

Leveraging autonomous vehicles crash narratives to understand the patterns of parking-related crashes

Boniphace Kutela^{1*} , Ricard Dzinyela² , Henrick Haule³ ,
Abbas Sheykhfard⁴ , Kelvin Msechu⁵ 

¹Texas A&M Transportation Institute, the United States of America

²Texas A&M University, the United States of America

³University of Arizona, the United States of America

⁴Babol Noshirvani University of Technology, Islamic Republic of Iran

⁵Atkins North America Inc., the United States of America

*Corresponding author: b-kutela@tti.tamu.edu

Handling editor: **Lai Zheng**, Harbin Institute of Technology, China

Reviewers: **Chuanyun Fu**, University of British Columbia, Canada
Petr Pokorny, Institute of Transport Economics, Norway

Received: 20 January 2023; Accepted: 20 June 2023; Published: 5 July 2023

Abstract: Autonomous vehicles (AVs) parking has been a subject of interest from various researchers; however, the focus has been on the parking demand, algorithm, and policies, while the safety aspect has received less attention, perhaps due to the lack of AV crash data. This study evaluated the magnitude and pattern of AV parking-related crashes that occurred between January 2017 and August 2022 in California. The study applied descriptive analysis, unsupervised text mining, and supervised text mining (Support Vector Machine, Naïve Bayes, Logitboost, Random Forest, and Neural network) with resampling techniques. It was indicated that parking-related crashes constitute about 16% of all AV crashes, most of them are likely to impact the AV on the rear or left side. The unsupervised text mining results showed that AVs in the conventional mode of operation, reversing, and parallel parking are among the key themes associated with parking-related crashes. The Support Vector Machine, Logitboost, Random Forest, and Neural network showed relatively high prediction accuracy. The important features from these supervised text mining approaches were conventional mode, reversing, passenger vehicle, parallel parking, which confirm the preliminary findings in the unsupervised text mining. The implications of the findings to operators and policymakers are included in the study. Findings from this paper could be used to introduce measures to reduce AV parking-related crashes.

Keywords: autonomous vehicles, crash narratives, machine learning, parking-related crashes, text network

1 Introduction

Autonomous vehicles (AVs) are becoming a reality, with various levels of automation being tested and operating in mixed traffic. Many cities and jurisdictions across the globe have approved AV testing. For instance, in America alone, there are over a hundred AV testing locations, including testing in mixed traffic in California, Arizona, New York, and Nevada, among others (NHSTA, 2022). In addition to the testing, individuals operate AVs of varying levels of automation, as indicated by the most recent statistics of AV penetration in the United States and globally (Statistica, 2022).

As the AVs penetration continues to increase, researchers have shown interest in various aspects, including their safety and operations (Khattak et al., 2020; Mahdavian et al., 2019; Morando et al., 2018; Novat et al., 2023; Parsa et al., 2021; Schoettle & Sivak, 2018; Favarò et al., 2018). Several approaches have been used to assess the safety implications of AVs, including traffic simulation, survey questionnaires, virtual reality, observational studies and analyses of disengagement reports, and AV-involved crash data. Some studies reported that AVs improve road safety by eliminating human errors associated with conventional vehicles (Mahdavian et al., 2019; Morando et al., 2018; Parsa et al., 2021). Other studies revealed that AVs have higher crash rates per million miles than conventional vehicles (Schoettle & Sivak, 2018; Favarò et al., 2018; Pokorný & Høye, 2022). Further analysis showed that AVs were mostly not at-fault, and the severity of crashes involving AVs was lower than the severity of crashes involving conventional vehicles (Mousavi et al., 2020; Schoettle & Sivak, 2018). Even when they were at fault, they were more likely to be in conventional mode (Pokorný & Høye, 2022; Kutela et al., 2022b). In recent years, the availability of the crash data involving AVs in California has opened a door for researchers to explore crash characteristics and associated factors (Boggs et al., 2020; Xu et al., 2019; Kutela et al., 2022a).

NHSTA (2022), Bahrami & Roorda (2022), Lee et al. (2009), Chan et al. (2021), Nakrani & Joshi (2022), and Hsieh & Özgüner (2008) discussed that although there is a significant interest in improving parking-related AV operations and safety, little is known about the scale of parking-related crashes, perhaps, due to the scarcity of actual crash data. However, the availability of crash data involving AVs in California opened

doors for numerous research DMV (n.d). Therefore, this study explores the scale and patterns of parking-related AV crashes. The study intends to answer the following questions: (a) What is the scale of parking-related AV crashes? (b) Do the characteristics of parking-related AV crashes differ from other crashes? (c) What recommendations could be provided from the critical patterns in parking-related AV crashes? Although parking-related crashes are likely to be property damage only, damaged AV features such as sensors may require substantial amount of funds to replace them. Thus, it is important to understand the associated factors in order to reduce the likelihood of such crashes to happen. The remaining part of the manuscript is organized as follows. The next section discusses the literature summary of the studies on the safety effects of AVs, followed by the details on the methodology and data used in the analysis. Results and discussions are then presented followed by the key takeaways and directions for future research.

2 Literature review

Various researchers have been interested in AV parking from different perspectives (Bahrami & Roorda, 2022; Chan et al., 2021; Lee et al., 2009; Nakrani & Joshi, 2022). A policy-based study, Bahrami & Roorda (2022), indicated that AV users would redirect their vehicles to park at their homes if they had to pay to use a parking facility. The same study revealed that AVs would travel between 12 and 47 minutes in the downtown setting to find cheaper parking. Lee et al. (2009) developed algorithms to solve three parallel-parking issues: trajectory planner, decision kernel, and trajectory tracking control. The algorithms could determine feasible parking locations, alternative parking routes, and re-performing parking routing if the targeted parking space has been occupied. Such operations would save not only time but also improve safety. The algorithms developed by Nakrani & Joshi (2022) would improve safety by avoiding static and moving obstacles. The developed model was simulated, and the results indicated a reasonable performance for parallel parking.

Although AV parking has been of interest, little has been studied using actual crash data involving AVs. A scan of studies that used real AVs crash data from California suggests mixed findings. However, only two studies, Ren et al. (2022) and Xu et al. (2019), have some details regarding parking-related crashes.

Almost all studies reported a low proportion of injury crashes for AV-involved crashes (Das et al., 2020; Song et al., 2021). Several studies have extensively studied collisions between AVs and conventional vehicles. Factors like vehicle/personal attributes (driving mode and non-motorized road users), road attributes like parking and lighting, and crash attributes like the manner of collision were considered in previous studies involving AVs. Results from these studies have been summarized in Table 1.

According to Ren et al. (2022), on-street parking is associated with an increased likelihood of severe AV crashes. However, the observation does not indicate that the AV was involved in parking-related crashes. Similarly, Xu et al. (2019) showed that 40% of AV crashes occurred at locations with roadside parking. The authors elaborated that such a high proportion may be due to a limited sight distance. However, this study also did not indicate whether the said proportion of crashes involved parking-related activities.

The literature review shows an attempt to understand various characteristics of AV-involved collisions, including a few studies on parking-related collisions. However, two significant limitations can be observed in the existing studies with regard to parking-related crashes: (i) relatively few studies focused on parking-related crashes applied descriptive analysis and did not evaluate the associated factors; (ii) studies that evaluated the associated factors for AV-involved crashes combined parking and non-parking related crashes. The combined analysis did not show the characteristics of parking-related crashes, making it hard to determine the appropriate measures to reduce such collisions and improve AV safety in parking locations. Therefore, this study investigated factors associated with parking-related crashes involving AVs using supervised and unsupervised text mining.

