ARTICLE IN PRESS

Accident Analysis and Prevention xxx (xxxx) xxx

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Mining patterns of autonomous vehicle crashes involving vulnerable road users to understand the associated factors

Boniphace Kutela [*], Subasish Das, Bahar Dadashova

Roadway Safety Program, Texas A&M Transportation Institute, 1111 RELLIS Parkway, Bryan, TX 77807, United States

ARTICLE INFO

Keywords:
Autonomous vehicles crashes
Vulnerable road users
Crash narratives
Text network
Text classifiers

ABSTRACT

Autonomous or automated vehicles (AVs) have the potential to improve traffic safety by eliminating majority of human errors. As the interest in AV deployment increases, there is an increasing need to assess and understand the expected implications of AVs on traffic safety. Until recently, most of the literature has been based on either survey questionnaires, simulation analysis, virtual reality, or simulation to assess the safety benefits of AVs. Although few studies have used AV crash data, vulnerable road users (VRUs) have not been a topic of interest. Therefore, this study uses crash narratives from four-year (2017–2020) of AV crash data collected from California to explore the direct and indirect involvement of VRUs. The study applied text network and compared the text classification performance of four classifiers - Support Vector Machine (SVM), Naïve Bayes (NB), Random Forest (RF), and Neural Network (NN) and associated performance metrics to attain the objective. It was found that out of 252 crashes, VRUs were, directly and indirectly, involved in 23 and 12 crashes, respectively. Among VRUs, bicyclists and scooterists are more likely to be involved in the AV crashes directly, and bicyclists are likely to be at fault, while pedestrians appear more in the indirectly involvements. Further, crashes that involve VRUs indirectly are likely to occur when the AVs are in autonomous mode and are slightly involved minor damages on the rear bumper than the ones that directly involve VRUs. Additionally, feature importance from the best performing classifiers (RF and NN) revealed that crosswalks, intersections, traffic signals, movements of AVs (turning, slowing down, stopping) are the key predictors of the VRUs-AV related crashes. These findings can be helpful to AV operators and city planners.

1. Introduction

Autonomous vehicles (AVs) have the potential to lower the number of crashes by eliminating human error, which contributes to 94% of roadway crashes (Combs et al., 2019; NHTSA, 2015). However, their implications for vulnerable roadway users (VRUs) such as pedestrians, bicyclists, scooters, and skateboard users are not clear. As the interest in AV deployment increases, there is an increasing need to assess and understand the expected implications of AVs on VRU safety. For instance, the interactions between the AVs and VRU in a mixed traffic setting have not been well explored. Furthermore, although it is understood that AVs are programmed to yield right of way to VRUs at the locations such as crosswalks and intersections, it is not clear how the other vehicles react to the AVs. The scarcity of AV crash data has been a major obstacle to evaluating AVs' safety implications on VRUs. Most of the previous studies have used alternative data sources such as survey data, simulation, and virtual reality to conduct the preliminary evaluation of AVs' impacts on VRU safety (Combs et al., 2019; Deb et al., 2017; Millard-Ball, 2018). Further, for a few studies that utilized the actual AV crash data (Boggs et al., 2020; Das et al., 2020b), the VRU involvement in AV crashes has not been given high priority. Also, crash narratives from AV crash data have rarely been utilized. It is argued that, while crashes due to direct interaction between VRU and AV can be traced using traditional crash data, crash narratives are necessary to extract details for the indirect involvement of VRU in a crash.

Therefore, this study explored the patterns of the VRU involvement in the AV crashes using text mining approaches to understand the associated factors. The study focuses on both- direct and indirect involvement of the VRUs in the AVs crashes. The rest of the manuscript is organized as follows. The next section presents the literature summary of the studies that focused on the safety aspects of the AVs. The methodology part is then presented, followed by the study data description. Results and discussion are then presented before the conclusion and future studies.

* Corresponding author.
E-mail addresses: b-kutela@tti.tamu.edu (B. Kutela), s-das@tti.tamu.edu (S. Das), b-dadashova@tti.tamu.edu (B. Dadashova).

https://doi.org/10.1016/j.aap.2021.106473
Received 31 May 2021; Received in revised form 17 August 2021; Accepted 29 October 2021

2. Literature review

Various studies have been performed to evaluate the safety implication of AV on VRUs. Among VRUs, pedestrians have been a focus for most AV-VRU studies. Millard-Ball (2018) used game theory to analyze the interaction between pedestrians and AVs. The study concluded that AVs would facilitate pedestrian movements in urban neighborhoods as they are risk-free, especially at the crosswalks. Further, Deb et al. (2017) analyzed nearly 500 survey responses to explore pedestrians' receptiveness on AVs using principal component analysis. The study found that male and young respondents were more receptive to AVs and believed that AVs would be involved in a smaller number of crashes. An interview-based study (Reig et al., 2018) involving 32 pedestrians exposed to AVs revealed a relationship between favorable perception and trust of AVs. Further, the trust in AV was associated with the AV company's brand. Brar and Caulfield (2018) used an online survey with over 250 responses to study the impact of AVs on pedestrians' safety. The study found that pedestrians have numerous concerns about AV safety, but a significant number of respondents were optimistic and expected that AV would improve traffic safety. A Vehicle-to-Pedestrian (V2P) communication study conducted by Gelbal et al., (2020), evaluated the collision avoidance of AV to pedestrians. The study presented various methods used to detect pedestrians in different environments for both stationary and moving pedestrians. It also presented the real-world experiment of communication between pedestrians and AVs. Furthermore, Kalatian and Farooq (2021) is the only study so far that used virtual reality to understand the interaction between pedestrians and AVs. They concluded that the AVs were one of the factors that affected the waiting time of pedestrians; other factors included lack of walking habits, wider lanes, and limited sight distance. All these studies did not use the actual AV crash data; therefore, the actual safety implication of AV to pedestrians in a mixed traffic cannot be clearly concluded.

Notably, a few studies have been dedicated to exploring the interactions between bicyclists and AVs (Das et al., 2020b; Fleskes and Hurwitz, 2019; Pyrialakou et al., 2020). Fleskes and Hurwitz, (2019) evaluated the influence of bicyclists' presence on AVs' autonomous mode to conventional mode transitions at right turn in their simulation-based study. The study used 43 participants subjected to distractions, each performing 18 right turn maneuvers, whereas the distance between bicycle and intersection and vehicle and intersection varied. The study found that the more time the drivers spent to interact with bicyclists, the safer the interactions were. Other studies used simulation analysis to understand the interaction between pedestrians and AVs (Burns et al., 2019; Mahadevan et al., 2019). A study by Burns et al., (2019) used virtual reality to assess the response of pedestrian to AV maneuvers. Participant showed less trust to AVs when they were faster and closer. Similarly, virtual reality was used by Mahadevan et al., (2019) to simulate the real-word interaction between pedestrians and AVs. However, both studies did not explain the indirect involvement of pedestrian in AV crashes. Another study (Das et al., 2020b) evaluated the technological perception of VRUs on AVs in Pittsburgh by using BikePGH survey data. Their study applied multiple correspondence analysis (MCA) to identify response pattern of over 1000 responses from residents with varying level of interaction with AVs. The study found that respondents who perceived that AVs were not safe were also against approving Pittsburgh as the AV's testing ground. BikePGH survey data has been used in several other studies to address different research topics such as understanding safety perceptions of AVs (Das, 2021) , understanding human–computer interaction patterns (Das and Zubaidi, 2021) , qualitative analysis of sharing the roads with AVs (Rahman et al., 2021). Moreover, a survey-based study (Pyrialakou et al., 2020) used data collected from Phoenix, Arizona, which revealed that residents perceived cycling near AV as least safe than walking and driving. Similar to pedestrian-AV studies presented earlier, the bicyclists-AV studies used either survey questionnaires or simulation data.

Relatively few studies have utilized actual AV crash data collected from the California Department of Motor Vehicle to evaluate the safety implications of AVs (Alambeigi et al., 2020; Boggs et al., 2020; Das et al., 2020a; Favarò et al., 2017; Goodall, 2021). However, the literature on the safety implication of AV on VRU is still limited. For instance, Das et al. (2020b) explored collision patterns of AV and CV, Favarò et al., (2017) and Boggs et al., (2020) focused on the frequency and type of collision, among others. Furthermore, in recent years, crash narratives have emerged as the alternative to traditional crash data for several safety studies (Arteaga et al., 2020; Boggs et al., 2020; Kwayu et al., 2021). Researchers are citing the added information that can only be extracted from crash narratives using text mining approaches as the main motive of utilizing text data (Wali et al., 2021).

It is further argued that traditional crash data without supplementing the narratives do not completely expose the scenarios during crashes (Rasouli and Tsotsos, 2020; Wali et al., 2021). For instance, the traditional crash data normally report the instances that the VRUs are directly involved, while the ones that are indirectly involved are rarely described. Therefore, this study intends to utilize AV crash narratives to identify patterns of the VRU involved crashes. The findings can help in understanding the associated factors for VRU involvement in AV crashes. To the authors' best knowledge, so far, no study has focused on evaluating the involvement of VRUs in AV crashes.

3. Methodology

As mentioned earlier, the objective of this study is to explore the patterns of VRU involved crashes in order to understand the associated factors. To attain the objectives, this study is divided into three main tasks: executed using various tools given certain input to produce the output. Fig. 1 shows the flowchart describing the study methodology.

