

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

Ahmed Hossain¹ , Xiaoduan Sun¹ , Shah Alam², Subasish Das³ , and Abbas Sheykfard⁴ 

Transportation Research Record

1–20

© National Academy of Sciences:

Transportation Research Board 2023

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0361198123120903

journals.sagepub.com/home/trr



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. Because of 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 12 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), the manner of collision (head-on), the 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 rail crossings. 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

The World Health Organization (WHO) estimates that road traffic crashes are the eighth leading cause of death, with more than 1.35 million fatalities globally each year (1). It is important to note, however, that crash prevalence and severity differ significantly between countries, with developing and underdeveloped countries reporting higher deaths than developed countries. Despite developing countries having fewer registered vehicles, they have a higher casualty rate because of factors such as inadequate road infrastructure, lowered safety standards, and riskier driving practices. According to WHO, 93% of all road fatalities occur in low- and middle-income countries (LMICs), despite these countries owning around 60% of the world's vehicles (2). Even so, developing countries must adopt

stricter traffic laws, promote safe driving habits, and enhance their road infrastructure to improve their road safety. In LMICs such as Bangladesh, fatalities and serious injuries because of traffic crashes are major health concerns. According to the 2018 WHO Global

¹Department of Civil Engineering, University of Louisiana at Lafayette, Lafayette, LA

²Department of Civil Engineering, Rajshahi Science & Technology University, Natore, Bangladesh

³College of Science of Engineering, Texas State University, San Marcos, TX

⁴Department of Civil Engineering, Babol Noshirvani University of Technology, Babol, Mazandaran, Iran

Corresponding Author:

Ahmed Hossain, ahmed.hossain1@louisiana.edu

Status Report (3), the WHO estimate of the number of 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 occurred in the economically productive age groups (15–64 years old), there were 15.3 fatalities occurred per 100K population, there were estimated serious injuries of 374,310, and the cost of fatalities and series injuries was US\$11,630 (5.3% of the country's gross domestic product [GDP]). Because of the underdeveloped crash reporting system in Bangladesh (4), it is challenging to collect comprehensive data on these collisions and their underlying causes.

Truck-involved crashes are one of the most significant types of crashes that need to be examined in depth, especially in developing countries. The vehicle mix and travel patterns are considerably different from those in developed countries, as most people rely on heavy vehicles (buses and trucks) for travel and goods movement (5). As a result of the greater exposure, crash involvement is higher for trucks and buses. For instance, a prior study found that the number of fatalities and injuries per 10,000 crashes was higher in Vietnam and Kenya than in the U.S.A., primarily as a result of the high frequency of collisions involving multi-passenger vehicles, including buses, minibuses, and trucks (6). There are several factors that can be attributed to truck-involved crashes (7–9). These factors are important to understand as these crashes pose a considerable risk to public safety, and analyzing the factors that contribute to them can help decrease the number of deaths and injuries. In addition, truck crashes usually cause property damage, medical expenses, and lost productivity, which can have a substantial impact on the economy. This cost can be reduced if policymakers and stakeholders in the industry understand the factors that lead to these crashes and implement follow-up actions (i.e., countermeasures) as well as measurements of those actions to achieve safety improvement. Furthermore, a deeper understanding of the factors behind fatal truck crashes can aid in policy development, including the development of regulations to enhance truck safety. Understanding the causes and patterns of crashes in this industry can support the development of best practices, such as driver training, vehicle maintenance, and other safety measures. Bangladesh has experienced a sharp rise in crashes involving heavy vehicles, particularly trucks, in recent years (10). Truck-involved crashes made up around 25% of all crashes in Bangladesh in 2022, which is second only to motorcycle crashes (29%) according to the Passenger Welfare Association of Bangladesh (11). Because of 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. For example, the fatal truck-involved crash that occurred on July 11, 2011, in Mirsarai, Chattogram, resulted in 44 fatalities (second row in the table), which is an extremely high number for a single road traffic crash.

Because of 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 (21). Das collected fatal crash reports (for all vehicles) 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 truck-involved crashes to address a specific safety concern. State-of-the-art text mining tools were used to identify insights and factors from reported text contents about fatal track-involved crashes. The findings of this study are expected to provide a better understanding of the crash contributing factors of truck-involved crashes in 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 other developing countries; and (3) studies on news media-related information extraction for transportation safety.

Major U.S. Studies

Truck-involved crashes are a key safety concern in the U.S.A. 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. (22) estimated large truck-related crash frequencies by developing a multivariate Poisson-lognormal (MVPLN) model. Comparisons among car, car–truck, and truck crash counts concerning different contributing factors were identified in the developed models. Cheung and Braver (23) explored the undercounting of large truck crash frequencies in federal and state crash

Table I. Some Major Truck-Involved Crashes in Bangladesh

Source	Location	Date	Crash scenario	Killed	Injured
bdnews24.com (12)	Mirsarai, Chattogram	July 11, 2011	A pick-up truck veered off the road and toppled into a roadside pond.	44	0
The Times of India (13)	Chunti area, Chittagong district	March 22, 2020	A truck rammed the passenger vehicle.	15	7
Gulf News (14)	Chuadanga, Khulna district	March 26, 2017	A collision between a truck and a mini truck.	10	12
Daily Star Digital Report (15)	Sherpur, Bogura district	February 21, 2021	A bus crashed into a stone-laden truck.	6	15
Daily Star Digital Report (16)	Chapainawabganj	May 15, 2021	A truck hit a human hauler from the opposite direction.	3	6
Daily Star Digital Report (17)	Rangpur	April 26, 2020	A speeding truck hit an auto-rickshaw from behind after the truck driver lost control of the steering.	3	5
Daily Star Digital Report (18)	Tangail, Dhaka	March 28, 2020	A cement-laden truck overturned.	5	11
Daily Star Digital Report (19)	Ashulia, Dhaka	January 30, 2019	A brick-laden truck fell into a 40-foot-deep ditch.	4	3
Prothom Alo English Desk (20)	Natore	January 12, 2022	A collision between a bus, a truck, and a rickshaw van.	2	5

Note: CNG = compressed natural gas.

Source: Online news.

databases. The results show that large truck involvement-related misclassification is significantly high. The Large Truck Crash Causation Study (LTCCS) provides in-depth information on crash data, which is usually unavailable in state-maintained traffic crash data. Using the LTCCS and naturalistic driving data, Hickman et al. (24) conducted a synthetic risk ratio analysis. Zheng et al. (25) applied gradient boosting and data mining methods to identify the impact of key contributing factors on crash severities associated with commercial trucks. This study examined 25 variables, and 22 of them were identified as significant variables contributing to injury severities. In Texas, truck-related crash frequencies increased by 82% from 2009 to 2012. Zhao et al. (26) identified hot spots and developed collision diagrams by collecting traffic crash data from Texas from 2011 to 2015. Rahimi et al. (27) applied a high-dimensional clustering approach to identify patterns of large truck crashes by using crash data from Florida between 2007 and 2016. 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. (28) applied an innovative dimension reduction method known as taxicab correspondence analysis (TCA) on fatal crash data involving a large truck. 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. Note that the current findings and recommendations in the context of the U.S.A. may not be transferable to developing countries such as Bangladesh because of the unique vehicle mix (slow- and fast-moving vehicles), traffic violation culture, level of education and awareness, failure to enforce traffic laws, and so on (29).

Major Truck-Involved Crash Analysis in Bangladesh and Other Developing Countries

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, two studies focused solely on truck-related crashes in Bangladesh. A study conducted by Sufian et al. (30) collected data on road traffic crashes in Bangladesh from 1998 to 2010. The EDA shows that truck drivers' activities contributed to 95% of truck crashes. Other factors were associated with pedestrians, vehicular properties, and the 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 (31). 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 highlights information (including the year, data source, location, methods, and key findings of 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 it is difficult to assess how basic human, roadway, or environmental factors are associated with truck-involved crashes.