3 Methodology

This study used crash narratives to uncover the key patterns among factors associated with parking-related crashes. It applied supervised and unsupervised text mining of crash narratives to understand and distinguish the critical patterns in parking and non-parking-related crashes. Unsupervised text mining involves the use of text network analysis. In contrast, supervised text mining used four classifiers, including Support Vector Machine (SVM), Naïve Bayes (NB), Random Forest (RF), and Neural Network (Nnet).

3.1 Text network analysis

Text network analysis (TNA), an unsupervised text mining tool, was utilized to express the interconnectedness between various keywords. Before its use in transportation, the TNA approach had been used extensively in literature and linguistics Hunter (2014). TNA has emerged as the preferred text mining approach in solving modern problems due to its ability to reveal and enable visualizations of different topic patterns (Jiang et al., 2020; Paranyushkin, 2012). Other areas where text analysis has been used include traffic safety analysis and operations (Boggs et al., 2020; Kutela et al., 2023; Lee et al., 2023; Kutela et al., 2022a) and bibliometrics of transportation studies (Jiang et al., 2020; Kutela et al., 2023, 2021b). In TNA, nodes representing keywords and edges representing connections are used to map unstructured text data. Figure 1 shows a typical text network topology.



Figure 1 An illustration of text network topology (Kutela et al., 2021b)

The TNA process starts with data cleaning and converting unstructured to structured data. This process includes removing redundant words and symbols peculiar to the study design and converting upper-case to lower-case letters. The unstructured data is further converted into a matrix of keywords and the frequency of occurrence included in the text data. The resulting text data, which is structured, is mapped to create a text network.

In the network (Figure 1) the node's size represents the frequency of the keyword, while the size of the edge represents the frequency of the co-occurrence of keywords (Kutela et al., 2021a). Keywords that appear close in the narratives also appear close in the network. Therefore, to better understand these text network patterns, measures such as 'keyword frequencies', 'co-

Table 1 Summary of studies on the safety implications of AVs

Factors	Category	Publication	Summary of results
Vehicle or personal attributes	Driving mode	Ashraf et al. (2021) Boggs et al. (2020) Kutela et al. (2021b)	Most common AV crashes occurred when autonomous mode is engaged. Higher likelihood of AV rear-end crash when autonomous driving system is engaged. Crashes involving non-motorized users indirectly are more likely to occur when AVs are in autonomous mode.
		Xu et al. (2019)	Autonomous driving mode reduced the risk of non-Property Damage Only (PDO) crashes.
Road attributes	Non-motorized road users	Kutela et al. (2021b)	Bicyclists and scooterists were more likely to be involved in AV crashes directly. Pedestrians were more likely to be involved in VA crashes indirectly.
	Intersections and roundabouts	Ashraf et al. (2021) Xu et al. (2019)	Most common AV crashes occurred at an intersection when AVs are stopped. Connected and autonomous vehicle (CAV) involved crashes at intersections were more likely to be PDOs.
		Morando et al. (2018)	AVs reduced the number of conflicts at intersections and roundabouts at specific market penetration rates.
	Parking	Xu et al. (2019) Gao et al. (2021)	Roadside parking increased the severity of CAV-involved crashes. On-street parking increased the severity of AV-involved rear-end crashes.
	Lighting	Das et al. (2020)	Dark lighting conditions with streetlights were associated with a higher proportion of injuries.
Land use		Gao et al. (2021) Boggs et al. (2020)	Mixed land use increased the severity of crashes involving AVs. AV-involved rear-end crash more likely in mixed land use. AV-involved rear-end crash less likely near schools.
	Built environment	Gao et al. (2021)	The presence of trees, were associated with a lower likelihood of severe AV-involved rear-end crashes.
Crash attributes	Types of collision	Ashraf et al. (2021)	The most common AV crashes were rear-end crashes.
		Liu et al. (2021) Xu et al. (2019)	The most frequent AV pre-crash scenarios were rear-end and lane change collisions. CAV-involved crashes were more severe than other crash types. CAV-involved crashes on one-way roads were more likely to be PDOs.
		Das et al. (2020)	Autonomous mode crashes were more likely to be rear-end. Turning, multi-vehicle, rear-end, and sideswipe collisions were associated with a higher proportion of injuries.
		Song et al. (2021) Chen et al. (2021)	The most common sequence involved a collision following AV stop. Most rear-end crashes involving conventional vehicles bumping into the rear of AVs.
		Mousavi et al. (2021)	An increase in the market penetration of AVs reduced rear-end and lane-changing conflicts.

occurrence frequency’, and ‘collocation frequency’ are used. The collocation frequency provides richer information than co-occurrence frequency as it presents the keywords that appear next to each other (Benoit et al., 2018; Blaheta & Johnson, 2011; Novat et al., 2023; Kutela et al., 2022b). In addition, the network topology was used to distinguish the patterns of the parking- and non-parking-related crashes.

3.2 Text classifiers

Five supervised machine learning algorithms, including SVM, NB, RF, Logitboost, and Nnet were used to classify parking-related crashes. The ability of these machine learning algorithms to classify parking AV-involved crashes was compared based on Accuracy, Precision, and F-1 score, described in Equation (1), Equation (2) and Equation (3) respectively:

$$Accuracy = \frac{TP + TN}{Total\ population}, \quad (1)$$

$$Precision = \frac{TP}{TP + FN}, \quad (2)$$

$$F - 1\ score = \frac{2 * Recall * Precision}{Recall + Precision}, \quad (3)$$

whereby

- TP: True Positive, actual positive values correctly classified as positive;
- TN: True Negative, actual negative values correctly classified as negative;
- FP: False Positive, actual negative values incorrectly classified as positive;
- FN: False Negative, actual positive values incorrectly classified as negative.

All the classifiers mentioned above use a different approach to classify the crash. Joachims (1998); Pranckevičius & Marcinkevičius (2017) and Yuan et al. (2019) provide detailed information about each classifier. Below is a brief description of each classifier used in this study:

- SVM is a classifier that uses a hyperplane to separate the classes of documents/data by maximizing the

margin between classes’ closest points. The algorithm plots all data items as points in n-dimensional space and then finds the hyperplane that best differentiates the two classes of the document/data.

- NB is a family of probabilistic algorithms based on Bayes theorem and is considered a simple classifier. The primary assumption for NB is that the presence of one feature does not affect other features. By utilizing NB one can obtain the probability of observing certain outcomes, given that specific predictors were observed.
- RF classifier works by establishing and aggregating predictions from several individual decision trees of varying depths, which works as an ensemble. It searches for the best feature from random decision trees to improve classification accuracy.
- Nnet algorithm is inspired by the biological neural network, which imitates the human brain learning process. The Nnet classifier uses documents as input nodes and assigns features weight to its input until the final classification is reached. Then, it uses backpropagation, which implies that misclassified cases are propagated back using neural networks trying to identify the node that caused the error. After the node has been identified, the weights are added to minimize the error.
- Logitboost is a boosting classification algorithm that strives to obtain the optimum classifiers by fitting an additive model that minimizes a logistic loss.

3.3 Data processing

The dataset used for this study is obtained from crash narratives. The dataset is further classified into parking- and non-parking-related AV crashes. The parking-related AV crashes included all crashes with at least one vehicle (both AV or at least one AV and one conventional car) involved in a crash when parking or already parked. Before classifying the dataset, all parking-related terms were removed. This process was necessary to ensure that the algorithm could easily classify parking-related crashes without the presence of crucial parking terms like parked, parking, and reversing. This method is consistent with an approach used by Arteaga et al. (2020) when analyzing the injury severity of crashes using an interpretable text-mining approach.