The first task is to identify the VRUs involved crashes using crash narratives. In this task, the keywords pedestrian, bicyclists/bicycle, scooter, and skateboard that are associated with VRUs were used. For a crash to qualify as VRU involved crash, it should have at least one VRU involved in a crash. Two types of VRU involvements – direct and indirect, were considered. The direct involvement is described as a crash between AV and VRU, while indirect involvement focused on crashes between AV and conventional vehicle (CV) whereby AV was yielding right of way to VRU. To identify the crashes that directly involved VRU, section 2 of the crash report was used (see Fig. 2(a)). In this section, a reporter needs to fill the AV crash's participants whereby either pedestrian, bicyclist, or other options are available (CaliforniaDMV, 2020).

On the other hand, the details for crashes that involved VRU indirectly can be found in section 5 of the AV crash report (see Fig. 2(b) and (c). In this section, the operator provides the sequence of events, the key players in the crashes, and the crash outcome. While the predefined section allows the operator to fill in the person, vehicle, object that was involved in the crash, it does not provide an opportunity to explain what happened just before a crash. For instance, while crash in Fig. 2(a) would be correctly coded as it involved a scooter, the pedestrians and bicyclists' detail in Fig. 2(a) and Fig. 2(b) do not normally appear in the predefined section since the final collision was between the AV and CV. Therefore, these narratives were reviewed to extract the pre-crash information. Further, a participant who *"made contact"* to the other participant in the respective crash was identified as being at fault.

The second task involves exploring patterns of the key terms for VRU and non-VRU crashes. In this task, the text network analysis (TNA) - an unsupervised text mining approach, was used to express the interconnectedness between different keywords. TNA is a relatively new approach in the transportation field but has been extensively used in areas such as in literature and linguistic (Hunter, 2014). In recent years, TNA has emerged as a preferred text mining approach due to its strength in revealing and enabling visualization of different topics' patterns (Jiang et al., 2020; Kutela et al., 2021a; Kutela et al., 2021b; Paranyushkin, 2011). It has been used for traffic safety analysis and operations (Kutela and Teng, 2021; Kwayu et al., 2021), and for bibliometrics

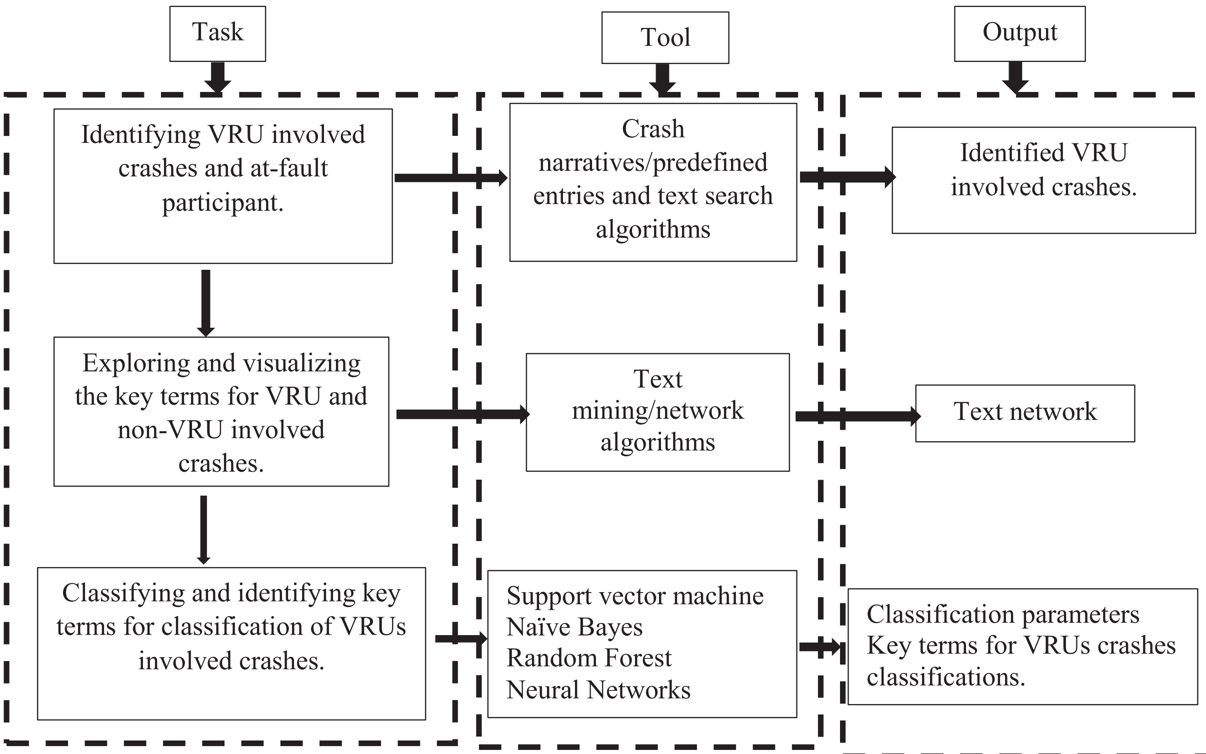
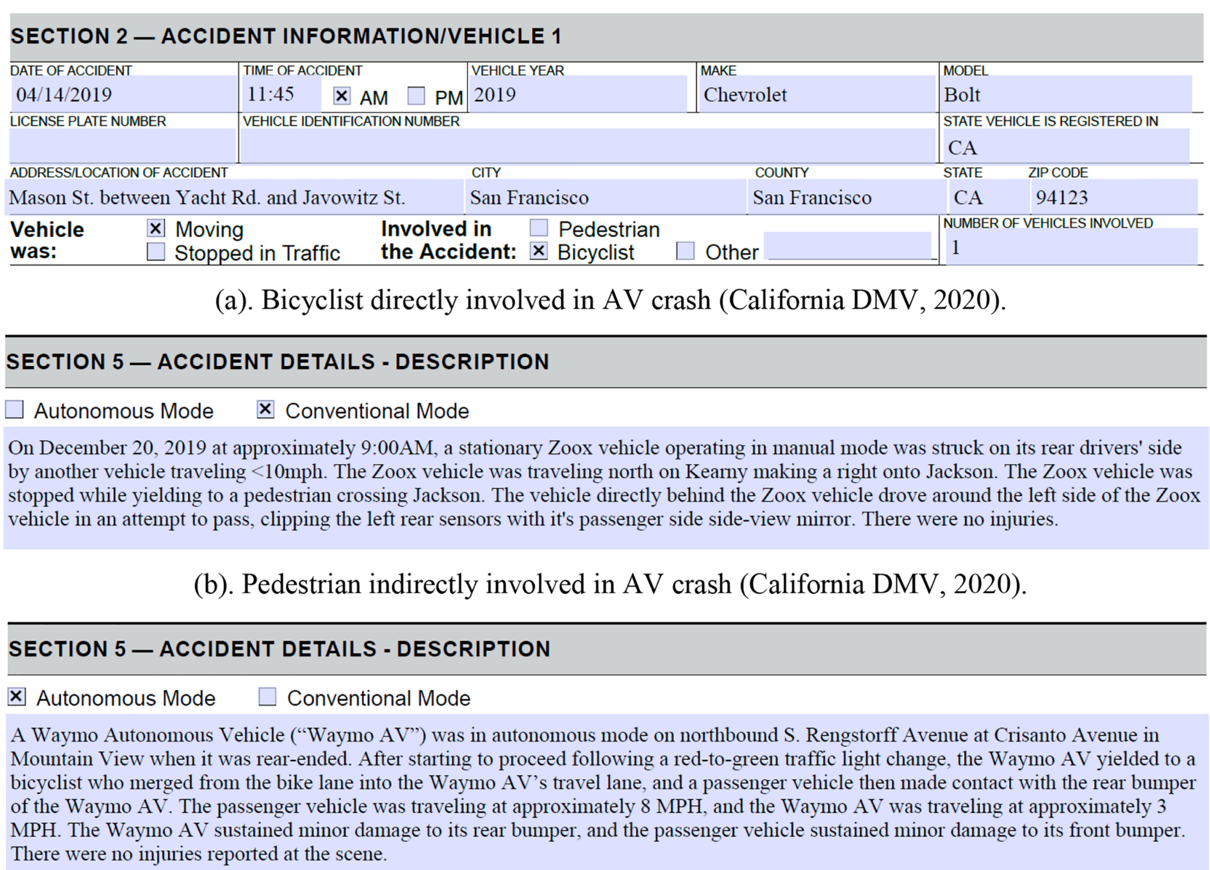


Fig. 1. Flowchart describing the methodological approach.



SECTION 2 — ACCIDENT INFORMATION/VEHICLE 1

DATE OF ACCIDENT	TIME OF ACCIDENT	VEHICLE YEAR	MAKE	MODEL
04/14/2019	11:45 ☒ AM ☐ PM	2019	Chevrolet	Bolt

LICENSE PLATE NUMBER	VEHICLE IDENTIFICATION NUMBER	STATE VEHICLE IS REGISTERED IN
		CA

ADDRESS/LOCATION OF ACCIDENT	CITY	COUNTY	STATE	ZIP CODE
Mason St. between Yacht Rd. and Javowitz St.	San Francisco	San Francisco	CA	94123

Vehicle was: ☒ Moving ☐ Stopped in Traffic **Involved in the Accident:** ☐ Pedestrian ☒ Bicyclist ☐ Other NUMBER OF VEHICLES INVOLVED 1

(a). Bicyclist directly involved in AV crash (California DMV, 2020).

SECTION 5 — ACCIDENT DETAILS - DESCRIPTION

☐ Autonomous Mode ☒ Conventional Mode

On December 20, 2019 at approximately 9:00AM, a stationary Zoox vehicle operating in manual mode was struck on its rear drivers' side by another vehicle traveling <10mph. The Zoox vehicle was traveling north on Kearny making a right onto Jackson. The Zoox vehicle was stopped while yielding to a pedestrian crossing Jackson. The vehicle directly behind the Zoox vehicle drove around the left side of the Zoox vehicle in an attempt to pass, clipping the left rear sensors with it's passenger side side-view mirror. There were no injuries.

