occur in collisions between a heavy truck and a motorcycle.
(*25*)	First Information Report (FIR), 2001-2002	Khulna	Descriptive analysis	• Trucks were involved in 26% of the accidents. • Trucks were involved in around 47% of the crashes resulting in pedestrian fatalities.
(*26*)	Highway police station, 2008 to 2015	Rajshahi-Sirajganj	Descriptive analysis	Involvement of truck was identified in about 39% of the crashes.
(*23*)	Accident Report Form (ARF)	Bangladesh	Quasi-Static Rollover Model	• Heavy trucks are involved in around 21% of overturning accidents. • Excessive speeding and reckless driving were identified as the prime causes of rollover-type ROR crashes.
(*27*)	Police report	Bangladesh	Descriptive analysis	Trucks were involved in around 33% of the of fatal crashes.
(*28*)	Police reported in Bangladesh, 1998-2005	Bangladesh	Descriptive analysis	• The involvement of buses/trucks was identified as most of the pedestrian-vehicle conflicts. • Approximately 27% of road accidents involve trucks.

Traffic Safety Analysis using News Media Mining

News media mining has been becoming increasingly popular among scientists when conventional data is limited. Das et al. applied different NLP tools to news media data to extract insights into fatal crashes in Bangladesh (*14*). In another study, Das et al. collected news media reports on the impact of speeding on crashes during the COVID-19 (*29*). This study applied text network analysis to identify the patterns of risk factors. This study developed topic models and interactive topic model web tools to explain the keywords and their significance in each topic. Yang et al. explored massive media reports to develop an e-scooter crash database (*30*). This study identified 169 e-scooter-related traffic crashes from the news reports during 2017-2019. This study also conducted an EDA on the developed crash datasets. Karpinski et al. identified 21 shared e-scooter fatalities in the U.S. from 2018 to 2020 by exploring media reports (*31*). This study explored the reports to identify potential risk factors. This study found that most crashes (86%) involved motor vehicles and 28% of these were hit-and-runs. Keliikoa et al. explored local media news coverage of non-motorist-related crashes (*32*). The results show that news article titles were usually non-agentive (77%) and focused on pedestrians or bicyclists (77%). This study also found that one-quarter of articles had implications towards making the non-motorists ‘at-fault’.

Research Gap and Study Objective

To the best of our knowledge, no prior study explored contributing factors and their patterns for truck-related crashes in the context of Bangladesh using the text mining approach. An in-depth analysis is thus needed to mitigate the current research gap. Therefore, the objective of this study is to identify the crash contributing factors and patterns associated with truck-involved crashes in Bangladesh by applying text mining tools. The findings of this research are expected to assist transportation experts and policymakers in identifying crash avoidance strategies to lower truck-related crashes in Bangladesh.

METHODS

The technique of extracting data from a collection of text is known as text mining or text data mining. In 1999, in research conducted by Hearst first used the term ‘text data mining’ and distinguished it from other ideas like natural language processing (*33*). The objective of text mining is to discover information and patterns from text data which can be unstructured or semi-structured (*34*). Along with applications in other research domains, text mining has become an increasingly popular approach in transportation safety research. Some of the recently published articles are related to heavy vehicle crashes (*35*), rail accidents (*36*), evaluating roadway crashes for road asset management (*37*), mining highway-rail grade crossing crash data (*38*), classification of roadway traffic injury collision characteristics (*39*), pedestrian violation behaviors (*40*), work zone crashes (*41*) and so on.

The basic methodology of text mining is to transform the text into a numeric dataset. To facilitate this, Term Document Matrix (TDM) method is utilized. In TDM, the text data is represented in the form of a matrix. Before creating TDM, basic preprocessing of the text is required including removal of punctuation, stopwords (common English words), white space, numeric numbers, and special characters. Also, all the text is converted into lowercase to reduce the variation of the same word. For example, ‘Accident’ and ‘accident’ are treated as the same words after transformation.

Term-Document Matrix (TDM)

TDM represents the document vector in matrix format. In this matrix, the rows correspond to the terms (or words) in the document, columns correspond to the documents in the corpus (complete collection of documents) and cells correspond to the weights of the terms. The weights are either 0 or 1. Here 1 indicates the presence and 0 indicates the absence of the term in a particular document. For example, let us consider the following three documents.

D1: An **elderly man** was **killed** in a **road accident**.
D2: The **truck driver lost control** and **hit** the **tree**.
D2: **Two men riding** a **motorcycle** have **died after** being **run over** by a **truck**.

These documents can be converted into TDM after basic preprocessing. The candidate terms are indicated in bold. **TABLE 3** represents the TDM for the three documents (D1, D2, D3).

TABLE 3 Example of Term Document Matrix (TDM)

Terms	Documents		
	D1	**D2**	**D3**
elderly	1	0	0
man	1	0	0
killed	1	0	0
road	1	0	0
accident	1	0	0
truck	0	1	1
driver	0	1	0
lost	0	1	0
control	0	1	0
hit	0	1	0
tree	0	1	0
two	0	0	1
men	0	0	1
riding	0	0	1
motorcycle	0	0	1
died	0	0	1
after	0	0	1
run	0	0	1
over	0	0	1

The word ‘elderly’ is present in document D1, which is why it is coded ‘1’ in document D1, but ‘0’ in the other two documents. The next important step in the text mining approach is the identification of Term Frequency (TF), Document Frequency (DF), and Inverse Document Frequency (IDF). Let, ‘t’ indicates terms (words), ‘d’ indicates documents (set of words), and ‘N’ is the count of the corpus. Note that the corpus is the total document set.

Term Frequency (TF)

The term frequency measures how frequently a term occurs in a document. The equation for TF is provided below.

$$TF\ (t,d) = \frac{Count\ of\ t\ in\ d}{Total\ number\ of\ terms\ in\ d}$$

Document Frequency (DF)

DF measures the importance of documents in the whole set of corpora. In other words, DF is the number of documents in which a specific word is present. For a specific word ‘t’, DF(t) is the occurrence of t in documents.

Inverse Document Frequency (IDF)

IDF is simply the inverse of DF. The IDF of a term indicates how frequently the term appears in a corpus that contains the term. The equation for IDF is provided below.

$$IDF\ (t,d) = \log[\frac{Total\ number\ of\ documents\ in\ the\ corpus}{Total\ number\ of\ documents\ in\ the\ corpus\ that\ contain\ the\ term}]$$

To avoid very low values of IDF, a logarithm is used. The TF-IDF of a term is calculated by multiplying TF and IDF scores. A hypothetical example is provided below:

Consider a document containing 100 words where the word ‘truck’ appears 5 times. The TF for the word ‘truck’ is 5 divided by 100 or 0.05. Now assume that out of a total of 10 million documents, 1,000 of them include the term ‘truck’. Then, the measure IDF can be calculated as the logarithm of (10,000,000 /1,000) = 4. Finally, the TF-IDF weight can be found as 0.05 multiplied by 4, which is 0.20.

Word Correlation

The measurement of word correlation determines whether certain words are found together. Let us consider two words ‘A’ and ‘Z’ and their appearance in the document is considered in a binary format – ‘1’ for presence, ‘0’ for absence. For example, N_{11} represents the number of documents where both word ‘A’ and word ‘Z’ appear, N_{00} is the number where neither word appears, and N_{10} and N_{01} are the cases where one appears without the other. To measure such a binary relationship, a phi coefficient (ϕ) is used. The equation for the ϕ co-efficient is provided below.

$$\varphi = \frac{N_{11}N_{00} - N_{10}N_{01}}{\sqrt{N_1 N_0 N_0 N_1}}$$

Rapid Automatic Keyword Extraction (RAKE)

RAKE is a method for extracting keywords from individual documents. The RAKE algorithm is unsupervised in nature, independent of both domain and language. The basic steps of RAKE are the determination of word degree, word frequency, and the ratio of the degree to frequency (also known as score) shown in **TABLE 4**. The degree of the word ‘XYZ’ can be found by counting the number of words that occur in candidate keywords containing ‘XYZ’, including ‘XYZ’ itself. Word frequency is simply the number of times the word occurs in the entire text. For illustration, the following sentence can be considered.

