Routledge
Taylor & Francis Group

International Journal of Urban Sciences

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rjus20

Exploration on prior driving modes for automated vehicle collisions

Subasish Das, Xiaoqiang "Jack" Kong & Md Mahmud Hossain

INTERNATIONAL JOURNAL OF URBAN SCIENCES
2023, VOL. 27, NO. 4, 622–644
https://doi.org/10.1080/12265934.2022.2142650

Exploration on prior driving modes for automated vehicle collisions

Subasish Das [a], Xiaoqiang "Jack" Kong [b] and Md Mahmud Hossain [c]

[a]Texas State University, San Marcos, TX, USA; [b]Texas A&M University, College Station, TX, USA; [c]Auburn University, Auburn, AL, USA

ABSTRACT

The emergence of automated vehicles (AV) has been occurring rapidly as these vehicles have the potential to reduce/eradicate human driving faults and related collisions. To enhance AV safety, the NHTSA recommends the continuous presence of a backup human driver that has a reasonable understanding of AV technologies to ensure disengagement when manual overtake is required. However, due to several AV-related traffic crashes during roadway testing and extensive media interest, AV safety has become a critical issue. This study collected 255 crash reports filed by different manufacturers testing AVs in California from September 2014 to April 2020. The crash dataset was analyzed using two data mining algorithms (association rule mining and text network analysis) to identify the key AV-related crash attributes and their associations based on the vehicle's prior driving mode (conventional or automated). The results show that the manner of collision and the prior movement of the testing vehicle are strongly connected with prior driving mode. For example, AV crashes in a manual driving mode often result in a sideswipe collision during moving status, whereas AV crashes in an automated driving mode are highly associated with rear-end collisions when AVs are stopped in traffic. The findings of this study can help policymakers and AV engineers improve AV deployment strategies to support the adoption of AVs and promote potential safety benefits for AV technologies.

ARTICLE HISTORY
Received 23 March 2022
Accepted 16 October 2022

KEYWORDS
Automated vehicle; crash pattern; crash data analysis; prior driving mode; data mining

Highlights

- Applied data mining on automated vehicle related crash reports in California.
- Generated 20 rules for each of the vehicle prior modes (automated vs. conventional).
- Automated driving modes are highly associated with rear-end collisions.
- Conventional driving models are highly associated with sideswipe collisions.

CONTACT Subasish Das subasish@txstate.edu Texas State University, San Marcos, TX, USA

1. Introduction

An automated vehicle (AV) is a vehicle that has been built with certain aspects of safety-control features to function without the direct involvement of drivers (Chang, Healy, & Wood, 2012). These vehicles use a combination of artificial intelligence (AI) and mechanics on hardware and software technologies to help human drivers control vehicles by observing the surrounding environment. As driver error persists as a leading cause of crash occurrences, the deployment of AVs is expected to eradicate human driving faults and related traffic collisions. In addition, emerging AVs have the potential to bring about a revolutionary change in the effectiveness of road safety improvement plans by increasing travel time reliability, alleviating congestion, and improving energy efficiency (Fagnant & Kockelman, 2015). The present pro-AV discussions also infer that AVs are safer than humans as human errors are essentially nullified with their use (NHTSA, 2020). The Society of Automotive Engineers (SAE) defined six levels of automation in AVs (SAE, 2019). The current AV research efforts mostly concentrate on Level 2 (Partial Driving Automation) to Level 5 (High Driving Automation), as the full driving automation ecosystem (level 6) is significantly centred on compatible roadway environments, which require a resolute collaboration of state and local transportation agencies.