(b). Pedestrian indirectly involved in AV crash (California DMV, 2020).

SECTION 5 — ACCIDENT DETAILS - DESCRIPTION

☒ Autonomous Mode ☐ Conventional Mode

A Waymo Autonomous Vehicle (“Waymo AV”) was in autonomous mode on northbound S. Rengstorff Avenue at Crisanto Avenue in Mountain View when it was rear-ended. After starting to proceed following a red-to-green traffic light change, the Waymo AV yielded to a bicyclist who merged from the bike lane into the Waymo AV’s travel lane, and a passenger vehicle then made contact with the rear bumper of the Waymo AV. The passenger vehicle was traveling at approximately 8 MPH, and the Waymo AV was traveling at approximately 3 MPH. The Waymo AV sustained minor damage to its rear bumper, and the passenger vehicle sustained minor damage to its front bumper. There were no injuries reported at the scene.

(c). Bicyclist indirectly involved in AV crash (California DMV, 2020).

Fig. 2. Description of VRU-involved AV crashes.

of transportation studies (Jiang et al., 2020), among others.

The TNA uses nodes and edges to map the unstructured text data. The nodes represent the keywords, while edges represent the connection between the keywords. Fig. 3 shows a typical text network topology.

To create a text network, the unstructured data are first converted to structured data by the normalization process. The process involves removing stopwords (i.e., redundant words based on the study design) and symbols, as well as converting capital case letters to lower case. The unstructured text data is converted to a matrix of keywords with their frequency of occurrence within the text data. The structured data is then mapped to create a text network- the high-frequency keywords result into the large nodes. Similarly, the keywords with high co-occurrence (occur together in a sentence) frequency result in a thick edge (Fig. 3). The closer the keywords on the network, the closer they appear in the text. Thus, keywords frequencies, co-occurrence frequency, and collocation frequency are the measures used to understand the patterns of the text network. Collocation frequency can provide richer information than co-occurrence frequency as it shows the keywords that appear next to each other (Benoit et al., 2018; Blaheta and Johnson, 2011). A group of keywords with similar theme, which most of the time are likely to be co-occurred keywords, form a community. For large networks, the communities can be identified by using statistical software (Boy, 2020). On the other hand, detection of the communities for small networks requires visual inspection (Boniphace Kutela et al., 2021). In addition, the degree centrality, which quantifies the extent of the connection between nodes that fall within the same community is used to explore the text network. The degree centrality is computed using equation (1) below.

$$Degree\ centrality\ (i) = \sum_{j=1}^{l} c_{ij} \tag{1}$$

whereby, c_{ij} takes a value of 1 if nodes i and j, are connected, and 0 otherwise.

The last task involves classifying and identifying key terms for the classification of VRUs involved crashes to understand the associated key factors for VRU-involved AV crashes. It should be noted that text classification is normally suitable for a large dataset. As the sample size for AV crashes is limited (California DMV, 2020) and the number of AV-VRU crashes is even smaller, this study can be considered a scoping study for future studies with a large dataset. In this task four supervised machine learning algorithms, which are Support Vector Machine (SVM), Naïve Bayes (NB), Random Forest (RF), and Neural Network (Nnet) were applied on the text data and their performance to classify VRUs involved crashes was compared. The best performing supervised machine learning algorithms were selected based on the three performance measures- Accuracy, Precision, and F-1 score, which are described in equation (3), 4 and 5, respectively.

$$Accuracy = \frac{TP + TN}{Total\ population} \tag{3}$$

$$Precision = \frac{TP}{TP + FN} \tag{4}$$

$$F - 1score = \frac{2 * Recall * Precision}{Recall + Precision} \tag{5}$$

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.

This study selected some of the top classification algorithms to perform this analysis. It is true that there are many other algorithms. The selection of these four algorithms is based on the findings of the other natural language processing (NLP) classification related studies (Arteaga et al., 2020; Das, Le, et al., 2020). All of these four models have been widely used in other NLP classification problems due to the high precision accuracy and less processing time. Each of the four classifiers uses different approach to classify the documents. The following section provides a brief description of each classifier. More information about these classifiers can be found in the literature (Joachims, 1998; Pranckevičius and Marcinkevičius, 2017; Yuan et al., 2019):

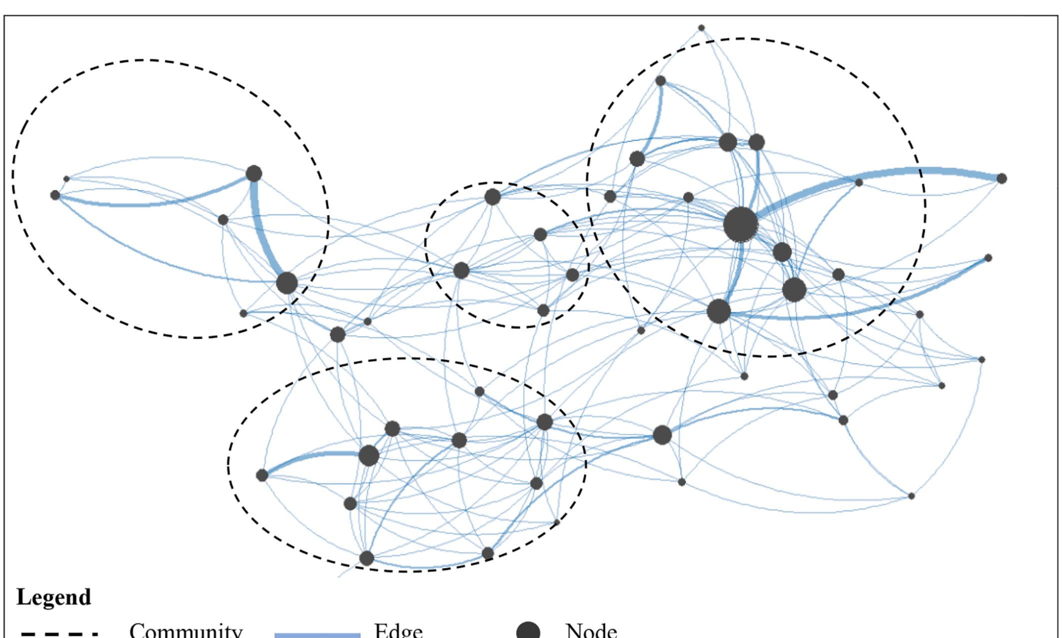


Fig. 3. A Skeleton of the Text Network (Kutela et al., 2021a).

- Support Vector Machine (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 then finds the hyperplane that best differentiate the two classes of the document/data.
- Naïve Bayes (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 certain predictors have been observed.
- Random Forest (RF) classifier works by establishing and aggregating predictions from several individual decision trees of varying depths which work as an ensemble. It searches for the best feature from random decision trees to improve classification accuracy.
- Neural Network (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. It uses back-propagation, 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.

The dataset used in this study had two classes– VRU involved and non-VRU involved crashes, obtained from crash narratives. The VRU involved crashes included all crashes associated with VRUs irrespective of either direct or indirect involvement. Prior to VRU involved crash classification, several processes were necessary to prepare the data for analysis. First, since the aim of the text classification was to determine the key features that are associated with the VRUs' involvement in AV crashes, all VRU involved terms were removed. This process was necessary to make sure that the algorithms can classify the VRUs crashes without the presence of the key VRUs' terms such as bicyclists, pedestrian, scooter, and skateboard. A similar approach was used by (Arteaga et al., 2020) when analyzing injury severity of crashes using an interpretable text mining approach. Second, since VRU involved crashes represent a small portion (about 14%) of all AV involved crashes, the study data suffers from class imbalance. Therefore, three resampling algorithms - SMOTE, over-sampling, and under-sampling (Kitali et al., 2019) were evaluated to take care of class imbalance. Further, cross-validation and bootstrap approaches were used for resampling to increase the sample size (Kitali et al., 2019). It was observed that a combination of under-sampling and bootstrap yielded models with reasonable classification accuracy. Also, different proportions of training and testing sets were examined to determine the proportion that yield high scores. At last, 70% of the data was used for training and the rest for testing to develop predictions.

The features/covariates necessary for VRU involved crash classification were developed from the text data using standard text conversion procedures. The procedures include narrative cleaning, generation of n-gram features, and generating a document matrix (Kwayu et al., 2020). Data cleaning involved the conversion of unstructured text into the corpus, removing punctuations, symbols, and numbers, and stemming (i.e., reducing a word-to-word root). 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 VRU involved crashes classification tasks.

4. Study data

In this study, the four-year (2017–2020) data of AV crashes were used. The data were collected from the California Department of Motor Vehicles (DMV) through their open-source data depository (California DMV, 2020). Since 2014, California DMV has been collecting AV crash data from the operators who have been licensed to test their AVs. Operators submit crash report to California DMV, who then post the PDF version of the reports on their website for public access. Crash narratives and other crash details were extracted and stored in excel files for further analysis.

For the past four years, a total of 252 AV crashes were reported. Most of the crashes occurred in San Francisco, Mountain View, and Palo Alto (Fig. 4). The three cities also experienced the highest number of AVs operated on their roadways (CaliforniaDMV, 2020). Therefore, crashes that occurred are proportional to the number of operating AVs. As observed, the number of AV crashes has been increasing since 2017 which can be explained by the increasing number of AVs on the road. Fewer crashes in 2020 can be explained by the travel restrictions due to the COVID-19 pandemic (Sutherland et al., 2020).