*“The **truck driver lost control** and **hit** the **tree**.”*

Here, the content words (total = 6) are shown in bold in the above sentence. Now, we need to define candidate keywords. Let us define three candidate keywords –

- truck driver
- lost control
- hit tree

In the next step, a word degree matrix is constructed where each row shows the number of times a given content word co-occurs with another content word in candidate keywords. For example, the word 'truck' appears with the word 'driver' and is coded as '1'. The diagonal portion of the table consists of only '1' because each word appears once in the text. Finally, the score for each individual word can be calculated as the ratio of the degree of the word and the word frequency. Now, the score for each candidate keyword is simply the combined sum of each individual score. For example, the score of the keyword 'truck driver' is 4.

TABLE 4 Example of Rapid Automatic Keyword Extraction process

Word Degree Matrix						
	truck	driver	lost	control	hit	tree
truck	1	1	0	0	0	0
driver	1	1	0	0	0	0
lost	0	0	1	1	0	0
control	0	0	1	1	0	0
hit	0	0	0	0	1	1
tree	0	0	0	0	1	1
Degree	**2**	**2**	**2**	**2**	**2**	**2**

Score Calculation			
Word	Degree of word	Word Frequency	Score
truck	2	1	2
driver	2	1	2
lost	2	1	2
control	2	1	2
hit	2	1	2
tree	2	1	2

Topic Modeling

Topic modeling is a machine learning technique that automatically analyzes text data to determine cluster words that frequently occur together within a set of documents (*42*). The research team utilized the LDA algorithm, which is the most popular topic modeling technique, to extract topics from a given corpus. Previous research conducted by Blei et al. provides a handy resource for the theoretical evolution of the LDA (*43*). There are two basic assumptions of LDA modeling– (a) every word is a combination of an underlying set of topics, and (b) every topic is a combination of a set of topic probabilities. The working flow diagram of the LDA algorithm is provided in the following **Figure 1**.

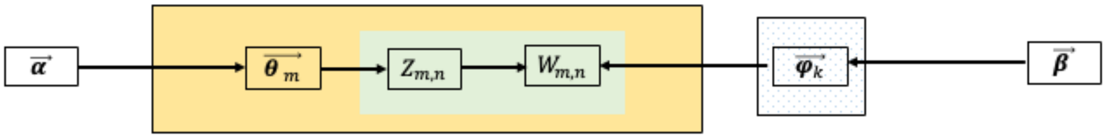

Figure 1. Working flow diagram of LDA model (*42*)

The interpretation of the LDA parameters are as follows: $\vec{\alpha}$ = Dirichlet parameter which controls per-document topic distribution, $\overrightarrow{\theta_m}$ = Document topic distribution, $Z_{m,n}$ = Word topic assignment, $W_{m,n}$ = Observed word, $\overrightarrow{\varphi_k}$ = Topic word distribution, and $\vec{\beta}$ = Dirichlet parameter which controls per topic word distribution.

Here, the yellow box refers to all the documents in the corpus and the green color box is the number of words in a document. According to LDA, every word is associated with a latent topic, which here is stated by Z. This assignment of Z to a topic word in these documents gives a topic word distribution present in the corpus that is represented by theta (θ). The LDA algorithm is an iterative process. The end goal of LDA is to find the most optimal representation of the Document-Topic matrix and the Topic-Word matrix to find the most optimized Document-Topic distribution and Topic-Word distribution.

The research team utilized R statistical software (version 4.2.0) to conduct the analysis. A wide range of open-source R software packages was utilized including ‘wordcloud2’, ‘topicmodels’, ‘tm’, ‘syuzhet’, ‘rapidraker’, and ‘quanteda’.

DATA PREPARATION

Working with Google Alerts

Google Alerts (google.com/alerts) is a notification service that allows users to get information such as web pages, newspaper articles, blogs, or scientific research that matches their search keywords. The research team used this tool to collect crash reports from online sources in Bangladesh from January 2021 to December 2021. To narrow down searches, the research team set the following keywords in the google alert service:

- Bangladesh accident
- Bangladesh road collision
- Bangladesh road crash
- Bangladesh road fatalities and injuries
- Bangladesh traffic accident
- Bangladesh traffic crash
- Highway crashes in Bangladesh

The articles were pulled from news agencies, most of which are local newspapers in Bangladesh. Every article was manually entered into the dataset. The same crash may occasionally be covered by multiple online newspapers. For this reason, in this step, every report was carefully examined to eliminate duplicates. The final prepared dataset contains a total of twelve variables including source, headline, date published, narratives, killed, injured, district, division, crash time, crash type, vehicle 1, and vehicle 2. The variable 'narratives' contains the crash reports collected from online sources and subsequently processed for applying text mining. The following **Figure 2** shows the database preparation and analysis flowchart in detail.

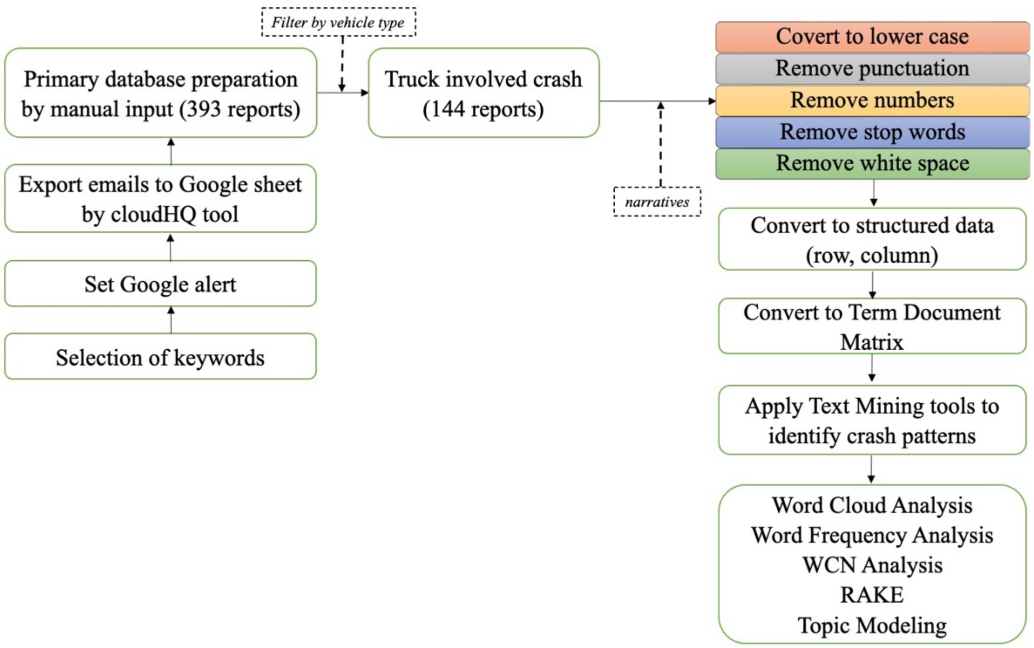

Figure 2. Database preparation and analysis flowchart

To demonstrate the structure of crash narratives, the one example of crash narratives below was randomly selected from the dataset.

"An elderly man was killed in a road accident in the Bejerdanga area on the Jashore-Khulna highway under Phultala upazila of Khulna district last night. The deceased was identified as Farazi Ashraf Hossain, 68, hailed from Abhainagar upazila of the district. Police said the accident occurred when a speeding truck hit Ashraf Hossain in the area as he was crossing the highway around 9.30 pm, leaving him dead on the spot. Police seized the truck and held its driver. A case was filed in this connection".

Exploratory Data Analysis (EDA)

The following **TABLE 5** provides a summary of the crash database used in this study. A total of 144 online news reports related to truck-involved accidents were collected during the study period. The mean value for each total word in each report was found as 106.3 with a 95% confidence interval of 6.5. Other truck-related crash statistics, including the number of fatalities and injuries, age, and gender, were also extracted from the news reports. For example, truck-involved crashes resulted in 627 fatalities and injuries in 2021 in Bangladesh, with 51.5% of those killed and 48.5% of those injured. Around 42.4% of the crash-involved individuals were male. Note that the age and gender of most of the crash-involved individuals were not reported in the online news.