Since AV parking-related crashes constitute a small proportion (about 14.4%) of all AV-involved crashes, the issue of class imbalance might arise. To address this, three resampling algorithms—SMOTE, over-sampling, and under-sampling (Kitali et al., 2019)—were considered to take care of class imbalance. In addition, resampling techniques like cross-validation and bootstrap approaches were used to increase the sample size (Kitali et al., 2019). From observations, it was realized that the combination of under-sampling and bootstrap yielded models with reasonable classification accuracy. Different proportions of training and testing data sets were examined in the quest for the ratio that yields a higher score. Finally, predictions were obtained based on a model developed using 70% of the data for training and 30% for testing. The features used for classifying parking-related crashes were developed from the text data using standard text conversion procedures. The conversion procedures involve cleaning, generating n -gram features, and generating a document matrix (Kwayu et al., 2020). Concerning cleaning the data, unstructured text was converted into the corpus. Also, punctuation, symbols, numbers, and stemming (i.e. reducing a word-to-word root) were removed. On the other hand, generating a document matrix covers determining each n -gram and converting the corpus to the document matrix. The document matrix was then used for AV parking-involved crash classification tasks.

4 Study data

The State of California has been testing AVs in mixed traffic since 2014. The AVs operating in California roadways vary by the levels of automation, whereby the lowest level is ADAS level 2, and the highest is level 4. The state requires any crash involving AV to be reported within ten days. Although crash data has been collected, crashes that occurred before 2017 were not well documented. Thus, this study used 460 AV crashes collected by the California Department of Motor Vehicles (CDMV) between January 2017 and June 2022.

Vehicle operators submit crash reports that contain the crash narratives to the CDMV in pdf format. Based on these narratives, several variables were manually extracted, including the location and time of crashes, the directional movement of the AV and CV (connected vehicle), and the type of driving mode during the

collision. Google Maps was also employed to get other roadway-related variables by following the address provided in the crash report.

To obtain crash location information, section 5 of the crash report (Figure 2) was used. The narratives clearly state the location of the crash. For instance, part of the narrative in Figure 2a states that ‘*On December 11, 2021, at 2:28 am PST a Waymo Autonomous Vehicle (Waymo AV) operating in San Francisco was in a collision involving a concrete post in the parking lot at 1900 Noriega Street...*’. The narrative suggests that the crash occurred at the parking lot. Thus, it is a parking-related crash. On the other hand, a portion of the narrative in Figure 2b states that ‘*...a Zoox vehicle in autonomous mode was stopped at a red traffic signal at the intersection of Harrison and 11th Streets in San Francisco with its left turn signal engaged. After the traffic signal changed to green and while waiting for an opening to turn left onto Harrison Street, the Zoox vehicle was struck by a passenger from behind...*’, suggesting that the crash was at an intersection.

5 Non-parking related crashes: descriptive summary

Overall, among 460 crashes, 15.7% (72 crashes) were parking-related crashes. Figure 3 shows the distribution of AV crashes during the period under consideration (2017–2022). The number of parking-related and non-parking-related crashes significantly differed over the years. The total number of AV crashes over the years shows an upward trend except in 2020. Also, the highest number of crashes was recorded in 2022 (23.9% of the total), while 2020 recorded the lowest number (9.3% of the total). This could be attributed to COVID-19 and the lower number of trips recorded nationally in 2020. The 12 parking-related crashes in 2020 represent the highest proportion (28%) of parking-related crashes per year, while in 2017, no parking-related crashes were observed. Although the year has not ended, the statistics suggest that 2022 might record the most significant number of parking-related crashes.

Figure 4 shows the number of crashes at different times of the day. For both parking and non-parking-related crashes, significant peaks were observed between 8:30 am and 9:30 am and between 3:30 pm and 4:30 pm. The parking-related crashes follow a similar trend to the non-parking-related crashes, except for the early night period (8:00 pm to midnight).

SECTION 5 — ACCIDENT DETAILS - DESCRIPTION

Autonomous Mode Conventional Mode

On December 11, 2021 at 2:28 AM PST a Waymo Autonomous Vehicle (“Waymo AV”) operating in San Francisco was in a collision involving a concrete post in the parking lot at 1900 Noriega Street.

At the time of the impact, the Waymo AV’s Level 4 ADS was not engaged and a test driver was operating the Waymo AV in manual mode. The operator of the vehicle was attempting to reverse into a parking spot when the right rear sensor casing of the vehicle struck a concrete post adjacent to the parking spot. The Waymo AV sustained minor damage to the right rear sensor housing.

(a) Parking related Crashes

SECTION 5 — ACCIDENT DETAILS - DESCRIPTION

Autonomous Mode Conventional Mode

A Zoox vehicle in autonomous mode was stopped at a red traffic signal at the intersection of Harrison and 11th Streets in San Francisco with its left turn signal engaged. After the traffic signal changed to green and while waiting for an opening to turn left onto Harrison Street, the Zoox vehicle was struck by a passenger vehicle from behind. The driver of that passenger vehicle stated that he was using his phone and not paying attention. No police or emergency services were called to the scene. The Zoox driver, front passenger, and rear passenger reported minor injuries for which they later sought medical attention. The Zoox vehicle incurred minor damage to its rear sensors.

(b) Non-Parking related Crashes

Figure 2 Typical narratives for parking and non-parking related crashes

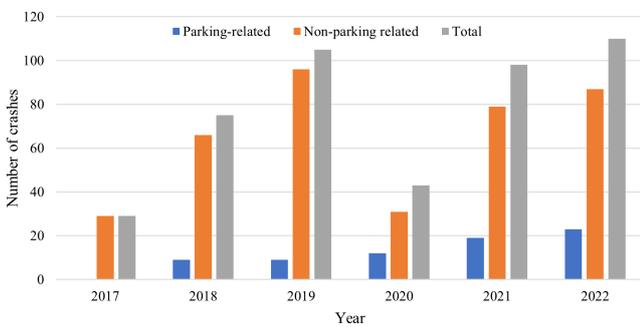


Figure 3 Distribution of AV crashes by year

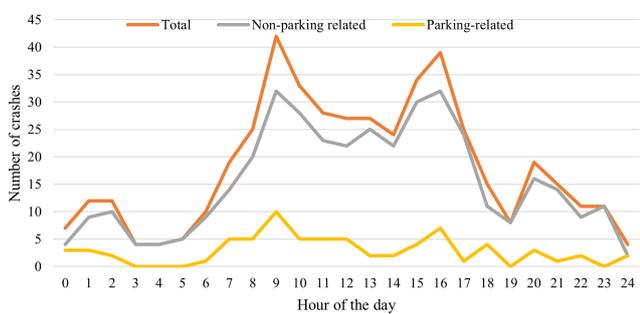


Figure 4 Distribution of AV crashes by time of the day

Collisions involving the rear of AVs constituted about 50.7% of the total crashes, as shown in Figure 5. This finding is in line with previous studies showing rear-end collisions as significant in AVs (Ashraf et al., 2021; Liu et al., 2021). Other hit sides are the left side (24.3%), right side (12.6%), and front side (12.4%). While most non-parking-related crashes

involved the rear side of the AVs, the distribution of parking-related crashes is relatively similar across the sides of the AVs. Such observation strengthens the importance of characterizing parking-related crashes because non-parking-related crashes can overshadow latent characteristics.

