Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Patterns Associated with Fatal Motorcycle-Involved Crashes in Bangladesh: Applying Text Mining Techniques and Structural Topic Modeling

Ahmed Hossain

ahmedhossain09@outlook.com

Arizona Department of Transportation https://orcid.org/0000-0003-1566-3993

Nazmus Sakib

Auburn University https://orcid.org/0000-0002-0525-6342

Abdullah Al Asif

Bangladesh University of Engineering & Technology https://orcid.org/0009-0007-2491-5976

Subasish Das

Texas State University https://orcid.org/0000-0002-1671-2753

Research Article

Keywords: Motorcycle, Speeding, Rural, Trucks, Safe System Approach

Posted Date: April 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9374992/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Patterns Associated with Fatal Motorcycle-Involved Crashes in Bangladesh: Applying Text Mining Techniques and Structural Topic Modeling

Abstract: Motorcycle crashes in Bangladesh pose a growing public health and transportation safety challenge, worsened by persistent gaps and availability in official crash data. This study uses a news media mining approach to examine 475 motorcycle-related crash reports from English-language Bangladeshi online newspapers published in 2022, forming a corpus of 50,190 words. Combining text mining with Structural Topic Modeling (STM) approach, the study identifies key crash patterns and latent themes within unstructured narratives. Preliminary analyses reveal frequent references to severe injuries, emergency response, young riders, fixed-object collisions, and nighttime crashes. Rapid Automatic Keyword Extraction and word co-occurrence network analyses underscore high-risk behaviors such as speeding and distraction, vulnerability of young male riders and students, and the influence of rural roads, highways, and urban environments. STM results indicate two dominant crash profiles: fatal single-vehicle crashes in rural settings linked to loss of control, and multiple-vehicle crashes in urban or highway areas often involving heavy vehicles. Guided by the Safe System Approach, the study proposes context-specific countermeasures, including infrastructure improvements, lane discipline enforcement, youth-focused education, improved vehicle visibility, and stronger emergency medical services. Overall, the findings demonstrate that online news media provide a scalable and informative source for motorcycle safety research in data-limited settings, supporting targeted strategies.

Keywords: Motorcycle, Speeding, Rural, Trucks, Safe System Approach

1. Introduction

Compared to other vehicle types, motorcycle riders face a disproportionate risk due to their unprotected nature in the event of a crash. Per vehicle mile traveled (VMT), motorcyclist riders are about 28 times more likely to be killed in a crash than car occupants, according to the data published by the National Highway Traffic Safety Administration (NHTSA) [1]. In addition, data collected from the Fatality Analysis and Reporting System (FARS) shows that motorcycle riders represent 23.7% of total driver fatalities in the US (more details in **Figure 1a**). The motorcycle riders' fatalities in 2023 are a 26.5% increase compared to 2019, clearly suggesting the growing vulnerability of motorcyclists on U.S. roadways. This upward trend highlights the urgent need for targeted interventions, including infrastructure planning, and behavioral strategies aimed specifically at reducing motorcycle-related fatalities.

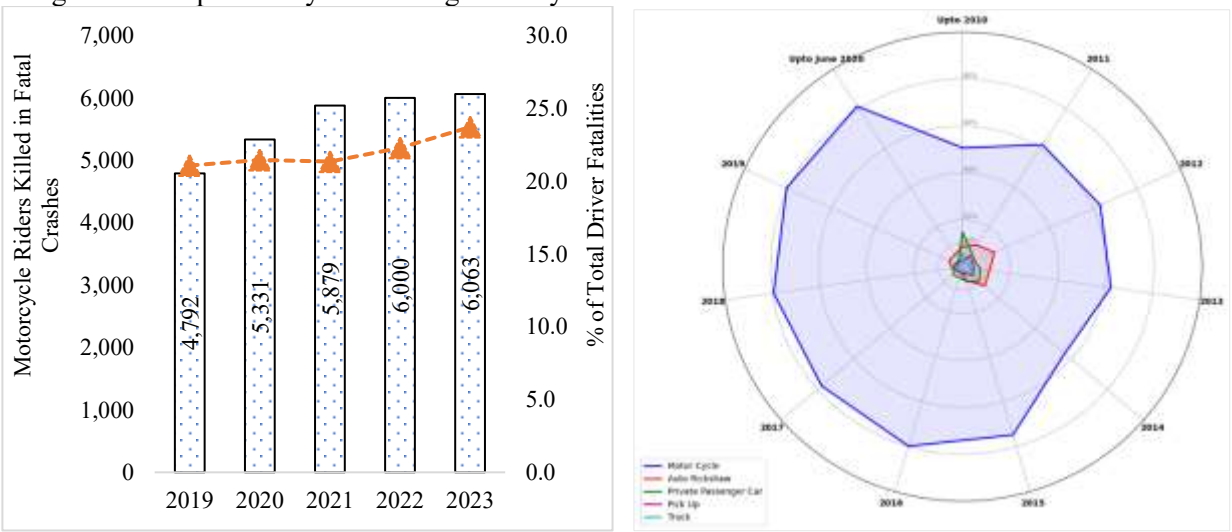


Figure 1. a) US Motorcycle Fatal Crash Statistics (left); b) Top Five Registered Vehicles in Bangladesh

From a general standpoint, previous research has predominantly addressed a wide range of contributing factors affecting the likelihood of crashes involving motorcyclists. These include contextual and situational factors, including riders' age and gender [2], [3], speeding behaviors [4], [5], [6], [7], helmet usage [8], curved segments [9], alcohol [6], [10], type of roads [11] and so on. These studies cover a wide spectrum of countries, including the US, China, Ghana, Malaysia, Thailand, India, Pakistan, and so on. In recent years, Bangladesh has witnessed a dramatic rise in motorcycle usage (more details in **Figure 1b**) driven by rapid urbanization, affordability, and the growth of informal ride-sharing services [12]. According to the Bangladesh Road Transport Authority (BRTA) latest available data, motorcycles account for over 81.3% of the total registered motor vehicles in the country, and this number continues to grow sharply. More alarmingly, data from the Accident Research Institute (ARI) and various road safety reports suggest that motorcyclists constitute a disproportionately high percentage of road traffic fatalities, often ranging between 25% and 30% annually [13]. These patterns underscore critical safety gaps in terms of enforcement, infrastructure, and rider behavior. Given the unique traffic dynamics and policy challenges in low and middle-income countries, a focused investigation into the factors contributing to motorcycle-involved fatal crashes in Bangladesh is urgently needed to inform context-appropriate interventions.

The motorcycle crash investigation paradigm in Asian countries like Bangladesh requires a fundamentally different analytical framework compared to other approaches, particularly those developed in the United States. While motorcycles in the US primarily serve recreational purposes with riders typically following established traffic patterns and lane discipline, motorcycles in Bangladesh constitute essential daily transportation for millions of commuters operating within complex, mixed-traffic environments characterized by fluid lane usage, informal traffic hierarchies, and adaptive riding behaviors. This operational context creates distinct crash causation patterns that cannot be adequately understood through investigation methodologies designed for homogeneous traffic conditions with strict lane adherence. Consequently, motorcycle crash investigations in Bangladesh should also account for the unique interplay between high-density mixed traffic, economic necessity driving risk acceptance, infrastructure limitations, and culturally adapted traffic negotiation behaviors, necessitating the development of region-specific investigation protocols that recognize these environmental and behavioral factors as integral components of crash causation rather than mere contributing circumstances.

The development of state-of-the-art modeling approaches for motorcycle crash analysis in Bangladesh is severely constrained by inadequate crash data collection systems, which lack the comprehensiveness and systematic documentation found in countries like the United States, where extensive police reports, hospital records, and standardized crash databases provide rich analytical datasets. This data scarcity presents a critical challenge that demands innovative methodological approaches, making news media mining an invaluable alternative data source that can capture crash incidents often unreported on official channels [14]. The application of text mining techniques to extract, categorize, and analyze crash-related information from English news reports represents a unique research approach that addresses both the data limitation challenge and the contextual specificity of Bangladesh's traffic environment. Focusing on this need, the objective of research is to investigate motorcycle involved crashes in Bangladesh using text data collected from online English news portals. The obtained result of this research is expected to provide an improved understanding of the core factors contributing to motorcycle involved crashes in Bangladesh and provide safety recommendations.

2. Literature Review

This section provides an in-depth review of the key contributing factors to motorcycle-involved crashes, including human, vehicle, roadway, environmental, and temporal elements, as identified by existing literature.

2.1 Human Factors

Several studies identified that not using helmets is associated with a higher mortality rate as well as an increased rate of severe head injuries [15], [16], [17], [18], [19], [20], [21], [22]. Pour-Rouholamin, Jalayer,

and Zhou (2017) found that older riders and those riding under the influence of drugs or alcohol were more likely to suffer severe injuries or fatalities [16]. Shaheed and Gkritza (2014) highlighted that impaired riding and older riders (over 25 years) increased injury severity [21]. Salum et al. (2019) reported that riding under the influence and reckless riding significantly increased the likelihood of fatalities [22]. Islam and Brown (2017) indicated that rider behavior, such as riding under the influence of alcohol, significantly influenced crash severity across both rural and urban roadways [19]. Tamakloe et al. (2022) found that full license status increased the risk of fatal injuries at signalized intersections, while inattentiveness and young riders increased the risk of fatal injuries at non-signalized intersections [23]. Wang et al. (2022) suggested that carrying vulnerable passengers, such as women, children, and elders, reduces the severity of injuries to riders in crashes, while these passengers themselves are more likely to suffer severe injuries [24]. Islam et al. (2021) indicated that age affects motorcyclists' safety perception and ability to assess risks, thereby influencing their involvement in risky behaviors [25]. Few studies have suggested that the role of traffic violations, such as speeding, in increasing crash severity [16], [20], [21], [22].

2.2 Vehicle and Collision Type

Vehicle-related factors play a vital role in the severity of motorcycle crashes. Several studies suggested that crashes involving multiple vehicles and crashes with fixed objects increased the likelihood of severe injuries [16], [21], [26], [27]. Jung, Xiao, and Yoon (2013) found that collisions involving trucks and head-on impacts were more likely to result in fatal injuries [20]. Wahab et al. (2020) found that collisions between motorcycles and heavy goods vehicles increase the likelihood of fatal injuries [28]. Li et al. (2021) showed that the pre-speed of the motorcycle was a critical factor in the severity of injuries in crashes involving multiple vehicles [29]. Farid et al. (2021) suggested that most collisions involving motorcycles and animals or barriers tend to result in severe outcomes compared to other single-motorcycle crashes [30].

2.3 Roadway Factors

Roadway characteristics significantly impact the severity of motorcycle crashes. Several studies emphasized that high posted speed limits and the presence of horizontal curves increased the odds of fatal crashes [22], [24], [27], [30]. Manan et al. (2018) found that crashes on rural roads, those involving curved road sections, lack of road markings, and smooth surfaces increased the probability of single-vehicle fatal crashes. For multiple-vehicle fatal crashes, factors such as expressways, primary and secondary roads, speed limits of more than 70 km/h, and roads with non-permissible markings (e.g., double lane lines were significant [27]. Wahab and Jiang (2020) identified that crashes occurring at junctions, on good road surfaces, and roads with poor shoulders or signage were associated with higher fatality rates [28]. Vajari et al. (2020) found that crashes at stop or give-way intersections, t-intersections, roundabouts, and uncontrolled intersections significantly increased the risk of severe injuries [26]. Wang, Yu, and Ma (2023) highlighted that various road characteristics, including junction controls, road obstacles, restricted lanes, and speed limits, contributed to crash severities [31]. Tamakloe et al. (2022) suggested that shoulder presence increased the risk of fatal injuries at signalized intersections, while good road surface and shoulder absence increased the risk of fatal injuries at non-signalized intersections [23].

2.4 Environmental and Temporal Factors

Several studies suggested that motorcycle crashes on weekends and off-peak hours increase the likelihood of crash severity [16], [20], [22], [26], [28]. Tamakloe et al. (2022) found that daytime/daylight increased the risk of fatal injuries at signalized intersections, while nighttime increased the risk of fatal injuries at non-signalized intersections [23]. Pour-Rouholamin, Jalayer, and Zhou (2017) indicated that crashes during summer and in darkness significantly increased the severity of injuries [16]. Salum et al. (2019) indicated that poor lighting conditions contributed to higher fatality risks [22]. Vajari et al. (2020) suggested that specific times of day, such as midnight to early morning, had a significant impact on the probability of severe crashes [26]. Islam and Brown (2017) identified that roadways without light at night and clear weather conditions were associated with severe injuries, particularly in rural areas [19]. Manan et al. (2018) suggested that crashes at late-night hours (00:00 am to 6:00 am) increased single-vehicle fatal crash risks,

and crashes on daylight conditions increased multiple-vehicle fatal crash risks. Additionally, the time of day between 7:00 pm to 12:00 am also had an increasing impact on the probability of fatal motorcycle crashes for both single and multiple vehicles [27].