TABLE 5 Descriptive statistics

Summary by news source		
Source Name	**Number of Crash Reports**	**Statistics for total words in each report**
Bangladesh Post	9	Mean = 106.3 Median = 101.5 Standard error = 3.3 Standard deviation = 39.2 95% confidence interval = 6.5 Minimum = 38 Maximum = 243 1^{st} Quartile = 77.75 2^{nd} Quartile = 101.5 3^{rd} Quartile = 127.25
bdnews24.com	13	
Dhaka Tribune	19	
Jagonews24	1	
New Age	1	
Risingbd	4	
The Asian Age	5	
The Daily Star	43	
The Financial Express	7	
The Independent	14	
United News of Bangladesh	28	
Total collected reports	**144**	

Division	Count of reports	Killed	Injured	Killed and injured
Barisal	1	3	0	3
Chattogram	25	56	36	92
Dhaka	35	71	41	112
Khulna	25	63	101	164
Mymensingh	6	21	6	27
Rajshahi	30	67	71	138
Rangpur	11	22	13	35
Sylhet	11	20	36	56
Grand Total	**144**	**323**	**304**	**627**

Summary by age group		
Age (years)	**Frequency**	**Percentage**
<15	22	3.5
15-30	93	14.8
31-45	78	12.4
>45	35	5.6
Not reported	399	63.6

Summary by gender		
Gender	**Frequency**	**Percentage**
Male	266	42.4
Female	53	8.5
Not reported	308	49.1

The following **Figure 3** shows the spatial distribution of truck-involved fatalities and injuries in Bangladesh. The top five districts in which truck-involved fatalities and injuries occurred were Jhenaidah (84), Sylhet (44), Bogra (40), Dhaka (36), and Jashore (35).

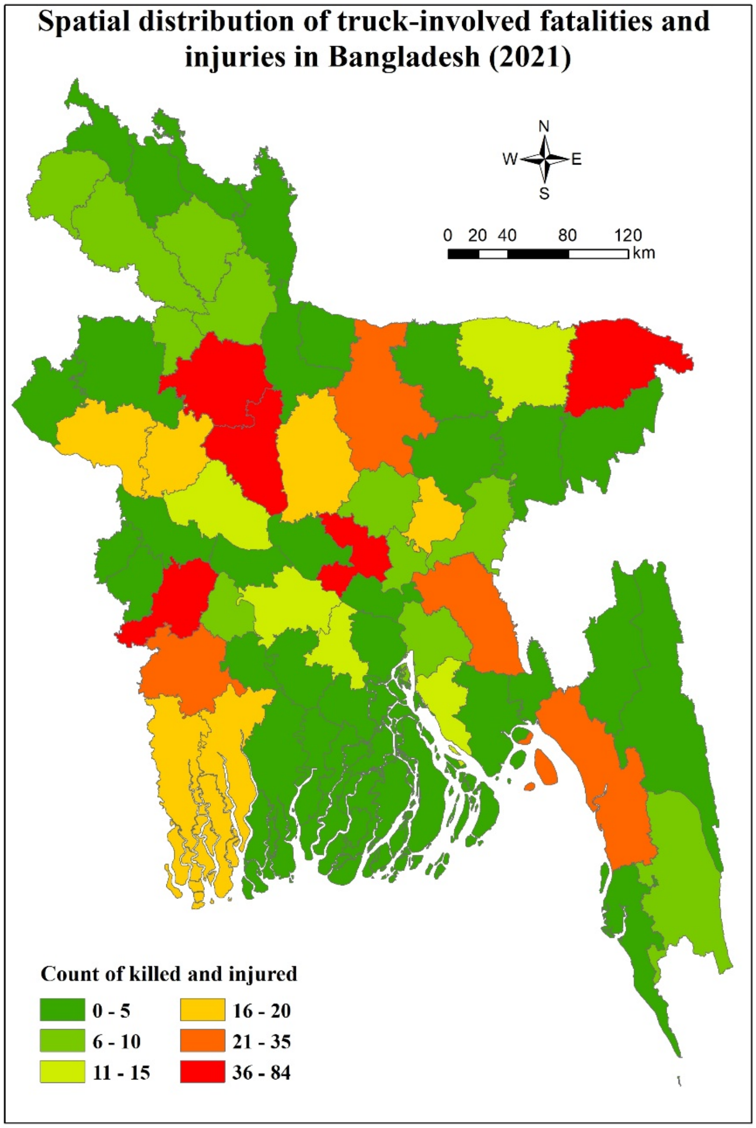

Figure 3. Spatial distribution of fatalities and injuries in truck-involved crashes in Bangladesh

To understand the temporal variation of truck-involved fatalities and injuries during 2021, the following time-series graph **(Figure 4)** is plotted. The term ‘ratio’ in the following graph indicates the proportion of the total killed and injured in all traffic crashes divided by the killed and injured in truck-involved crashes. For total crashes, peaks are observed in July and September. Since the Eid (the biggest Muslim celebration) holiday is celebrated in July, there are more crashes that occur throughout that month. This is in line with a recent investigation by Aljazeera (*44*). Additionally, more crashes are reported at the beginning and end of the year, probably because of the foggy weather during those times of the year.

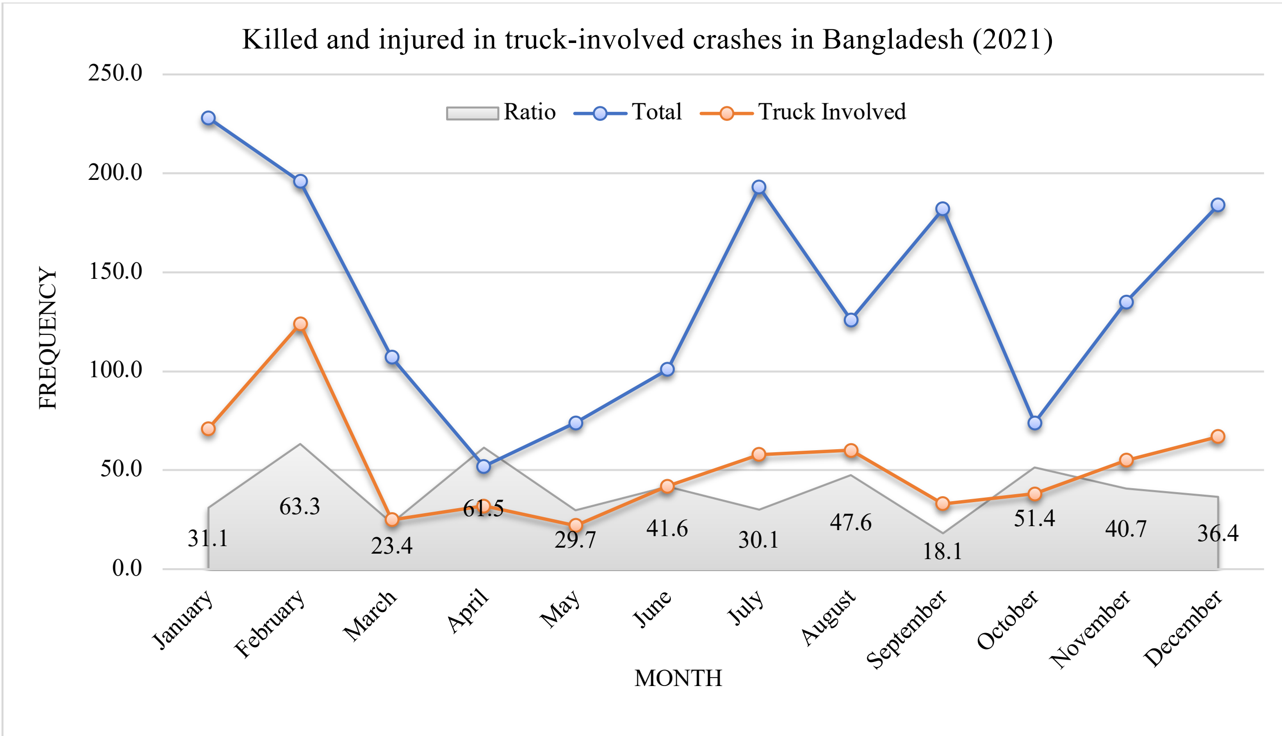

Figure 4. Temporal variation of truck-involved and total crashes

RESULTS AND DISCUSSION

TF-IDF Results

In this analysis, a total of 144 crash reports (i.e., documents) were utilized which consists of a total of 15,300 words (i.e., terms). After preprocessing, only 1,950 words remain. Therefore, the final matrix used for the analysis consists of 144 rows and 1,950 columns with a total of 280,800 elements or cells. Only 7,234 out of these total 280,800 cells had non-zero entries. Therefore, the sparsity of this TDM was found to be 97.42%, suggesting that 97.42% of the 280,800 cells had zero entries.