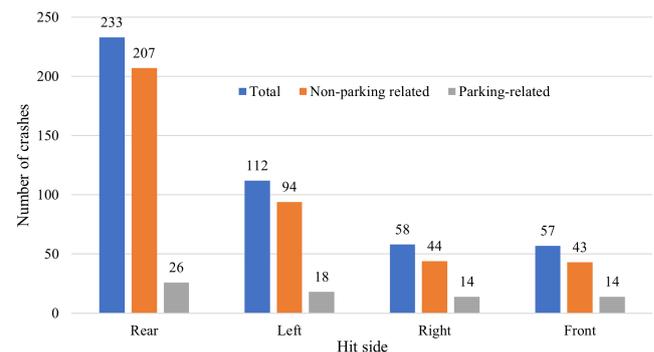


Figure 5 Distribution of AV crashes by the side hit

6 Results and discussion

This section presents the results and discussion of this study. It is divided into two major subsections: text mining and discussion, and model results and discussion. The text mining part presents the results and discussion from the text mining, while the model section avails the results and discussion from the statistical models.

6.1 Text networks results

The networks results and discussion covers the discussion on parking-related and non-parking-related crashes. The discussion is based on text networks' metrics, including keywords frequency, co-occurrences, and collocations

6.1.1 Parking-related crashes

Table 2 and Figure 6 show the text network and main keywords associated with AV crashes at parking spaces. It was found that the significant keywords include vehicle, mode, parked, passenger, damage, street, and rear. The keywords connected to vehicle indicated that crashes were likely to involve passenger vehicles. Also, the keyword vehicle was connected to the keyword unoccupied using thick lines, meaning that there were no occupants in parked vehicles involved in the crashes. Unlike non-parking related crashes, the node for keyword conventional was more significant than for keyword automated. This implied that AVs involved in crashes at parking locations were more likely to be in conventional than autonomous mode. As expected, parking and parked were other major keywords in the network, alluding to crashes in parking zones or parking lots. The size of the node for keyword parallel was more significant than the keyword lot, indicating a higher likelihood of crashes involving AVs in on-street parking spaces than in parking lots.

The keyword damage was highly connected to keywords side, mirror, door, minor, and no injuries. Contrary to the non-parking related crashes, it was revealed that the AV crashes in parking locations were likely to cause damage to the side mirrors. The network revealed that the parking crashes might also involve door damage. Moreover, the keywords minor and no-injuries indicate the severity of these crashes. As a result, police are not called or engaged in most of these crashes. The rear and its related keywords, including made contact, bumper, front, and left, indicate the nature of the collisions in parking lots. They are also supported by keyword reversing, which suggests a common vehicle movement when parking. It could therefore mean crashes are more likely to involve contact between the front and rear bumpers when a driver reverses to get into or out of the parking spot. However, it is unclear what the keywords left and right imply in crashes involving AVs at parking locations. It is important to note that crashes at parking locations were less likely to affect non-motorized users, as

proven by the absence of keywords such as bicyclists or pedestrians. Moreover, lighting was not among the network keywords indicating the insignificance effect of lighting on AV crashes at parking locations.

6.1.2 Non-parking related crashes

Figure 7 and Table 3 present the text network and associated main keywords for crashes involving AVs at non-parking locations (intersections and non-intersections). It can be observed that eight primary keywords, vehicle, mode, rear, bumper, damage, passenger, autonomous, and street stand out from Figure 7. The network suggests that passenger vehicles were crucial participants in non-parking-related crashes involving AVs. This might be true since autonomous vehicles usually operate in urban locations where passenger vehicles dominate. The non-parking text network indicated that AVs operated more in autonomous mode than in conventional mode. This finding is opposite to results found in parking-related crashes. This inference was made by comparing the node sizes of autonomous and traditional keywords in text networks of both parking- and non-parking-related crashes. The thick line between the keywords autonomous and mode indicates that the vehicle was operating in autonomous mode.

The node size of the keyword rear was more significant than the keywords front and left. The keyword rear appeared in 375 observations, while front and left were reported in 205 and 306 observations, respectively. The keyword rear was mainly connected to the keywords bumper, damage, made-contact, and minor bumper, damage, and minor. These keywords are indicative of most crashes involving AVs at non-parking locations being rear-ended collisions. The thick lines between these keywords reveal that damage associated with rear-end crashes is likely minor when contacts are being made with the rear bumper. In addition, it was revealed that rear bumper and rear damage were among the most co-occurring words suggesting that crashes at non-parking locations were mainly rear-ending collisions. Similar to parking-related crashes, thick lines between the keywords no-injuries, not-called, police, and scene imply that police were likely not called in crashes not involving injuries. In contrast to parking-related crashes, the presence of the keyword bicyclist reveals that AVs are likely to be involved in crashes with non-motorized vehicles in non-parking locations.

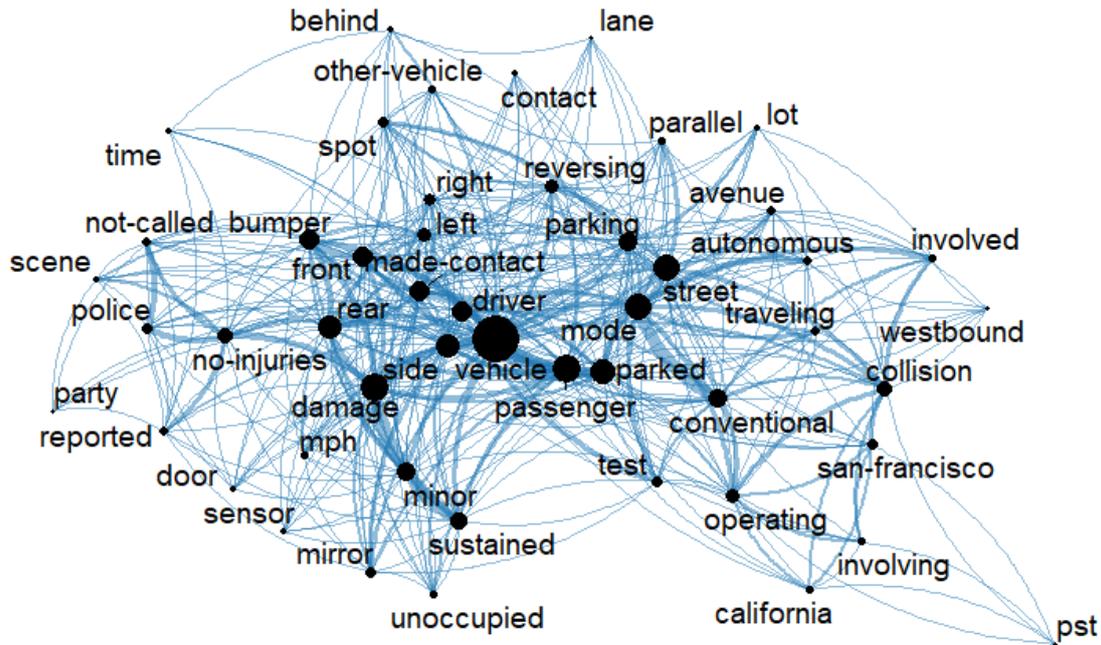


Figure 6 Text network for parking-related crashes

Table 2 Extracted keywords for parking-related crashes

	Frequency	Co-occurrences		Collocations		Z-score		
		freq*	docfreq	count	count			
1	Mode	80	64	Passenger vehicle	64	Passenger vehicle	48	15.3
2	Damage	65	52	Conventional mode	54	Minor damage	35	13.7
3	Vehicle	120	48	Parked vehicle	49	No-injuries police	22	12.9
4	Conventional	54	47	Minor damage	39	Operating conventional	18	11.6
5	Made-contact	47	45	Damaged vehicle	37	No-injuries reported	18	11.6
6	No-Injuries	44	44	Vehicle made contact	35	Operating san-francisco	17	11.3
7	Street	79	41	Vehicle bumper	30	Rear bumper	18	11.3
8	Parked	62	38	Front bumper	30	Sustained minor	32	11.2
9	Rear	56	38	Police not-called	29	Test driver	25	11.0
10	Collision	39	37	Front vehicle	28	Police not-called	29	10.8
11	Minor	42	35	Damage bumper	28	Front bumper	15	10.6
12	Driver	52	33	Rear bumper	27	Reported scene	10	10.4
13	Parking	48	32	Rear damage	27	Involved collision	19	9.8
14	Not-Called	32	32	Autonomous mode	27	San-Francisco California	13	9.6
15	Bumper	45	31	Parked street	26	Vehicle sustained	17	9.3
16	Operating	42	31	Test driver	25	Parked vehicle	20	8.9
17	Police	31	31	Side vehicle	25	California collision	13	8.9
18	Reversing	34	29	Parking spot	25	Mode involved	12	8.7
19	Side	51	28	Parked side	25	pm PST	6	8.7
20	Front	45	28	Parking reversing	23	Damage front	12	8.6