The rapid advancement in AV technologies along with comprehensive research needs have increased the testing of AVs in real traffic conditions, mainly conducted by large tech companies like Uber Technologies Inc. Assuring appropriate safety performance of automation systems in complex driving situations having a greater number of potential conflict points is a challenging task (Campbell, Egerstedt, How, & Murray, 2010). In 2012, the National Highway Traffic Safety Administration (NHTSA) offered critical recommendations to ensure the safe implementation of AV technologies while testing and operating automated vehicles on public roadways (EUA, 2012). They strongly recommended the continuous presence of a backup human driver that has a reasonable understanding of AV technologies to ensure disengagement when manual overtake is required. The study further suggested AV testing on limited-access highways that have light traffic or dense traffic at low posted speed limits. However, crashes have still been occurring during these AV roadway tests. In 2018, Uber Technologies Inc. conducted massive testing with its fleet of prototype AVs in Arizona. These experimental Uber vehicles, when operating in automated mode, were not programmed to brake or warn the test operator to take control in case of an emergency. During these tests, a crash occurred; a pedestrian was walking her bicycle across a four-lane arterial road around 10 pm when she was hit by an Uber instrumented vehicle (2017 Volvo XC90) running in automated mode (NTSB, 2018). After this incident, safety concerns associated with AV deployment have been an area of interest among manufacturers and safety practitioners.

From September 2014, the California Department of Motor Vehicles (CA DMV) authorized accredited companies to conduct pilot experiments on public roadways to examine AV technologies to stimulate their transparent deployment. Following the NHTSA guidelines, the CA DMV mandates that a trained human driver must remain behind the wheel during any AV testing to facilitate speedy transmission to manual mode if the automated mode faces any technical difficulties in unfamiliar driving circumstances. This enforcement is put in place to promote traffic safety. The testing companies must submit a crash report to CA DMV if any collision happens, and CA DMV commissions that reported automated car crashes will be made public. The first type of report included in this mandate is a brief list of all AV disengagements (during failure or difficulty in control events, a human driver will take control by putting the automated feature of the car disengaged) (Favarò, Eurich, & Nader, 2018). The second type of report supplies a thorough summary of events in which a collision, damage to property, and/or injuries take place. Many recent studies have used AV collision reports and disengagement reports from the California Department of Motor Vehicles (Banerjee, Jha, Cyriac, Kalbarczyk, & Iyer, 2018; Boggs, Arvin, & Khattak, 2020a; Favarò, Nader, Eurich, Tripp, & Varadaraju, 2017; Favarò et al., 2018; Das, Dutta, & Tsapakis, 2020; Kutela, Das, & Dadadhova, 2021; Song, Chitturi, & Noyce, 2021; Wang & Li, 2019a; Zhang & Xu, 2021).

Ensuring traffic safety on all roadways is a complicated issue, and it is affected by a wide range of contributing factors featuring driver, environment, and vehicle characteristics (Das, Jha, Fitzpatrick, Brewer, & Shimu, 2019a, 2022; M. M. Hossain, Rahman et al., 2021). AVs and associated technologies require a critical knowledge pool to improve the design and assure safety. California-based AV crash data offers unique research opportunities to explore the corresponding safety challenges by providing crucial information about AV crashes, including contributing factors such as location, weather, lighting condition, crash type, and movement prior to collision. For the automated industry, it is important to know the differences in AV crash patterns in relation to the driving mode (conventional or automated) before the crash incidence. This paper intends to investigate AV crash data and shed further light on the heterogeneity effects of roadway geometric features, human interactions, and other attributes with respect to the prior driving modes. This study analyzed the AV crashes to (1) identify the key crash attributes and (2) discover the associations of contributing factors resulting in collisions based on the vehicle's prior driving modes. As per the authors' knowledge, the previous AV crash-related studies did not explore the significance of the prior driving mode in the related crash patterns. This study aims to mitigate this research gap by using two data analysis methods (association rules mining and text network analysis).

2. Literature review

Several studies have explored AV-related crash data to explore the corresponding crash characteristics. In this study, the literature review section has been divided into three subsections: (1) key contributing factors of AV crashes identified by previous studies, (2) the complexities of transforming the driving mode in AV, and (3) the existing literature gap and the study scope.