Word Cloud Analysis

The quantitative analysis of keywords was done using word cloud analysis to provide a visual representation of crash narratives. **Figure 5** shows the word cloud with the 150 most frequently used words in the reported truck-involved crashes on online news in Bangladesh. Note that the bigger the letters of the word in the picture, the more often it occurs in the text. For example, ‘police’, ‘said’, ‘upazila’, ‘injured’ ‘accident’, ‘highway’, ‘deceased’, ‘killed’, ‘station’, and ‘hospital’ are some of the most frequent words in the reports. The general findings of this word cloud analysis are – truck accidents causing injuries and fatalities, police involvement in the crash scene for investigation, and transportation of crash victims to the medical college hospital.

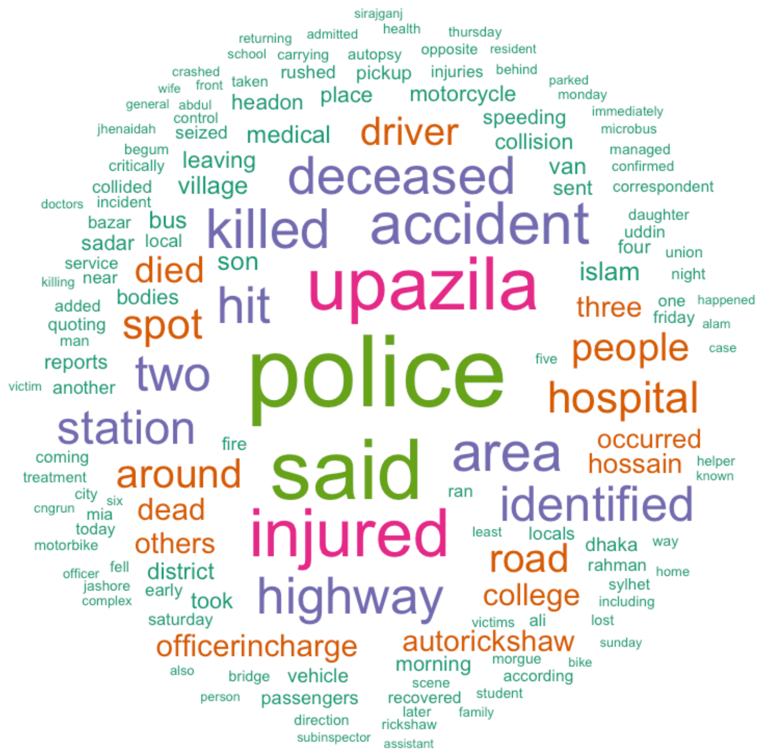

Figure 5. Word cloud analysis

Word Frequency Analysis

To identify the frequency associated with words, the following bar plot is provided **(Figure 6).**

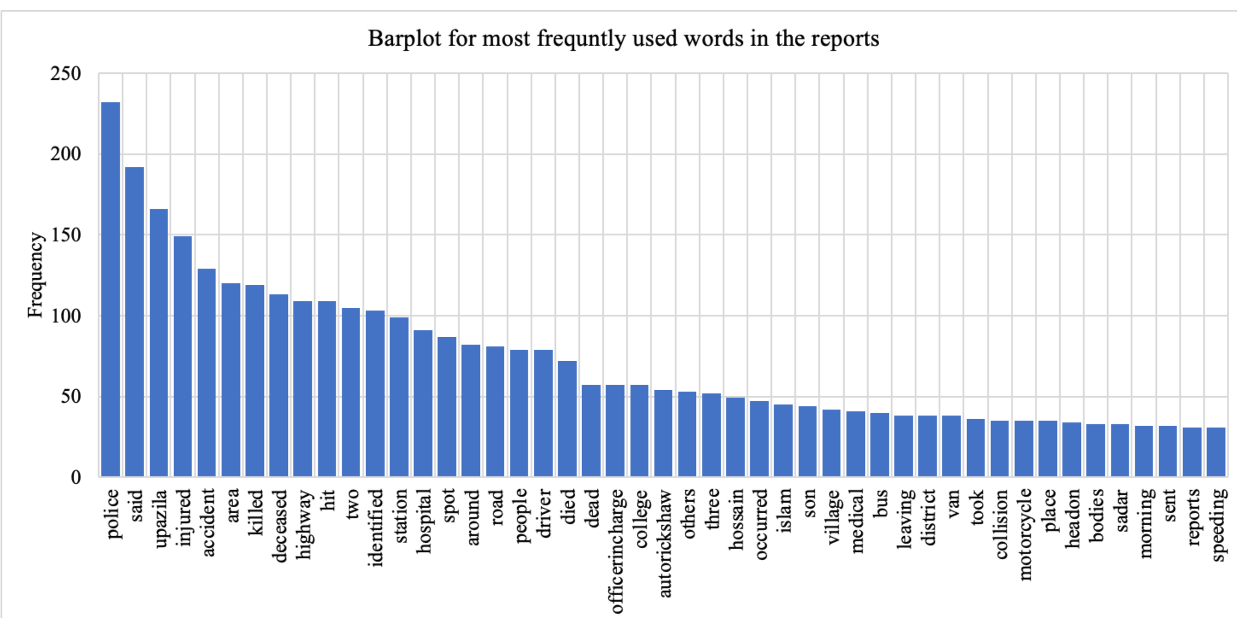

Figure 6. Word frequency plot

Note that, a minimal frequency level is specified over which the words are included in the bar plot, allowing for the discovery of intriguing words related to truck-involved crashes. For example, the minimum frequency threshold was set as 30. Therefore, any words that appeared at least 30 times in the bag of words were included in the barplot.

There are a total of 47 words in the above barplot. The top five most frequent words in the above bar plot are police (232 times), said (192 times), upazila (166 times), injured (149 times), and accident (129 times). The barplot also contains some useful information about vehicle type, manner of collision, and time of the day. For example, four different vehicle types appeared in the above bar plot including auto-rickshaw (54 times), bus (40 times), van (38 times), and motorcycle (35 times). This implies that the majority of fatal truck-involved crashes occurred due to collisions with buses, motorcycles, auto-rickshaw, and vans. The term 'head-on' was used 34 times, indicating that head-on collisions were the primary cause of the majority of truck-related accidents. The word 'speeding' appeared 31 times, suggesting that speeding is one of the crash contributing factors to truck-involved crashes. The word 'morning' appeared 32 times, which suggests that morning hours were the most common time for truck-related crashes.

Word Co-occurrence Network (WCN) Analysis

WCN is a key tool for visualizing the relationship among words that appear together in a sentence. The following **Figure 7** shows the WCN plot for truck-involved crashes.