Truck-involved crashes have also received research attention in other developing countries. For example, a study conducted in Iran analyzed 4359 single-vehicle truck crashes (March 2011–March 2012) and identified several significant factors contributing to crashes, including drivers’ education status, advanced braking system deployment, curved alignment, and high posted speed limit (37). Another study conducted in Iran evaluated the factors influencing the severity of heavy truck-involved incidents and revealed unsafe lane changing as a major contributor to crashes (7). The study also emphasized unrealistic schedules and expectations of truck companies as a source of stress for large truck drivers. Another study utilized police-reported pedestrian crashes in Addis Ababa, Ethiopia, from 2009 to 2012 (38). According to these findings, if a truck is involved in the crash, the likelihood of fatal and severe pedestrian injuries increases by around 11% and 2.1%, respectively. A previous study suggested that overloaded truck transport is an inevitable outcome of fast economic growth in developing countries and can make up as high as 80% of the total number of trucks on the highway (39). Again, fatal crashes involving heavy trucks are correlated to the amount of overloading, as suggested by a few of the previous studies (40–42).

Traffic Safety Analysis Using News Media Mining

News media mining has been becoming increasingly popular among researchers when conventional data is limited. Das (21) applied different NLP tools to news media data to extract insights into fatal crashes in Bangladesh. In another study, Das (43) collected news media reports on the impact of speeding on crashes during COVID-19. 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. (44) explored massive media reports to develop an e-scooter crash database. This study identified 169 e-scooter-related traffic crashes from news reports during 2017–2019. This study also conducted an EDA on the developed crash datasets. Karpinski et al. (45) identified 21 shared e-scooter fatalities in the U.S.A. from 2018 to 2020 by exploring media reports. 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. (46) explored local media news coverage of non-motorist-related crashes in Hawaii in 2019. The content analysis shows that language patterns in news article titles were usually non-agentive (77%) and focused on pedestrians or bicyclists (77%) without mentioning drivers or vehicles (69%).

Research Gap, Objectives, and Novelty of Study

Traditional data collection methods such as crash reports by police may be insufficient in providing a comprehensive understanding of the factors contributing to road crashes. Digital data sources such as news articles, social media posts, and other textual data are becoming increasingly available, making text mining an effective tool for analyzing road crash factors and patterns, as well as extracting insights that can be applied to improving road safety. To the best of our knowledge, no prior study has explored contributing factors and their patterns for truck-related crashes in developing countries via the text mining approach. An in-depth analysis is thus needed to mitigate the current research gap. Thus, the goal of this study is to apply text mining tools to identify the factors contributing to truck crashes in Bangladesh, along with the patterns associated with them. Analyzing fatal truck-involved crashes using a text mining approach can be highly effective for numerous reasons. Firstly, it enables the analysis of large amounts of unstructured data, such as police reports and crash records, to identify the key factors and patterns associated with these types of crashes. As a second benefit, text mining can uncover insights and patterns that were not immediately visible in the data, revealing new factors or combinations of factors that contribute to truck crashes. Moreover, real-time monitoring of news articles, social media, and other sources can help identify trends or factors contributing to fatal truck crashes. Lastly, text mining can provide policymakers, industry stakeholders, and safety professionals with insights into contributing factors and patterns of truck crashes. Last but not least, prioritizing interventions and solutions can be accomplished by identifying the most common and significant

Table 2. Findings from Previous Studies

Ref.	Author, year	Title	Data source	Location	Methods	Key findings on the truck-involved crash
(32)	Siddique, 2018	Accident severity analysis on national highways in Bangladesh using ordered probit model	Bangladesh police report from 2004–2015	Bangladesh	Ordered probit model	<ul style="list-style-type: none">• There is a greater likelihood of a car-truck rear-end crash when the brakes are applied suddenly to a moving vehicle.• Serious injuries occur in collisions between a heavy truck and a motorcycle.
(33)	Quazi et al., 2005	Road traffic accident situation in Khulna City, Bangladesh	First Information Report (FIR), 2001–2002	Khulna	Descriptive analysis	<ul style="list-style-type: none">• Trucks were involved in 26% of the crashes.• Trucks were involved in around 47% of the crashes resulting in pedestrian fatalities.
(34)	Islam et al., 2019	Road accident analysis and prevention measures of Rajshahi-Sirajganj Highway in Bangladesh	Highway police station, 2008–2015	Rajshahi-Sirajganj	Descriptive analysis	<ul style="list-style-type: none">• Involvement of trucks was identified in about 39% of the crashes.
(31)	Hasanat E-Rabbi et al., 2014	Heavy truck rollover model for single-vehicle run-off-road crashes in Bangladesh	Accident Report Form (ARF)	Bangladesh	Quasi-Static Rollover Model	<ul style="list-style-type: none">• Heavy trucks are involved in around 21% of overturning crashes.• Excessive speeding and reckless driving were identified as the prime causes of rollover-type ROR crashes.
(35)	Ahsan et al., 2012	Heavy vehicle aggressivity in Bangladesh: Case study on large truck	Police report	Bangladesh	Descriptive analysis	<ul style="list-style-type: none">• Trucks were involved in around 33% of fatal crashes.
(36)	Hoque and Mahmud, 2009	Road safety engineering challenges in Bangladesh	Police reports in Bangladesh, 1998–2005	Bangladesh	Descriptive analysis	<ul style="list-style-type: none">• The involvement of buses/trucks was identified as most of the pedestrian-vehicle conflicts.• Approximately 27% of road crashes involve trucks.

Note: ROR = run-off-road.

factors contributing to fatal truck crashes. Through targeted interventions and better decision-making, text mining can provide a more comprehensive understanding of the factors and patterns that contribute to these crashes.

In summary, understanding the contributing factors and patterns associated with fatal truck-involved crashes is essential for improving public safety, reducing economic costs, focusing on policy development, and promoting best practices in the industry. Therefore, 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.

Scope

The scope of this research is limited to online news articles published in the English language in Bangladesh. Other news articles published in the local languages, such as “Bangla,” are not included in this paper.

Methods

The technique of extracting data from a collection of texts is known as text mining or text data mining. In 1999, Hearst (47) first used the term “text data mining” and distinguished it from other ideas, like NLP. The objective of text mining is to discover information and patterns from text data, which can be unstructured or semi-structured (48). There are several steps involved in text mining, including data collection, data preprocessing, text analysis, and interpretation. Preprocessing removes noise, such as stop words and punctuation, and transforms the text into a structured format that can be analyzed. The data is collected from various sources and then preprocessed to remove noise. Afterward, patterns, relationships, and trends are identified by analyzing the text using a variety of techniques, such as NLP, machine learning, and statistical analysis. As a final step, the results are interpreted to gain meaningful insights and knowledge. 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 (49), rail crashes (50), evaluating roadway crashes for road asset management (51), mining highway–rail grade crossing crash data (52), classification of roadway traffic injury collision characteristics (53), pedestrian violation behaviors (54), work zone crashes (55), and so on.

The basic methodology of text mining is to transform the text into a numeric dataset. To facilitate this, the term document matrix (TDM) method is utilized. In the TDM, the text data is represented in the form of a matrix. Before creating the TDM, basic preprocessing of the text is required, including the removal of

Table 3. Example of a 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
crash	1	0	0
truck	0	—	—
driver	0	—	0
lost	0	—	0
control	0	1	0
hit	0	—	0
tree	0	—	0
two	0	0	—
men	0	0	—
riding	0	0	—
motorcycle	0	0	—
died	0	0	—
after	0	0	—
run	0	0	—
over	0	0	—

punctuation, stop words (common English words), white space, numeric numbers, and special characters. Also, all of the text is converted into lowercase to reduce variations of the same word. For example, “Accident” and “accident” are treated as the same words after transformation. It is worth noting that the majority of online news sites in Bangladesh use the term “accident” instead of “crash” when reporting, despite the two words having different meanings.

Term Document Matrix

The TDM represents the document vector in matrix format. In this matrix, 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 crash**.