*freq = frequency; docfreq = document frequency

6.2 Machine learning results

The default performance measures of the five classifiers that were considered without resampling and imbalance treatments are shown in Table 4. Among the used classifiers, Nnet had the highest prediction accuracy (90.8%) but was weak in predicting only 40% of parking-related crashes. Conversely, the Nnet classifier predicted 97.4% of non-parking-related crashes. RF and SVM had 100% specificity, meaning they could predict all non-parking-related crashes. Like Nnet, RF and SVM also predicted low parking-related crashes, 6.7%, and 0%, respectively. This study, however, intends to understand the parking-related crashes, therefore, in addition to the overall prediction accuracy, the sensitivity score is the most critical performance metric. The NB classifier has the maximum sensitivity result, but its overall prediction accuracy is lower than other classifiers. This finding calls for the need for data imbalance treatment.

Table 4 Performance metrics for text classifiers (without resampling or imbalance treatments)

	Accuracy	Sensitivity	Specificity
SVM	88.5%	0.0%	100.0%
NB	40.8%	73.3%	36.5%
Logitboost	86.2%	33.3%	93.0%
RF	89.2%	6.7%	100.0%
Nnet	90.8%	40.0%	97.4%

Several data imbalances and resampling treatments are available (Gao et al., 2021; Morris & Yang, 2021; Mujalli et al., 2016; Zhou et al., 2016). In this study, a total of two resampling techniques were used. These include bootstrap and cross-validation. Further, three class imbalance treatments were used, Up, Down, and SMOTE. Table 5 and Table 6 present the best resampling techniques with the three class imbalance treatments. For each case, the best-performing classifier is noted. It can be observed that Bootstrap & Up and Cross-Validation & Up are the best two resampled and class imbalance-treated classifiers using the NB model. The sensitivity score associated with Bootstrap & Up and Cross-validation & Up is 80%. For the same models, the overall accuracy was 42.3% (Bootstrap and Up) and 44.6% (Cross-validation & Up). Therefore, these two classifiers are used to infer the key features associated with parking-related crashes.

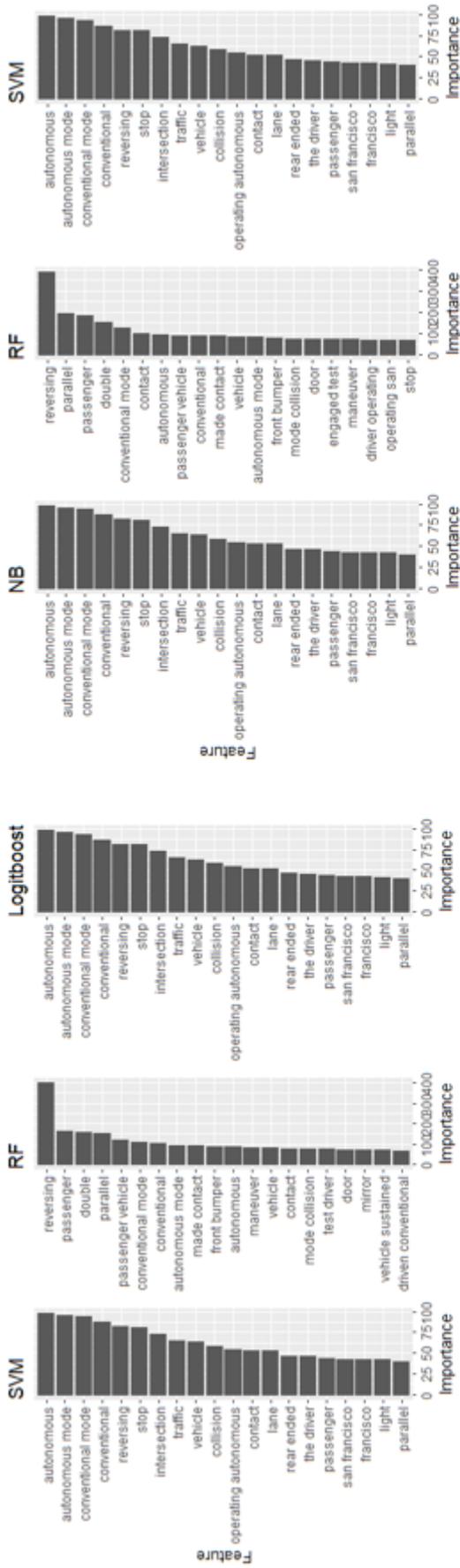
Figure 8 presents the top 20 critical features for three classifiers with high specificity, prediction accuracy, and sensitivity scores. Overall, the feature ‘reversing’ appears to be the most critical feature. The observation implies that parking-related crashes likely involve either AV or conventional vehicles reversing from the parking lot. 25 parking-related crashes involved either AV or conventional vehicle reversing.

One of the typical narratives states that ‘...the Waymo AV came to a stop behind a stalled bus blocking the roadway. The driver transitioned the system to manual mode and began to reverse out of the narrow road. While reversing, a van approached from behind and the Waymo AV made contact with the front left fender of the van...’. The narrative indicates that the AV involved in the parking-related crash was in conventional mode and reversing to bypass the stalled bus.

Further, the feature ‘parallel’ appears to be ranked the highest by RF. This observation can imply that parallel parking is also associated with a high likelihood of parking-related crashes. The raw data indicates that 19 of the parking-related crashes parallel parking vehicles. For instance, one of the narratives stated that ‘...while maneuvering to avoid a parallel parking vehicle, the driver of the Cruise AV made contact with another vehicle that was pulling out of a driveway on the opposite side of the street...’. This narrative shows that AV was involved in a crash when it was performing parallel parking. Another key observation among the top 20 features is that ‘conventional mode’ is ranked higher than ‘automated’. This feature’s high ranking implies that AVs were likely to be involved in parking-related crashes when manually operated. However, the feature ‘automated’ in the top 20 most essential features suggests that several crashes involved AVs in autonomous mode. Other features such as front passenger, mirror, front bumper, injury, etc. explain more of what happens after a collision than what causes a collision.