Out of 252 crashes that occurred within four years, 35 crashes involved VRUs as the key participants either in the final collision or in the intermediate events. Further, among 35 crashes that involved VRUs, 22 crashes (63%) were direct involvement whereby for almost all crashes (21 crashes) VRUs were at fault. The narratives for the crashes that VRUs were at fault revealed that three of the crashes involved some sort of intentional vandalism to the AVs. For instance, part of the GMCruise_010218 crash report states that "……..*A different pedestrian from the southwest corner of Valencia and 16th ran across Valencia Street, against the "do not walk" symbol, shouting, and struck the left side of the Cruise AV's rear bumper and hatch with his entire body…….*". On the other hand, the remaining 13 crashes (37%) involved VRUs indirectly. Such information can only be determined if the crash narratives are used for analysis, as previously described. Furthermore, out of 252 crashes, over a half (59%) occurred on signalized intersections followed by 18% on non-intersection locations, whereas stop on minor and all-way stop intersections had about 10% of all crashes each. Also, 18% of crashes occurred when it was dark with streetlights, while only 3% occurred during dusk/dawn condition. Crashes in clear weather accounted for 92% of all crashes, while dry surface condition accounted for 97%.

5. Results and discussion

This section of the manuscript presents the results and discussion. It covers the unsupervised and supervised machine learning results produced using text networks and other four classifiers- SVM, RF, NB, and Nnet.

5.1. Text network analysis results

Two R-packages *quanteda* and *igraph* (Benoit et al., 2018; Csárdi, 2020). were used to map the text and extract the performance parameters. Since the analysis resulted in a large number of keywords, only the top 50 keywords were used for text network mapping. This section presents the text network results which covers the network topology and the associated metrics such as keywords frequency, co-occurrences, collocations, and degree centralities. Three text networks – two networks for crashes that directly and indirectly involved VRUs, and one network for crashes that did not involve VRUs were developed and compared. The comparison of the text network results is based on the topology of the text network, the pattern of the topics, and the ranking of the co-occurred and collocated keywords, and degree centralities. It should be noted that the obvious keywords such as names of the AV companies (Cruise, Waymo, Zoox etc., were removed from the text during text network development as they do not add value in the study.

5.1.1. VRU directly involved in AV crashes

Fig. 5 presents the text network and the associated metrics for AV crashes that directly involved VRUs. The topology of the text networks (Fig. 5(a)) shows that it is centered at the keyword *bicyclist*, and it has four main communities. These communities describe the key participants in a crash, injury and law enforcement involvement, damage related, and direction of operating mode during the crash.

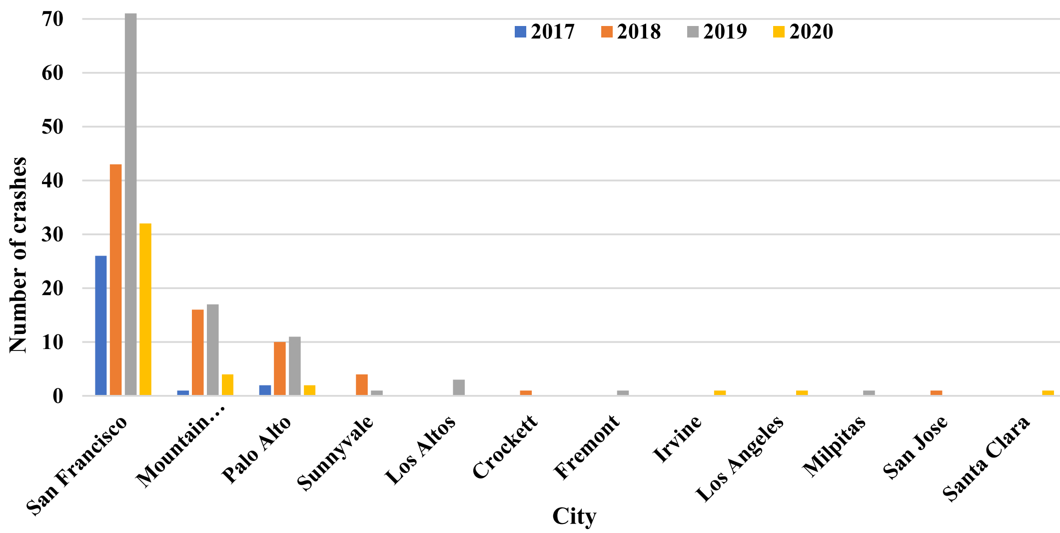


Fig. 4. Distribution of AV Crashes in California.

Fig. 5(a) shows that *bicyclists* and *electric scooters* were the key participants in the AV crashes that directly involved VRUs. This is depicted by the size of the nodes of bicyclists and scooters compared to pedestrians. The node's size, which is translated to the frequency of the keywords shown in Fig. 5(b), shows that bicyclists are the top VRUs involved in AV crashes. The frequency results show that the keyword *bicyclist* occurs 39 times in the narrative and is linked to 40 keywords as revealed by the keywords frequency and degree centrality, respectively. On the other hand, *scooter* occurs 17 times but is linked to 32 times to other keywords, while pedestrians are not among the top 20 highly ranked keywords. This observation implies that bicyclists and electric scooter (e-scooter) users are more involved in direct AV-VRU crashes than pedestrians. The keyword *bicyclist* is strongly linked to *right* and *made-contact,* among others, which makes the first community. The co-occurrence of these keywords shows the participant that hit the other participant (*made-contact*) and the side that was hit (*right*). In fact, according to the crash narratives, ten out of 22 crashes involved bicyclists hitting the AV, while eight out of 22 crashes involved electric scooters hitting AVs (California DMV, 2020).

The second community shows the injuries and the police's involvement after crashes as indicated by the keyword *police, not-called, scene,* and *no-injuries*. The community shows that most crashes did not result in any injuries, as shown by the keyword no-injuries; as a result, police were not called. Further, the same community shows that for a relatively small number of crashes, the participants left without exchanging information, as shown by co-occurred keywords *without information*. In fact, about 27% (six out of 22) of crashes involved bicyclists riding away without exchanging information with the AV operators. The literature shows that, on average, 8.1% of crashes in California are hit and run (Tay et al., 2009). Thus, the hit and run statistics for AV crashes are relatively higher than that of conventional vehicles presented in the literature. Moreover, the traditional hit and run crashes involve vehicle hitting VRU; conversely, for the AV crashes, the narratives reported by the operators show that VRU hit the AVs.

The third community that describes the AV operating modes and direction during crashes is centered on the keywords *autonomous* and *mode*. Other keywords in this community include *traveling, intersection, operating, conventional southbound,* and *westbound,* to mention a few. A comparison between the operating mode shows that most of the VRU directly involved crashes occurred when AVs were in *autonomous mode* as compared to the *conventional mode* as depicted by the ranking of these keywords in the co-occurrence results (see Fig. 5(b)).

This observation is against the conventional understanding of AV operations, as AVs are programmed to yield to VRUs when they are in autonomous mode. However, the fact that most of the crashes involved at fault VRUs can explain the unconventional scenario. Furthermore, most of these crashes occurred at the intersections whereby the AVs vehicles were stopped during the red light. This is revealed by the presence of the keywords *intersection, stop, red,* and *light.*

The fourth community describes the damage-related themes presented by keywords *damage, rear, left, side,* and *made-contact.* It can be observed that most of the VRU directly involved crashes involved the rear-left side and rear-right side of the AV. A relatively small number of crashes involved damaged vehicles, as revealed by the low frequency of the keyword damage. Similarly, a relatively small percentage of crashes involved the bumper of the AV, as indicated by the size of the keyword *bumper* in the text network.

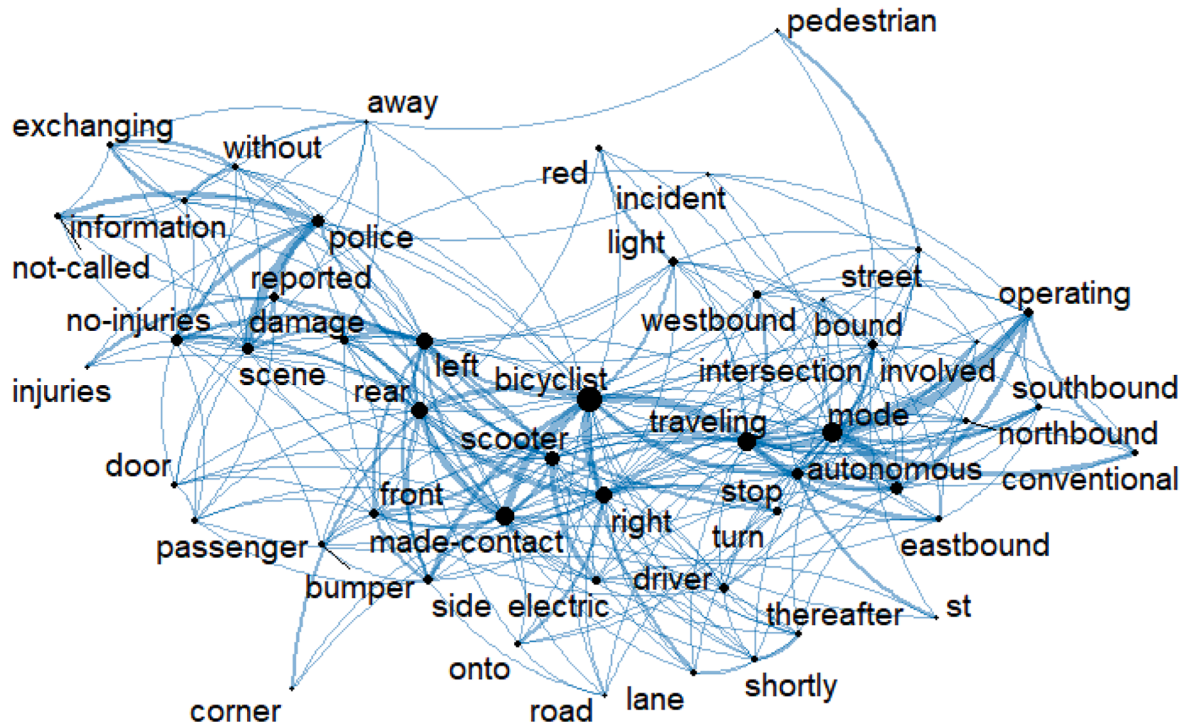
5.1.2. *VRU indirectly involved in AV crashes*

Fig. 6 presents the text network and the associated metrics for AV crashes that indirectly involved VRUs. Similar to Fig. 5, the communities describing the injury and law enforcement, location and direction, as well as key participants, are presented in text network in Fig. 6(a).