D2: The **truck driver** **lost control** and **hit the tree**.

D3: **Two men** **riding a motorcycle** have **died** **after** being **run over** by a **truck**.

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

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 the term frequency (TF), document frequency (DF), and inverse document frequency (IDF). Let “ t ” indicate terms (words), “ d ” indicate documents (set of words), and “ N ” indicate the count of the corpus. Note that the corpus is the total document set.

Term Frequency

The TF measures how frequently a term occurs in a document. The equation for TF is provided below (56):

$$TF(t, d) = \frac{\text{Count of } t \text{ in } d}{\text{Total number of terms in } d} \quad (1)$$

Document Frequency

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 the documents.

Inverse Document Frequency

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 (56):

$$IDF(t, d) = \log \left[\frac{\text{Total number of documents in the corpus}}{\text{Total number of documents in the corpus that contain the term}} \right] \quad (2)$$

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

Consider a document containing 100 words where the word “truck” appears five 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, 1000 of them include the term “truck.” Then, the measure IDF can be calculated as the logarithm of $(10,000,000/1000) = 4$. 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 let their appearance in the document be considered in a binary format: “1” for presence, “0” for absence. For example, N_{11} represents the number of documents where both the word “A” and the 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 coefficient is provided below (57):

$$\phi = \frac{N_{11}N_{00} - N_{10}N_{01}}{\sqrt{N_1N_0N_0N_1}} \quad (3)$$

Rapid Automatic Keyword Extraction

Rapid automatic keyword extraction (RAKE) is a method for extracting keywords from individual documents. The RAKE algorithm is unsupervised and 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 the score), as 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:

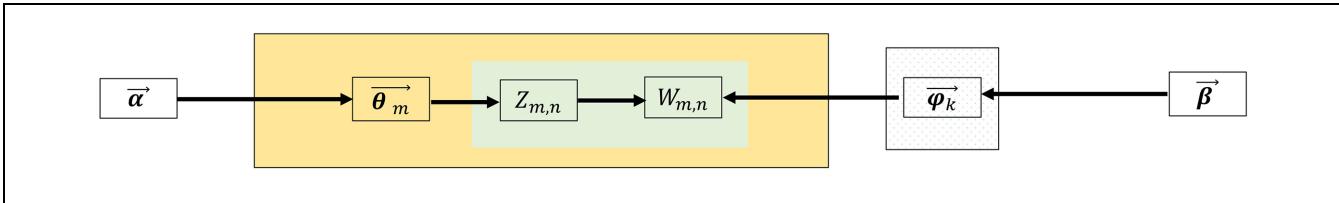
- truck driver;
- lost control;
- hit tree.

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 word can be calculated as the ratio of the degree of the word and the word frequency. Now, the score for each

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

**Figure 1.** Working flow diagram of the latent Dirichlet allocation model (59) (color online only).

candidate keyword is simply the combined sum of each score. For example, the score of the keyword “truck driver” is 4.

Topic Modeling

Topic modeling is an unsupervised machine learning technique that automatically analyzes text data to determine cluster words that frequently occur together within a set of documents (58). The research team utilized the latent Dirichlet allocation (LDA) algorithm, which is the most popular topic modeling technique to extract topics from a given corpus and performs better than other methods, such as non-negative matrix factorization (NMF), latent semantic analysis (LSA), and the Pachinko allocation model (PAM) (59–61). Previous research conducted by Blei et al. (62) provides a handy resource for the theoretical evolution of LDA. 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 Figure 1.

The interpretation of the LDA parameters is as follows: $\vec{\alpha}$ is the Dirichlet parameter, which controls per-document topic distribution, $\vec{\theta}_m$ is the document topic distribution, $Z_{m,n}$ is the word topic assignment, $W_{m,n}$ is the observed word, $\vec{\varphi}_k$ is the topic word distribution, and $\vec{\beta}$ is the 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.

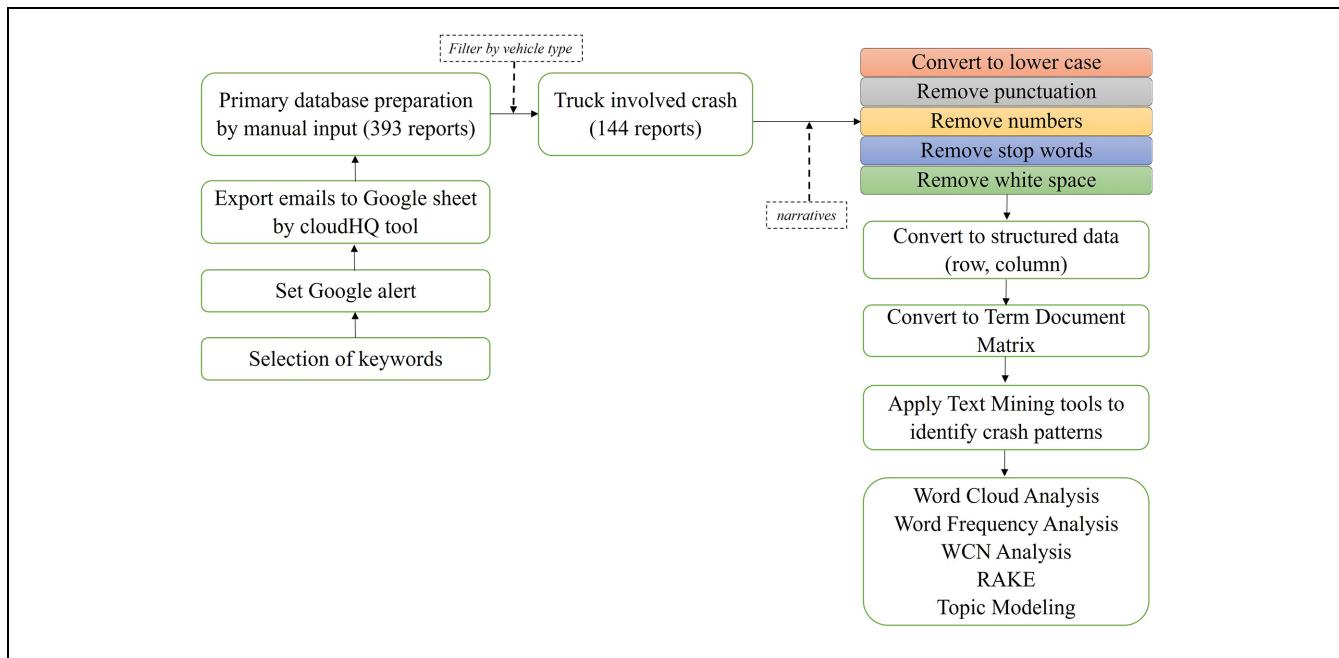


Figure 2. Database preparation and analysis flowchart.

Note: RAKE = rapid automatic keyword extraction; WCN = word co-occurrence network.

The research team utilized R statistical software (version 4.2.0) to conduct the text mining analysis. A wide range of open-source R software packages was utilized, including "wordcloud2," "topic models," "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 used the following keywords in the Google Alerts 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. Note that the study considered only online news articles written in

English, and did not include articles published in any other languages (e.g., Bangla). Each 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 12 variables, namely 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. Figure 2 shows the database preparation and analysis flowchart in detail.

To demonstrate the structure of crash narratives, one example of the crash narratives 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, who 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

Table 5 provides a summary of the crash database used in this study. A total of 144 online news reports related

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		
bdnews24.com	13	Median = 101.5		
Dhaka Tribune	19	Standard error = 3.3		
Jagonews24	1	Standard deviation = 39.2		
New Age	1	95% confidence interval = 6.5		
Risingbd	4	Minimum = 38		
The Asian Age	5	Maximum = 243		
The Daily Star	43	1st Quartile = 77.75		
The Financial Express	7	2nd Quartile = 101.5		
The Independent	14	3rd Quartile = 127.25		
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
		Frequency	Percentage	
Summary by age (years)				
<15		22	3.51	
15–30		93	14.83	
31–45		78	12.44	
>45		35	5.58	
Not reported		399	63.64	
Total		627	100	
Summary by gender				
Male		266	42.42	
Female		53	8.45	
Not reported		308	49.12	
Total		627	100	

to truck-involved crashes 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% killed and 48.5% injured. Note that the age (63.64% unknown) and gender (49.12% unknown) information of most of the crash-involved individuals were not reported in the online news. This is identified as a potential shortcoming of online crash news reporting in Bangladesh.