2.5 Major Motorcycle Crash Investigations in Bangladesh

Table 1 provides information from a few previous studies in the context of Bangladesh, including the year, data source, location, methods, and key findings of motorcycle-involved crashes. Note that the findings of these studies are limited, and it is difficult to assess how human, roadway, or environmental factors are interrelated with motorcycle-involved crashes.

Table 1. Findings from Key Selected Investigations in Bangladesh

Ref.	Author, Year	Data Source	Location, Methods	Key findings on the motorcycle-involved crash
[32]	Rahman et al., 2021	Accident Research Institute (ARI), 2006-2015	Dhaka, Binary logit	• The motorcycle crash death rate in Bangladesh at 28.4 deaths per 10,000 motorcycles. • Crashes occurring on weekends, rainy season, dawn and night periods, non-intersections, straight and flat roads, highways, pedestrian-related incidents, defect-free motorcycles, heavier vehicles, helmet non-usage, and drivers with alcohol suspicion increase the likelihood of fatal motorcycle crashes.
[33]	Pervaz et al., 2020	Accident Research Institute (ARI) from 1998-2017, Newspaper and Questionnaire Survey Data	Bangladesh, Descriptive analysis	• Approximately 74% of motorcycle fatalities occurred in rural locations, and 22% of victims were 26-30 years old. • In 49% of motorcycle-involved fatal crashes were head-on, and 88% of the motorcyclists were without helmets.
[34]	Akter and Pervaz, 2019	Accident Research Institute (ARI), 2000-2014	Dhaka, Descriptive analysis and graphical representation	• Rear-end collisions accounted for a significant majority, 53.5%, in motorcycle-involved crashes • In 94% of accidents, motorcyclists involved were not wearing helmets
[35]	Hossain et al., 2022	Questionnaire Survey Data	Dhaka City, Heckman probit model	• Motorcycles involved in road accidents accounted for 42.1%.
[36]	Das et al., 2023	Questionnaire Survey Data	Dhaka city, t-test and one-way analysis of variance (ANOVA)	• Only 54.6% of motor bikers demonstrated good knowledge of road signs and safety regulations, and 50.6% of respondents followed good practices of road safety measures.
[37]	Pervaz et al., 2022	Accident Research Institute (ARI), 2010-2015	Bangladesh, Descriptive analysis	• The percentage of motorcycle crashes has nearly tripled over a 17-year period (1998-2015).

2.6 Traffic Safety Analysis Using News Media Mining

Recent studies have increasingly utilized news media as an alternative data source for analyzing traffic safety, particularly when traditional crash records are limited or inconsistent. Das and Sarkar (2024) studied the effect of the COVID-19 pandemic on how international news media portrayed speed-related crashes and found that reckless driving, fatality surges, and equity issues were dominant themes across multiple outlets [38]. Yang et al. (2020) used data from U.S.-based news articles to compile 169 e-scooter crash reports and identified skewed patterns in rider demographics, crash severity, and contributing behaviors such as riding under the influence and limited helmet use [39]. Goddard et al. (2019) examined the relationship between editorial framing and public perception of crash responsibility in a controlled experiment and found that shifting from pedestrian-focused to driver-focused narratives significantly reduced victim-blaming and increased support for systemic safety interventions such as infrastructure reform [40].

Yankson et al. (2010) reviewed 240 newspaper articles from Ghana and found that while traffic crashes were frequently reported, pedestrian injuries and key preventative topics such as speed control and seatbelt use were underrepresented, particularly in state-run media [41]. Das (2021) used data from 419 Bangladeshi news articles to analyze fatal crash coverage and found substantial inconsistencies in how crash causes and geometric features were documented [42]. Bhuiyan et al. (2022) developed crash severity prediction models using newspaper-derived data in Bangladesh and found that factors such as license type, seatbelt use, road classification, and crash timing significantly influenced severity levels [43]. The study highlighted the value of media-reported crash data in capturing details often missing from official records. Moreover, these studies developed methodologies to quantify risk factors using media narratives and affirmed that news media mining constitutes a viable and methodologically robust approach to traffic safety analysis, particularly in developing countries where reliable crash data are often limited.

2.7 Structural Topic Modeling (STM) in Previous Research

STM has increasingly been used to extract latent themes from crash narrative data. A previous research applied STM to 463 crash reports of autonomous vehicles and identified 15 latent topics [44] . The study revealed that rear-end crashes are most frequent but declining, alongside a rise in pedestrian and bicycle-related incidents. Another investigation leveraged STM on fatal crash narratives from the Michigan traffic crash facts database [45]. Several themes are uncovered, including run-off-road events, lane-change failures, intersection violations, and speeding. They also linked these patterns to driver demographics. In another study, STM was applied to analyze cyclists' near-miss and collision reports and identified high-risk topics such as car dooring, driver violations at crosswalks, and bike lane encroachment [46].

Xu et al. (2025) used STM to explore passenger injury incidents on public buses and identified 33 themes of non-collision bus injuries [47]. The research highlighted that standing passenger imbalance, skidding, and boarding incidents were most associated with severe injuries. Das et al. (2021) applied STM to uncover patterns in motorcycle crashes and found factors such as 'unsafe speed,' 'intersection,' and 'fracture' were linked to fatal crashes, while terms like 'headlight,' 'lane,' and 'broken' were associated with non-fatal outcomes [48]. Another study by Hossain et al. (2024) used STM to extract patterns from Bangladesh truck-involved fatal crashes. This study identified seven crash-related topics from news reports and revealed key factors like head-on collisions, speeding, and overtaking. They also found a notable shortage of licensed truck drivers in Bangladesh [14].

STM has also been adopted in other domains of the transportation safety context. For example, studies by Kuhn [49] and Xing et al. [50] used STM to identify latent themes in aviation safety and incident reports. Another study by Hong et al. proposed a framework using topic modeling for automated analysis of railway accident reports [51]. STM has also been employed in the wind turbine accident news analysis [52]. These studies collectively demonstrate the versatility and explanatory power of STM in uncovering latent insights across multiple transport domains.

2.8 Research Gap, and Study Objectives

Based on the extensive investigation of previous research on motorcycle crashes and a methodological standpoint, there is a need to identify the core factors contributing to motorcycle involved crashes in Bangladesh. To address this gap, there is a pressing need to explore innovative methodological approaches capable of extracting crash insights from alternative data sources. In this regard, the integration of text mining and STM offers a promising avenue. By utilizing unstructured crash narratives from English-language news media, this study aims to systematically extract and analyze latent themes and contributing factors associated with motorcycle crashes in Bangladesh. The objective is to uncover data-driven insights that can inform evidence-based safety interventions and policymaking, particularly in data-scarce settings where conventional methods prove inadequate.

2.9 Scope

This study focuses exclusively on online news articles written in English and published within Bangladesh. News content published in local languages, such as Bangla, falls outside the scope of this research and is therefore not included in the analysis.

3. Methods

Text mining, or text data mining, is the process of extracting meaningful insights from unstructured textual data by converting it into a structured format suitable for analysis. This multi-stage process includes data collection, preprocessing, transformation, analysis, interpretation, and visualization. Sources such as documents, news articles, and social media feeds serve as input. During preprocessing, the text is cleaned through techniques like tokenization (splitting text into words or phrases), stop-word removal (eliminating common words such as 'the' or 'and'), lowercasing, and stemming (reducing words to their root form, e.g., 'running' to 'run'). The cleaned text is then transformed into numerical representations using methods like Term Frequency-Inverse Document Frequency (TF-IDF), which highlights the importance of words in individual documents relative to the entire corpus. Subsequent analysis may involve techniques like topic modeling or classification to uncover patterns, followed by interpretation and visualization for informed decision-making.

A common technique for representing structured text data is the Term Document Matrix (TDM), where rows correspond to terms and columns to documents. Before constructing the TDM, standard preprocessing steps such as converting text to lowercase and removing punctuation, numbers, special characters, stopwords, and extra spaces are crucial to ensure consistency. For example, 'Accident' and 'accident' are treated as the same term after preprocessing. In the context of Bangladeshi online news media, it is important to note that the term 'accident' is widely used in reporting, even though 'crash' may be more accurate in many cases.

3.1 Term-Document Matrix (TDM)

A Term-Document Matrix (TDM) is a matrix representation of a collection of text documents. It displays the frequency of terms (i.e., words) occurring in the documents. Here, each row represents a document, and each column represents a unique term. The values in the matrix are the frequencies of each term in a document. The values 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 -

Document 1: A deadly **motorcycle crash** occurred in **Dhaka**. Many people were **injured**.
Document 2: **Motorcycle accidents** in Bangladesh have increased. **Authorities** are concerned.
Document 3: **Speeding motorcyclists** are a major cause of **crashes** in **Dhaka**.

Term	Doc 1	Doc 2	Doc 3
motorcycle	1	1	1
crash	1	0	0

Dhaka	1	0	1
accidents	0	1	0
speeding	0	0	1
authorities	0	1	0
injured	1	0	0

3.2 Term Frequency (TF)

Term Frequency (TF) measures the frequency of a word in a document. It is the number of times a term appears in a document divided by the total number of terms in the document. TF helps determine how important a word is in a particular document. Here, the assumption is that the more often a word appears, the more significant it is. For a term *t* in a document *d*, TF is defined as:

$$TF(t,d) = \frac{Number\ of\ times\ term\ t\ appears\ in\ document\ d}{Total\ number\ of\ terms\ in\ document\ d} \quad (1)$$

For example, in document 1 ('A deadly motorcycle crash occurred in Dhaka. Many people were injured'), the word 'motorcycle' appears once, and the total word count of the document is 9. The TF for the term 'motorcycle' would be:

$$TF(motorcycle, Doc\ 1) = \frac{1}{9} \approx 0.11$$

3.3 Document Frequency (DF)

Document Frequency (DF) is the count of how many documents in a corpus contain a particular term. It provides a sense of how widespread or rare a word is across all documents in the dataset. For example, the term 'motorcycle' appears in all three documents, so the DF for 'motorcycle' is 3. On the other hand, the term 'injured' appears only in Document 1, so the DF for 'injured' is 1.

3.4 Inverse Document Frequency (IDF)

Inverse Document Frequency (IDF) measures the importance of a term within the whole corpus. Words that appear in many documents are less informative (because they are common) and get a low IDF score. IDF helps to emphasize rare words that appear in fewer documents, as they are likely more meaningful for distinguishing documents. The equation for IDF is provided below:

$$IDF(t) = log\left(\frac{N}{df(t)}\right) \quad (2)$$

Here, *N* is the total number of documents in the corpus, and *df(t)* is the number of documents containing the term. If we have 3 documents in total and the term 'motorcycle' appears in all of them (DF = 3), the IDF would be $log\left(\frac{3}{3}\right) = 0$. Since 'motorcycle' is very common across the documents, it's IDF value is low, suggesting it's not very informative. On the other hand, a rare term like 'injured' (which appears in only 1 document) would have a higher IDF, $log\left(\frac{3}{1}\right) \approx 1.1$

3.5 Rapid Automatic Keyword Extraction (RAKE)

RAKE is an algorithm used to extract important keywords or key phrases from a text document. It focuses on identifying words or phrases that occur frequently and are relevant to the content of the document. RAKE uses the co-occurrence of words and their relationships to identify these keywords. The advantage of RAKE is that it does not require labeled data or any supervised learning.

3.6 Structural Topic Modeling (STM)

STM is a method used in natural language processing (NLP) for discovering topics within a collection of documents. Unlike traditional topic modeling approaches such as Latent Dirichlet Allocation (LDA), STM incorporates metadata (such as author, time, or any other covariates) into the modeling process. This inclusion of metadata helps to reveal how these factors influence topic distribution.

STM assumes each document is a mixture of topics, and each topic is a distribution over words. The advantage of STM is its ability to model with topic prevalence and topic content. Topic prevalence tells how frequently a topic appears in a document as a function of the document-level covariate. Topic content focused on how the vocabulary is used to express a topic based on covariates. Let x_d and y_d denote document-level covariates for topic prevalence and content, respectively. The following generative process defines the STM model:

3.6.1 Topic Proportions

Topic proportion is the share of each topic in a document. For each document *d*, the topic proportion vector $\theta_d \in \Delta^{K-1}$ (where *K* is the number of topics) is drawn from a logistic normal distribution, which is parameterized by a mean function of the covariates x_d, and can be expressed as:

$$\theta d \sim LogisticNormal\,(\,\mu(x_d), \Sigma\,) \tag{3}$$

Here, $\mu(x_d)$ is a linear function of the covariates, and Σ is the covariance matrix that captures topic correlations.