Contrary to the text network in Fig. 5(a), Fig. 6(a) shows that among VRUs, *pedestrians* occur more frequently and have the largest degree centrality score among VRUs. The observations suggest that pedestrians are more involved in AV crashes indirectly. In fact, according to the crash narratives, ten out of 13 AV crashes that indirectly involved VRU were the AVs hit by another vehicle when yielding to pedestrians, while bicyclist and scooter accounted for two and one crash, respectively. Further, while the collocation results in Fig. 6(b) do not show the keywords *conventional mode,* the two keywords are ranked third in the crashes that directly involved VRUs (Fig. 5(b). The observation implies that fewer crashes that indirectly involved VRUs occurred when the AVs were in conventional mode.

Furthermore, the text network for the indirect involvement of the VRUs shows that keywords and connections between keywords damage, rear, and bumper have a higher frequency than the crashes that directly involve VRUs. The observation implies that crashes that indirectly involve VRUs are more likely to result in damages, especially on the rear bumper, compared to the ones involving VRUs directly. This is because the VRU indirectly involved crashes are more likely to be between AV and CV. However, according to the raw narratives, the proportion of the crashes that resulted in minor damage is relatively small (two out of 13 crashes).

Contrary to crashes that directly involved VRUs, the results in Fig. 6 (b) show that the topmost keyword and co-occurred keyword are about the areas where the crashes happen. Co-occurred keywords such as rear-



(a). Text Network Topology

(b). Text Network Metrics

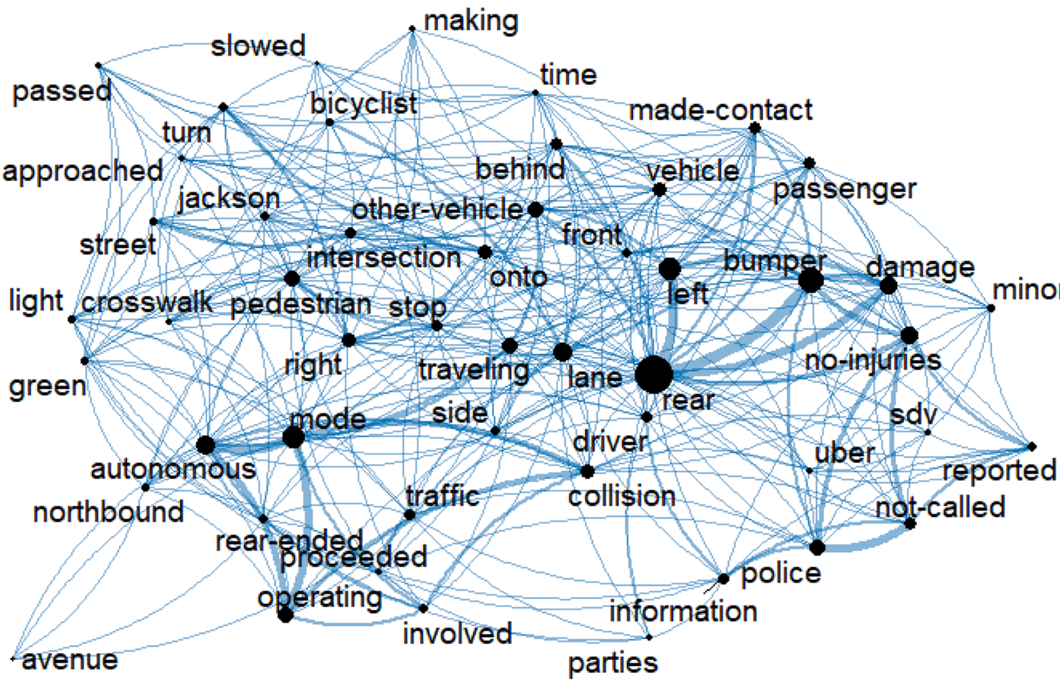
Fig. 5. Text Network Topology and Associated Metrics for VRUs' Directly Related Crashes.

bumper, left-rear, lane. However, the collocated keyword continues to describe the mode and police involvement after the crash. The keyword rear appears to have the highest frequency and degree centrality. The observation implies that most of these crashes affected the rear side of the vehicle.

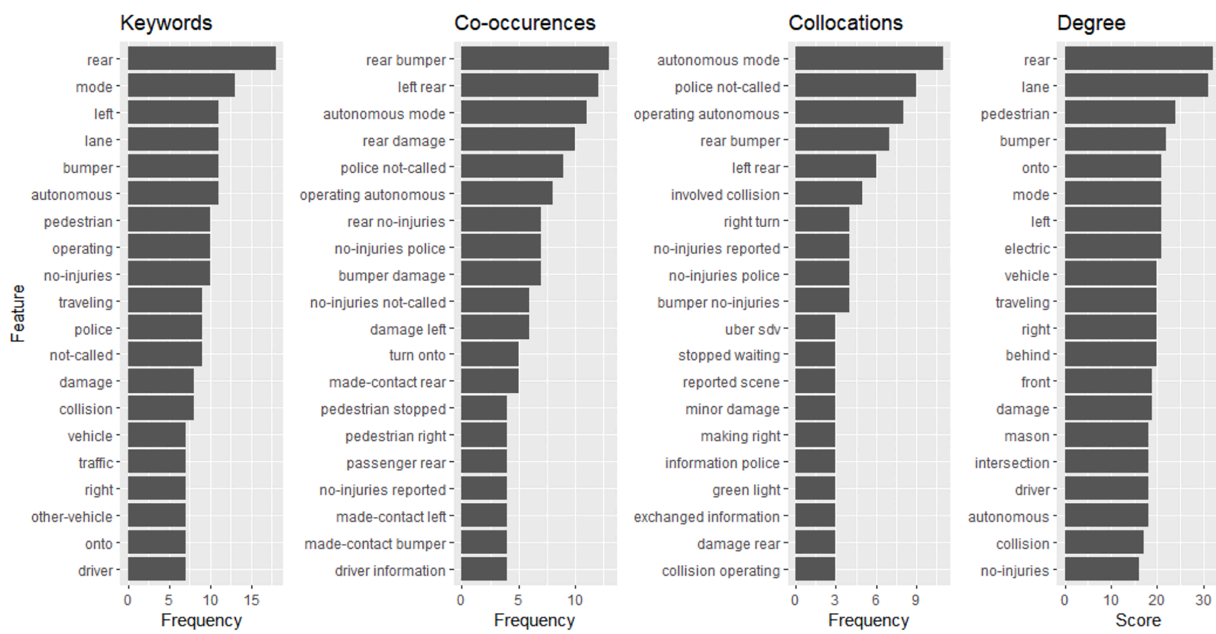
5.1.3. *Non-VRU involved crashes*

To better assess the patterns for VRU involved crashes, a comparison of VRU and non-VRU involved crashes was necessary. Fig. 7 presents the text network and associated metrics for non-VRU involved crashes. Compared to the previous two networks, a number of similarities and differences can be observed. The three communities, which are police involvement, mode of operation, and damaged vehicle position, are similar to the previous figures.

Contrarily, the disengagements from autonomous mode to conventional mode were observed for some crashes as revealed by the rare but



(a). Text Network Topology



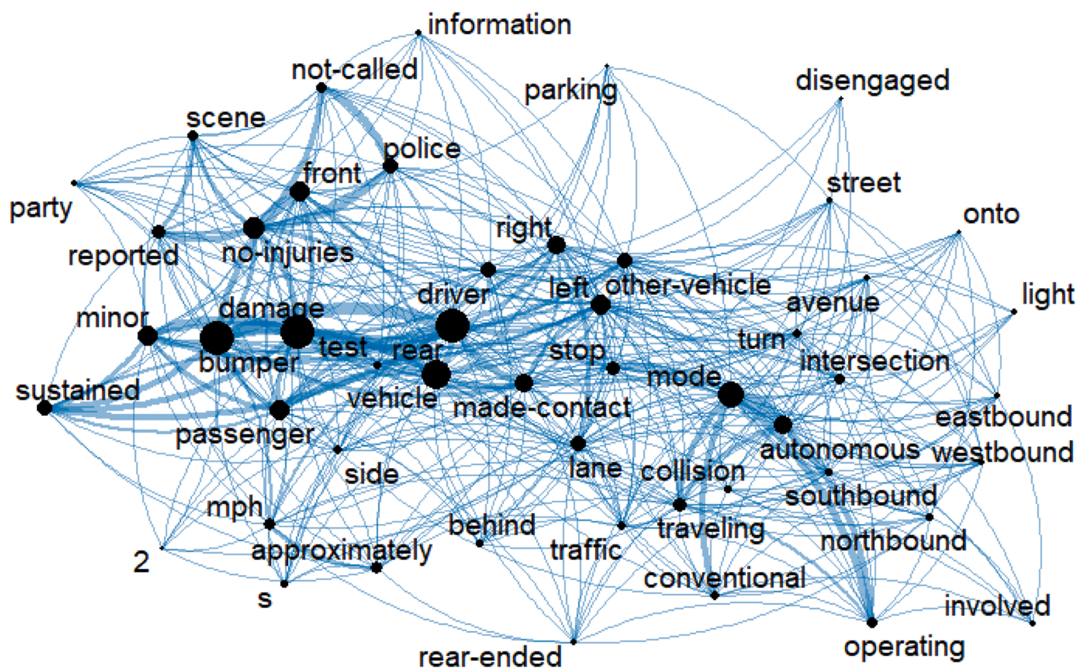
(b). Text Network Metrics

Fig. 6. Text Network Topology and Associated Metrics for VRUs' Indirectly Related Crashes.