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).

To understand the temporal variation of truck-involved fatalities and injuries during 2021, a time-series graph is plotted (Figure 4). The term “ratio” in the graph indicates the proportion of the total killed and injured in all traffic crashes divided by the killed and injured in truck-involved crashes. The ratio of truck-involved crashes was found to be higher during February (63.3%), April (61.5%), and October (51.4%). For total crashes, peaks are observed in July and September. Since the Eid (the biggest Muslim celebration) holiday is celebrated in July, more crashes occur throughout this month. This is in line with a recent investigation by the Al Jazeera news

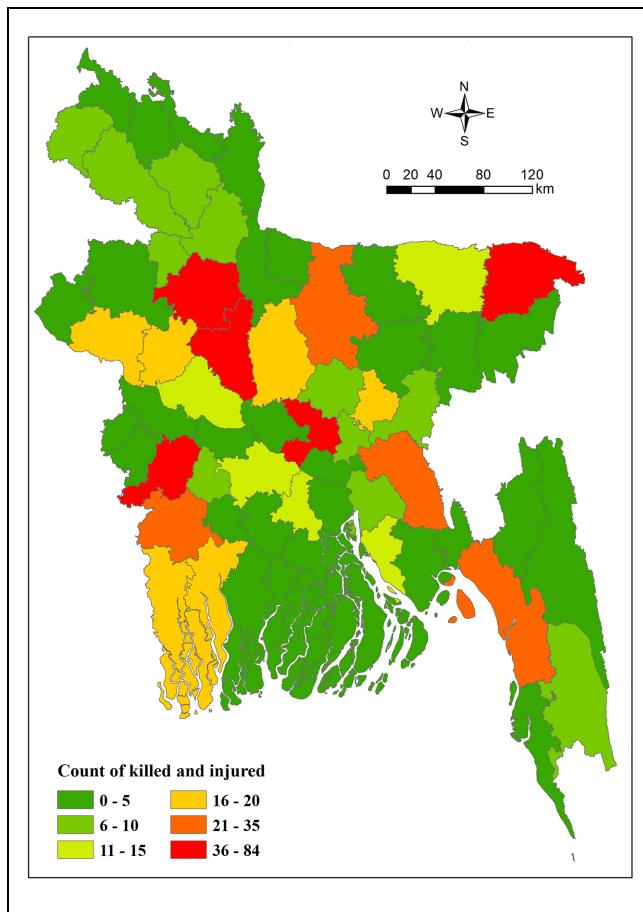


Figure 3. Spatial distribution of fatalities and injuries in truck-involved crashes in Bangladesh (2021).



Figure 4. Temporal variation of truck-involved and total crashes (reported in online news).

network (63). In addition, more crashes are reported at the beginning and end of the year, probably because of the foggy weather conditions during those times of the year. This is consistent with a recent study in Bangladesh,

which identified that adverse weather conditions significantly increase the likelihood of fatalities and severity of crashes (64). One possible explanation is that fog obscures drivers' vision of the road, making it difficult to assess the distance between vehicles ahead (65).

Results and Discussion

TF-IDF Results

In this analysis, a total of 144 crash reports (i.e., documents) were utilized, which consist of a total of 15,300 words (i.e., terms). After preprocessing, only 1950 words remained. Therefore, the final matrix used for the analysis consists of 144 rows and 1950 columns with a total of 280,800 elements or cells. Only 7234 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 in 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-involved crashes causing injuries and fatalities, police involvement in the crash scene for investigation, and transportation of crash victims to the medical college hospital.

Word Frequency Analysis

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

Note that a minimum 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 bar plot.

There are a total of 29 words in the above bar plot. The top five most frequent words in the above bar plot are police (232 times), injured (149 times), accident (129 times), killed (119 times), and deceased (113 times). The bar plot also contains some useful information about vehicle type, manner of collision, and time of the day.



Figure 5. Word cloud analysis.

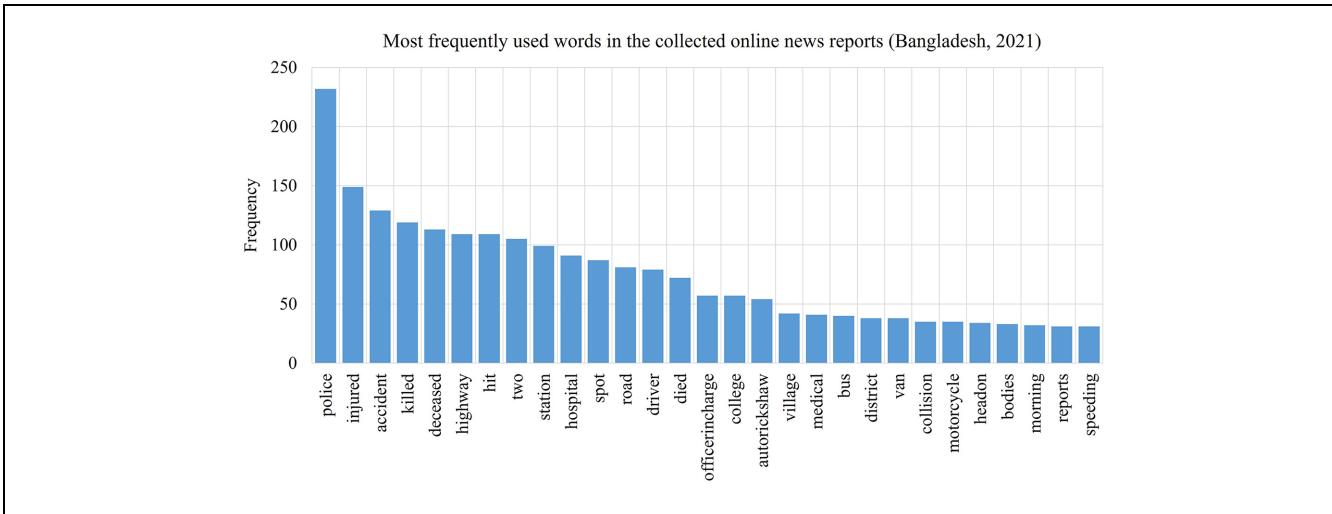


Figure 6. Word frequency bar plot (arranged in ascending order).

For example, four different vehicle types appeared in the above bar plot, namely 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 because of collisions with buses, motorcycles, auto-rickshaws, 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 crashes. The