3.6.2 Topic Assignment

In the topic assignment, the model chooses a topic for each word in a document. For each word *n* in document *d,* a topic assignment $Z_{dn} \in \{1, \ldots, K\}$ is drawn from the document's topic proportions, and can be expressed as:

$$Z_{dn} \sim Multinomial(\,\theta_d\,) \tag{4}$$

3.6.3 Word Generation

In this step, the model selects a word for each word position in a document from the assigned topic's word distribution. Given the assigned topic Z_{dn}, the observed word W_{dn} is drawn from a topic-specific word distribution $\beta_{Z_{dn}}(y_d)$ that is influenced by content covariates, and can be expressed as:

$$W_{dn} \sim Multinomial\,[\,\beta_{Z_{dn}}\,(y_d)] \tag{5}$$

This structure allows STM to estimate how topic usage (prevalence) and expression (content) are systematically related to covariates such as time, author, or source. In contrast to LDA, which assumes fixed topic distributions across documents, STM enables dynamic, context-sensitive topic modeling.

3.7 Software Details

The research team utilized R statistical software (version 4.5.1) with the RStudio environment to conduct the analysis. Several R packages were utilized to facilitate text mining tasks. The ‘tm’ [53] and ‘quanteda’ [54] packages were employed for text preprocessing, including tokenization, ‘stopword’ removal, and the creation of document-feature matrices. The ‘tidytext’ package [55] supported the transformation of text data into a tidy format, enabling seamless integration with other ‘tidyverse’ tools [56]. Package ‘stringr’ [57] was used for efficient string manipulation, while ‘dplyr’ assisted in data wrangling and filtering operations. Visualization of word frequencies and topic patterns was achieved through ‘ggplot2’ and ‘wordcloud’, with enhanced color schemes provided by ‘RColorBrewer’. For unsupervised topic modeling, the ‘stm’ (Structural Topic Modeling) package [58] enabled the identification of latent themes across

documents while accounting for metadata. The 'rainette' package was used for hierarchical clustering and textual segmentation to uncover coherent content clusters. Additionally, 'igraph' facilitated network-based visualization and analysis of term co-occurrence patterns, supporting a deeper understanding of textual relationships.

4. Data

4.1 Google Alerts

Google Alerts (google.com/alerts), a web-based notification service, was utilized to systematically collect crash-related news articles from online sources in Bangladesh over a 12-month period (January 2022 – December 2022). This tool allows users to receive automated notifications from various digital platforms, such as web pages, news outlets, and blogs, based on user-defined keywords. To ensure the relevance of retrieved content, the research team configured Google Alerts with a set of targeted search terms, including: 'Bangladesh accident,' 'Bangladesh road collision,' 'Bangladesh road crash,' 'Bangladesh road fatalities and injuries,' 'Bangladesh traffic accident,' 'Bangladesh traffic crash,' and 'Highway crashes in Bangladesh.' Most of the collected articles originated from local English-language news organizations in Bangladesh.

It is important to note that the study focused exclusively on articles published in English, excluding reports written in Bangla or other local languages. All retrieved articles were manually reviewed and entered into the dataset. Given the likelihood of multiple news agencies reporting on the same incident, each report was carefully examined to identify and eliminate duplicates. Over the study period, a total of 9,300 English-language news articles were collected. From this pool, the research team conducted a thorough screening process to identify reports specifically involving motorcycle-related crashes. This process resulted in a final set of 475 unique motorcycle crash reports, which were selected for further analysis.

The curated dataset includes several key variables: source, headline, publication date, narrative, number killed, number injured, district, division, time of crash, crash type (single or multiple), and details of the two involved vehicles (vehicle 1 and vehicle 2). Of particular importance is the 'narrative' variable, which contains the full textual description of each crash. These narratives served as the primary input for the text mining and STM components of the study. The following **Figure 2** shows the database preparation and analysis flowchart in detail.

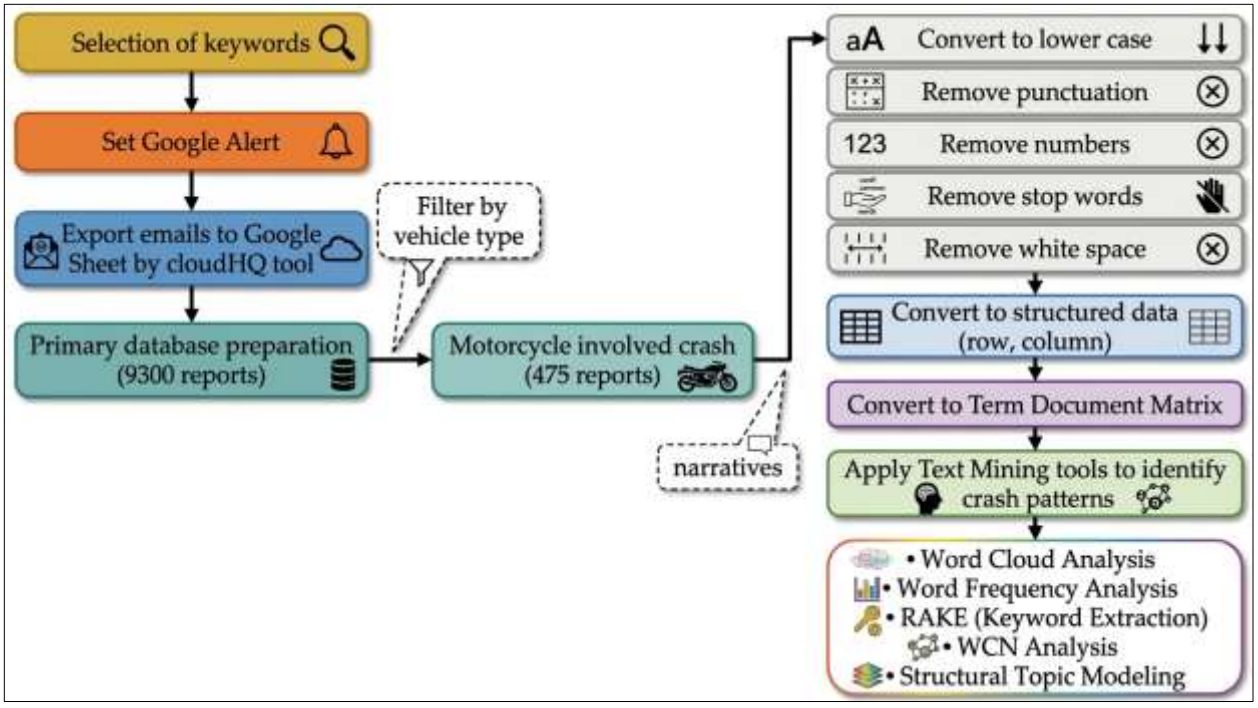


Figure 2. Database preparation and analysis flowchart

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

"Two motorcyclists were killed in a tragic road accident at Harirampur area on Manikganj-Jhitka road in Manikganj on Thursday night. The deceased were identified as Zahirul Islam, 23, son of late Jasim. Mollah and Shah Alam, 22, son of Yunus Ali of Gala village under Harirampur upazila in Manikganj. Officer-in-Charge of Harirampur Police Station Mueed Chowdhury said, Zahirul along with his friend auto driver Shah Alam were returning home from Manikganj district town around 10 pm on Thursday. Due to fast speed, their motorcycle lost control and hit a roadside tree. Both of them died on the spot. The bodies of the deceased were recovered at night and sent to the morgue of Manikganj Sadar Hospital for autopsy. A case has been filed with Harirampur police station in this regard, added the OC".

4.2 Exploratory Data Analysis (EDA)

The final dataset comprises 475 unique motorcycle crash reports. Descriptive statistics of the word count for each narrative reveal key insights into the structure and variability of the reports. The mean word count is approximately 105.66, with a standard deviation of 44.61, indicating a moderate level of variation in report length. The median value (106) is close to the mean, suggesting a symmetrical distribution, though the positive skewness (1.10) and kurtosis (4.82) indicate a longer right tail and a distribution more peaked than normal. The word counts range from a minimum of 18 to a maximum of 368 words, with a total word sum of 50,190 across all narratives. The 95% confidence interval for the mean is ±4.02 words, reinforcing the reliability of the average word count estimate.

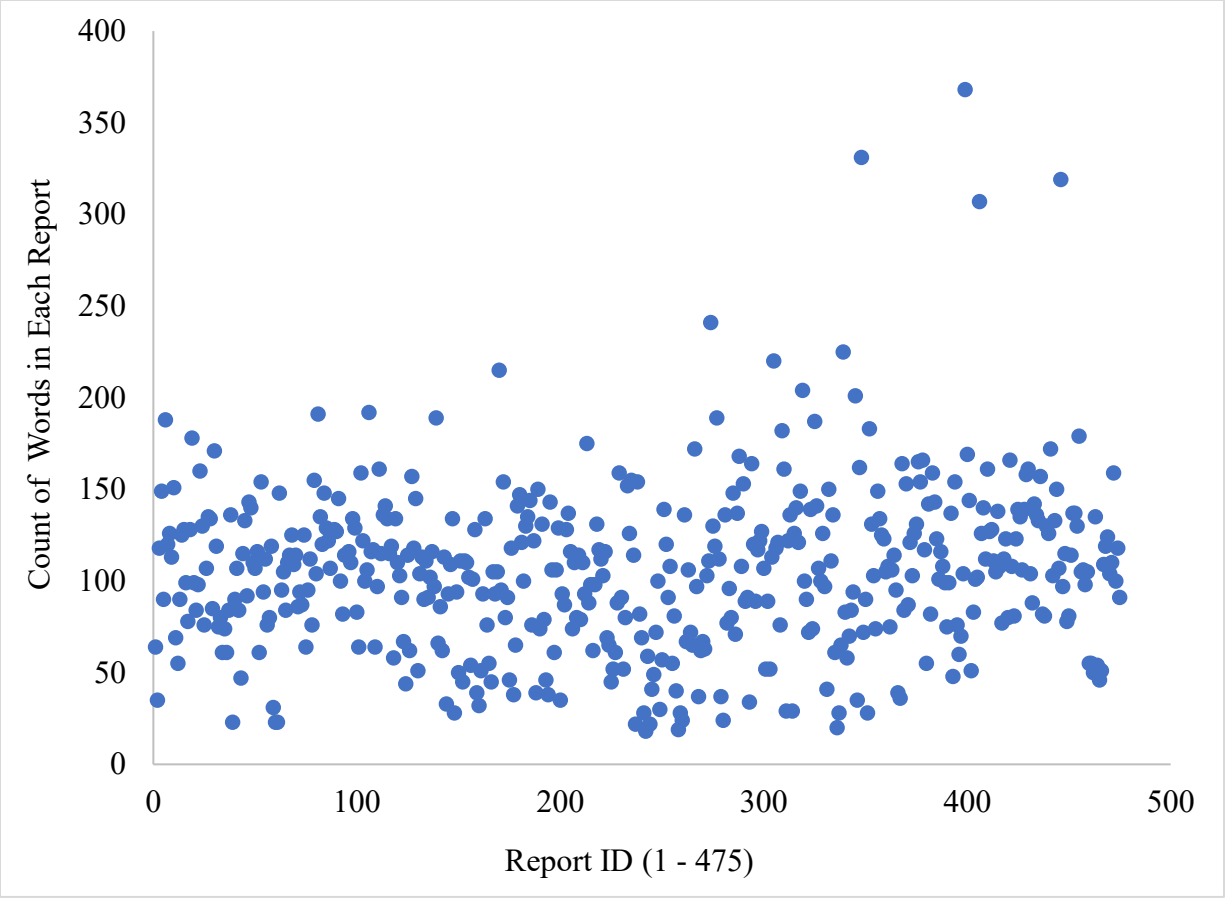


Figure 3. Count of Words for Each Collected Report

In this analysis, a total of 475 crash reports (i.e., documents) were utilized, which consist of a total of 50,190 words (i.e., terms). The Term Document Matrix (TDM) constructed from the 475 unique motorcycle crash narratives contains 4,307 distinct terms, reflecting the rich vocabulary used across the dataset. Each row in the matrix represents a unique term, while each column corresponds to an individual crash report. The matrix entries capture the term frequency (TF), indicating how often each term appears in each document. Out of a total of 2,045,825 possible entries (4,307 terms × 475 documents), only 22,599 are non-zero, resulting in a sparsity level of approximately 99%. This high sparsity is typical in text mining applications and indicates that most terms appear in only a small fraction of documents, which reflects the diversity of language and context in the narratives.