2.1. Key contributing factors of AV crashes

A number of studies have used CA DMV's open-source AV crash data to explore the relationship among numerous contributory factors pertaining to AV crashes. Favarò et al. (2017) generated a detailed assessment on four years of AV crash records to discern the most prevalent collision classes and their relative effects. The majority of the reports exhibited rear-end collisions where AV was in front of a regular vehicle. In parallel with conventional vehicles, the number of AV crashes was positively correlated with vehicle miles travelled. In a follow-up study, Favarò et al. (2018) analyzed the safety-critical functions of AV disengagement that required the timely and safe return of vehicle control to the human driver. The findings inferred that the modality of disengagements, such as failure detection and safety operation, had no significant impact on the number of crash occurrences. Later, Khattak, Fontaine, Smith, and Board (2019) further investigated the connection between disengagement and AV crashes by applying a nested logit model on an AV dataset blended with crash and disengagement information. In contrast with Favarò et al. (2018), the study concluded that the transmission of driving modality was an important part of AVs' safety performance to avoid critical failure of adopted technologies. Concentrating on collision type and crash severity, Xu, Ding, Wang, and Li (2019) employed both descriptive statistics and binary logistic regression to investigate the trend and features of CAV-related crash risk factors. The findings indicated that the influential variables of CAV crash severity levels were the CAV driving mode, road-side parking, one-way road, and rear-end collision. In relation to vehicle movements, two crash scenarios were frequently visible: (1) passing of following regular vehicles over the preceding CAV vehicles and (2) stopped CAVs and proceeding straight movement of traditional vehicles. In the same year, Wang and Li (2019a) adopted a combined approach of logistic regression and classification tree algorithms to better understand the AV crash mechanism, taking into account multiple roadways, environment, and traffic-related factors. The results indicated that severe injury crashes were associated with automated driving mode, state highways, and high travel speed.

Boggs et al. (2020a) comprehensively examined the safety performances of AVs and found that AVs on roadways other than interstates were less likely to initiate disengagement. Wang, Zhang, Huang, and Zhao (2020) statistically analyzed AV crashes to connect the corresponding understandings in the context of other road users. They claimed that the vigilance of conventional vehicle drivers, pedestrians, and motorcyclists was important to ensure the safety effectiveness of AV. Das et al. (2020) demonstrated a variational inference algorithm for Bayesian latent class models to identify the correlation between different crash variables and collision traits. The Bayesian latent class model identified six classes of collision patterns. The classes that included turning, multi-vehicle collisions, dark-lighted conditions, and sideswipe/rear-end collisions were found to have a higher proportion of injury severity levels. In addition, weather-related crashes were more likely when the AVs were stopped and in automated mode. Sinha, Vu, Chand, Wijayaratna, and Dixit (2021) utilized seven years of AV crash information and six years of disengagement reports from CA DMV to identify the road characteristics affiliated with the degree of severity. The outcomes pointed out that intersection geometry, signalization, and road type played a critical role in the AV crash severity. Most recently, Ashraf, Dey, Mishra, and Rahman (2021) applied decision tree and association rule techniques to discern the patterns of AV-involved crashes. Consistent with previous studies, the study implied the prevalence of rear-end collisions at intersections when non-AVs attempted to make any turn or to pass the AVs. Moreover, head-on, broadside, and sideswipe collisions had occurred due to conventional vehicles' unsafe movement and disobeying traffic rules at intersections.

2.2. Complexities of transforming the driving mode in AV

Perceived safety and the value attributed to AVs are associated with the intention of adopting the technology. However, failure of those technologies often transforms the vehicle into non-automated or manual driving mode immediately prior to the crash occurrence (Montoro et al., 2019). In addition, automated driving is conjoined with substantial inconveniences (Taeihagh & Lim, 2019). For example, if an animal suddenly jumps in front of an AV, does the vehicle hit the animal, lead towards lane departure, or run off the road? If multiple vehicle collision occurs, should AVs give more priority to their passengers in minimizing the potential injuries or other vehicle occupants? (Tscharaktschiew & Evangelinos, 2019). Therefore, the automated system demands a human to observe and supervise the existing system and regain manual control when essential. However, automated to manual transmission could increase drivers' risky driving behaviours. Brandenburg and Skottke (2014) examined drivers' pre-automation and post-automation behaviour after staying in the automated driving mode for 20 min. The results showed a notable decline in the car following distances in the post-automation periods. Also, vehicle operating speed and lateral vehicle control significantly differed between the two periods. Later, Leanne, Yang, and Kuo (2018) found that drivers exhibited slower responses during the transition from automated to manual driving mode and took a long time to resume the control under the optimal workload. One possible reason can be that the driver is required to follow a process of adaptation in this transition from a state of low situational awareness to a higher one (Morales-Alvarez, Sipele, Léberon, Tadjine, & Olaverri-Monreal, 2020; Russell et al., 2016). Thus, drivers' ability to safely resume manual control is one of the primary safety concerns in automated driving (Cayeux et al., 2021; Eriksson & Stanton, 2017).