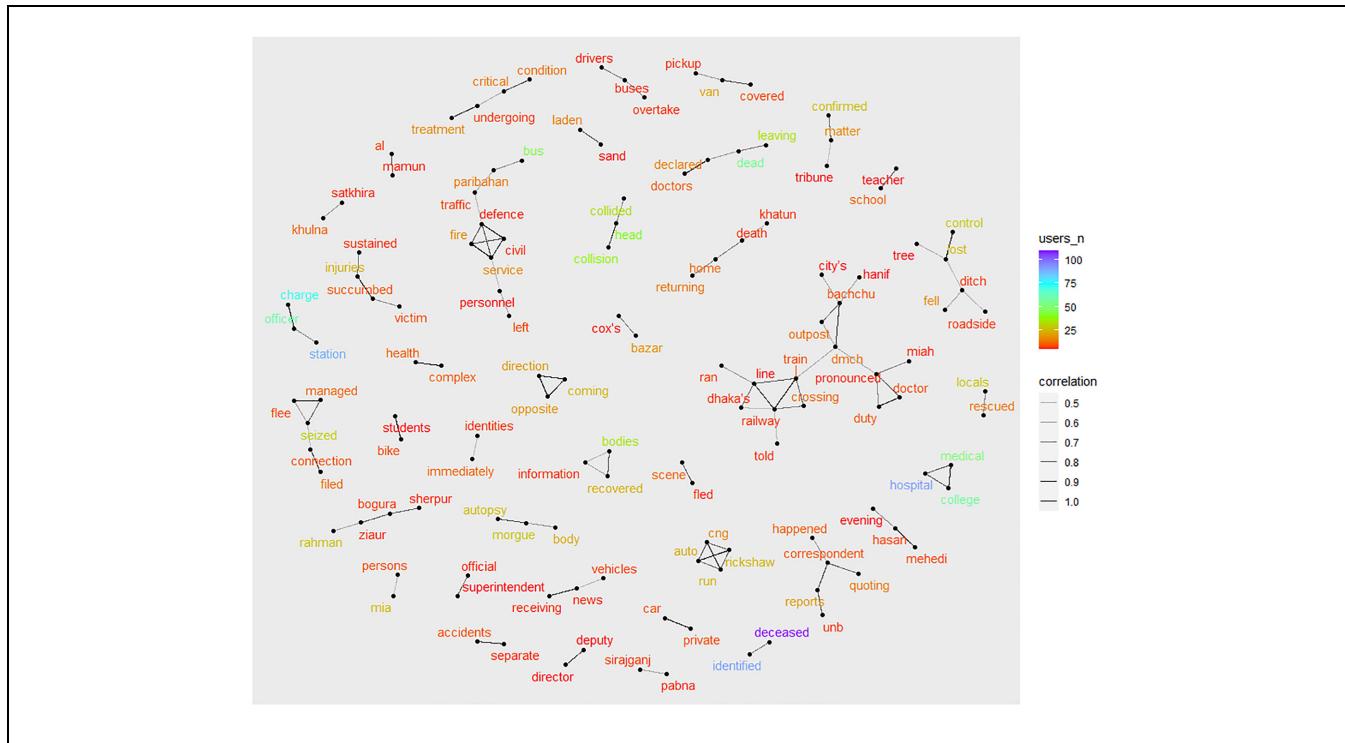


Figure 7. Word co-occurrence network (WCN) plot for truck-involved crashes.

word “speeding” appeared 31 times, suggesting that speeding is one of the crash contributing factors for 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 Analysis

The word co-occurrence network (WCN) is a key tool for visualizing the relationships among words that appear together in a sentence. Figure 7 shows the WCN plot for truck-involved crashes.

The 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 ” represents 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 $\phi = 0.40$ after several trial-and-error runs. 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.” This indicates that the

majority of truck-involved crashes in Bangladesh occur because of 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.” This indicates a crash scene involving a train and truck at the railway crossing. In Bangladesh, fatal crashes 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 ($\phi = 0.56$), possibly pointing to a hit-and-run crash 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.

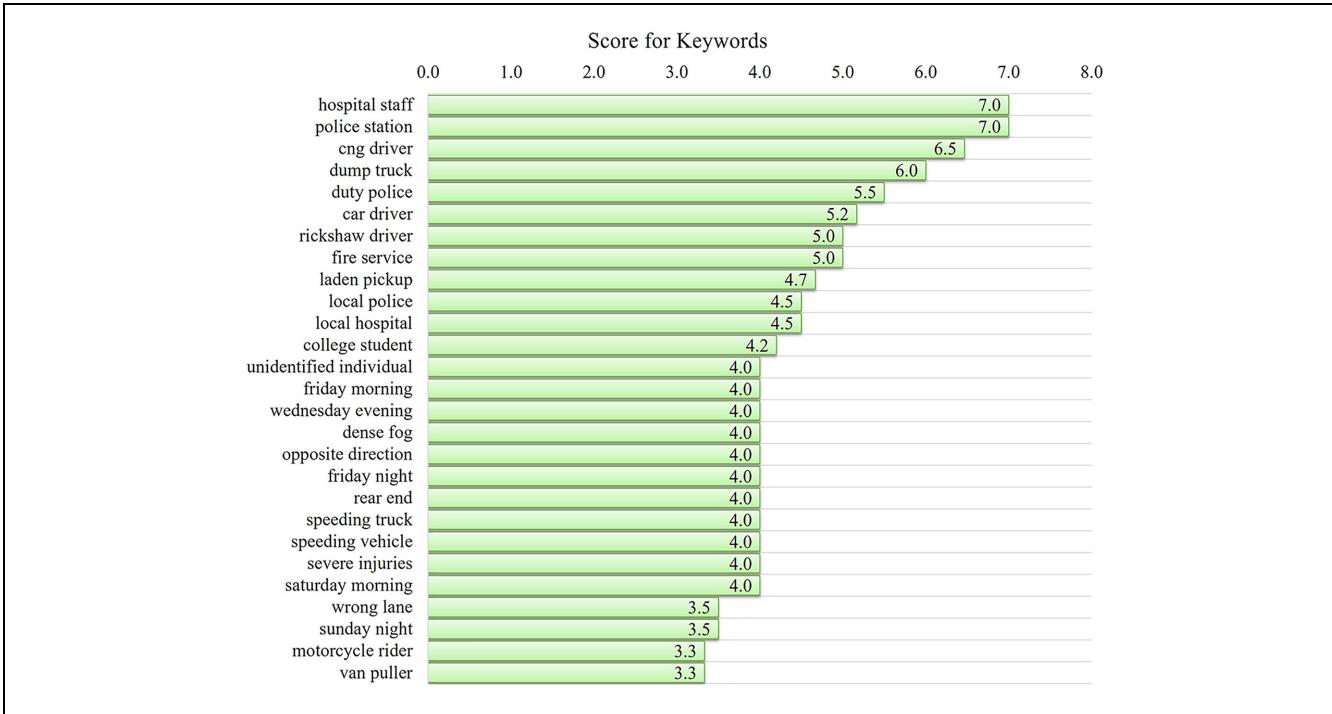


Figure 8. Rapid automatic keyword extraction for truck-involved crashes.

- The correlation coefficient between the words “bike” and “students” was found to be 0.73. This implies a collision involving students on motorcycles and trucks.
- The network plot also offers some important details with respect to the type of vehicles involved in truck-related crashes in Bangladesh. Some of these crash-involved vehicles are compressed natural gas (CNG) auto-rickshaws (a link is observed among “cng,” “auto,” “rickshaw,” and “run”), covered vans/pickups (a link is observed among “covered,” “van,” and “pickup”), sand-laden trucks (a link is observed between “sand” and “laden”), and private cars (a link is observed between “private” and “car”).
- 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).

Rapid Automatic Keyword Extraction

The RAKE algorithm identified a total of 4102 keywords. The distribution of these keywords is as follows: 1-gram (2500), 2-gram (1131), 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) provided the most useful information. The research team reviewed all of the 1131 keywords and

identified key factors that were most likely to have played a role or were associated with the truck crashes. Figure 8 shows keywords (selected 2 grams) extracted by the RAKE process.

Note that the notation “ s ” represents the “score” of the keywords. 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 drivers of different vehicle types were found to be associated with the truck-involved crashes, including cng drivers ($s = 6.5$), car drivers ($s = 5.2$), rickshaw drivers ($s = 5$), and van pullers ($s = 3.3$). In addition, the involvement of other vehicles (dump trucks and laden pickups) was 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 ($s = 4$).