5. Results and Discussions

5.1 Word Cloud Analysis

The quantitative analysis of keywords was conducted using word cloud visualization to highlight prominent terms in the crash narratives. Figure 4 shows the word cloud featuring the 150 most frequently mentioned words in the reported motorcycle-involved crashes from online news sources in Bangladesh. The larger the font size of a word, the more frequently it appears in the dataset. For instance, ‘police’, ‘motorcycle’, ‘upazila’, ‘injured’, ‘dead’, ‘deceased’, ‘hit’, ‘hospital’, ‘station’, ‘died’, ‘killed’, ‘truck’, ‘highway’, ‘accident’, ‘spot’, ‘identified’, and ‘officer-in-charge’ are among the most common words found in the reports. The general findings of this word cloud analysis suggest that motorcycle crashes often lead to injuries or fatalities, involve collisions with trucks on highways or in upazila areas, and typically require

police presence and investigation at the crash spot, with victims being transported to nearby hospitals for treatment and identification.



Figure 4. Word Cloud Analysis

5.2 Word Frequency Analysis

To explore the distribution of key terms within crash narratives, a word frequency bar plot was illustrated (Figure 5). A minimum frequency threshold of 60 (decided by trial and error) was applied to include only the most significant and recurring words in the dataset. The top five most frequent words were ‘police’ (693 times), ‘upazila’ (537 times), ‘motorcycle’ (529 times), ‘hit’ (377 times), and ‘killed’ (363 times). High-frequency words like ‘injured’ (348), ‘killed’ (365), ‘deceased’ (360), ‘died’ (267), spot’ (259) and ‘dead’ (242) suggest that most motorcycle-involved crashes resulted in serious or fatal injuries. Additionally, Words like ‘hospital’ (357), ‘medical’ (178), and ‘morgue’ (84) reflect the post-crash outcomes and emphasize the role of healthcare facilities in the crash aftermath. Terms such as ‘upazila[1]’ (537), ‘sadar[2]’ (151), ‘highway’ (240), and ‘dhaka’ (121), indicating the crash sites are often on highways and near administrative or densely populated areas. Besides ‘motorcycle’ (377), ‘bike’ (152), and ‘motorbike’ (144), the inclusion of terms such as ‘truck’ (349) and ‘bus’ (160) suggest that motorcycle-crashes involving heavy vehicles are frequently reported. Moreover, the frequent occurrence of the term ‘night’ (102 times) suggests that many motorcycle-involved crashes took place during nighttime hours. Finally, the high frequency of terms like ‘police’ (693), ‘station’ (324), and ‘officerincharge’ (176) indicates that motorcycle-involved crashes are often reported through law enforcement narratives in newspapers.

[1] Bangla word. The term ‘*upazila’* indicates a sub-district administrative region in Bangladesh.
[2] Bangla word. Within each district, the main upazila that contains the district headquarters is often referred to as the *‘sadar’* or *‘sadar upazila’.*

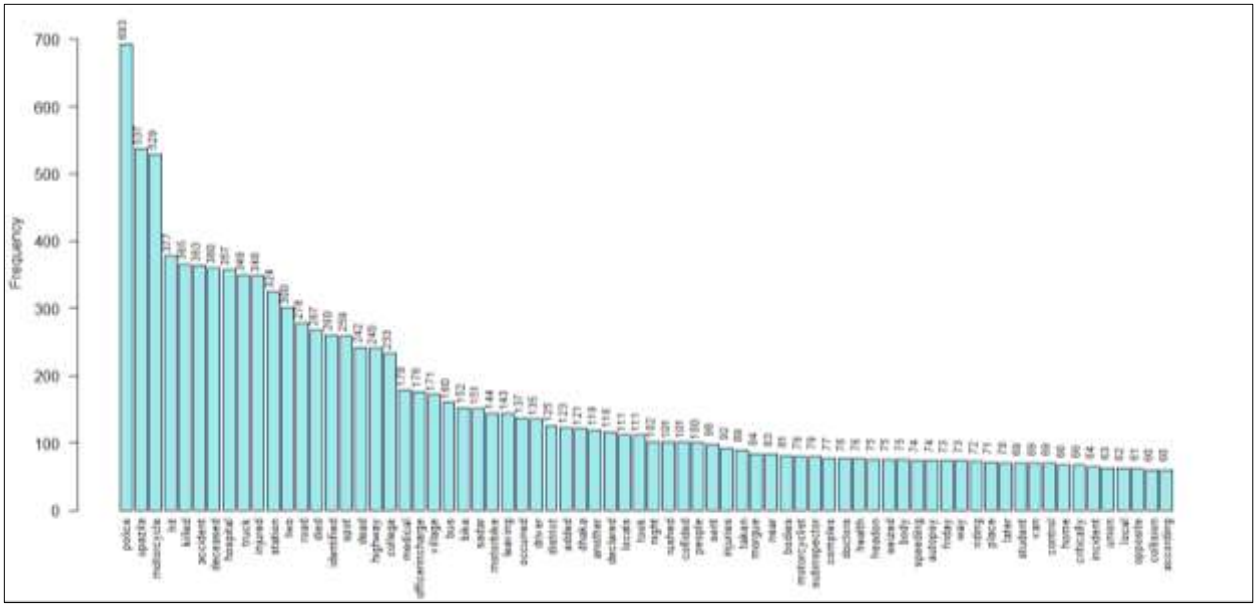


Figure 5. Word frequency bar plot (arranged in ascending order)

5.3 Rapid Automatic Keyword Extraction (RAKE)

The RAKE algorithm identified a total of 10,034 keywords. The distribution of these keywords is as follows: 1-gram (6,016), 2-gram (2,655), 3-gram (885), 4-gram (340), 5-gram (85), 6-gram (35), 7-gram (13), 8-gram (4), and 9-gram (1). In general, 2-gram (consisting of two words) provides the most useful information. The research team reviewed all the 2655 keywords and identified key factors after careful consideration that are most likely to have played a role or are associated with the motorcycle-involved crashes. **Figure 6** shows keywords (selected 2 grams) extracted by the RAKE process. Note that the notation 's' stands for the 'score' of the keywords.

In Bangladesh, most individuals involved in motorcycle-involved crashes were younger riders, particularly students, as suggested by the frequent appearance of terms such as 'college students' (s = 7.3), 'college student' (s = 5.0), 'teenage son' (s = 4.5), and 'young men' (s = 4.0). Several location-based and environmental keywords, such as 'electricity pole' (s = 5.0), 'hilly area' (s = 4.5), and 'roadside tree' (s = 4.0), indicate that fixed roadside objects often played a significant role in motorcycle-involved crashes in Bangladesh. Temporal patterns were also evident, with terms like 'Friday morning' (s = 4.0), 'Thursday night' (s = 4.0), and 'Friday midnight' (s = 4.0) pointing to weekend and nighttime crash occurrences.

Speed-related behaviors appeared prominently in keywords such as 'speeding truck' (s = 4.0), 'speeding bus' (s = 4.0), and 'high speed' (s = 4.0), indicating that excessive speed contributes significantly to motorcycle crash severity. Descriptions like 'rear side' (s = 4.0) and 'motorbike overturned' (s = 4.0) reflect the common collision types observed in such crashes. Notably, terms such as 'opposite direction' (s = 4.0) and 'absconding driver' (s = 4.0) suggest behavioral violations and tendencies toward hit-and-run incidents. Finally, the recurring presence of phrases like 'single motorcycle' (s = 4.0), 'unidentified motorcycle' (s = 4.0), and 'mobile phone' (s = 4.0) implies frequent solo-vehicle involvement and highlights mobile phone use while riding as a potential contributing factor in motorcycle crashes.

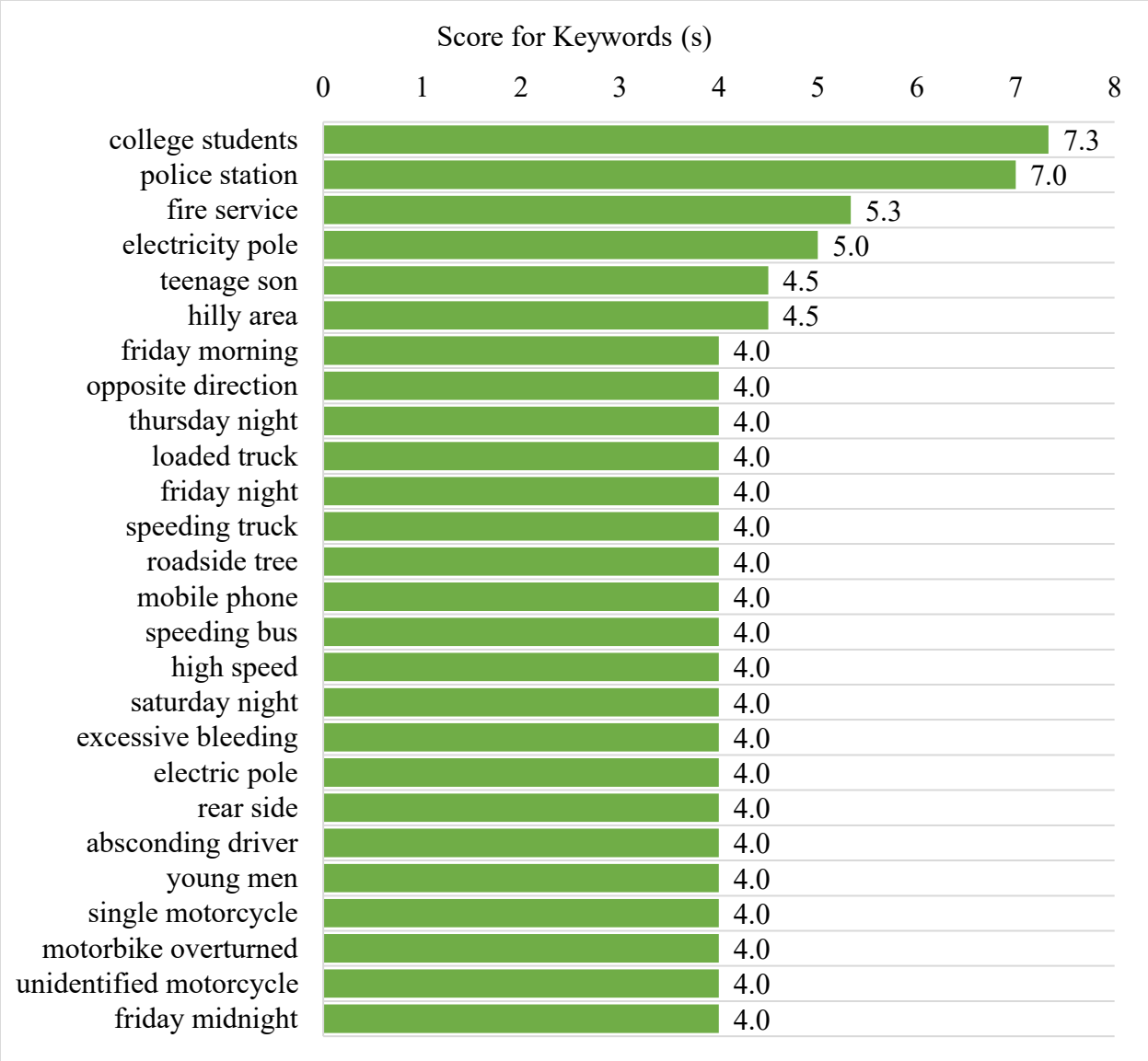


Figure 6. RAKE for selected 2-grams

5.4 Word Co-occurrence Network (WCN) Analysis

WCN is a key tool for visualizing the relationships among words that appear together in a sentence. **Figure 7** shows the WCN plot for motorcycle-involved crashes. The Fruchterman-Reingold ('fr') layout was selected for the WCN analysis, as it offers an effective balance between computational efficiency and interpretability. As a force-directed algorithm, it spatially arranges nodes based on simulated physical forces, drawing frequently co-occurring terms closer together while pushing unrelated terms apart. This results in well-separated clusters that reflect meaningful thematic groupings within the crash narratives. Given the medium-sized nature of the dataset, the 'fr' layout was found to be optimal for identifying patterns related to crash types, profiles, and contributing factors in motorcycle-involved crashes in Bangladesh. Compared to alternative layouts, 'fr' provides a clearer and more intuitive visualization of semantic relationships among words.