2.3. Literature gap and research scope

Although AV experiments produce a vast amount of data, these data cannot be attained due to data privacy issues. Also, artificial data generated by microsimulation and other safety assessment models do not necessarily represent real-world AV scenarios. Since the AV-related crash database is not as common as the traditional crash database, there is a lack of a thorough and rigorous understanding of AV crashes and their associated risk factors' interconnections with respect to specific driving conditions. Therefore, a number of previous studies utilized CA DMV open-access AV crash data to comprehend the causes of unprecedented crash situations. Both conventional and automated driving modes have safety issues in separate driving circumstances. However, none of the previous studies try to mine AVs crash patterns and compare the risk factors associated with the prior driving mode of AV crashes. Given this critical gap, the current study aims to explore an array of key contributing factors by dissecting both structured AV crashes and supported crash narratives. This study employed association rule mining (ARM) and text network analysis (TNA) to analyze the AV crashes that occurred in California during 2014–2020. Relying on predefined support and confidence threshold, association rules can explain the relationships between variables in multiple circumstances without confining the nature of the variables. On the contrary, TNA can explore the patterns of AV crashes from unstructured crash narrative data. It is anticipated that the current research design can provide an overall picture of AV crashes in terms of prior driving modes and can contribute to AV safety improvement.

3. Methodology

3.1. Data collection

This study developed a database that provides descriptive and detailed reports of AV crashes in California during 2014–2020. The total number of reported crashes used in this study was 255. This study used a comparatively larger sample than other AV collision-related publications.

3.2. Exploratory data analysis

Table 1 lists the yearly number of AV crashes by AV companies in California. The companies with the highest total number of collisions are Cruise (128 collisions), Google/Waymo (81 collisions), and Zoox (23 collisions), followed by Lyft, Aurora, Apple, UATC, and Pony.AI. The remaining companies each only had one crash on the record. The duration of the records goes back to 2014. Moreover, Cruise and Google/Waymo are the companies with the greatest total number of AV crashes. These higher counts might be associated with the number of AVs deployed or total AV mileages by these companies.

The heat map polygons (Figure 1(a)) represent the Voronoi Tessellation (VT) around the AV crash locations in California. Under the assumption of there being no interactions among the occurrences of AV crashes, the population of polygon sizes generated by VT follows a Poisson-Voronoi (PV) uniform random distribution that allows the determination of the natural neighbours for each AV crash and distance among the crash locations. The colours of the cells indicate the name of the city. The locations of the AV crashes are shown as black dots. It is important to note that the area presented on the map is not the boundary of these cities. The boundaries were developed based on the density of the spatial locations of the AV crashes. Instead of showing conventional

Table 1. Number of yearly crashes by the AV companies.[a]

Company	2014	2015	2016	2017	2018	2019	2020 (January–April)	Total
Cruise	–	–	1	22	37	61	7	128
Google/Waymo	–	9	13	4	25	24	6	81
Zoox	–	–	1	1	6	7	8	23
Lyft	–	–	–	–	–	5	–	5
Aurora	–	–	–	–	2	1	1	4
Apple	–	–	–	–	2	1	–	3
UATC	–	–	–	3	–	–	–	3
Pony.AI	–	–	–	–	–	2	–	2
Aimotive	–	–	–	–	–	1	–	1
Delphi	1	–	–	–	–	–	–	1
Drive.ai	–	–	–	–	1	–	–	1
Jingchi.ai	–	–	–	–	1	–	–	1
Nissan	–	–	1	–	–	–	–	1
Toyota	–	–	–	–	1	–	–	1
Grand Total	1	9	16	30	75	102	22	255

[a]https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/autonomous-vehicle-collision-reports/.