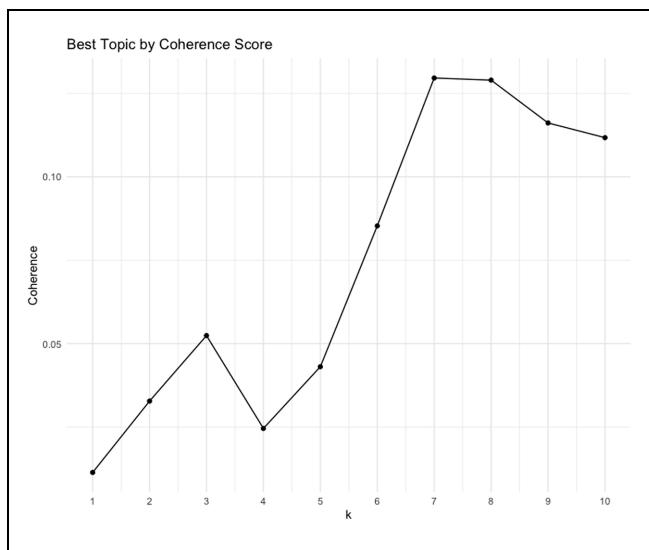


Figure 9. Selection of the optimum number of topics by coherence score.

Topic Modeling Results

The research team utilized the LDA 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).

The plot reveals that $k = 7$ provides the highest coherence score. The results obtained from topic modeling after setting $k = 7$ are provided in 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 a 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 a 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 autorickshaw and truck and the injured passengers who were taken to the hospital.

Discussion

The findings of the present study on truck-related crashes in Bangladesh provide valuable insights into the contributing factors and patterns of these crashes. In addition, studies conducted in the U.S.A. have indicated that a variety of factors contribute to truck-involved crashes, including driver fatigue, distracted driving, speeding, and inadequate training and supervision of drivers (66–69). The Federal Motor Carrier Safety Administration (FMCSA) found that driver fatigue contributed to approximately 13% of large truck crashes in the U.S.A. (70). It has also been reported that trucks with higher safety ratings are less likely to be involved in crashes (71). Despite that the specific contributing factors may differ between countries, the overall implications of these studies are similar. The authors emphasize the importance of addressing driver behavior, vehicle safety, and infrastructure concerns to reduce the number of truck crashes. As part of efforts to address these issues, regulations have been enacted in the U.S.A., including the Hours of Service (HOS) rule, which limits the number of hours a driver can work each day or week, and the Electronic Logging Device (ELD) mandate, which requires drivers to electronically document their hours of driving. There is some evidence in Bangladesh that addressing the issue of drivers’ licenses versus the number of registered vehicles may be the key to reducing the number of truck-related crashes.

The results of research on truck-involved crashes in developed and developing countries suggest that to develop effective strategies and policies to reduce the occurrence of these crashes, it is essential to understand the contributing factors and patterns of these crashes. The following is a summary of a few of the major findings of this study.

- EDA 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. This can be attributed to several factors, such as there being more social events and celebrations during the holidays, which result in driving under the influence, as well as adverse weather conditions during the winter, fatigue from vacations, and the rush to meet deadlines at the end of the year. The risk of road crashes increases when these factors are combined.
- The word cloud analysis identified some of the most frequent words in the reports, including “police,” “injured,” “accident,” “highway,” “deceased,” “killed,” “station,” “hospital,” and so

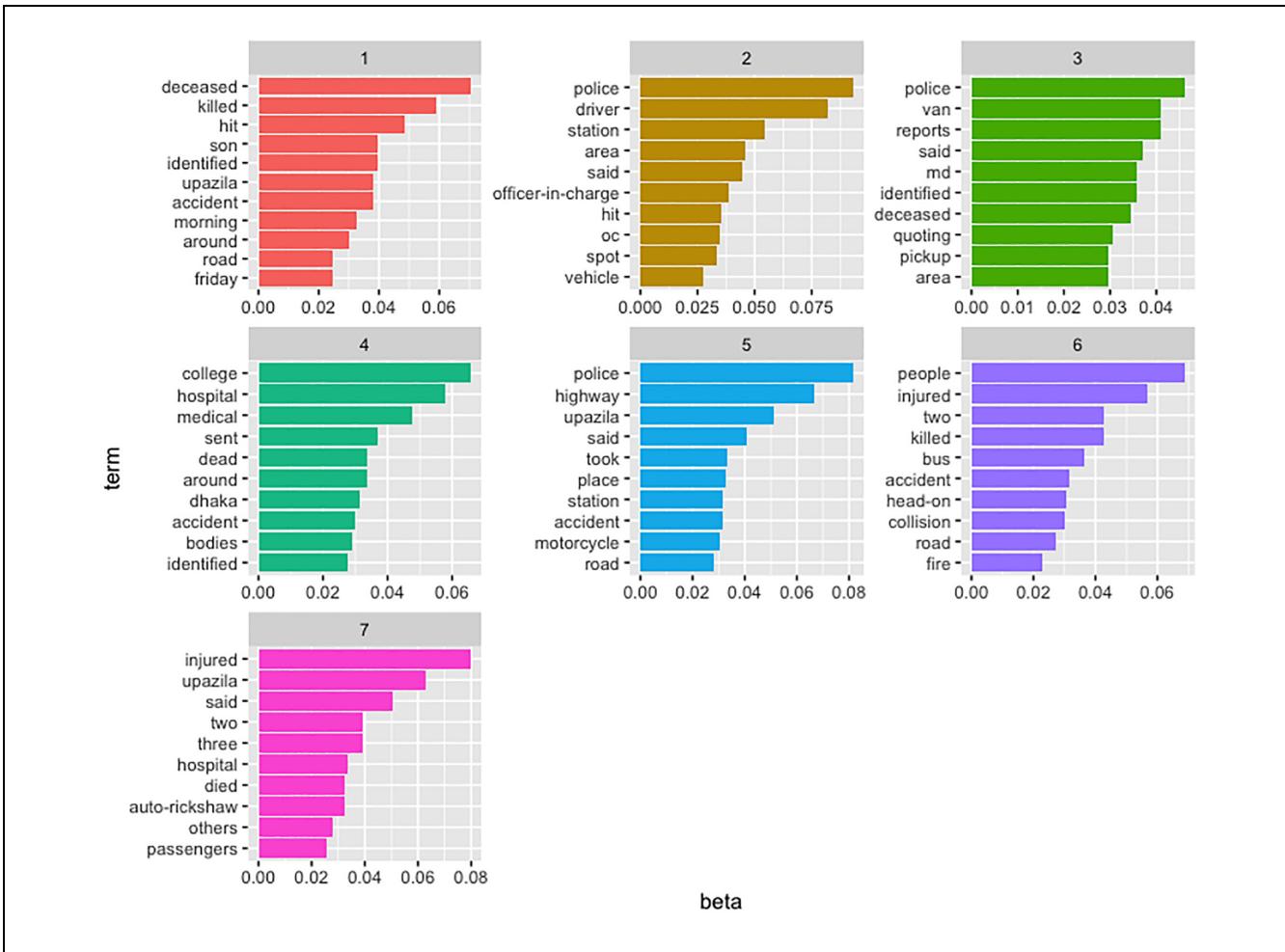


Figure 10. Identified topic models for truck-involved crashes.

on. The general findings of this word cloud analysis are truck-involved crashes 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 provided some useful information, such as the kinds of vehicles involved (auto-rickshaw, bus, van, and motorcycle), the manner of collision (head-on, speeding), and the 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.” This indicates that vulnerable roadway user risk and crossing-related issues need 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. This is consistent with a recent study (64) that emphasized inclement weather as a major factor in the severity of collisions during the winter season.
- Wrong-way driving (WWD) was identified as a factor contributing to truck-involved crashes. 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). Transportation agencies can focus on these specific routes.
- Topic modeling suggested several important crash scenarios, including “a truck-involved crash

occurring on Friday morning resulting in fatalities,” a “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,” and “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.

Conclusion

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, including word cloud analysis, word frequency analysis, WCN analysis, RAKE, and topic modeling. 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 (72), there are currently 292,000 drivers with heavy and medium driving licenses in the country. However, 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 crashes. Drivers without a legal license are more likely to drive recklessly and pay less attention to traffic laws and regulations. To address road crashes, the Bangladeshi government must first concentrate on this license issue. Furthermore, future research on truck-related crashes in Bangladesh can go extensively into these licensing challenges.