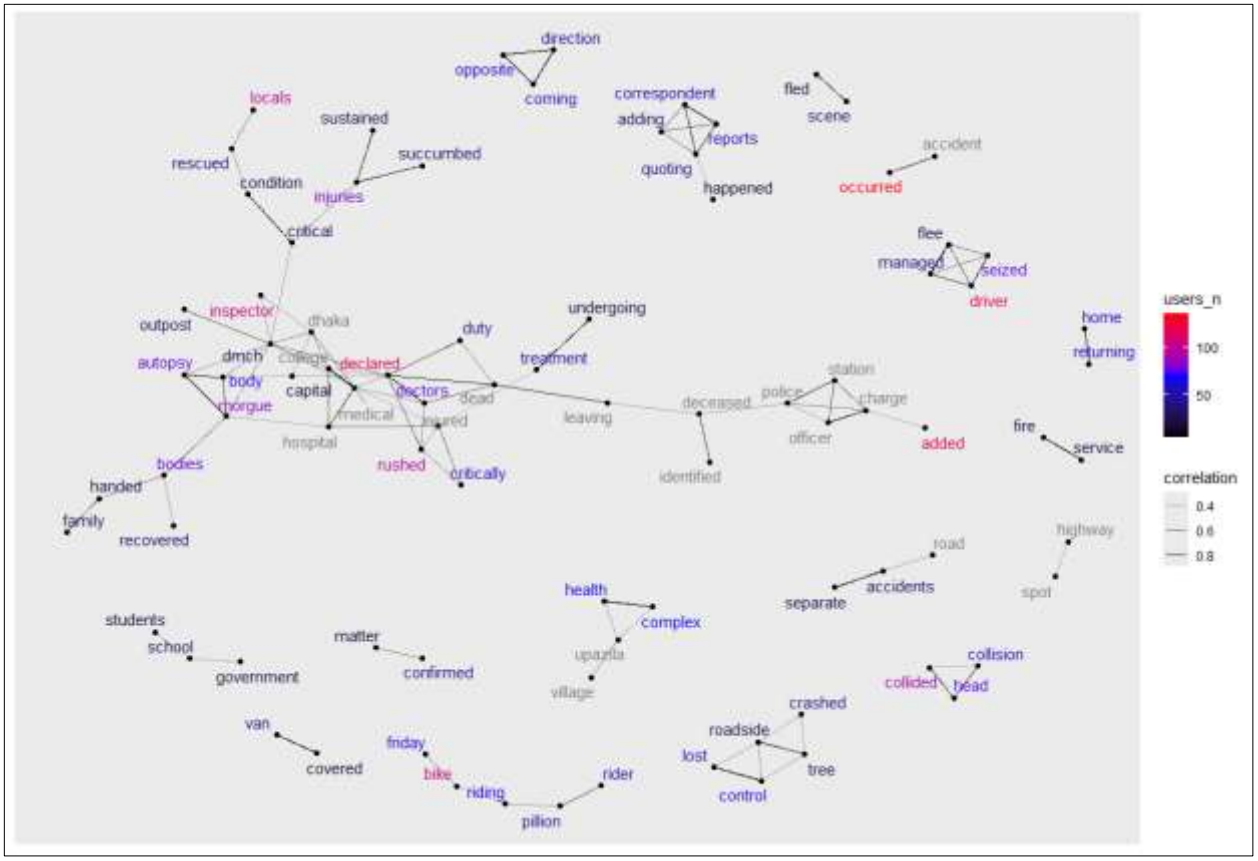


Figure 7. Word Co-Occurrence Network (WCN) plot for truck-involved crashes

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 $n = 20$ and $\varphi = 0.25$ 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 visible link is observed around the word - 'collision', 'head', and 'collided'. This indicates that head-on collisions and direct impacts are common in motorcycle-involved crashes in Bangladesh. Another link formed between 'lost', 'control', 'roadside', 'tree', and 'crashed' suggests that many of these crashes involved riders losing control and hitting fixed roadside objects such as trees.
- The connection between 'doctors', 'injured', 'declared', 'dead', and 'critically' highlights the severe medical outcomes and emergency response following such incidents.
- A cluster can be seen around the word - 'capital', 'college', and 'students', suggesting that many of the crash victims were young riders, particularly students in urban areas like Dhaka.
- The words 'bike', 'riding', 'pillion', and 'Friday' form a connection that implies weekend riding patterns, often involving multiple riders on a single motorcycle.
- A link is also observed between 'fled' and 'scene', suggesting hit-and-run crashes are common. Another cluster formed by 'managed', 'flee', and 'seized' implies cases where drivers either escaped or were caught during flight from the crash location.
- The triad 'coming', 'opposite', and 'direction' reflects a recurring pattern of head-on crashes caused by vehicles approaching from the opposite direction.

- Co-occurrences like ‘reports’, ‘correspondent’, ‘adding’, and ‘quoting’ indicate the journalistic nature of crash documentation in news articles, often referencing statements from police or eyewitnesses.
- The grouping of ‘autopsy’, ‘body’, ‘morgue’, and ‘dmch’[3] indicates detailed post-mortem and hospital reporting for fatal crash cases.
- The cluster with the words - ‘health’, ‘complex’, and ‘upazila’, highlights rural locations where crashes may occur near health facilities.
- A connection is observed between the words ‘returning’ and ‘home’, suggesting that many motorcycle-involved crashes occurred while victims were en route to their residences. This implies that crash events frequently take place during return journeys, possibly after school, work, or evening outings.
- A prominent cluster centers around the word ‘critical’, which connects to ‘condition’, ‘rescued’, and ‘local’, suggesting that many victims were critically injured, rescued by local residents, and reported in serious condition. Additionally, the term ‘injuries’ branches into ‘sustained’ and ‘succumbed’, indicating that victims frequently suffered severe trauma, some surviving with injuries while others succumbed to them, highlighting the grave nature of motorcycle-involved crashes in Bangladesh.

This WCN analysis reveals recurring patterns in motorcycle-involved crash narratives in Bangladesh, emphasizing crash types such as head-on collisions, behavioral factors like fleeing the scene (hit and run), the involvement of young urban riders, and detailed medical and police responses following severe or fatal incidents.

5.5 Structural Topic Modeling Results

The most critical step in STM is the selection of the optimal number of topics. For this purpose, several metrics, including held-out likelihood, residuals, semantic coherence, and lower bound, were evaluated with the number of topics (see **Figure 8** for more details). In addition, semantic coherence and exclusivity metrics were also evaluated (see **Figure 9** for more details).

Among the tested values (K = 5 to 10), the model with K = 7 offers a balanced performance across all evaluation criteria (**refer to Figure 8**). It demonstrates a meaningful improvement in semantic coherence compared to models with fewer topics, indicating enhanced interpretability of topics. At the same time, residuals remain within an acceptable range, and the lower bound shows stable convergence, suggesting no significant trade-offs in model fit or optimization. While models with higher K values (e.g., K = 10) achieve slightly better predictive accuracy and coherence, they do so at the cost of increased residuals and complexity, raising the risk of overfitting.

Furthermore, the trade-off between semantic coherence and exclusivity supports the selection of K = 7 as optimal (**refer to Figure 9**). While coherence improves with more topics, exclusivity a measure of how distinct and non-overlapping the topics are begins to decline slightly. The K = 7 model captures a clear improvement in coherence while maintaining higher exclusivity than models with K = 8 or above, ensuring that the resulting topics are both interpretable and well-differentiated. This balance makes K = 7 the most suitable choice, offering a parsimonious and meaningful topic structure for analyzing motorcycle crash narratives in the dataset.

[3] dmch is the short form of Dhaka Medical College Hospital, located in the capital of Bangladesh

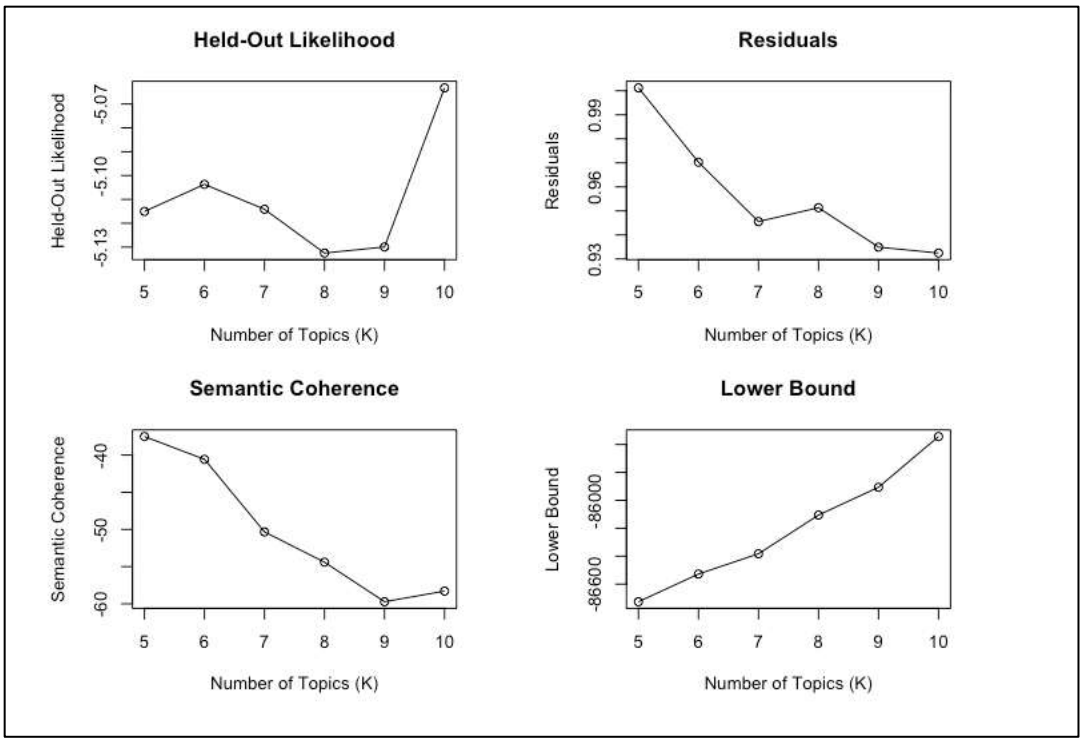


Figure 8. Diagnostic Values by Number of Topics

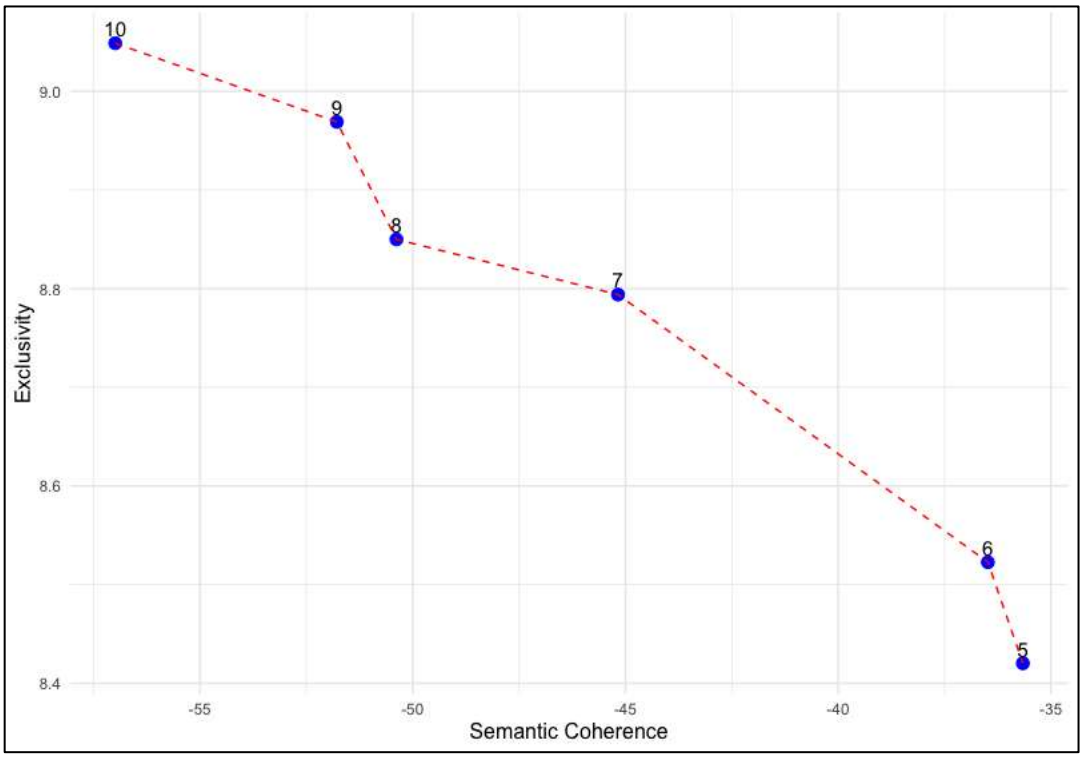


Figure 9. Semantic Coherence and Exclusivity for each Topic

The identified 7 topics with individual word clouds are provided below (Figure 10):