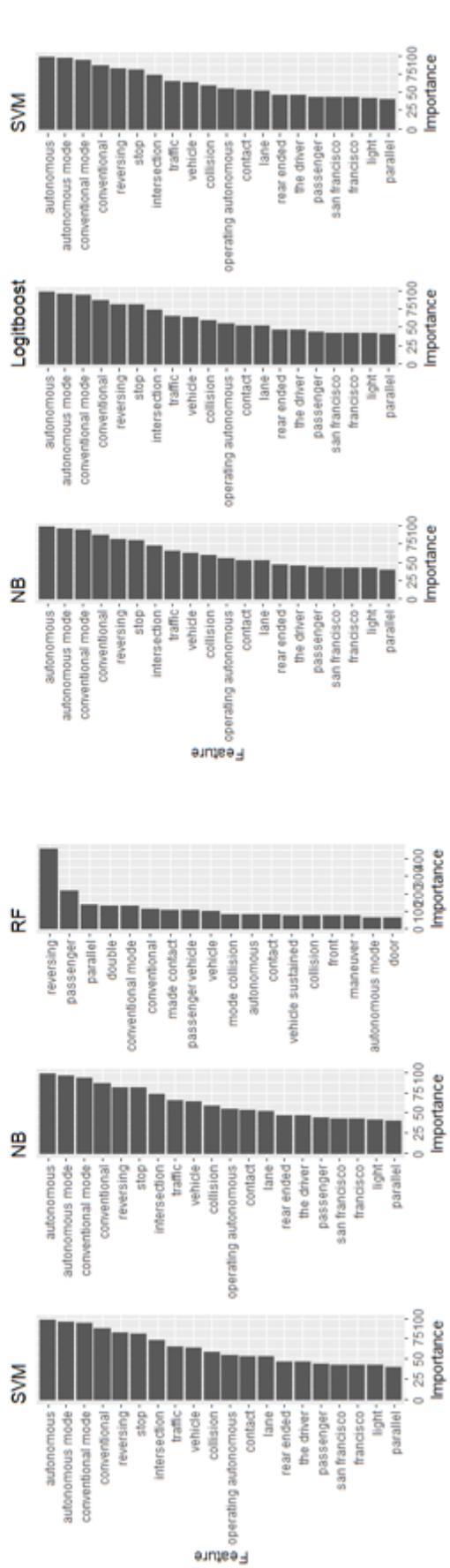
7 Conclusions

This study investigated factors associated with parking-related crashes involving AVs. A total of 460 AVs crashes collected by the CDMV within six years (2017–2022) were analysed. Text network analysis was applied to identify factors associated with parking- and non-parking-related crashes in unstructured narrative data. Using text network topology, keyword frequency, co-occurrence, and collocations, features peculiar to



(a) Bootstrap & Down sample

(b) Bootstrap & Up sample



(c) CV & Down sample

(d) CV & Up sample

Figure 8 Top 20 features for non-parking related crashes

Table 5 Performance metrics for text classifiers (with bootstrap resampling and imbalance treatments)

	Bootstrap					
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
	Bootstrap			Bootstrap & Down		
SVM	88.5%	0.0%	100.0%	68.5%	93.3%	65.2%
NB	40.8%	73.3%	36.5%	42.3%	66.7%	39.1%
Logitboost	87.7%	33.3%	94.8%	83.1%	80.0%	83.5%
RF	89.2%	6.7%	100.0%	83.1%	80.0%	83.5%
Nnet	90.0%	40.0%	96.5%	79.2%	60.0%	81.7%
Best model	Nnet	NB	SVM&RF	Logitboost & RF	SVM	Logitboost & RF
	Bootstrap & Up			Specificity		
SVM	80.0%	0.0%	83.5%	83.8%	66.7%	86.1%
NB	40.8%	73.3%	37.4%	48.5%	73.3%	45.2%
Logitboost	86.9%	13.3%	98.3%	87.7%	53.3%	92.2%
RF	89.2%	6.7%	96.5%	87.7%	33.3%	94.8%
Nnet	90.8%	40.0%	93.9%	89.2%	46.7%	94.8%
Best model	RF	NB	Logitboost	Nnet	NB	RF & Nnet

Table 6 Performance metrics for text classifiers (with cross-validation and imbalance treatments)

	Cross-validation					
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
	Cross-validation			Cross-validation & Down		
SVM	88.5%	0.0%	100.0%	60.0%	93.3%	55.7%
NB	40.8%	73.3%	36.5%	46.2%	73.3%	42.6%
Logitboost	86.9%	13.3%	96.5%	73.8%	46.7%	77.4%
RF	89.2%	6.7%	100.0%	75.4%	73.3%	75.7%
Nnet	90.8%	40.0%	97.4%	79.2%	46.7%	83.5%
Best Model	Nnet	NB	RF	Nnet	SVM	Logitboost
	Cross-validation & Up			Cross-validation & SMOTE		
SVM	78.5%	66.7%	80.0%	83.8%	66.7%	86.1%
NB	44.6%	80.0%	40.0%	51.5%	73.3%	48.7%
Logitboost	87.7%	80.0%	88.7%	92.3%	60.0%	96.5%
RF	90.0%	46.7%	95.7%	87.7%	33.3%	94.8%
Nnet	88.5%	46.7%	93.9%	88.5%	46.7%	93.9%
Best Model	Nnet	NB & Logitboost	RF	Logitboost	NB	Logitboost

parking- and non-parking-related crashes could be identified. For example, AVs operated more in a conventional mode in parking spaces than in non-parking spaces. In addition, it was identified that AVs are likely to collide with non-motorized vehicles in non-parking areas as compared to parking spaces. However, rear-end collisions were found to be shared in both parking and non-parking areas.

Furthermore, the study used five text classifiers—SVM, NB, Logitboost, RF, and Nnet—to predict

parking-related crashes. Performance metrics like accuracy, sensitivity, and specificity were used to compare the classifiers. In addition, two resampling techniques—bootstrap and cross-validation were employed to account for data imbalances and resampling treatments. Results from classifiers indicated that parking-related crashes are likely to involve either AV or conventional vehicles reversing from the parking lot. In addition, parallel parking was associated with a high likelihood of parking-related

crashes.

Based on the study results, AV operators need to improve AV operation during parking especially when reversing in the parallel parking. The narratives indicated that reverse was among the keywords that appear more frequently and associated with parking-related crashes. Furthermore, the additional considerations should be given to the parking procedures. The fact that most of AVs are in conventional mode during parking shows that AVs are not performing well in that aspect. This can be solved by having dedicated sensor for parking maneuvers, which will be calibrated using parking environment to improve its accuracy and reduce parking-related crashes.

8 Study limitations

This study has some limitations which need to be addressed in future research. Firstly, while traditional crash data are collected by the police, this study used crash data reported by vehicle owners. That being the case, the accuracy of the data is not guaranteed, thus, future research might obtain police crash reports and compare crash narratives. This study used data from California, however, currently various states are testing AVs and are collecting crash data. Future studies may consider crash data from all states testing AVs. Lastly, text classifications work better with large data. The sample size of 460 crashes is relatively small, and the results obtained need to be carefully interpreted. Future studies can use large sample sizes for broader analysis.

CRedit contribution statement

Boniphace Kutela: Conceptualization, Methodology, Data curation, Investigation, Writing—original draft.

Ricard Dzinela: Conceptualization, Methodology, Data curation, Investigation, Writing—original draft.

Henrick Haule: Data curation, Investigation, Writing—original draft.

Abbas Sheykhfard: Writing—original draft.

Kelvin Msechu: Data curation, Investigation, Writing—original draft.

Declaration of competing interests

Authors have no competing interest of any kind.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

We would like to thank the California Department of Motor Vehicles for the open-source autonomous vehicle crashes used in this study.