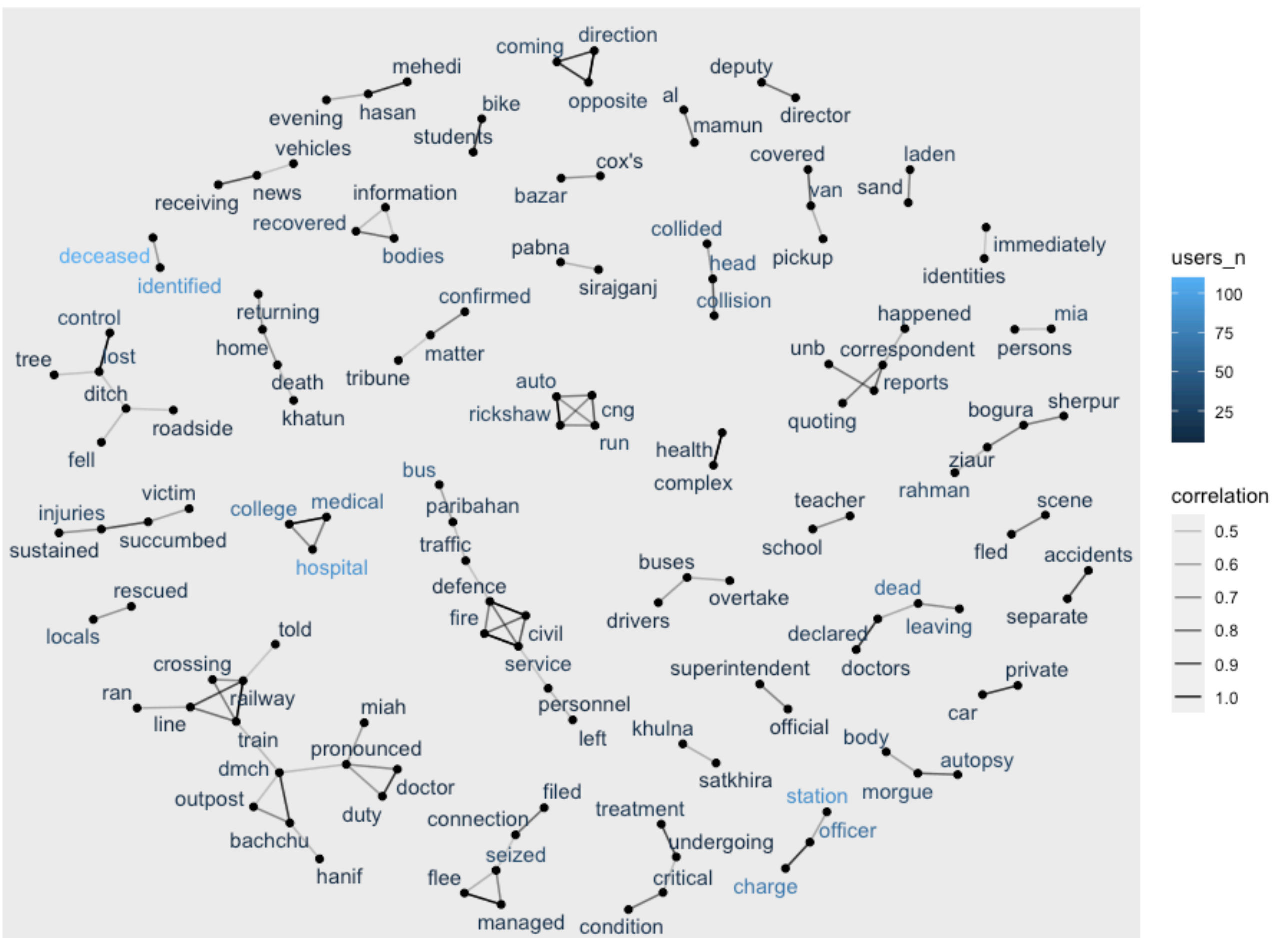

Figure 7. Word Co-Occurrence Network (WCN) plot for truck-involved crashes

WCN is created by joining the vertices of n consecutive words in a sentence. Two important parameters (n, ϕ) are required to produce a WCN plot. The parameter ‘n’ stands for the minimum number of users (reports in this case) who used these words. The parameter ‘ϕ’ indicates the minimum correlation among consecutive words. The value of these parameters was set as n = 5 and ϕ = 0.40 after several trials and errors. The selection of these parameters is based on subject matter experience and the identification of meaningful patterns. Some of the meaningful co-occurrences of terms selected from the above WCN plot are explained below.

- A link is visible around the words ‘head’, ‘collision’, and ‘collided’. It indicates that the majority of truck-involved crashes in Bangladesh occur due to head-on collisions with other vehicles. Another link is observed among the words ‘coming’, ‘opposite’, and ‘direction’. Therefore, ‘coming from opposite direction’ and ‘head-on collision’ are two important sequences of events in truck-involved crashes.
- Another link is spotted around the words ‘train’, ‘line’, ‘railway’ ‘crossing’, ‘ran’, and ‘told’. It indicates a crash scene involving a train and truck at the railway crossing. In Bangladesh, fatal accidents at railroad crossings occur frequently as most of them are left unauthorized or unattended.
- The connection among the words ‘lost’, ‘control’, ‘tree’, ‘fell’, ‘roadside’, and ‘ditch’ indicates a series of events in truck-involved crashes.
- A link is observed among the words ‘buses’, ‘drivers’, and ‘overtake’, suggesting that ‘overtaking’ is an important crash contributing factor in truck-bus crashes.
- The words ‘fled’ and ‘scene’ are connected (ϕ = 0.56), possibly pointing to a hit-and-run accident involving truck drivers. The words ‘managed’, ‘flee’, and ‘seized’ are linked similarly in another instance, implying two distinct scenarios: (a) truck drivers who managed to flee and (b) truck drivers who were seized while attempting to flee.
- The correlation coefficient between the words ‘bike’ and ‘students’ was found as 0.73. It implies a collision involving students on motorcycles and trucks.
- The network plot also offers some important details regarding the type of vehicles involved in truck-related accidents in Bangladesh. Some of these crash-involved vehicles are CNG Auto Rickshaws (link is observed among ‘cng’, ‘auto’, ‘rickshaw’, ‘run’), Covered Vans/Pickups (link is observed among ‘covered’, ‘van’, ‘pickup’), sand-laden trucks (link is observed between ‘sand’, and ‘laden’), and private cars (link is observed between ‘private’, ‘car’).
- Some of the crash locations/routes were also identified in the network plot, including Cox’s Bazar (ϕ = 0.53), Bogura-Sherpur (ϕ = 0.55), Pabna-Sirajganj (ϕ = 0.43), and Khulna-Satkhira (0.44).

Rapid Automatic Keyword Extraction (RAKE)

The RAKE algorithm identified a total of 4,102 keywords. The distribution of these keywords is as follows – 1-gram (2,500), 2-gram (1,131), 3-gram (292), 4-gram (118), 5-gram (27), 6-gram (26), 7-gram (6), 8-gram (1), and 9-gram (1). In general, 2-gram (consisting of two words) provides the most useful information. The research team reviewed all of the 1,131 keywords and identified key factors that are most likely to have played a role or were associated with the truck crashes. The following **Figure 8** shows keywords (selected 2-grams) extracted by the RAKE process.