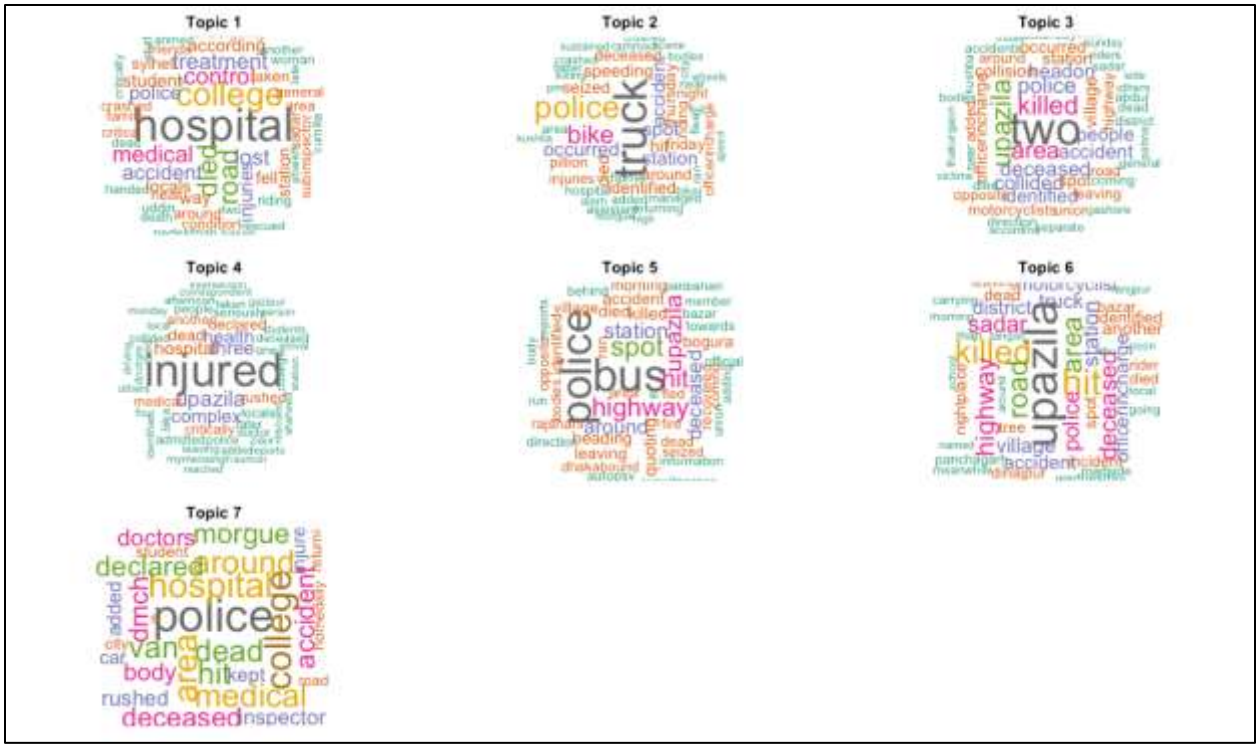


Figure 10. Word Cloud for each Individual Topics (K1 – K7)

Since the number of vehicles involved in a crash, particularly whether it is a single-vehicle or multi-vehicle incident, is a critical contextual factor strongly associated with crash outcomes, we further examined its influence in the context of motorcycle-related crashes. Motorcycle crashes often differ significantly in nature and severity depending on whether the motorcyclist was the sole party involved or collided with another vehicle [21], [59], [60]. To explore how this distinction shapes the narrative themes, we estimated the effect of crash type (single vs. multiple vehicle) on the prevalence of each of the seven identified topics in the STM framework. This analysis, based on metadata from the 475 crash reports, reveals how specific topics are differentially emphasized in news narratives depending on crash type (details in **Table 2**).

Table 2. Results of STM for Crash Type (Single vs. Multple)

Topic	1	2	3	4	5	6	7
Intercept Estimate	0.222	0.088	0.089	0.152	0.039	0.266	0.144
Intercept (S.E.)	0.018	0.017	0.018	0.017	0.017	0.020	0.022
Intercept (t-value)	12.480	5.238	4.967	8.703	2.282	13.310	6.485
Intercept (p-value)	<2e-16	2.45e-07	9.50e-07	<2e-16	0.0229	<2e-16	2.23e-10
Multiple-vehicle (Estimate)	-0.126	0.076	0.087	-0.037	0.105	-0.103	-0.001
Multiple-vehicle (S.E.)	0.020	0.019	0.021	0.020	0.020	0.022	0.025
Multiple-vehicle (t-value)	-6.410	4.022	4.198	-1.914	5.399	-4.597	-0.032
Multiple-vehicle (p-value)	3.53e-10	6.72e-05	3.22e-05	0.0563	1.06e-07	5.5e-06	*0.974*

**Note: Single vehicle is set as a reference in the model*

The topics are discussed under the following sub-heading:

Topic 1: Severe Injuries and Hospitalization
Multi-vehicle crashes show a 12.6% lower association ($p<0.001$) with this topic compared to single-vehicle crashes. The key terms 'hospital,' 'died,' and 'treatment' indicate that this topic represents the most severe crash outcomes. The strong negative coefficient suggests single-vehicle motorcycle crashes (like losing control) more frequently result in critical injuries requiring extensive medical care. This may reflect the vulnerability of motorcyclists in solo crashes where there are no other vehicles to absorb impact forces.

Topic 2: Truck/Motorcycle Collisions
For topic 2, multiple vehicle crashes show a 7.6% greater association ($p<0.001$). The prominence of 'truck,' 'bike,' and 'speeding' clearly shows that this topic captures collisions between motorcycles and large trucks. The positive coefficient confirms that these interactions primarily occur in multiple vehicle scenarios, likely due to size/speed disparities and poor road sharing on Bangladesh's congested highways.

Topic 3: Head-On Collisions
Multi-vehicle crashes are 8.7% more associated ($p<0.001$) with this topic featuring 'head-on,' 'motorcyclists,' and 'collided.' This clearly represents dangerous opposing-direction crashes, particularly between motorcycles and other vehicles. The significant positive effect suggests inadequate lane discipline or passing attempts as key contributors to these often-fatal crashes.

Topic 4: Rural Health Complex Cases
Showing borderline significance (3.7% lower for multi-vehicle, $p = 0.056$), this topic with 'upazila' and 'health complex' terms reflect rural crash outcomes. The negative trend indicates single-vehicle crashes (like motorcycles hitting trees) dominate in rural areas, where victims are taken to local health centers rather than major hospitals.

Topic 5: Bus-Related Crashes
Multi-vehicle crashes show a 10.5% stronger association ($p<0.001$). The terms 'bus,' 'highway,' and 'fled' reveal that this topic captures motorcycle-bus collisions, often with hit-and-run incidents. The large positive coefficient highlights how motorcycles are particularly vulnerable when sharing roads with buses, especially on high-speed routes.

Topic 6: Rural Motorcyclist Fatalities
Multi-vehicle crashes are 10.3% less associated ($p<0.001$) with this 'motorcyclist,' 'tree,' and 'rural' topic. This strong negative effect confirms that single-vehicle crashes (e.g., run-off-road) account for most rural motorcycle fatalities, likely due to poor road conditions or rider error on rural routes.

Topic 7: Urban Death Declarations
No significant difference between crash types ($p = 0.974$). The 'hospital' and 'DMCH' terms show that this topic covers formal death pronouncements in urban hospitals. The null result suggests both crash types contribute equally to urban fatalities requiring official documentation.

Overall, the STM results reveal two distinct crash patterns:

- Single-vehicle crashes are more frequent in rural areas (Topics 4,6), typically involving loss of control or fixed object impacts. These likely stem from poor road conditions, rider errors, or a lack of safety infrastructure.
- Multi-vehicle crashes prevail in urban/highway settings (Topics 2,3,5), characterized by collisions with larger vehicles (trucks/buses) and head-on impacts. These reflect dangerous vehicle interactions and traffic mix issues.

Notably, single-vehicle crashes show stronger links to severe outcomes (Topic 1), possibly because unprotected riders take the full impact. Multi-vehicle crashes more often involve legal consequences (Topics 2,5), suggesting they are more likely to involve traffic violations or hit-and-run incidents.

6. Conclusions

To address the persistent data limitations in official crash reporting systems in Bangladesh, this study adopted an alternative data-driven approach by leveraging unstructured text from online English-language news media. Over a 12-month period spanning January to December 2022, a total of 9,300 news articles were collected from multiple national and regional online sources. Following a rigorous manual screening and filtering process, 475 unique articles (containing a bag of 50,190 words) specifically reporting motorcycle-involved crashes were identified for further analysis. These articles served as the foundation for applying text mining and STM techniques to uncover latent themes and contextual crash factors. This methodological shift toward media-based data mining demonstrates a scalable and context-sensitive approach to investigating motorcycle crash patterns in data-scarce environments like Bangladesh.

The preliminary text mining analyses, including word cloud and frequency distribution, provide a intuitive overview of motorcycle crash reporting in Bangladesh. The narratives are dominated by references to severe or fatal incidents, often involving emergency responders and medical intervention. The frequent mention of administrative regions, particularly upazila (sub-districts) and sadar (district centers), alongside major urban centers like Dhaka, suggests that these crashes commonly occur in central locations or along key transportation routes. The involvement of heavier vehicles (e.g., buses, trucks) alongside motorcycles points to the vulnerability of two-wheeler riders in mixed traffic conditions. This is consistent with previous literature (similar context in Pakistan) which suggest that, motorcycle riders are 2.9 times more likely to be involved in fatal crashes for collision with heavy vehicles mainly due to: a) larger momentum of heavy vehicles, and b) less attention to motorcycle riders [61]. Temporal patterns are also identified, with the word ‘night’ indicating that many motorcycle crashes happen after dark, possibly linked to reduced visibility and higher risk-taking behaviors of riders [62], [63]. The recurrent use of terms related to police stations and officers highlights the dependence on official law enforcement accounts in documenting these events.

Further insights into the contributing factors and contextual dynamics of motorcycle-involved crashes were identified using RAKE and WCN analyses. The RAKE algorithm identified over 10,000 keywords, with bigram phrases offering particularly valuable detail on behavioral, temporal, demographic, and environmental crash dimensions. Notably, keywords such as college students, teenage son, and young men point to a disproportionate involvement of young riders, especially students, in motorcycle crashes in Bangladesh. This essentially points to the role of inexperience [64], risk raking behaviors [65], and unsafe rider actions [66]. Environmental risk factors were also evident, with recurring terms like hilly area, electricity pole, and roadside tree suggesting frequent collisions with roadside fixed objects and specific geographics. Motorcycle crashes with fixed object is also well documented in previous literature and some of the potential factors are – loss of control, speeding, and poor road conditions, and impairment [67], [68], [69], [70]. Temporal patterns emerged through phrases such as Friday morning, Thursday night, and Friday midnight, which indicate a higher crash prevalence during weekends and nighttime hours. Behavioral aspects, mainly excessive speed and reckless driving were identified by phrases like speeding truck, high speed, and opposite direction, while terms such as absconding driver and hit-and-run reflect the frequent occurrence of post-crash behavior. Descriptions like rear side, motorbike overturned, single motorcycle capture common collision types and the frequent involvement of solo riders. Furthermore, mentions of mobile phones suggest distracted driving may be a contributing factor in crashes. This is also consistent with previous research that indicates a higher likelihood of being involved in a collision or fall results from using a cell phone to text or look up information while riding [71].