References

- Arteaga, C., A. Paz, J. W. Park (2020), ‘Injury severity on traffic crashes: A text mining with an interpretable machine-learning approach’, *Safety Science*, 132, 104988, <https://doi.org/10.1016/j.ssci.2020.104988>.
- Ashraf, M. T., K. Dey, S. Mishra, M. T. Rahman (2021), ‘Extracting Rules from Autonomous-Vehicle-Involved Crashes by Applying Decision Tree and Association Rule Methods’, *Transportation Research Record: Journal of the Transportation Research Board*, 2675(11), 522–533, <https://doi.org/10.1177/03611981211018461>.
- Bahrami, S., M. Roorda (2022), ‘Autonomous vehicle parking policies: A case study of the City of Toronto’, *Transportation Research Part A: Policy and Practice*, 155, 283–296, <https://doi.org/10.1016/j.tra.2021.11.003>.
- Benoit, K., K. Watanabe, H. Wang, P. Nulty, A. Obeng, S. Müller, A. Matsuo (2018), ‘quanteda: An R package for the quantitative analysis of textual data’, *Journal of Open Source Software*, 3(30), 774, <https://doi.org/10.21105/joss.00774>.
- Blaheta, D., M. Johnson (2011), ‘Unsupervised learning of multi-word verbs’, *Proceedings of the ACL Workshop on Collocations*, 54–60, <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=6bb607eea1875141b4fa89f2c2d361335026c592>.
- Boggs, A. M., B. Wali, A. J. Khattak (2020), ‘Exploratory analysis of automated vehicle crashes in California: A text analytics & hierarchical Bayesian heterogeneity-based approach’, *Accident Analysis & Prevention*, 135, 105354, <https://doi.org/10.1016/j.aap.2019.105354>.
- Chan, T. K., C. S. Chin, Z. Vale, J. Ball, M. Ricco (2021), ‘Review of Autonomous Intelligent Vehicles for Urban Driving and Parking’, *Electronics* 2021, 10(9), 1021, <https://doi.org/10.3390/electronics10091021>.
- Chen, H., H. Chen, R. Zhou, Z. Liu, X. Sun (2021), ‘Exploring the Mechanism of Crashes with Autonomous Vehicles Using Machine Learning’, *Mathematical Problems in Engineering*, 2021, 1–10, <https://doi.org/10.1155/2021/5524356>.
- Das, S., A. Dutta, I. Tsapakis (2020), ‘Automated vehicle collisions in California: Applying Bayesian latent class model’, *IATSS Research*, 44(4), 300–308, <https://doi.org/10.1016/j.iatssr.2020.03.001>.
- DMV (n.d), ‘Autonomous Vehicle Collision Report’, State of California, Department of Motor Vehicles,

- <https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/autonomous-vehicle-collision-reports/>, accessed 2023-06-30.
- Favarò, F., S. Eurich, N. Nader (2018), 'Autonomous vehicles' disengagements: Trends, triggers, and regulatory limitations', *Accident Analysis & Prevention*, 110, 136–148, <https://doi.org/10.1016/j.aap.2017.11.001>.
- Gao, L., P. Lu, Y. Ren (2021), 'A deep learning approach for imbalanced crash data in predicting highway-rail grade crossings accidents', *Reliability Engineering & System Safety*, 216, 108019, <https://doi.org/10.1016/j.res.2021.108019>.
- Hsieh, M. F., U. Özgüner (2008, September), 'A parking algorithm for an autonomous vehicle', *2008 IEEE Intelligent Vehicles Symposium* (Eindhoven, Netherlands: IEEE), 4 - 6 June 2008, <https://doi.org/10.1109/ivs.2008.4621317>.
- Hunter, S. (2014), 'A Novel Method of Network Text Analysis', *Open Journal of Modern Linguistics*, 04(02), 350–366, <https://doi.org/10.4236/ojml.2014.42028>.
- Jiang, C., C. R. Bhat, W. H. K. Lam (2020), 'A bibliometric overview of Transportation Research Part B: Methodological in the past forty years (1979-2019)', *Transportation Research Part B: Methodological*, 138, 268–291, <https://doi.org/10.1016/j.trb.2020.05.016>.
- Joachims, T. (1998), 'Text categorization with Support Vector Machines: Learning with many relevant features', in and others (ed.), *Machine Learning: ECML-98*, Springer Link, , pp. 137–142, <https://doi.org/10.1007/bfb0026683>.
- Khattak, Z. H., M. D. Fontaine, B. L. Smith (2020), 'Exploratory Investigation of Disengagements and Crashes in Autonomous Vehicles Under Mixed Traffic: An Endogenous Switching Regime Framework', *IEEE Transactions on Intelligent Transportation Systems*, 22(12), 7485–7495, <https://doi.org/10.1109/tits.2020.3003527>.
- Kitali, A. E., P. Alluri, T. Sando, W. Wu (2019), 'Identification of Secondary Crash Risk Factors using Penalized Logistic Regression Model', *Transportation Research Record: Journal of the Transportation Research Board*, 2673(11), 901–904, <https://doi.org/10.1177/0361198119849053>.
- Kutela, B., S. Das, B. Dadashova (2022a), 'Mining patterns of autonomous vehicle crashes involving vulnerable road users to understand the associated factors', *Accident Analysis & Prevention*, 165, 106473, <https://doi.org/10.1016/j.aap.2021.106473>.
- Kutela, B., C. Kadeha, R. T. Magehema, R. E. Avelar, P. Alluri (2023), 'Leveraging text mining approach to explore research roadmap and future direction of wrong-way driving crash studies. Data and Information Management', *Data and Information Management*, 100044, <https://doi.org/10.1016/J.DIM.2023.100044>.
- Kutela, B., N. Langa, S. Mwendu, E. Kidando, A. E. Kitali, P. Bansal (2021a), 'A text mining approach to elicit public perception of bike-sharing systems', *Travel Behaviour and Society*, 24, 113–123, <https://doi.org/10.1016/j.tbs.2021.03.002>.
- Kutela, B., R. T. Magehema, N. Langa, F. Steven, R. J. Mwekh'iga (2022b), 'A comparative analysis of followers' engagements on bilingual tweets using regression-text mining approach. A case of Tanzanian-based airlines', *International Journal of Information Management Data Insights*, 2(2), 100123, <https://doi.org/10.1016/j.jjime.2022.100123>.
- Kutela, B., N. Novat, N. Langa (2021b), 'Exploring geographical distribution of transportation research themes related to COVID-19 using text network approach', *Sustainable Cities and Society*, 67, 102729, <https://doi.org/10.1016/j.scs.2021.102729>.
- Kwayu, K. M., V. Kwigizile, J. Zhang, J.-S. Oh (2020), 'Semantic N-Gram Feature Analysis and Machine Learning-Based Classification of Drivers' Hazardous Actions at Signal-Controlled Intersections', *Journal of Computing in Civil Engineering*, 34(4), [https://doi.org/10.1061/\(asce\)cp.1943-5487.0000895](https://doi.org/10.1061/(asce)cp.1943-5487.0000895).
- Lee, C. K., C. L. Lin, B. M. Shiu (2009), 'Autonomous Vehicle Parking Using Hybrid Artificial Intelligent Approach', *Journal of Intelligent and Robotic Systems*, 56(3), 319–343, <https://doi.org/10.1007/s10846-009-9319-9>.
- Lee, S., R. Arvin, A. J. Khattak (2023), 'Advancing investigation of automated vehicle crashes using text analytics of crash narratives and Bayesian analysis', *Accident Analysis & Prevention*, 181, 106932, <https://doi.org/10.1016/j.aap.2022.106932>.
- Liu, Q., X. Wang, X. Wu, Y. Glaser, L. He (2021), 'Crash comparison of autonomous and conventional vehicles using pre-crash scenario typology', *Accident Analysis & Prevention*, 159, 106281, <https://doi.org/10.1016/j.aap.2021.106281>.
- Mahdavian, A., A. Shojaei, A. Oloufa (2019), 'Assessing the long-and mid-term effects of connected and automated vehicles on highways', *International Conference on Sustainable Infrastructure 2019: Leading Resilient Communities through the 21st Century*, Los Angeles, CA, , Nov 6-9, pp. 263–273, <https://doi.org/10.1061/9780784482650.027>.
- Morando, M. M., Q. Tian, L. T. Truong, H. L. Vu (2018), 'Studying the safety impact of autonomous vehicles using simulation-based surrogate safety measures', *Journal of Advanced Transportation*, 6135183, <https://doi.org/10.1155/2018/6135183>.
- Morris, C., J. J. Yang (2021), 'Effectiveness of resampling methods in coping with imbalanced crash data: Crash type analysis and predictive modeling', *Accident Analysis & Prevention*, 159, 106240, <https://doi.org/10.1016/j.aap.2021.106240>.
- Mousavi, M., S. Lord, D. Dadashova, B. Mousavi, S (2020), 'Can Autonomous vehicles enhance traffic safety at unsignalized intersections?', *International Conference on Transportation and Development 2020*, Seattle,