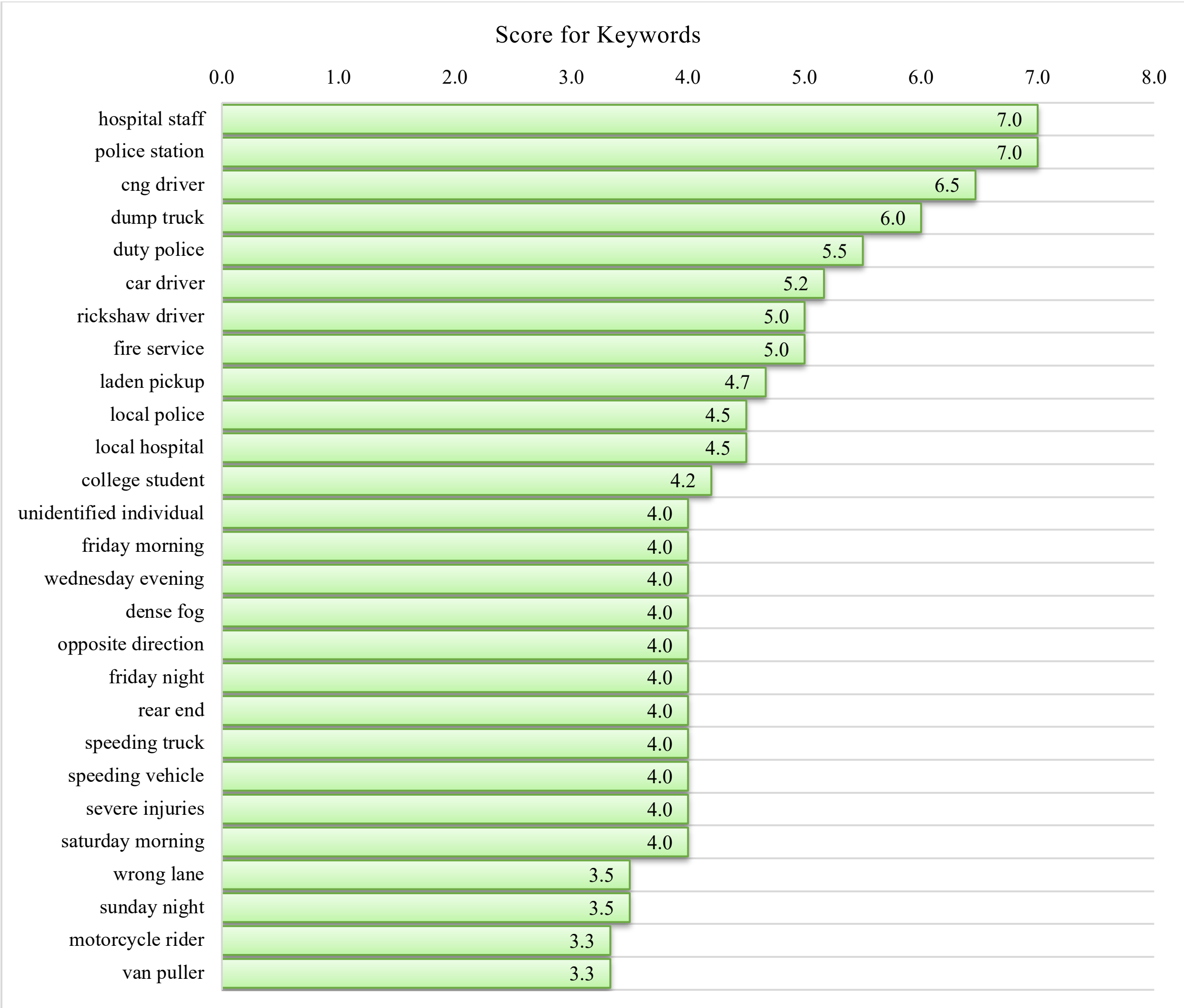

Figure 8. RAKE for truck-involved crashes

In Bangladesh, police, medical, and fire service personnel frequently respond to crash scenes. Some of the keywords that support this statement include hospital staff (s = 7), police station (s = 7), duty police (s = 5.5), fire service (s= 5), local police (s = 4.5), and local hospital (s = 4.5). The driver of different vehicle types was found to be associated with the truck-involved crashes, including cng driver (s = 6.5), car driver (s = 5.2), rickshaw driver (s = 5), and van puller (s = 3.3). Additionally, the involvement of other vehicles (dump trucks, laden pickups) was also observed in crashes involving trucks. The time of the day was also identified as an important factor in truck-involved crashes. Some of the examples of different times of the day were Friday morning (s = 4), Wednesday evening (s = 4), Friday night (s = 4), Saturday morning (s = 4), and Sunday night (s = 3.5). The 'dense fog' keyword (s=4) was identified, suggesting that poor environmental conditions play a role in truck-involved crashes. In crashes involving trucks, driver behavior also makes a substantial contribution. For example, 'speeding truck' (s=4), 'speeding vehicle' (s=4), and 'wrong lane' (s=3.5) keywords were identified. The majority of truck-involved crashes occur with vehicles coming from the opposite direction (s=4).

Topic Modeling Results

The research team utilized the LDA (Latent Dirichlet Allocation) algorithm for topic modeling. One important step in the LDA technique is to select the number of topics. The research team used the 'coherence score' as a measure to select the optimum number of topics. The plot of the coherence score against the number of topics is provided below **(Figure 9).**

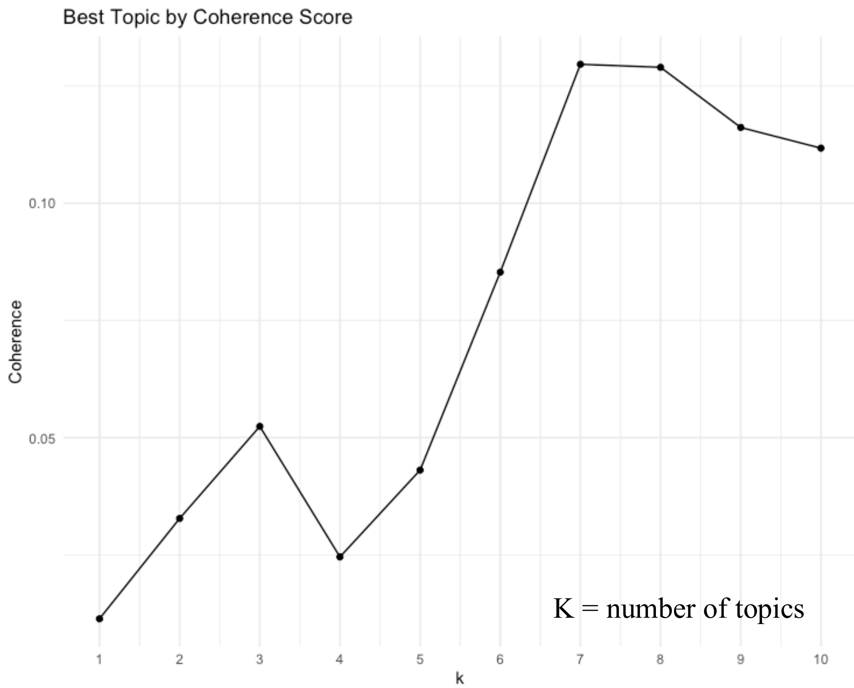

Figure 9. Selection of the optimum number of topics by coherence score

The plot reveals that k = 7 provides the highest coherence score. The results obtained from topic modeling after setting k= 7 are provided below **(Figure 10).** Note that most of the topics include the word 'police', suggesting their involvement in the investigation of truck-involved crashes. Some other useful findings from the topic modeling are provided below.

Topic 1: Describes truck-involved crash scenario on Friday morning resulting in fatalities.
Topic 2: Describes the involvement of police such as the officer-in-charge at the crash spot.
Topic 3: Describes a crash scene involving a van with a pickup.
Topic 4: Describes truck-involved crash scenario in Dhaka where injured and killed persons were taken to the medical college hospital for treatment.
Topic 5: Describes the involvement of motorcycles in truck-involved crashes.
Topic 6: Describes crash scenarios between a bus and truck involved in a head-on collision.
Topic 7: Describes crash scenarios between an auto-rickshaw and truck and the injured passengers who were taken to the hospital.