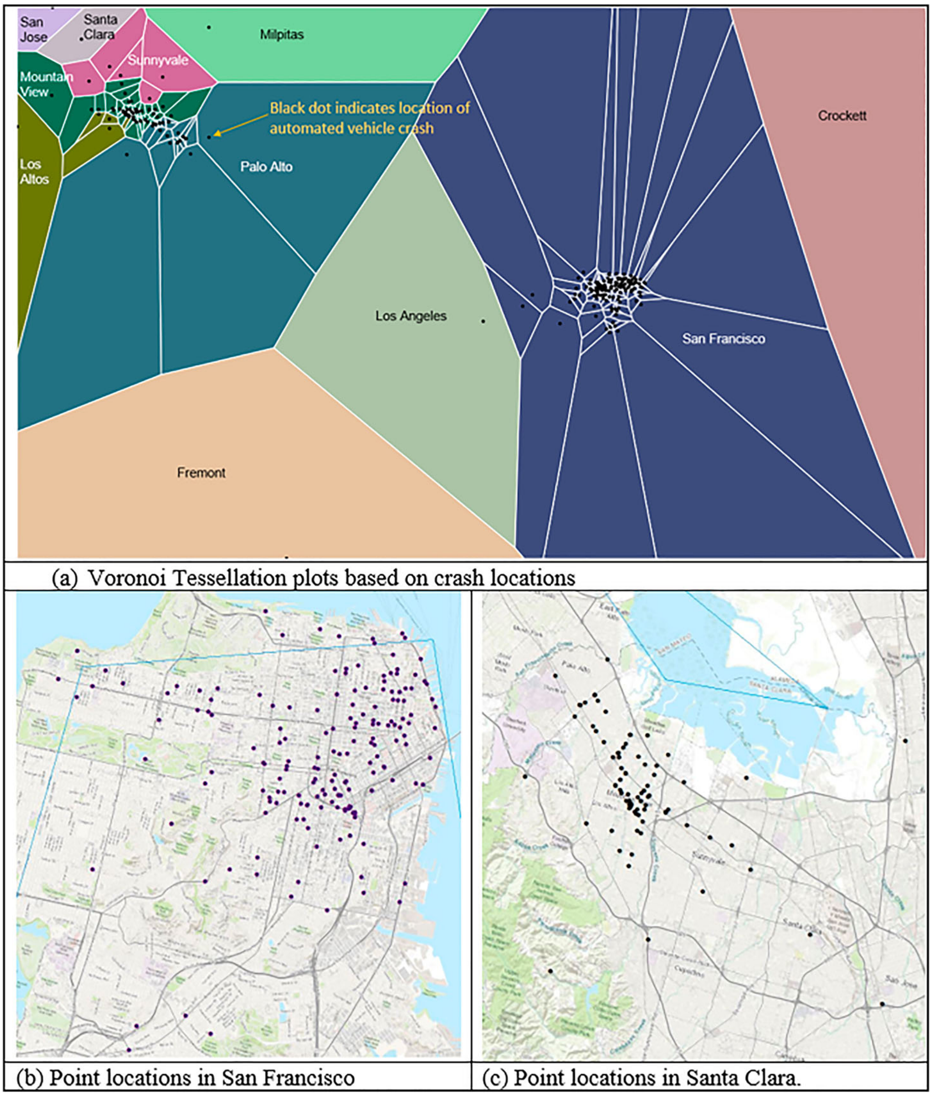

Figure 1. Location of AV crashes in San Francisco and Santa Clara. (a) Voronoi Tessellation plots based on crash locations. (b) Point locations in San Francisco. (c) Point locations in Santa Clara.

maps, this map shows crash density surrounded by two locations: San Francisco (180 collisions) and Santa Clara (96 collisions). The VT plot exhibits that the AV trips occurred closer to the city centres, and non-major locations and roadways were less explored. For general interpretation, point crash locations of San Francisco and Santa Clara are shown in Figure 1(b,c).