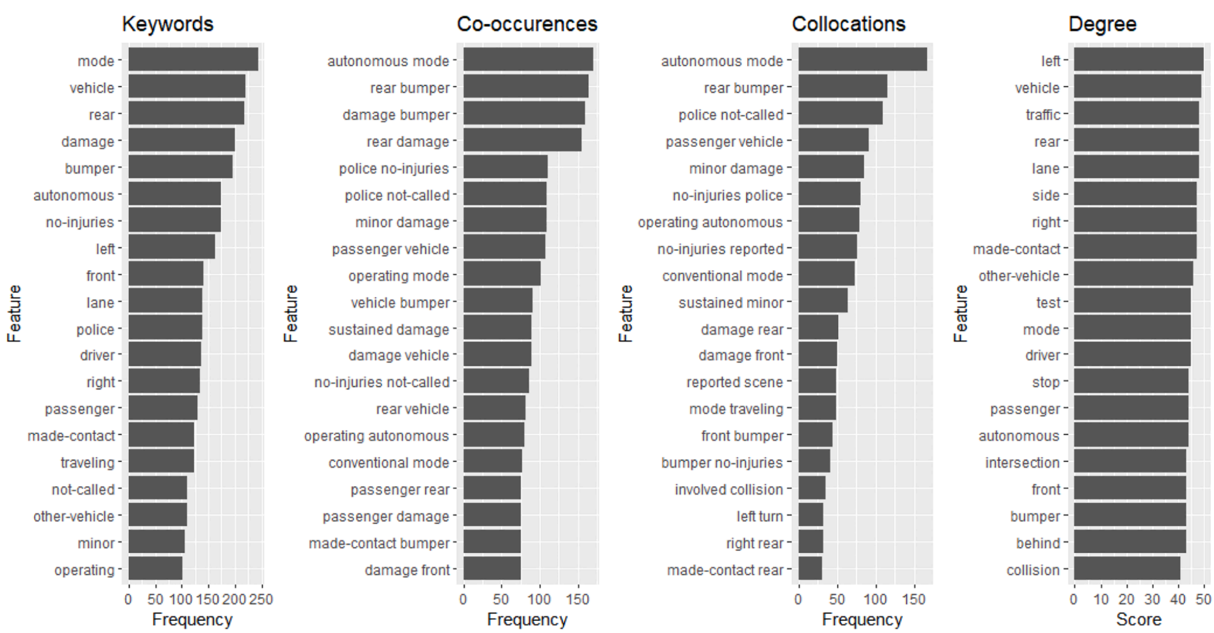
important keyword *disengaged*. Further, the keyword *left* emerged as the top-ranked in terms of degree centrality, meaning that it has the highest number of connections with other keywords. In this context, the keyword *left* has three meanings; the first is that the other vehicle's driver left the scene (hit and run crash), while the second is the location where the contact was made (i.e., left side). The third meaning is the turn made by either AV or CV during the crash (e.g., left turn). The raw data shows that 108 out of 252 crashes involved either contact made on the left side or left-turning vehicles (California DMV, 2020). On the other hand, about 10% of non-VRU involved crashes, drivers of the other vehicles left the scene without exchanging contact information.

Although all three networks display injury-related and police involvement topics, it can be observed that more crashes involved minor damages on the rear bumper for non-VRU involved crashes compared to VRU involved crashes. The text network for non-VRUs involved crashes Fig. 7 (a) revealed that the nodes for keywords *minor, damage, rear,* and *bumper* are significantly larger than corresponding nodes in Fig. 5(b) and Fig. 6(b). A similar pattern is observed for crashes that indirectly involve VRUs, but slightly different in terms of co-occurrence frequency for crashes that directly involved VRUs.

The text network analysis facilitated visualizing the patterns of most frequent keywords and communities for VRU and non-VRU involved



(c). Text Network Topology



(d). Text Network Metrics

Fig. 7. Text Network Topology and Associated Metrics for non-VRUs' Related Crashes.

crashes. Through text networks analysis, the patterns for the VRU involved crashes were identified using the keyword frequency, colocations, and degree centrality measures. However, such measures do not indicate the important features for the classification of the VRU involved crashes. Thus, the next section presents results from supervised machine learning classifiers to predict the important features for the classification of VRU involved crashes.

5.2. VRUs' involved crash classification

Among 252 crashes, 35 crashes involved VRU crashes, while 217 were non-VRU involved crashes. The dataset was divided into two sets- training data (70%) and testing data (30%). By utilizing bootstrapped and under-sampled data, four classifiers were compared on their capabilities to perform classification for the VRUs involved crashes. Three performance metrics- accuracy, recall, and F-1 score were used to determine the best performing model. Table 1 presents the confusion matrices, and Fig. 8 summarizes performance measures for balanced and imbalanced data for each algorithm.

According to the results in Fig. 8, classification accuracy from imbalanced data is relatively high for SVM, RF, and Nnet; however, both SVM and RF have no recall and F-1 scores. This is due to the reason that

Table 1
Confusion Matrices for Imbalanced and Balanced Data by Algorithm.

Data	Algorithm	Prediction	Reference class: VRU related	Reference class: Non-VRU related
Imbalanced data	SVM	VRU related	0	0
Imbalanced data	SVM	Non-VRU related	13	62
Imbalanced data	NB	VRU related	6	31
Imbalanced data	NB	Non-VRU related	7	31
Imbalanced data	RF	VRU related	0	0
Imbalanced data	RF	Non-VRU related	13	62
Imbalanced data	Nnet	VRU related	4	3
Imbalanced data	Nnet	Non-VRU related	9	59
Balanced data	SVM	VRU related	12	34
Balanced data	SVM	Non-VRU related	1	28
Balanced data	NB	VRU related	8	27
Balanced data	NB	Non-VRU related	5	35
Balanced data	RF	VRU related	10	16
Balanced data	RF	Non-VRU related	3	46
Balanced data	Nnet	VRU related	9	16
Balanced data	Nnet	Non-VRU related	4	46

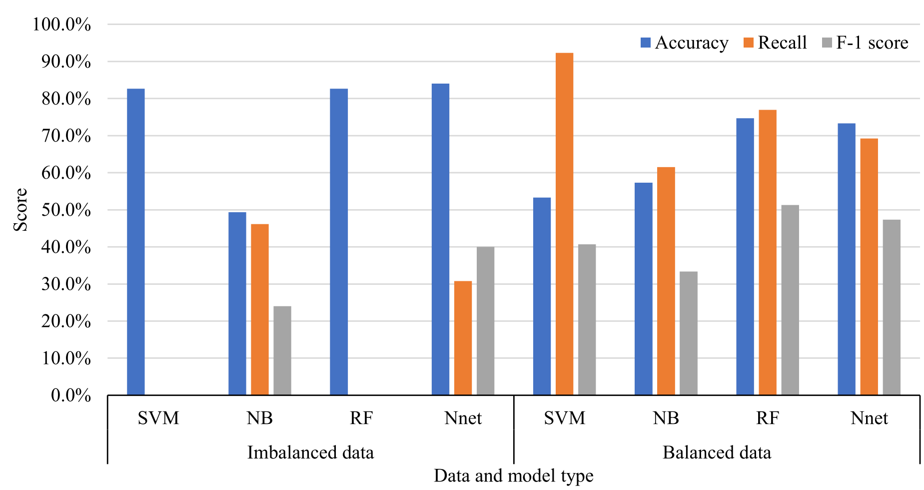


Fig. 8. Comparable Performance Metrics for Imbalanced and Balanced Data.

the two classifiers could not correctly classify any VRU related crashes (see Table 1). On the other hand, there are several improvements observed with the use of balanced data. Results from balanced data show that SVM has the highest capability to classify the VRU related crashes, as revealed by the recall score of 92.3%. However, the same algorithm has the lowest overall classification accuracy (53.3%). On the other hand, the NB algorithm has the lowest recall score (61.5%) and F-1 score (33.3%). The RF classifier performed relatively better than the rest as it has the highest prediction accuracy and F-1 scores, and a relatively high Recall score. The next to RF, in terms of overall performance, is the Nnet.

5.3. Importance features

As a result of text processing, over 300 features were created. With that large number of features, it is important to extract a few features that are significant for the classification of VRU involved crashes.