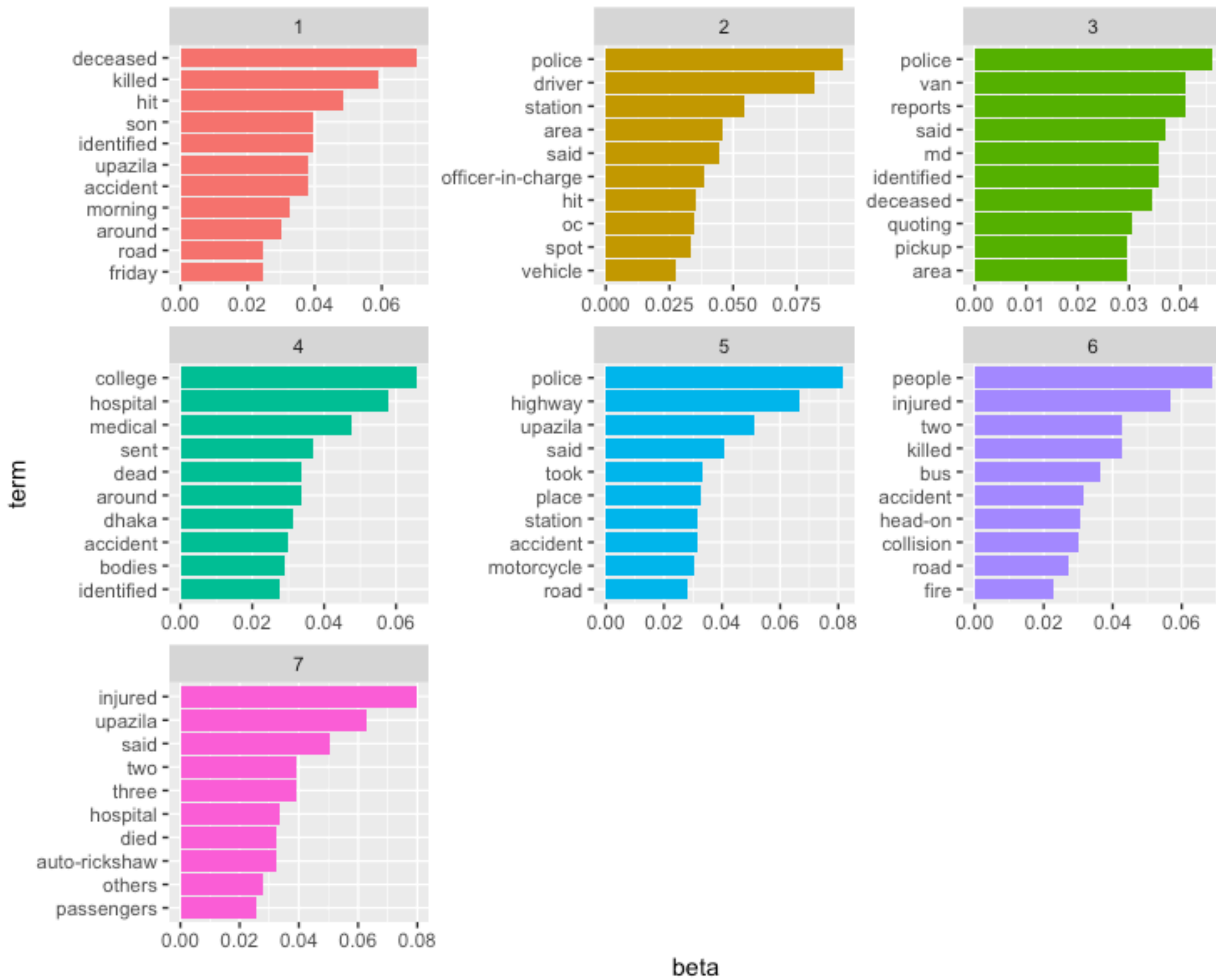

Figure 10. Top seven (k=7) topic models

CONCLUSIONS

The study analyzed fatal truck-involved crashes in Bangladesh using crash reports collected from several online news portals. The database consists of a total of 144 fatal truck-involved crash reports (bag of 15,300 words) collected between January 2021 and December 2021. Several text mining tools were utilized for this purpose including word cloud analysis, word frequency analysis, word co-occurrence network analysis, rapid automatic keyword extraction, and topic modeling. Some of the major findings from this research are as follows:

- Exploratory data analysis suggests that most of the fatal truck-involved crashes occurred in the Khulna division and Jhenaidah district. Temporal patterns suggest that a higher number of crashes occur at the beginning and end of a year.
- Word cloud analysis identified some of the most frequent words in the reports including ‘police’, ‘said’, ‘upazila’, ‘injured’ ‘accident’, ‘highway’, ‘deceased’, ‘killed’, ‘station’, ‘hospital’ and so on. The general findings of this word cloud analysis are – truck accidents causing injuries and fatalities, police involvement in the crash scene for investigation, and transportation of crash victims to the medical college hospital.

- The word frequency analysis provides some useful information, such as the kind of vehicle involved (auto-rickshaw, bus, van, and motorcycle), manner of collision (head-on, speeding), and time of the day (morning).
- 'Coming from opposite direction' and 'head-on collision' are two important sequences of events in truck-involved crashes. 'Overtaking' is an important crash contributing factor in truck-bus crashes.
- Several other crash scenarios were identified including 'a train and truck at the railway crossing' and 'students in bike and truck'. It indicates that vulnerable roadway user risk and crossing-related issues are needed to be addressed with appropriate guidance and effective countermeasures, respectively.
- The 'dense fog' keyword was identified, suggesting that poor environmental conditions play a role in truck-involved crashes.
- Wrong-way driving (WWD) was identified as a factor contributing to truck-involved accidents. To tackle this, 'Wrong Way' signs can be a useful tool to alert truck drivers.
- Some of the crash locations/routes were also identified in the network plot, including Cox's Bazar ($\phi = 0.53$), Bogura-Sherpur ($\phi = 0.55$), Pabna-Sirajganj ($\phi = 0.43$), and Khulna-Satkhira (0.44). The transportation agency can focus on these specific routes.
- Topic modeling suggests several important crash scenarios – 'a truck-involved crash occurring on Friday morning resulting in fatalities', 'crash scene involving a van with a pickup truck', 'involvement of motorcycles in truck-involved crashes', 'a crash scenario between bus and truck involving in a head-on collision', 'a crash scenario between an auto-rickshaw and truck and the injured passengers were taken to the hospital.' These crash scenarios can be utilized as an 'exposure pattern' in educational campaigns.

In summary, most news reports briefly mention the contributing factors or patterns leading to fatal truck-involved crashes in Bangladesh. Along with the findings in this study, the identification of in-depth crash contributing factors can assist in appropriate countermeasure selection and policy/regulation development. According to Bangladesh Road Transport Authority (BRTA) statistics, there are currently 292,000 drivers with heavy and medium driving licenses in the country. But the total number of registered heavy and medium vehicles is now 423,000 in the country. The huge discrepancy between the number of registered vehicles and the number of driving licenses potentially suggests the root cause of the truck-involved accidents. Drivers without a legal license are more likely to drive recklessly and pay less attention to traffic laws and regulations. To address road accidents, the Bangladeshi government must first concentrate on this license issue.

Research Contribution

This research has two major contributions. In absence of a reliable and current crash database in Bangladesh, the idea of using Google news alert and text mining tools would help other researchers to investigate crash patterns. Also, the identified crash patterns and crash contributing factors for truck-involved crashes would help policymakers to identify crash countermeasures.

Limitation

The research team utilized crash reports collected for a 12-month period. Additional periods of data collection would yield more intriguing crash patterns. Additionally, the database is only available for English-language-based online websites. The inclusion of the Bangla news portal is recommended for future studies.

Acknowledgment

The research team would like to acknowledge the assistance of internet news portals from which the database was gathered.

Author Contribution

The authors confirm their contribution to the paper as follows: study conception and design: A. Hossain, data collection: S. Alam; analysis and interpretation of results: A. Hossain, S. Das; draft manuscript preparation: A. Hossain, X. Sun, S. Alam, S. Das. All authors reviewed the results and approved the final version of the manuscript.

Conflict of Interest

The authors state that they have no known competing financial interests or personal relationships that could have influenced the findings of this study.

REFERENCES

1. Global Status Report on Road Safety 2018. https://www.who.int/publications-detail-redirect/9789241565684. Accessed Jul. 22, 2022.
2. Adhikary, T. S. Poor Data Frustrates Road Safety Measures. *The Daily Star*. https://www.thedailystar.net/news/bangladesh/news/poor-data-frustrates-road-safety-measures-2945701. Accessed Jul. 25, 2022.
3. Ahsan, H. Road Safety in Bangladesh: Key Issues and Countermeasures. In *Forum, monthly publication of Daily Star*, No. 6, 2012.
4. bdnews24.com. 43 School-Goers Die in Truck Plunge. *bdnews24.com*. https://bdnews24.com/bangladesh/43-school-goers-die-in-truck-plunge. Accessed Jul. 22, 2022.
5. 15 Killed in Road Accident in Bangladesh. *The Times of India*, Mar 22, 2020.
6. 10 Die in Bangladesh Accident. https://gulfnews.com/world/asia/10-die-in-bangladesh-accident-1.2000436. Accessed Jul. 22, 2022.
7. Report, S. D. Bus-Truck Collision Leaves 6 Dead, 13 Injured in Bogura. *The Daily Star*. https://www.thedailystar.net/country/news/bus-truck-collision-leaves-6-dead-13-injured-bogura-2048637. Accessed Jul. 22, 2022.
8. Report, S. D. 3 Killed in Chapainawabganj Road Crash. *The Daily Star*. https://www.thedailystar.net/country/news/3-killed-chapainawabganj-road-crash-2093013. Accessed Jul. 22, 2022.
9. Report, S. O. 3 Killed as Truck Hits Auto-Rickshaw in Rangpur. *The Daily Star*. https://www.thedailystar.net/country/bangladesh-road-accident-3-killed-in-rangpur-1896907. Accessed Jul. 22, 2022.
10. Report, S. O. 5 Day Labourers Die as Truck Overturns in Tangail. *The Daily Star*. https://www.thedailystar.net/truck-overturns-in-tangail-5-die-1886908. Accessed Jul. 22, 2022.
11. Correspondent, O. and Savar. 4 Killed as Truck Falls into Ditch. *The Daily Star*. https://www.thedailystar.net/country/2-die-2-other-go-missing-as-truck-falls-ditch-in-dhaka-ashulia-1694434. Accessed Jul. 22, 2022.
12. Correspondent, S. and Ctg. 5 Die as Truck Ploughs Thru' Auto Stand. *The Daily Star*. https://www.thedailystar.net/country/news//bangladesh-road-accident-in-chattogram-2-killed-truck-ploughs-1638409. Accessed Jul. 22, 2022.
13. Desk, P. A. E. 2 Killed, 5 Injured in Natore Road Accident. *Prothomalo*. https://en.prothomalo.com/bangladesh/accident/2-killed-5-injured-in-natore-road-accident. Accessed Jul. 22, 2022.
14. Das, S. Understanding Fatal Crash Reporting Patterns in Bangladeshi Online Media Using Text Mining. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2675, No. 10, 2021, p. pp 960-971. https://doi.org/10.1177/03611981211014200.
15. Dong, C., D. B. Clarke, S. H. Richards, and B. Huang. Differences in Passenger Car and Large Truck Involved Crash Frequencies at Urban Signalized Intersections: An Exploratory Analysis. *Accident Analysis & Prevention*, Vol. 62, 2014, pp. 87–94. https://doi.org/10.1016/j.aap.2013.09.011.
16. Cheung, I., and E. R. Braver. Undercounting of Large Trucks in Federal and State Crash Databases: Extent of Problem and How to Improve Accuracy of Truck Classifications.