With crashes being a complex phenomenon with an abundance of categorical variables, it is often difficult to perceive the contexts with the usage of one or two variables. Alluvial tools are excellent data visualization tools that provide a simplistic view of multiple variables in a two-dimensional space. The black bars indicate the ratios of the categories in each variable. These bars are sorted in descending order of the proportions in

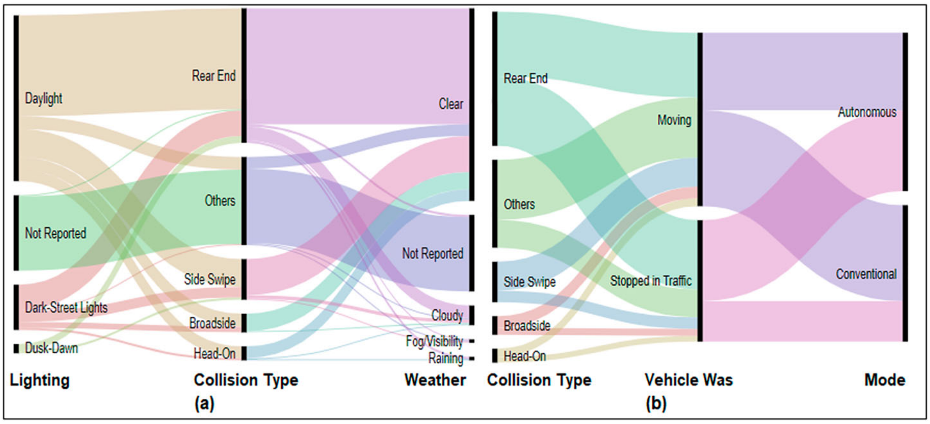

Figure 2. Alluvial plots showing the association between several key variables.

each variable. The width of the links between the variables indicates the in-between proportions. From a quick glance, it is found that lighting condition = daylight, weather = clear, and collision type = rear-end are the major attributes in these variables. Figure 2 (a) shows the association between three factors: lighting condition, collision type, and weather. Daylight, rear-end collision, and clear weather are the dominant categories in each of these factors. The rear-end collision is the most frequent type of collision across the database, and it could occur under different lighting and weather conditions. Daylight and clear weather are two dominant factors associated with it. Dark-street lights and cloudy weather are the secondary dominant factors associated with rear-end collisions. Attributes as 'not reported' are mostly associated with collision type = others. This also implies that trivial crash data reports have missing information. Other AV collision studies have found similar results. For example, Wang and Li (2019b) have identified driving mode, collision location, roadside parking, rear-end collision, and one-way road are the significant influencing factors from the regression analysis. Figure 2(b) shows the association between three factors: collision type, vehicle movement, and prior driving mode. Rear-end collision, the vehicle was moving, and automated as prior driving mode are the dominant categories in each of these factors. The stopped in traffic condition is associated with rear-end collisions and automated as prior driving mode. Moving vehicles have a high likelihood with conventional as the prior driving mode.

A simple Chi-square test was performed to examine the difference between the attributes by prior modes (conventional or automated). Statistical analyses were performed on 'R' software using the package 'compareGroups' (Salvador, 2019). This study defined statistical significance as a p-value <.05 (statistically significant values are shown with an asterisk mark in Table 2). Consider for a given data matrix with the response and predictor variables that the mean parameter vector is $\bar{\beta}$. Using the odds of the combination x,

$$\mathrm{Odd}(\bar{\beta}x) = \frac{\pi(\bar{\beta}x)}{1 - \pi(\bar{\beta}x)} = e^{(x'\bar{\beta})}$$