- Washington, USA, 26–29 May 2020, <https://doi.org/10.1061/9780784483138.018>.
- Mousavi, S. M., O. A. Osman, D. Lord, K. K. Dixon, B. Dadashova (2021), ‘Investigating the safety and operational benefits of mixed traffic environments with different automated vehicle market penetration rates in the proximity of a driveway on an urban arterial’, *Accident Analysis & Prevention*, 152, 105982, <https://doi.org/10.1016/j.aap.2021.105982>.
- Mujalli, R. O., G. López, L. Garach (2016), ‘Bayes classifiers for imbalanced traffic accidents datasets’, *Accident Analysis & Prevention*, 88, 37–51, <https://doi.org/10.1016/j.aap.2015.12.003>.
- Nakrani, N. M., M. M. Joshi (2022), ‘A human-like decision intelligence for obstacle avoidance in autonomous vehicle parking’, *Applied Intelligence*, 52(4), 3728–3747, <https://doi.org/10.1007/s10489-021-02653-3>.
- NHSTA (2022), ‘AV TEST Initiative | Automated Vehicle Tracking Tool’, <https://www.nhtsa.gov/automated-vehicle-test-tracking-tool>, accessed 2023-06-24.
- Novat, N., E. Kidando, B. Kutela, A. E. Kitali (2023), ‘A comparative study of collision types between automated and conventional vehicles using Bayesian probabilistic inferences’, *Journal of Safety Research*, 84, 251–260, <https://doi.org/10.1016/j.jsr.2022.11.001>.
- Paranyushkin, D. (2012), ‘Visualization of Text’s Polysingularity Using Network Analysis’, NODUS LABS, <https://noduslabs.com/research/visualization-text-polysingularity-network-analysis/>, accessed 2023-06-27.
- Parsa, A. B., R. Shabanpour, A. Mohammadian, J. Auld, T. Stephens (2021), ‘A data-driven approach to characterize the impact of connected and autonomous vehicles on traffic flow’, *Transportation Letters*, 13(10), 687–695, <https://doi.org/10.1080/19427867.2020.1776956>.
- Pokorny, P., A. Høye (2022), ‘Descriptive analysis of reports on autonomous vehicle collisions in California: January 2021–June 2022’, *Traffic Safety Research*, 2, 000011, <https://doi.org/10.55329/xydm4000>.
- Pranckevičius, T., V. Marcinkevičius (2017), ‘Comparison of Naïve Bayes, Random Forest, Decision Tree, Support Vector Machines, and Logistic Regression Classifiers for Text Reviews Classification’, *Baltic Journal of Modern Computing*, 5(2), 221–232, <https://doi.org/10.22364/bjmc.2017.5.2.05>.
- Ren, W., B. Yu, Y. Chen, K. Gao (2022), ‘Divergent Effects of Factors on Crash Severity under Autonomous and Conventional Driving Modes Using a Hierarchical Bayesian Approach’, *International Journal of Environmental Research and Public Health*, 19(18), 11358, <https://doi.org/10.3390/ijerph191811358>.
- Schoettle, B., M. Sivak (2018, April), ‘A Preliminary Analysis of Real-World Crashes Involving Self-Driving Vehicles’, Transportation Research Institute, The University of Michigan, UMTRI-2015-34, <http://websites.umich.edu/~umtriswt/PDF/UMTRI-2015-34.pdf>.
- Song, Y., M. Chitturi, D. A. Noyce (2021), ‘Automated vehicle crash sequences: Patterns and potential uses in safety testing’, *Accident Analysis & Prevention*, 153, 106017, <https://doi.org/10.1016/j.aap.2021.106017>.
- Statista (2022), ‘Worldwide - AV market penetration 2030’, <https://www.statista.com/statistics/875080/av-market-penetration-worldwide-forecast/>, accessed 2023-06-24.
- Xu, C., Z. Ding, C. Wang, Z. Li (2019), ‘Statistical analysis of the patterns and characteristics of connected and autonomous vehicle involved crashes’, *Journal of Safety Research*, 71, 41–47, <https://doi.org/10.1016/j.jsr.2019.09.001>.
- Yuan, J., M. Abdel-Aty, Y. Gong, Q. Cai (2019), ‘Real-Time Crash Risk Prediction using Long Short-Term Memory Recurrent Neural Network’, *Transportation Research Record: Journal of the Transportation Research Board*, 2673(4), 314–326, <https://doi.org/10.1177/0361198119840611>.
- Zhou, B., A. M. Roshandeh, S. Zhang, Z. Ma (2016), ‘Analysis of Factors Contributing to Hit-and-Run Crashes Involved with Improper Driving Behaviors’, *Procedia Engineering*, 137, 554–562, <https://doi.org/10.1016/j.proeng.2016.01.292>.

About the authors



Boniphace Kutela is an Associate Research Engineer at the Texas A&M Transportation Institute (TTI). He holds a PhD in Transportation Engineering from University of Nevada Las Vegas.

Dr. Kutela has over nine years of research experience in traffic safety and operations, intelligent transportation systems, railroad network analysis and safety, and connected and autonomous vehicles. His major areas of expertise include Roadway Data Analysis Methodologies, Machine Learning, Predictive Modeling, and Natural Language Processing with a focus on traffic safety and operations.



Ricard Dzinyela is a graduate student in Texas A&M University, college Station. He holds a bachelor’s degree in civil engineering from Kwame Nkrumah University of Science and Technology, Ghana. Mr. Dzinyela has over three

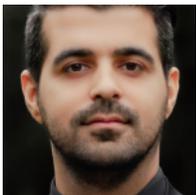
years of research experience in traffic safety and operations. His major areas of expertise include crash data modeling using Econometric, Spatial and Machine Learning Models.



All contents are licensed under the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



Henrick Haule is a Research Assistant Professor at the University of Arizona (UA) and a Manager for the UA Center for Applied Transportation Sciences (CATS). He worked as Postdoctoral Research Fellow at Florida Atlantic University. He received his Ph.D. in Civil Engineering from Florida International University (FIU), an M.S. in Civil Engineering from the University of North Florida, and a B.S. in Civil Engineering from the University of Dar es Salaam in Tanzania. His research focuses on Transportation Systems Management and Operations (TSMO) strategies, traffic incident management, highway safety and crash modeling, and applications of real-time traffic data in transportation.



Abbas Sheykhfard is a researcher at the Department of Civil Engineering, Babol Noshirvani University of Technology, Iran. He received his Ph.D. degree in road and transportation engineering at the same university in 2021. He carried out part of his PhD thesis work at Delft University of Technology, the Netherlands. His research interests include road safety and road user behavioral analysis.



Kelvin Msechu is an Intelligent Transportation Systems (ITS)/Traffic Engineer at Atkins North America Inc. He holds a Master of Civil Engineering from the University of Tennessee at Chattanooga. Kelvin's areas of interests and experience include roadway safety, drone deployment for roadway data collection and safety analysis, traffic, and ITS designs. At Atkins, Kelvin has reviewed hundreds of police narratives on roadway crash reports, he has performed numerous traffic mobility analyses, ITS inspections, and traffic signal design. Kelvin is also involved in journal publications of existing roadway safety issues and emerging technologies.