The WCN visualization offered relational insights by clustering semantically linked terms that frequently co-occurred in crash reports. For instance, the prominent triad of collision, head, and collided points to the high prevalence of head-on impacts, while another cluster linking lost, control, roadside, and tree reinforces the pattern of crashes involving loss of control and impact with fixed objects. Medical

outcomes were identified through associations like doctors, critically, declared, and dead, underscoring the severe and often fatal nature of injuries. Sociodemographic patterns were reflected in the grouping of capital, college, and students, pointing again to young urban riders as a high-risk group. Clusters involving bike, riding, pillion, and Friday suggest a behavioral trend of multi-passenger riding during weekends, while the frequent linkage between fled, scene, and seized indicates that hit-and-run scenarios and attempted escapes are recurring phenomena. Patterns of head-on crashes emerged through the combination of coming, opposite, and direction. Lastly, clusters like autopsy, morgue, DMCH, and health complex highlighted the post-crash handling and documentation of fatalities, particularly in both urban and rural health facilities.

STM analysis provided insights into the distinct thematic patterns underlying motorcycle-involved crashes in Bangladesh. The results revealed two broadly contrasting crash profiles: one dominated by single-vehicle incidents, and the other by multi-vehicle collisions. Single-vehicle crashes were more frequently associated with severe injury outcomes and medical emergencies. These cases often involved riders losing control and colliding with fixed roadside objects, particularly in rural settings, and typically required immediate attention from local health facilities. In contrast, multi-vehicle crashes were more commonly linked to collisions involving larger vehicles, such as trucks and buses, which often occurred on highways or in urban areas. These scenarios reflected complex traffic interactions, speed disparities, and a lack of structured lane discipline - factors that contributed to frequent head-on collisions and a higher incidence of hit-and-run behavior. The contextual factor in hit-and-run crashes in developing countries like Bangladesh is driven by the lack of effective law enforcement, adequate surveillance infrastructure, and public trust in the legal system, which often leads drivers to flee the scene to avoid legal consequences or social retaliation [72], [73]. Moreover, while single-vehicle crashes appeared more closely tied to fatal or critical injury outcomes, multi-vehicle collisions were more often embedded within narratives of legal response and police involvement, underscoring the broader societal and enforcement dimensions of these crashes.

Based on the study's findings, context-specific safety recommendations are proposed through the lens of the Safe System Approach (SSA), which emphasizes proactive and system-wide interventions to reduce crash severity and prevent fatalities. In the context of Bangladesh, where single-vehicle motorcycle crashes are prevalent in rural areas, infrastructure upgrades are urgently needed. These include the installation of roadside barriers, removal of fixed roadside hazards (e.g., trees, poles), improved lighting, and clearer signage on curved and hilly roads. In urban and highway settings, where multi-vehicle collisions involving buses and trucks are more frequent, the introduction of Exclusive MotorCycle Lane (EMCL) [74], stricter enforcement of lane discipline, and better speed management, particularly for heavy vehicles, are critical. Targeted behavioral interventions are also necessary. Public education campaigns tailored to young riders, especially students, should emphasize safe riding practices, helmet use, and the dangers of reckless overtaking and nighttime riding. To reduce crash risk, motorcycles should be equipped with high-visibility features such as daytime running lights and reflective markings. Post-crash care must be improved by strengthening emergency medical services at upazila health complexes and ensuring faster referral systems to district-level hospitals. Finally, the high incidence of hit-and-run cases and police-reported legal consequences in multi-vehicle crashes indicates the need for tighter regulation of commercial vehicles, mandatory driver training, and better mechanisms for crash investigation and legal follow-up.

6.1 Research Contribution

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

6.2 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 interesting 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.

6.3 Acknowledgment
The research team would like to acknowledge the assistance of internet news portals from which the primary database was gathered. The research team would also like to acknowledge Shahrin Islam, Shah Alam, and Mahir Shahrier Simanto for their contribution to data collection.

6.4 Declaration of competing interest
The authors declare no conflict of interest related to this study. This research is based on data from the online news media articles (published in Bangladesh) and is intended solely to advance understanding in the field of motorcycle safety. The analysis and findings presented are those of the authors and do not reflect the views or positions of any affiliated agencies or organizations.

6.5 Funding
The authors disclosed no financial support for the research, authorship, and/or publication of this article.

REFERENCES

[1] "Motorcycle Safety | NHTSA." Accessed: Jul. 31, 2025. [Online]. Available: https://www.nhtsa.gov/road-safety/motorcycles

[2] T. Allen *et al.*, "Contributing factors to motorcycle injury crashes in Victoria, Australia," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 45, pp. 157–168, Feb. 2017, doi: 10.1016/j.trf.2016.11.003.

[3] S.-J. Chen, C.-Y. Chen, and M.-R. Lin, "Risk factors for crash involvement in older motorcycle riders," *Accident Analysis & Prevention*, vol. 111, pp. 109–114, Feb. 2018, doi: 10.1016/j.aap.2017.11.006.

[4] J. Oxley *et al.*, "Commuter motorcycle crashes in Malaysia: An understanding of contributing factors," *Ann Adv Automot Med*, vol. 57, pp. 45–54, Sep. 2013.

[5] M. T. Yousif, A. F. M. Sadullah, and K. A. A. Kassim, "A review of behavioural issues contribution to motorcycle safety," *IATSS Research*, vol. 44, no. 2, pp. 142–154, Jul. 2020, doi: 10.1016/j.iatssr.2019.12.001.

[6] C. Ding, M. Rizzi, J. Strandroth, U. Sander, and N. Lubbe, "Motorcyclist injury risk as a function of real-life crash speed and other contributing factors," *Accident Analysis & Prevention*, vol. 123, pp. 374–386, 2019.

[7] E. R. Teoh and M. Campbell, "Role of motorcycle type in fatal motorcycle crashes," *Journal of Safety Research*, vol. 41, no. 6, pp. 507–512, Dec. 2010, doi: 10.1016/j.jsr.2010.10.005.

[8] W. H. Schneider, P. T. Savolainen, D. Van Boxel, and R. Beverley, "Examination of factors determining fault in two-vehicle motorcycle crashes," *Accident Analysis & Prevention*, vol. 45, pp. 669–676, Mar. 2012, doi: 10.1016/j.aap.2011.09.037.

[9] D. Wu, M. Hours, and J.-L. Martin, "Risk factors for motorcycle loss-of-control crashes," *Traffic Injury Prevention*, vol. 19, no. 4, pp. 433–439, May 2018, doi: 10.1080/15389588.2017.1410145.

[10] A. Moskal, J.-L. Martin, and B. Laumon, "Risk factors for injury accidents among moped and motorcycle riders," *Accident Analysis & Prevention*, vol. 49, pp. 5–11, Nov. 2012, doi: 10.1016/j.aap.2010.08.021.

[11] A. Ariannezhad, H. Razi-Ardakani, and M. Kermanshah, "Exploring factors contributing to crash severity of motorcycles at suburban roads," in *Proc., 93rd Annual Meeting of the Transportation Research Board. Washington, DC: Transportation Research Board*, 2014. Accessed: Jul. 30, 2024. [Online]. Available: https://www.researchgate.net/profile/Hesamoddin-Razi-Ardakani/publication/340105983_Exploring_factors_contributing_to_crash_severity_of_motorcycles_at_Suburban_roads/links/64e79c7840289f7a0fb162a2/Exploring-factors-contributing-to-crash-severity-of-motorcycles-at-Suburban-roads.pdf

[12] "Bangladesh Road Transport Authority - Government of Peoples Republic of Banglaesh." Accessed: Jul. 31, 2025. [Online]. Available: https://brta.gov.bd/

[13] S. Hossain, "Bangladesh counts highest death rates from bike accidents," Prothomalo. Accessed: Jul. 31, 2025. [Online]. Available: https://en.prothomalo.com/bangladesh/accident/bangladesh-counts-highest-death-rates-from-bike-accidents

[14] A. Hossain, X. Sun, S. Alam, S. Das, and A. Sheykhfard, "Crash Contributing Factors and Patterns Associated with Fatal Truck-Involved Crashes in Bangladesh: Findings from the Text Mining Approach," *Transportation Research Record*, vol. 2678, no. 7, pp. 706–725, Jul. 2024, doi: 10.1177/03611981231209031.

[15] L. A. Brockhus, P. Liasidis, M. Lewis, D. A. Jakob, and D. Demetriades, "Injury patterns and outcomes in motorcycle driver crashes in the United States: the effect of helmet use," *Injury*, vol. 55, no. 3, p. 111196, 2024.

[16] M. Pour-Rouholamin, M. Jalayer, and H. Zhou, "Modelling single-vehicle, single-rider motorcycle crash injury severity: an ordinal logistic regression approach," *International Journal of Urban Sciences*, vol. 21, no. 3, pp. 344–363, Sep. 2017, doi: 10.1080/12265934.2017.1311801.

[17] C. Rappole, M. Canham-Chervak, B. Taylor, and B. H. Jones, "Factors associated with motorcycle traffic crash fatalities among active duty U.S. Army personnel," *Traffic Injury Prevention*, vol. 20, no. 2, pp. 174–181, Feb. 2019, doi: 10.1080/15389588.2018.1532082.
[18] J. V. Ouellet and V. Kasantikul, "Motorcycle Helmet Effect on a Per-Crash Basis in Thailand and the United States," *Traffic Injury Prevention*, vol. 7, no. 1, pp. 49–54, Mar. 2006, doi: 10.1080/15389580500338652.
[19] S. Islam and J. Brown, "A comparative injury severity analysis of motorcycle at-fault crashes on rural and urban roadways in Alabama," *Accident Analysis & Prevention*, vol. 108, pp. 163–171, 2017.
[20] S. Jung, Q. Xiao, and Y. Yoon, "Evaluation of motorcycle safety strategies using the severity of injuries," *Accident Analysis & Prevention*, vol. 59, pp. 357–364, Oct. 2013, doi: 10.1016/j.aap.2013.06.030.
[21] M. S. Shaheed and K. Gkritza, "A latent class analysis of single-vehicle motorcycle crash severity outcomes," *Analytic Methods in Accident Research*, vol. 2, pp. 30–38, Apr. 2014, doi: 10.1016/j.amar.2014.03.002.
[22] J. H. Salum, A. E. Kitali, H. Bwire, T. Sando, and P. Alluri, "Severity of motorcycle crashes in Dar es Salaam, Tanzania," *Traffic Injury Prevention*, vol. 20, no. 2, pp. 189–195, Feb. 2019, doi: 10.1080/15389588.2018.1544706.
[23] R. Tamakloe, S. Das, E. Nimako Aidoo, and D. Park, "Factors affecting motorcycle crash casualty severity at signalized and non-signalized intersections in Ghana: Insights from a data mining and binary logit regression approach," *Accident Analysis & Prevention*, vol. 165, p. 106517, Feb. 2022, doi: 10.1016/j.aap.2021.106517.
[24] S. Wang, F. Li, Z. Wang, and J. Wang, "A random parameter bivariate probit model for injury severities of riders and pillion passengers in motorcycle crashes," *Journal of Transportation Safety & Security*, vol. 14, no. 8, pp. 1289–1306, Jul. 2022, doi: 10.1080/19439962.2021.1916666.
[25] M. Islam, "The effect of motorcyclists' age on injury severities in single-motorcycle crashes with unobserved heterogeneity," *Journal of Safety Research*, vol. 77, pp. 125–138, Jun. 2021, doi: 10.1016/j.jsr.2021.02.010.
[26] M. Abrari Vajari, K. Aghabayk, M. Sadeghian, and N. Shiwakoti, "A multinomial logit model of motorcycle crash severity at Australian intersections," *Journal of Safety Research*, vol. 73, pp. 17–24, Jun. 2020, doi: 10.1016/j.jsr.2020.02.008.
[27] M. M. A. Manan, A. Várhelyi, A. K. Çelik, and H. H. Hashim, "Road characteristics and environment factors associated with motorcycle fatal crashes in Malaysia," *IATSS research*, vol. 42, no. 4, pp. 207–220, 2018.
[28] L. Wahab and H. Jiang, "Severity prediction of motorcycle crashes with machine learning methods," *International Journal of Crashworthiness*, vol. 25, no. 5, pp. 485–492, Sep. 2020, doi: 10.1080/13588265.2019.1616885.
[29] J. Li, S. Fang, J. Guo, T. Fu, and M. Qiu, "A Motorcyclist-Injury Severity Analysis: A Comparison of Single-, Two-, and Multi-Vehicle Crashes Using Latent Class Ordered Probit Model," *Accident Analysis & Prevention*, vol. 151, p. 105953, Mar. 2021, doi: 10.1016/j.aap.2020.105953.
[30] A. Farid and K. Ksaibati, "Modeling severities of motorcycle crashes using random parameters," *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 8, no. 2, pp. 225–236, Apr. 2021, doi: 10.1016/j.jtte.2020.01.001.
[31] S. Wang, J. Yu, and J. Ma, "Identifying the heterogeneous effects of road characteristics on Motorcycle-Involved crash severities," *Travel Behaviour and Society*, vol. 33, p. 100636, Oct. 2023, doi: 10.1016/j.tbs.2023.100636.
[32] Md. H. Rahman, N. M. Zafri, T. Akter, and S. Pervaz, "Identification of factors influencing severity of motorcycle crashes in Dhaka, Bangladesh using binary logistic regression model," *International Journal of Injury Control and Safety Promotion*, vol. 28, no. 2, pp. 141–152, Apr. 2021, doi: 10.1080/17457300.2021.1878230.