Traffic Injury Prevention, Vol. 17, No. 2, 2016, pp. 202–208. https://doi.org/10.1080/15389588.2015.1034273.

17. Hickman, J. S., R. J. Hanowski, and J. Bocanegra. A Synthetic Approach to Compare the Large Truck Crash Causation Study and Naturalistic Driving Data. *Accident Analysis & Prevention*, Vol. 112, 2018, pp. 11–14. https://doi.org/10.1016/j.aap.2017.12.006.
18. Zheng, Z., P. Lu, and B. Lantz. Commercial Truck Crash Injury Severity Analysis Using Gradient Boosting Data Mining Model. *Journal of Safety Research*, Vol. 65, 2018, pp. 115–124. https://doi.org/10.1016/j.jsr.2018.03.002.
19. Zhao, Q., T. Goodman, M. Azimi, and Y. Qi. Roadway-Related Truck Crash Risk Analysis: Case Studies in Texas. *Transportation Research Record*, Vol. 2672, No. 34, 2018, pp. 20–28. https://doi.org/10.1177/0361198118794055.
20. Rahimi, A., G. Azimi, H. Asgari, and X. Jin. Clustering Approach toward Large Truck Crash Analysis. *Transportation Research Record*, Vol. 2673, No. 8, 2019, pp. 73–85. https://doi.org/10.1177/0361198119839347.
21. Das, S., M. Islam, A. Dutta, and T. H. Shimu. Uncovering Deep Structure of Determinants in Large Truck Fatal Crashes. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2674, No. 9, 2020, p. pp 742-754. https://doi.org/10.1177/0361198120931507.
22. Sufian, A., I. Ahmed, and S. Khan. *A Study on the Factors Involved in Truck Accidents in Bangladesh*. 2014.
23. Hasanat-E-Rabbi, S., I. Ahmed, and M. S. Hoque. Heavy Truck Rollover Model for Single Vehicle Run–off–Road Crashes in Bangladesh. *Jurnal Teknologi*, Vol. 70, No. 4, 2014, pp. 21–26. https://doi.org/10.11113/jt.v70.3484.
24. Siddique, M. T. Accident Severity Analysis on National Highways in Bangladesh Using Ordered Probit Model. *Scientific Research and Essays*, Vol. 13, No. 14, 2018, pp. 148–157.
25. Quazi, S. H., S. K. Adhikary, W. I. Wan Ibrahim, and R. B. Rezaur. Road Traffic Accident Situation in Khulna City, Bangladesh. *Proceedings of the Eastern Asia Society for Transportation Studies, Vol 5*, Vol. 5, 2005, pp. 65–74.
26. Islam, M., R. Bin Ali, and F. K. Chowdhury. Road Accident Analysis and Prevention Measures of Rajshahi - Sirajganj Highway in Bangladesh. Vol. 126, No. April, 2019, pp. 209–221.
27. Ahsan, H., S. Mahmud, and M. Bhuiyan. Heavy Vehicle Aggressivity in Bangladesh: Case Study on Large Truck. *Buet.Ac.Bd*, No. March, 2012, pp. 8–10.
28. Hoque, M., and S. M. S. Mahmud. Road Safety Engineering Challenges in Bangladesh.
29. Das, S. News Media Mining to Explore Speed-Crash-Traffic Association during COVID-19. *Transportation Research Record*, 2022.
30. Yang, H., Q. Ma, Z. Wang, Q. Cai, K. Xie, and D. Yang. Safety of Micro-Mobility: Analysis of E-Scooter Crashes by Mining News Reports. *Accident Analysis & Prevention*, Vol. 143, 2020, p. 105608. https://doi.org/10.1016/j.aap.2020.105608.
31. Karpinski, E., E. Bayles, L. Daigle, and D. Mantine. Characteristics of Early Shared E-Scooter Fatalities in the United States 2018–2020. *Safety Science*, Vol. 153, 2022, p. 105811. https://doi.org/10.1016/j.ssci.2022.105811.
32. Keliikoa, L. B., M. D. Thompson, C. J. Johnson, S. L. Cacal, C. M. Pirkle, and T. L. Sentell. Public Health Framing in Local Media Coverage of Crashes Involving Pedestrians or Bicyclists in Hawaiʻi, 2019: A Content Analysis. *Transportation Research*

Interdisciplinary Perspectives, Vol. 13, 2022, p. 100525. https://doi.org/10.1016/j.trip.2021.100525.
33. Hearst, M. A. Untangling Text Data Mining. Presented at the Proceedings of the 37th Annual meeting of the Association for Computational Linguistics, 1999.
34. Cai, Y., and J.-T. Sun. Text Mining. In *Encyclopedia of Database Systems* (L. Liu and M. T. Özsu, eds.), Springer US, Boston, MA, pp. 3061–3065.
35. Arteaga, C., A. Paz, and J. Park. Injury Severity on Traffic Crashes: A Text Mining with an Interpretable Machine-Learning Approach. *Safety Science*, Vol. 132, 2020, p. 104988.
36. Brown, D. E. Text Mining the Contributors to Rail Accidents. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 17, No. 2, 2015, pp. 346–355.
37. Nayak, R., N. Piyatrapoomi, and J. Weligamage. Application of Text Mining in Analysing Road Crashes for Road Asset Management. In *Engineering Asset Lifecycle Management*, Springer, pp. 49–58.
38. Soleimani, S., A. Mohammadi, J. Chen, and M. Leitner. Mining the Highway-Rail Grade Crossing Crash Data: A Text Mining Approach. Presented at the 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA), 2019.
39. Giummarra, M. J., B. Beck, and B. J. Gabbe. Classification of Road Traffic Injury Collision Characteristics Using Text Mining Analysis: Implications for Road Injury Prevention. *PloS one*, Vol. 16, No. 1, 2021, p. e0245636.
40. Ghomi, H., and M. Hussein. An Integrated Text Mining, Literature Review, and Meta-Analysis Approach to Investigate Pedestrian Violation Behaviours. *Accident Analysis & Prevention*, Vol. 173, 2022, p. 106712.
41. Sayed, M. A., X. Qin, R. J. Kate, D. Anisuzzaman, and Z. Yu. Identification and Analysis of Misclassified Work-Zone Crashes Using Text Mining Techniques. *Accident Analysis & Prevention*, Vol. 159, 2021, p. 106211.
42. Alghamdi, R., and K. Alfalqi. A Survey of Topic Modeling in Text Mining. *Int. J. Adv. Comput. Sci. Appl.(IJACSA)*, Vol. 6, No. 1, 2015.
43. Blei, D. M., A. Y. Ng, and M. I. Jordan. Latent Dirichlet Allocation. *Journal of machine Learning research*, Vol. 3, No. Jan, 2003, pp. 993–1022.
44. Mahmud, F. Eid Holidays in Bangladesh Saw Record Road Accident Deaths: Group. https://www.aljazeera.com/news/2022/7/24/eid-holidays-in-bangladesh-saw-record-road-accident-deaths-group. Accessed Oct. 29, 2022.