There are several practical solutions and implications that can be considered to reduce the number of fatal truck-involved crashes in Bangladesh. Some of these solutions include location-specific interventions, seasonal interventions, crash scenario-specific interventions, and education campaigns. To implement location-specific interventions, interventions can be tailored to the Khulna division and the Jhenaidah district, where the majority of fatal truck crashes occur. Several measures can be taken to accomplish this, such as enforcing traffic rules more strictly, improving road infrastructure, and promoting awareness campaigns for truck drivers and other road users. To address the factors that contribute to crashes during the holiday season, seasonal interventions should be implemented. Several countermeasures can be implemented to address specific crash scenarios, including addressing vulnerable road users’ risk at railway

crossings, addressing crossing-related issues, and utilizing “wrong way” signs to address WWD. Disseminating information about specific crash scenarios and ways to avoid them can be used in education campaigns to promote safe driving practices and create awareness. Further research should investigate the effectiveness of these interventions, identify additional factors contributing to truck-involved crashes, and use machine learning algorithms for predictive analysis to inform targeted interventions. Lastly, it is important to examine the economic implications of truck-related crashes, including medical treatment costs, property damage, and lost productivity.

Research Contribution

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

Limitation

The research team’s reliance on crash reports collected over a 12-month period limits the scope of their findings. Conducting data collection for extended periods would likely reveal more intriguing crash patterns. Furthermore, the availability of the database is restricted to online websites that are based on the English language. To enhance future studies, it is recommended to include the Bangla news portal. To get a clear understanding of the context, the research team further reviewed a few Bangla and English news articles and focused on reported accidents. We found similarities between the crashes reported by both language newspapers. Interestingly, the newspapers in Bangladesh have both Bangla and English versions to cover both national and international audiences. For example, the newspaper named “Prothom Alo” ([link: https://www.prothomalo.com/](https://www.prothomalo.com/)) has both Bangla and English versions. Therefore, the same crash is expected to be reported by both the Bangla and English versions of the newspaper. The inclusion of the Bangla language newspaper is expected to be a redundant process of data collection and may end up reporting the same crash in both versions. Based on this observation, the research team decided not to include any Bangla version of the newspaper. It is worth noting that online news portals generally offer generic information with respect to crashes without delving into the in-depth technical details of the contributing factors. The identification of generic crash reporting practices in online news portals in Bangladesh suggests the need for additional research to explore the

underlying reasons and potential avenues for improvement.

Acknowledgments

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

Author Contributions

The authors confirm 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, A. Sheykhfard. All authors reviewed the results and approved the final version of the manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Ahmed Hossain  <https://orcid.org/0000-0003-1566-3993>
 Xiaoduan Sun  <https://orcid.org/0000-0001-7282-1340>
 Subasish Das  <https://orcid.org/0000-0002-1671-2753>
 Abbas Sheykhfard  <https://orcid.org/0000-0002-9536-3108>

References

1. Dutta, U., X. Zhong, and N. Gsouda. Analysis of Global Road Traffic Death Data Using a Clustering Approach. *Current Urban Studies*, Vol. 10, No. 2, 2022, pp. 275–292. <https://doi.org/10.4236/cus.2022.102017>.
2. WHO. Road Traffic Injuries. <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries>. Accessed May 31, 2023.
3. WHO. Global Status Report on Road Safety 2018. <https://www.who.int/publications-detail-redirect/9789241565684>. Accessed July 22, 2022.
4. 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 July 25, 2022.
5. Sperling, D., and D. Salon. *Transportation in Developing Countries: An Overview of Greenhouse Gas Reduction Strategies*. Pew Center on Global Climate Change, Arlington, VA, 2002.
6. Nantulya, V. M., and M. R. Reich. The Neglected Epidemic: Road Traffic Injuries in Developing Countries. *BMJ*, Vol. 324, No. 7346, 2002, pp. 1139–1141.
7. Hosseinzadeh, A., A. Moeinaddini, and A. Ghasemzadeh. Investigating Factors Affecting Severity of Large Truck-Involved Crashes: Comparison of the SVM and Random Parameter Logit Model. *Journal of Safety Research*, Vol. 77, 2021, pp. 151–160. <https://doi.org/10.1016/j.jsr.2021.02.012>.
8. Yuan, Y., M. Yang, Y. Guo, S. Rasouli, Z. Gan, and Y. Ren. Risk Factors Associated with Truck-Involved Fatal Crash Severity: Analyzing Their Impact for Different Groups of Truck Drivers. *Journal of Safety Research*, Vol. 76, 2021, pp. 154–165. <https://doi.org/10.1016/j.jsr.2020.12.012>.
9. Chang, L.-Y., and J.-T. Chien. Analysis of Driver Injury Severity in Truck-Involved Accidents Using a Non-Parametric Classification Tree Model. *Safety Science*, Vol. 51, No. 1, 2013, pp. 17–22. <https://doi.org/10.1016/j.ssci.2012.06.017>.
10. Ahsan, H. Road Safety in Bangladesh: Key Issues and Countermeasures. *Forum, Monthly Publication of Daily Star*, No. 6, 2012.
11. Road Accidents Kill 9,951 in 2022. *New Age. The Most Popular Outspoken English Daily in Bangladesh*. <https://www.newagebd.net/article/190603/road-accidents-across-bangladesh-kill-9951-in-2022-report>. Accessed May 24, 2023.
12. bdnews24.com. 43 School-Goers Die in Truck Plunge. *bdnews24.com*. <https://bdnews24.com/bangladesh/43-school-goers-die-in-truck-plunge>. Accessed July 22, 2022.
13. The Times of India. 15 Killed in Road Accident in Bangladesh. *The Times of India*, March 22, 2020.
14. Gulf News. 10 Die in Bangladesh Accident. <https://gulf-news.com/world/asia/10-die-in-bangladesh-accident-1-2000436>. Accessed July 22, 2022.
15. Daily Star Digital Report. 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 July 22, 2022.
16. Daily Star Digital Report. 3 Killed in Chapainawabganj Road Crash. *The Daily Star*. <https://www.thedailystar.net/country/news/3-killed-chapainawabganj-road-crash-2093013>. Accessed July 22, 2022.
17. Daily Star Digital Report. 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 July 22, 2022.
18. Daily Star Digital Report. 5 Day Labourers Die as Truck Overturns in Tangail. *The Daily Star*. <https://www.thedailystar.net/truck-overturns-in-tangail-5-die-1886908>. Accessed July 22, 2022.
19. Daily Star Digital Report, 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 July 22, 2022.
20. Prothom Alo English Desk. 2 Killed, 5 Injured in Natore Road Accident. *Prothomalo*. <https://en.prothomalo.com/bangladesh/accident/2-killed-5-injured-in-natore-road-accident>. Accessed July 22, 2022.
21. Das, S. Understanding Fatal Crash Reporting Patterns in Bangladeshi Online Media Using Text Mining.