[33] S. Pervaz, M. Rahman, S. Hasanat-E-Rabbi, I. Uddin, and M. Rahman, "A review of motorcycle safety situation in Bangladesh," in *Proceedings of the 5th International conference on civil engineering for sustainable development (ICCESD 2020), KUET, Khulna, Bangladesh*, 2020, pp. 7–9.
[34] T. Akter and S. Pervaz, "Assessing motorcycle accident and injury characteristics in Dhaka Metropolitan City," in *International Conference on Urban and Regional Planning*, 2019, pp. 5–6. Accessed: Feb. 23, 2024. [Online]. Available: https://www.researchgate.net/profile/Shahrior-Pervaz/publication/337563331_ASSESSING_MOTORCYCLE_ACCIDENT_AND_INJURY_CHARACTERISTICS_IN_DHAKA_METROPOLITAN_CITY/links/5de2119992851c8364549c30/ASSESSING-MOTORCYCLE-ACCIDENT-AND-INJURY-CHARACTERISTICS-IN-DHAKA-METROPOLITAN-CITY.pdf
[35] S. Hossain, E. Maggi, and A. Vezzulli, "Factors associated with crash severity on Bangladesh roadways: empirical evidence from Dhaka city," *International Journal of Injury Control and Safety Promotion*, vol. 29, no. 3, pp. 300–311, Jul. 2022, doi: 10.1080/17457300.2022.2029908.
[36] S. K. Das *et al.*, "Exploring the knowledge and practices on road safety measures among motorbikers in Dhaka, Bangladesh: a cross-sectional study," *Injury prevention*, 2023, Accessed: Feb. 24, 2024. [Online]. Available: https://injuryprevention.bmj.com/content/early/2023/11/28/ip-2023-045071.abstract
[37] S. Pervaz, S. M. S. Mahmud, M. A. Raihan, and M. I. Uddin, "Road Crash in Bangladesh: Where We Were, Where We Are, and Where We Will Be," in *Advances in Civil Engineering*, vol. 184, S. Arthur, M. Saitoh, and S. K. Pal, Eds., in Lecture Notes in Civil Engineering, vol. 184. , Singapore: Springer Singapore, 2022, pp. 301–312. doi: 10.1007/978-981-16-5547-0_29.
[38] S. Das and S. Sarkar, "News Media Mining to Explore Speed-Crash-Traffic Association During COVID-19," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2678, no. 12, pp. 349–366, Dec. 2024, doi: 10.1177/03611981221121261.
[39] H. Yang, 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, p. 105608, 2020.
[40] T. Goddard, K. Ralph, C. G. Thigpen, and E. Iacobucci, "Does news coverage of traffic crashes affect perceived blame and preferred solutions? Evidence from an experiment," *Transportation research interdisciplinary perspectives*, vol. 3, p. 100073, 2019.
[41] I. K. Yankson *et al.*, "Reporting on road traffic injury: content analysis of injuries and prevention opportunities in Ghanaian newspapers," *Injury prevention*, vol. 16, no. 3, pp. 194–197, 2010.
[42] S. Das, "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, pp. 960–971, Oct. 2021, doi: 10.1177/03611981211014200.
[43] H. Bhuiyan *et al.*, "Crash severity analysis and risk factors identification based on an alternate data source: a case study of developing country," *Scientific reports*, vol. 12, no. 1, p. 21243, 2022.
[44] P. Li, S. Chen, L. Yue, Y. Xu, and D. A. Noyce, "Analyzing relationships between latent topics in autonomous vehicle crash narratives and crash severity using natural language processing techniques and explainable XGBoost," *Accident Analysis & Prevention*, vol. 203, p. 107605, Aug. 2024, doi: 10.1016/j.aap.2024.107605.
[45] K. M. Kwayu, V. Kwigizile, K. Lee, and J.-S. Oh, "Discovering latent themes in traffic fatal crash narratives using text mining analytics and network topology," *Accident Analysis & Prevention*, vol. 150, p. 105899, Feb. 2021, doi: 10.1016/j.aap.2020.105899.
[46] K. M. Kwayu, V. Kwigizile, K. Lee, J.-S. Oh, and T. Nelson, "Automatic topics extraction from crowdsourced cyclists near-miss and collision reports using text mining and Artificial Neural Networks," *International Journal of Transportation Science and Technology*, vol. 11, no. 4, pp. 767–779, Dec. 2022, doi: 10.1016/j.ijtst.2021.10.005.

[47] P. Xu, Q. Wang, Y. Ye, S. C. Wong, and H. Zhou, "Text as data: Narrative mining of non-collision injury incidents on public buses by structural topic modeling," *Travel Behaviour and Society*, vol. 39, p. 100981, Apr. 2025, doi: 10.1016/j.tbs.2024.100981.
[48] S. Das, A. Dutta, and I. Tsapakis, "Topic Models from Crash Narrative Reports of Motorcycle Crash Causation Study," *Transportation Research Record*, vol. 2675, no. 9, pp. 449–462, Sep. 2021, doi: 10.1177/03611981211002523.
[49] K. D. Kuhn, "Using structural topic modeling to identify latent topics and trends in aviation incident reports," *Transportation Research Part C: Emerging Technologies*, vol. 87, pp. 105–122, Feb. 2018, doi: 10.1016/j.trc.2017.12.018.
[50] Y. Xing *et al.*, "Discovering latent themes in aviation safety reports using text mining and network analytics," *International Journal of Transportation Science and Technology*, vol. 16, pp. 292–316, Dec. 2024, doi: 10.1016/j.ijtst.2024.02.009.
[51] W.- Hong, G. Clifton, and J. D. Nelson, "A conceptual framework for the automated analysis of railway accident reports," presented at the Australasian Transport Research Forum (ATRF), 44th, 2023, Perth, Western Australia, Australia, Nov. 2023. Accessed: Jul. 21, 2025. [Online]. Available: https://trid.trb.org/View/2367010
[52] G. Ertek and L. Kailas, "Analyzing a Decade of Wind Turbine Accident News with Topic Modeling," *Sustainability*, vol. 13, no. 22, Art. no. 22, Jan. 2021, doi: 10.3390/su132212757.
[53] I. Feinerer, K. Hornik, and M. I. Feinerer, "Package 'tm,'" *Corpus*, vol. 10, no. 1, 2015, Accessed: Aug. 01, 2025. [Online]. Available: https://mirror.linux.duke.edu/cran/web/packages/tm/tm.pdf
[54] K. Benoit *et al.*, "quanteda: An R package for the quantitative analysis of textual data," *Journal of Open Source Software*, vol. 3, no. 30, pp. 774–774, 2018.
[55] C. Fay, "Text mining with R: a tidy approach," *Journal of Statistical Software*, vol. 83, pp. 1–3, 2018.
[56] H. Wickham *et al.*, "Welcome to the Tidyverse," *JOSS*, vol. 4, no. 43, p. 1686, Nov. 2019, doi: 10.21105/joss.01686.
[57] H. Wickham and M. H. Wickham, "Package 'stringr,'" *Website: http://stringr. tidyverse. org, https://github. com/tidyverse/stringr*, 2019, Accessed: Aug. 01, 2025. [Online]. Available: https://mirror.linux.duke.edu/cran/web/packages/stringr/stringr.pdf
[58] M. E. Roberts, B. M. Stewart, and D. Tingley, "Stm: An R package for structural topic models," *Journal of statistical software*, vol. 91, pp. 1–40, 2019.
[59] M.-H. Wang, "Investigating the Difference in Factors Contributing to the Likelihood of Motorcyclist Fatalities in Single Motorcycle and Multiple Vehicle Crashes," *International Journal of Environmental Research and Public Health*, vol. 19, no. 14, p. 8411, Jan. 2022, doi: 10.3390/ijerph19148411.
[60] M. M. Haque, H. C. Chin, and A. K. Debnath, "An investigation on multi-vehicle motorcycle crashes using log-linear models," *Safety science*, vol. 50, no. 2, pp. 352–362, 2012.
[61] A. Pervez, J. Lee, and H. Huang, "Identifying Factors Contributing to the Motorcycle Crash Severity in Pakistan," *Journal of Advanced Transportation*, vol. 2021, pp. 1–10, Feb. 2021, doi: 10.1155/2021/6636130.
[62] M. S. B. Shaheed, W. Zhang, K. Gkritza, and Z. Hans, "Differences in motorcycle conspicuity-related factors and motorcycle crash severities in daylight and dark conditions," in *3rd International Conference on Road Safety and Simulation. Indianapolis*, 2011, pp. 1–22. Accessed: Oct. 16, 2025. [Online]. Available: https://smarter-usa.org/wp-content/uploads/2019/07/Differences-in-Motorcycle-Conspicuity-Related-facotrs-and-Motorcycle-crash-Severity-in-Daylight-and-Dark-Conditions.pdf
[63] C. J. Robbins and S. Fotios, "Motorcycle safety after-dark: The factors associated with greater risk of road-traffic collisions," *Accident Analysis & Prevention*, vol. 146, p. 105731, 2020.
[64] A. H. Goodwin, Y. C. Wang, R. D. Foss, and B. Kirley, "The role of inexperience in motorcycle crashes among novice and returning motorcycle riders," *Journal of safety research*, vol. 82, pp. 371–375, 2022.

[65] K. Sumit, V. Ross, K. Brijs, G. Wets, and R. A. C. Ruiter, "Risky motorcycle riding behaviour among young riders in Manipal, India," *BMC Public Health*, vol. 21, no. 1, p. 1954, Dec. 2021, doi: 10.1186/s12889-021-11899-y.
[66] S. Dubois, D. Antoniazzi, R. Klein, and M. Bédard, "Age and engine displacement are associated with unsafe motorcycle rider actions," *Journal of Safety Research*, vol. 74, pp. 119–124, Sep. 2020, doi: 10.1016/j.jsr.2020.04.004.
[67] S. H. Tung, S. V. Wong, T. H. Law, and R. S. Radin Umar, "Crashes with roadside objects along motorcycle lanes in Malaysia," *International Journal of Crashworthiness*, vol. 13, no. 2, pp. 205–210, Mar. 2008, doi: 10.1080/13588260701788534.
[68] S. Das, R. Dzinyela, J. Liu, B. Dadashova, and C. Silvestri-Dobrovolny, "Understanding patterns of factor influences in motorcycle crashes with fixed objects," *Journal of Transportation Safety & Security*, vol. 17, no. 5, pp. 483–509, May 2025, doi: 10.1080/19439962.2024.2429077.
[69] M. R. Bambach, R. H. Grzebieta, J. Olivier, and A. S. McIntosh, "Fatality Risk for Motorcyclists in Fixed Object Collisions," *Journal of Transportation Safety & Security*, vol. 3, no. 3, pp. 222–235, Sep. 2011, doi: 10.1080/19439962.2011.587940.
[70] A. Daniello and H. C. Gabler, "Fatality risk in motorcycle collisions with roadside objects in the United States," *Accident Analysis & Prevention*, vol. 43, no. 3, pp. 1167–1170, 2011.
[71] L. T. Truong, H. T. T. Nguyen, and C. De Gruyter, "Mobile phone use while riding a motorcycle and crashes among university students," *Traffic Injury Prevention*, vol. 20, no. 2, pp. 204–210, Feb. 2019, doi: 10.1080/15389588.2018.1546048.
[72] A. A. Chowdhury and H. K. Fahim, "A Legal Approach to Prevent Road Accident in Bangladesh: Drawbacks and Recommendations," *JL Pol'y & Globalization*, vol. 81, p. 61, 2019.
[73] M. Ahmad, F. N. Rahman, M. Ali, and M. A. S. Al Azad, "Road Traffic Injury among Pedestrians in Dhaka City," *Journal of Armed Forces Medical College, Bangladesh*, vol. 12, no. 1, pp. 3–7, 2016.
[74] H. K. Saini, S. S. Chouhan, and A. Kathuria, "Exclusive motorcycle lanes: A systematic review," *IATSS Research*, vol. 46, no. 3, pp. 411–426, Oct. 2022, doi: 10.1016/j.iatssr.2022.05.004.