Fig. 9 shows the 20 most important features as classified by the four classifiers. Across classifiers, there are several similarities and differences as numerous key features appear in each algorithm. Such features include *crosswalk, street, vehicle,* among others. However, their positions in terms of their importance differ significantly except for the two algorithms, SVM and NB, which have a similar list of important variables. The SVM and NB show that *vehicle* is the most important feature in classifying the VRUs' crashes. On the other hand, RF and Nnet show that *street* and *crosswalk*, respectively, are the topmost key features for the classification of VRUs' crashes. These are the locations with higher exposure between VRUs and AV; therefore, it makes sense that they have been identified as the topmost features for the classification of the VRU involved crashes. In fact, the raw data shows that about eight out of 35 VRU involved crashes occurred at signalized intersections where the AV

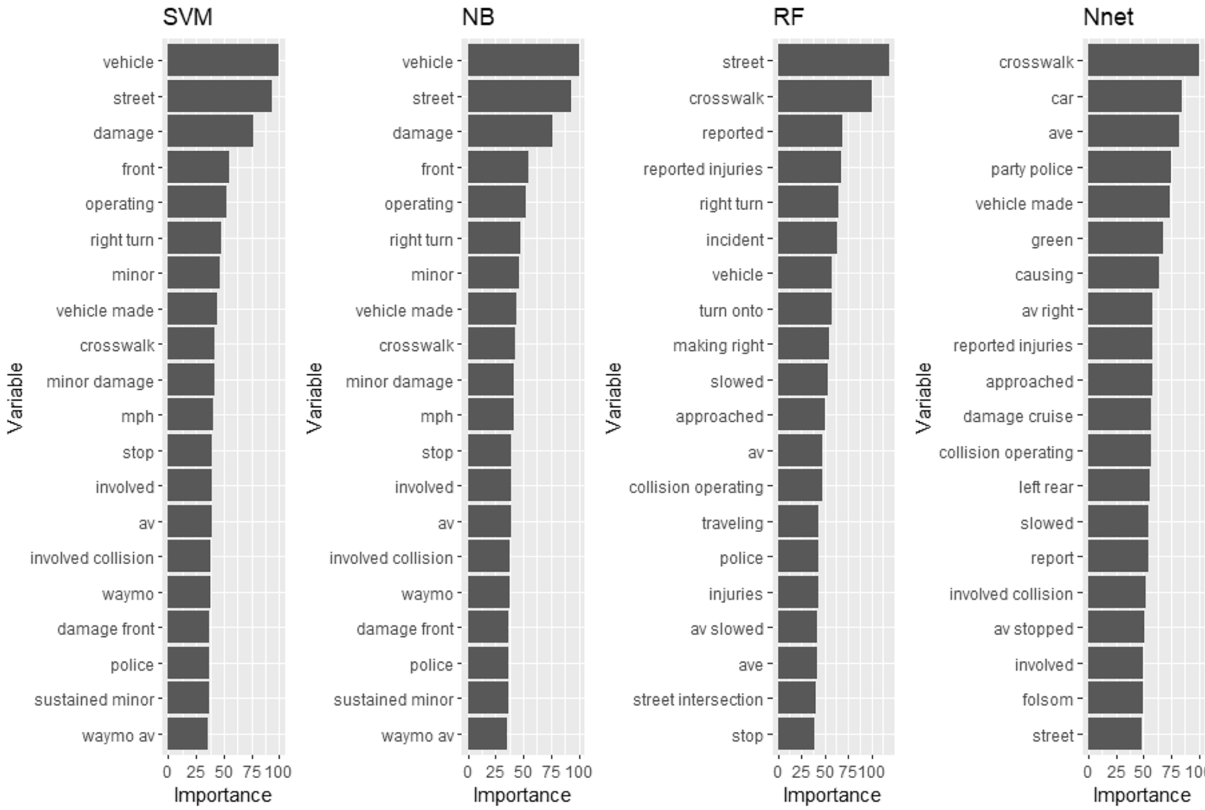


Fig. 9. Top 20 Important Features.

yielded right of way to VRUs.

Moreover, the RF revealed that several turning-related features performed well in the classification of VRU- involved crashes. Terms such as *AV right, right turn, turn onto, street intersection,* and *making right,* show that VRUs were involved in the crashes where either AV or CV was turning. This is true, as revealed by the row data, that 20% of VRU-AV crashes involved AV turning left/right.

Furthermore, the important features from Nnet and RF algorithms revealed several AV and CV-related actions that are associated with the VRUs' involvement in crashes. The features such as *AV stopped, AV slowed* show that VRUs are likely to be involved in crashes when the AVs are either slowing or stopping to allow VRUs to cross. This is true, as revealed by 49% of VRU-AV crashes (17 out of 35 crashes) that AVs were rear-ended after slowing/stopping for VRUs.

It is worthy to mention that variable importance measures are strongly related to the underlying model and the optimization techniques as different machine learning models uses a variable in a different way. For example, some algorithms may recursively partition it while other algorithms can fit a spline in the mechanism. However, some of the important variables and features could be common in different models.

6. Conclusions

This study explored the patterns of VRU-involved AV crashes to understand the associated factors. We used four years (2017–2020) of AV crash data collected by the California Department of Motor Vehicles, whereby a total of 252 crashes were recorded within the four-year period. The study applied the text network technique and four classifiers- SVM, NB, RF, and Nnet to extract the VRU-involved AV crash patterns and compare the VRU and non-VRU crashes. The text network topology, keyword frequency, collocations, and degree centrality were used to explore the VRUs and non-VRU related crashes. Further, classification accuracy, recall, and F-1 scores were used to assess the performance of the four classifiers. The key features associated with VRU involved crashes were extracted using the variable importance algorithm. To this end, several conclusions can be made.

This study presented the text mining approach to explore the involvement of the VRUs in AV crashes using unstructured narrative data. The use of the narratives helped to obtain additional information that led to identifying the involvement of VRUs in AV crashes. Using the crash narrative data, we found that VRUs' role can be either directly or indirectly involved in AV crashes. Direct involvement includes crashes that involve direct contact between VRU to the AV, while indirect involvement covers crashes whereby the AV was accommodating VRU's needs but ended up in a crash with CV. The latter is common when the AV yields right of way to VRUs, then CV crash on the AV. The indirect involvement of VRUs in the AV crashes could not be identified using traditional crash data whereby the crash participants are entered in the predefined fields.

According to crash reports, the study found that the VRUs were at fault in a relatively large percent of crashes that directly involved VRUs. The fact that the AVs were at fault for only one out of 22 crashes indicates a great potential for safety improvement associated with the AVs. However, this observation should be interpreted with a great caution as the crash narratives are reported by the operators and not the police.

Further, the study found that CVs were at fault for all 13 crashes that indirectly involved VRUs. This is because the AVs are programmed to yield to VRUs, while for CV to yield, it depends on the driver's safety perception. This area needs more research and improvements, especially when AVs are operating in mixed traffic. The improvement would include enabling the AVs to learn from the surrounding CVs regardless of the presence of VRUs who wants to cross the roadway. For example, if the AV is in the middle of CVs, and all CVs do not yield right of way to VRUs, the AV should also not yield to avoid collision with other CVs.

The text network analysis results revealed several similarities for VRU and non-VRU involved crashes. The key differences displayed were on the injury and damage patterns. The topology and associated keywords revealed that VRU involved crashes are less severe compared to non-VRU involved crashes. Furthermore, among VRUs, we found that bicyclists are more likely to be directly associated with AV crashes. Compared to scooters (36%) and pedestrians (14%), bicyclists were involved in a large proportion of crashes (50%). This includes 45% that bicyclists were at fault and 18% hit and run crashes involving bicyclists.

The extracted variables using RF and Nnet classifiers pointed out crosswalk, and intersection-related features as the key variables for the classification of VRU involved crashes. The keyword crosswalk emerged as the top feature, while the keyword green, which indicates green light at the signalized intersections, was among the top 10 key features. Furthermore, intersection-related terms, especially the turning terms such as right turn, making the turn, AV right, AV stopped, AV slowed, emerged as the key terms for classification of the VRUs crashes. Such findings point out more research needs on the areas with high exposure to VRUs.

7. Limitations and future directions

This study has several limitations which are needed be addressed in future studies. First, the data used were reported by the operators. The accuracy of the utilized data at this moment is not known; thus, the reliability of the findings would depend on the trustworthiness of the reporter. To overcome this limitation, future studies might identify the police reports of the same crash and compare the narratives.

Secondly, this study used descriptive analysis, text network, and supervised machine learning. The descriptive analysis was used to quantify the nature of the problem, while text network and associated metrics were used to reveal the pattern associated with the VRU involved crashes. Although these methodologies have facilitated understanding the nature of VRU involved crashes, the available sample size could not facilitate extensive analysis of supervised machine learning. Similarly, the sample size issue, especially for the number of VRU involved crashes, could not facilitate traditional regression approaches or Bayesian regression. Therefore, in the future, studies can use machine learning, Bayesian regression, and traditional regression approaches to extract more insights from AV-VRU crashes.

Thirdly, the acquired data is limited in size. It is worthy to mention that AV-VRU collision data is limited. This study conducted this analysis based on the latest available data. Results from the current analysis require careful interpretation due to this data size limitation. Future studies can further explore this topic with a larger dataset.

Funding

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

CRediT authorship contribution statement

Boniphace Kutela: Conceptualization, Data curation, Formal analysis, Methodology, Software,Validation, Writing – original draft. Subasish Das: Formal analysis, Methodology, Writing – review & editing. Bahar Dadashova: Formal analysis, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to acknowledge the California Department of Motor Vehicles for the open-source AV-involved crashes used in this study.

References

Alambeigi, H., McDonald, A.D., Tankasala, S.R., 2020. Crash Themes in Automated Vehicles: A Topic Modeling Analysis of the California Department of Motor Vehicles Automated Vehicle Crash Database. http://arxiv.org/abs/2001.11087.

Arteaga, C., Paz, A., Park, J.W., 2020. Injury severity on traffic crashes: A text mining with an interpretable machine-learning approach. Saf. Sci. 132, 104988. https://doi.org/10.1016/j.ssci.2020.104988.

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., Matsuo, A., 2018. quanteda: An R package for the quantitative analysis of textual data. J. Open Source Software 3 (30). https://doi.org/10.21105/joss.00774.