- Transportation Research Record: Journal of the Transportation Research Board*, 2021. 2675: 960–971.
22. 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>.
 23. 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>.
 24. 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>.
 25. 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>.
 26. Zhao, Q., T. Goodman, M. Azimi, and Y. Qi. Roadway-Related Truck Crash Risk Analysis: Case Studies in Texas. *Transportation Research Record: Journal of the Transportation Research Board*, 2018. 2672: 20–28.
 27. Rahimi, A., G. Azimi, H. Asgari, and X. Jin. Clustering Approach toward Large Truck Crash Analysis. *Transportation Research Record: Journal of the Transportation Research Board*, 2019. 2673: 73–85.
 28. 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*, 2020. 2674: 742–754.
 29. Mahmud, S. S., I. Ahmed, and M. S. Hoque. Road Safety Problems in Bangladesh: Achievable Target and Tangible Sustainable Actions. *Jurnal Teknologi*, Vol. 70, No. 4, 2014, pp. 43–49.
 30. Sufian, A., I. Ahmed, and S. Khan. A Study on the Factors Involved in Truck Accidents in Bangladesh. *Proc., 2nd International Conference on Civil Engineering for Sustainable Development (ICCESD-2014)*, February 14–16, 2014, KUET, Khulna, Bangladesh.
 31. 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>.
 32. 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.
 33. 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, 2005, pp. 65–74.
 34. Islam, M., R. Bin Ali, and F. K. Chowdhury. Road Accident Analysis and Prevention Measures of Rajshahi - Sirajganj Highway in Bangladesh. Vol. 126, 2019, pp. 209–221.
 35. Ahsan, H., S. Mahmud, and M. Bhuiyan. Heavy Vehicle Aggressivity in Bangladesh: Case Study on Large Truck. *Proc., 1st International Conference on Civil Engineering for Sustainable Development (ICCESD-2012)*, March 2–3, 2012, KUET, Khulna, Bangladesh, pp. 8–10. Buet.Ac.Bd.
 36. Hoque, Md. M., and S. M. S. Mahmud. Road Safety Engineering Challenges in Bangladesh. *Proc., 13th Conference of the Road Engineering Association of Asia and Australasia (REAAA)*, Songdo Convensia, Incheon, Korea, 2009.
 37. Rahimi, E., A. Shamshiripour, A. Samimi, and A. K. Mohammadian. Investigating the Injury Severity of Single-Vehicle Truck Crashes in a Developing Country. *Accident Analysis & Prevention*, Vol. 137, 2020, p. 105444.
 38. Tulu, G. S., S. Washington, Md., M. Haque, and M. J. King. Injury Severity of Pedestrians Involved in Road Traffic Crashes in Addis Ababa, Ethiopia. *Journal of Transportation Safety & Security*, Vol. 9, Supplement 1, 2017, pp. 47–66. <https://doi.org/10.1080/19439962.2016.1199622>.
 39. Chan, Y. C. M. *Truck Overloading Study in Developing Countries and Strategies to Minimize Its Impact*. Queensland University of Technology, Brisbane, Australia, 2008.
 40. Wen, H., Y. Du, Z. Chen, and S. Zhao. Analysis of Factors Contributing to the Injury Severity of Overloaded-Truck-Related Crashes on Mountainous Highways in China. *International Journal of Environmental Research and Public Health*, Vol. 19, No. 7, 2022, p. 4244.
 41. Chen, S., S. Zhang, Y. Xing, and J. Lu. Identifying the Factors Contributing to the Severity of Truck-Involved Crashes in Shanghai River-Crossing Tunnel. *International Journal of Environmental Research and Public Health*, Vol. 17, No. 9, 2020, p. 3155.
 42. Aliakbari, M., and S. Moridpoore. Management of Truck Loading Weight: A Critical Review of the Literature and Recommended Remedies. *MATEC Web of Conferences*, No. 81, 2016, p. 03007.
 43. Das, S. News Media Mining to Explore Speed-Crash-Traffic Association during COVID-19. *Transportation Research Record: Journal of the Transportation Research Board*, 2022, p. 03611981221121261.
 44. 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>.
 45. 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>.
 46. 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 Hawai'i, 2019: A Content Analysis. *Transportation Research Interdisciplinary Perspectives*, Vol. 13, 2022, p. 100525. <https://doi.org/10.1016/j.trip.2021.100525>.
 47. Hearst, M. A. Untangling Text Data Mining. *Proc., 37th Annual meeting of the Association for Computational Linguistics*, Maryland, 1999.

48. Cai, Y., and J.-T. Sun. Text Mining. In *Encyclopedia of Database Systems* (L. Liu, and M. T. Özsü, eds.), Springer, Boston, MA, 2009, pp. 3061–3065.
49. 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.
50. Brown, D. E. Text Mining the Contributors to Rail Accidents. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 17, No. 2, 2015, pp. 346–355.
51. Nayak, R., N. Piyatrapoomi, and J. Weligamage. Application of Text Mining in Analysing Road Crashes for Road Asset Management. In *Engineering Asset Lifecycle Management* (D. Kiritsis, C. Emmanouilidis, A. Koronios, and J. Mathew, eds.), Springer, London, pp. 49–58.
52. Soleimani, S., A. Mohammadi, J. Chen, and M. Leitner. Mining the Highway-Rail Grade Crossing Crash Data: A Text Mining Approach. *Proc., 18th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Boca Raton, FL, 2019.
53. Giumentra, 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.
54. 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.
55. 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.
56. Qaiser, S., and R. Ali. Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. *International Journal of Computer Applications*, Vol. 181, No. 1, 2018, pp. 25–29.
57. Silge, J., and D. Robinson. *Text Mining with R: A Tidy Approach*. O'Reilly Media, Inc., Sebastopol, CA, 2017.
58. Alghamdi, R., and K. Alfalqi. A Survey of Topic Modeling in Text Mining. *International Journal of Advanced Computer Science and Applications (IJACSA)*, Vol. 6, No. 1, 2015, pp. 147–153.
59. Kherwa, P., and P. Bansal. Topic Modeling: A Comprehensive Review. *EAI Endorsed Transactions on Scalable Information Systems*, Vol. 7, No. 24, 2019.
60. Goyal, A., and I. Kashyap. Latent Dirichlet Allocation - An Approach for Topic Discovery. *Proc., International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)*, Faridabad, India, 2022.
61. Egger, R., and J. Yu. A Topic Modeling Comparison between Lda, Nmf, Top2vec, and Bertopic to Demystify Twitter Posts. *Frontiers in Sociology*, Vol. 7, 2022, p. 886498.
62. Blei, D. M., A. Y. Ng, and M. I. Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, Vol. 3, 2003, pp. 993–1022.
63. 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 October 29, 2022.
64. Haque, F., A. S. Huq, Z. S. Ishmam, and M. M. Fuad. Visualizing the Hot Spots of Adverse Weather Induced Traffic Accidents in Bangladesh. Presented at 101st Annual Meeting of the Transportation Research Board, Washington, D.C., 2022.
65. Pourroostaei Ardakani, S., X. Liang, K. T. Mengistu, R. S. So, X. Wei, B. He, and A. Cheshmehzangi. Road Car Accident Prediction Using a Machine-Learning-Enabled Data Analysis. *Sustainability*, Vol. 15, No. 7, 2023, p. 5939. <https://doi.org/10.3390/su15075939>.
66. Dong, C., S. H. Richards, B. Huang, and X. Jiang. Identifying the Factors Contributing to the Severity of Truck-Involved Crashes. *International Journal of Injury Control and Safety Promotion*, Vol. 22, No. 2, 2015, pp. 116–126.
67. Newnam, S., D. Blower, L. Molnar, D. Eby, and S. Koppel. Exploring Crash Characteristics and Injury Outcomes among Older Truck Drivers: An Analysis of Truck-Involved Crash Data in the United States. *Safety Science*, Vol. 106, 2018, pp. 140–145. <https://doi.org/10.1016/j.ssci.2018.03.012>.
68. Khattak, A. J., and F. Targa. Injury Severity and Total Harm in Truck-Involved Work Zone Crashes. *Transportation Research Record: Journal of the Transportation Research Board*, 2004. 1877: 106–116.
69. Islam, M., and S. Hernandez. Large Truck-Involved Crashes: Exploratory Injury Severity Analysis. *Journal of Transportation Engineering*, Vol. 139, No. 6, 2013, pp. 596–604.
70. Federal Motor Carrier Safety Administration. Large Truck and Bus Crash Facts 2014. U.S. Department of Transportation, Washington, D.C., 2014.
71. Large Trucks. *IIHS-HLDI Crash Testing and Highway Safety*. <https://www.iihs.org/topics/large-trucks>. Accessed March 21, 2023.
72. Most Bus, Truck Drivers Reluctant to Upgrade Licence in Bangladesh. *New Age. The Most Popular Outspoken English Daily in Bangladesh*. <https://www.newagebd.net/article/128192/most-bus-truck-drivers-reluctant-to-upgrade-licence-in-bangladesh>. Accessed May 29, 2023.