Blaheta, D., Johnson, M., 2011. Unsupervised learning of multi-word verbs *. In: Proceedings of the ACL Workshop on Collocations, pp. 54–60.

Boggs, A.M., Wali, B., Khattak, A.J., 2020. Exploratory analysis of automated vehicle crashes in California: a text analytics & hierarchical Bayesian heterogeneity-based approach. Accid. Anal. Prev. 135, 105354 https://doi.org/10.1016/j.aap.2019.105354.

Boy, J.D., 2020. textnets: A Python package for text analysis with networks. J. Open Source Software 5 (54). https://doi.org/10.21105/joss.02594.

Brar, J. S., & Caulfield, B. (2018). Impact of autonomous vehicles on pedestrians' safety. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, 2018-March, 714–719. doi: 10.1109/ITSC.2017.8317963.

Burns, C.G., Oliveira, L., Hung, V., Thomas, P., Birrell, S., 2019. Pedestrian Attitudes to Shared-Space Interactions with Autonomous Vehicles – A Virtual Reality Study. Adv. Intelligent Systems Computing 964, 307–316. https://doi.org/10.1007/978-3-030-20503-4_29.

CaliforniaDMV. (2020). Autonomous Vehicle Collision Reports – California DMV. https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/autonomous-vehicle-collision-reports/.

Combs, T.S., Sandt, L.S., Clamann, M.P., McDonald, N.C., 2019. Automated vehicles and pedestrian safety: exploring the promise and limits of pedestrian detection. Am. J. Prev. Med. 56 (1), 1–7. https://doi.org/10.1016/j.amepre.2018.06.024.

Csárdi, G. (2020). "igraph"; Network Analysis and Visualization.

Das, S., 2021. Autonomous vehicle safety: understanding perceptions of pedestrians and bicyclists. Transp. Res. Part F: Traffic Psychol. Behav. 81, 41–54. https://doi.org/10.1016/J.TRF.2021.04.018.

Das, S., Dutta, A., Fitzpatrick, K., 2020a. Technological perception on autonomous vehicles: perspectives of the non-motorists. Technol. Anal. Strategic Management 32 (11), 1335–1352. https://doi.org/10.1080/09537325.2020.1768235.

Das, S., Dutta, A., Tsapakis, I., 2020b. Automated vehicle collisions in California: applying Bayesian latent class model. IATSS Res. 44 (4), 300–308. https://doi.org/10.1016/j.iatssr.2020.03.001.

Das, S., Le, M., Dai, B., 2020c. Application of machine learning tools in classifying pedestrian crash types: a case study. Transp. Safety Environ. 2 (2), 106–119. https://doi.org/10.1093/TSE/TDAA010.

Das, S., & Zubaidi, H. (2021). Autonomous Vehicles and Pedestrians: A Case Study of Human Computer Interaction. 226–239. doi: 10.1007/978-3-030-78358-7_15.

Deb, S., Strawderman, L., Carruth, D.W., DuBien, J., Smith, B., Garrison, T.M., 2017. Development and validation of a questionnaire to assess pedestrian receptivity toward fully autonomous vehicles. Transp. Res. Part C: Emerging Technol. 84, 178–195. https://doi.org/10.1016/j.trc.2017.08.029.

Favarò, F.M., Nader, N., Eurich, S.O., Tripp, M., Varadaraju, N., 2017. Examining accident reports involving autonomous vehicles in California. PLoS ONE 12 (9), e0184952. https://doi.org/10.1371/journal.pone.0184952.

Fleskes, K., Hurwitz, D.S., 2019. Influence of bicyclist presence on driver performance during automated vehicle take-over requests. Transp. Res. Part F: Traffic Psychol. Behav. 64, 495–508. https://doi.org/10.1016/j.trf.2019.06.007.

Gelbal, S.Y., Aksun-Guvenc, B., Guvenc, L., 2020. Collision avoidance of low speed autonomous shuttles with pedestrians. Int. J. Automot. Technol. 21 (4), 903–917. https://doi.org/10.1007/s12239-020-0087-7.

Goodall, N.J., 2021. Comparison of automated vehicle struck-from-behind crash rates with national rates using naturalistic data. Accid. Anal. Prev. 154, 106056. https://doi.org/10.1016/j.aap.2021.106056.

Hunter, S., 2014. A novel method of network text analysis. Open J. Modern Linguistics 04 (02), 350–366. https://doi.org/10.4236/ojml.2014.42028.

Jiang, C., Bhat, C.R., Lam, W.H.K., 2020. A bibliometric overview of Transportation Research Part B: Methodological in the past forty years (1979–2019). Transp. Res. Part B: Methodol. 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 (pp. 137–142). doi: 10.1007/bfb0026683.

Kalatian, A., Farooq, B., 2021. Decoding pedestrian and automated vehicle interactions using immersive virtual reality and interpretable deep learning. Transp. Res. Part C: Emerging Technol. 124, 102962. https://doi.org/10.1016/j.trc.2020.102962.

Kitali, A.E., Alluri, P., Sando, T., Wu, W., 2019. Identification of secondary crash risk factors using penalized logistic regression model. Transp. Res. Record: J. Transp. Res. Board 2673 (11), 901–914. https://doi.org/10.1177/0361198119849053.

Kutela, B., Novat, N., Langa, N., 2021a. 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.

Kutela, B., Langa, N., Mwende, S., Kidando, E., Kitali, A.E., Bansal, P., 2021b. A text mining approach to elicit public perception of bike-sharing systems. Travel Behav. Soc. 24, 113–123. https://doi.org/10.1016/j.tbs.2021.03.002.

Kutela, B., Teng, H., 2021. The use of dynamic message signs (DMSs) on the freeways: an empirical analysis of DMSs logs and survey data. J. Transp. Technol. 11 (01), 90–107. https://doi.org/10.4236/jtts.2021.111006.

Kwayu, K.M., Asce, S.M., Kwigizile, V., Asce, M., Zhang, J., Asce, A.M., Oh, J.-S., 2020. Semantic N-Gram Feature Analysis and Machine Learning-Based Classification of Drivers' Hazardous Actions at Signal-Controlled Intersections. doi: 10.1061/(ASCE)CP.1943-5487.0000895.

Kwayu, K.M., Kwigizile, V., Lee, K., Oh, J.-S., Oh, J.-S., 2021. Discovering latent themes in traffic fatal crash narratives using text mining analytics and network topology. Accid. Anal. Prev. 150, 105899 https://doi.org/10.1016/j.aap.2020.105899.

Mahadevan, K., Sanoubari, E., Somanath, S., Young, J.E., Sharlin, E., 2019. AV-Pedestrian Interaction Design Using a Pedestrian Mixed Traffic Simulator. In: Proceedings of the 2019 on Designing Interactive Systems Conference. https://doi.org/10.1145/3322276.

Millard-Ball, A., 2018. Pedestrians, autonomous vehicles, and cities. J. Planning Educ. Res. 38 (1), 6–12. https://doi.org/10.1177/0739456X16675674.

NHTSA, 2015. Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey.

Paranyushkin, D., 2011. Identifying the Pathways for Meaning Circulation using Text Network Analysis. Venture Fiction Practices, 2(4). www.noduslabs.com.

Pranckevičius, T., Marcinkevičius, V., 2017. Comparison of naïve bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification. Baltic J. Modern Computing 5 (2), 221–232. https://doi.org/10.22364/bjmc.2017.5.2.05.

Pyrialakou, V.D., Gkartzonikas, C., Gatlin, J.D., Gkritza, K., 2020. Perceptions of safety on a shared road: driving, cycling, or walking near an autonomous vehicle. J. Saf. Res. 72, 249–258. https://doi.org/10.1016/j.jsr.2019.12.017.

Rahman, M.T., Dey, K., Das, S., Sherfinski, M., 2021. Sharing the road with autonomous vehicles: a qualitative analysis of the perceptions of pedestrians and bicyclists. Transp. Res. Part F: Traffic Psychol. Behav. 78, 433–445. https://doi.org/10.1016/J.TRF.2021.03.008.

Rasouli, A., Tsotsos, J.K., 2020. Autonomous vehicles that interact with pedestrians: a survey of theory and practice. IEEE Trans. Intell. Transp. Syst. 21 (3), 900–918. https://doi.org/10.1109/TITS.697910.1109/TITS.2019.2901817.

Reig, S., Norman, S., Morales, C. G., Das, S., Steinfeld, A., & Forlizzi, J. (2018). A field study of pedestrians and autonomous vehicles. Proceedings – 10th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2018, 198–209. doi: 10.1145/3239060.3239064.

Sutherland, M., Mckenney, M., & Elkbuli, A. (2020). Vehicle related injury patterns during the COVID-19 pandemic: What has changed? doi: 10.1016/j.ajem.2020.06.006.

Tay, R., Barua, U., Kattan, L., 2009. Factors contributing to hit-and-run in fatal crashes. Accid. Anal. Prev. 41 (2), 227–233. https://doi.org/10.1016/j.aap.2008.11.002.

Wali, B., Khattak, A.J., Ahmad, N., 2021. Injury severity analysis of pedestrian and bicyclist trespassing crashes at non-crossings: a hybrid predictive text analytics and heterogeneity-based statistical modeling approach. Accid. Anal. Prev. 150, 105835. https://doi.org/10.1016/j.aap.2020.105835.

Yuan, J., Abdel-Aty, M., Gong, Y., Cai, Q., 2019. Real-time crash risk prediction using long short-term memory recurrent neural network. Transp. Res. Record: J. Transp. Res. Board 2673 (4), 314–326. https://doi.org/10.1177/0361198119840611.