Table 2. Comparison of contributing factors based on prior driving mode

Variables	Conventional $N = 118$	Automated $N = 137$	OR	*p* ratio	*p*-Value
DOW (Day of the week)					.038*
FSS (Fri-Sat-Sun)	51 (43.2%)	41 (29.9%)	Ref.	Ref.	
MTWT (Mon-Tue-Wed-Thu)	67 (56.8%)	96 (70.1%)	1.78 [1.06;2.99]	0.029	
Veh1 Make (Vehicle 1 make)					.241
Chevrolet	54 (45.8%)	72 (52.6%)	Ref.	Ref.	
Chrysler	19 (16.1%)	38 (27.7%)	1.49 [0.78;2.92]	0.228	
Google	2 (1.69%)	6 (4.38%)	2.14 [0.45;16.6]	0.354	
Lexus	7 (5.93%)	12 (8.76%)	1.27 [0.47;3.68]	0.637	
Others	19 (16.1%)	2 (1.46%)	0.09 [0.01;0.31]	<0.001	
Toyota	17 (14.4%)	7 (5.11%)	0.31 [0.11;0.79]	0.013	
Location City (Location)					.244
Mountain View	19 (16.1%)	35 (25.5%)	Ref.	Ref.	
Others	12 (10.2%)	9 (6.57%)	0.41 [0.14;1.16]	0.094	
Palo Alto	12 (10.2%)	15 (10.9%)	0.68 [0.26;1.78]	0.431	
San Francisco	75 (63.6%)	78 (56.9%)	0.57 [0.29;1.07]	0.082	
Vehicle Was (Prior movement)					.001*
Moving	83 (70.3%)	67 (48.9%)	Ref.	Ref.	
Stopped in Traffic	35 (29.7%)	70 (51.1%)	2.46 [1.47;4.17]	0.001	
No Vehicle Involved (Involved vehicles)					.198
One	17 (14.4%)	12 (8.76%)	Ref.	Ref.	
Two	98 (83.1%)	124 (90.5%)	1.78 [0.81;4.02]	0.149	
Three	3 (2.54%)	1 (0.73%)	0.52 [0.02;5.10]	0.599	
Vehicle was OtherParty (Prior movement of other party)					<.001*
Moving	93 (78.8%)	131 (95.6%)	Ref.	Ref.	
Not Reported	17 (14.4%)	5 (3.65%)	0.21 [0.07;0.57]	0.001	
Stopped in Traffic	8 (6.78%)	1 (0.73%)	0.10 [0.00;0.57]	0.006	
Weather					.02*
Clear	88 (74.6%)	78 (56.9%)	Ref.	Ref.	
Cloudy	6 (5.08%)	11 (8.03%)	2.04 [0.73;6.28]	0.177	
Fog/Visibility	1 (0.85%)	2 (1.46%)	2.11 [0.17;67.3]	0.564	
Not Reported	21 (17.8%)	45 (32.8%)	2.40 [1.33;4.46]	0.004	
Raining	2 (1.69%)	1 (0.73%)	0.60 [0.02;7.56]	0.698	
Lighting					.003*
Dark-Street Lights	20 (16.9%)	19 (13.9%)	Ref.	Ref.	
Daylight	77 (65.3%)	66 (48.2%)	0.90 [0.44;1.85]	0.778	
Dusk-Dawn	1 (0.85%)	7 (5.11%)	6.42 [0.97;175]	0.054	
Not Reported	20 (16.9%)	45 (32.8%)	2.34 [1.03;5.42]	0.042	

Collision Type (Type of Collisions)					<.001*
Broadside	9 (7.63%)	7 (5.11%)	Ref.	Ref.	
Head-On	11 (9.32%)	1 (0.73%)	0.14 [0.00;1.02]	0.052	
Others	29 (24.6%)	47 (34.3%)	2.06 [0.68;6.46]	0.2	
Rear End	44 (37.3%)	72 (52.6%)	2.08 [0.71;6.32]	0.178	
Side Swipe	25 (21.2%)	10 (7.30%)	0.52 [0.15;1.85]	0.31	
Business (Vehicle Operator)					<.001*
Cruise	55 (46.6%)	73 (53.3%)	Ref.	Ref.	
Google	7 (5.93%)	17 (12.4%)	1.80 [0.72;5.01]	0.216	
Others	21 (17.8%)	2 (1.46%)	0.08 [0.01;0.28]	<0.001	
Waymo	19 (16.1%)	38 (27.7%)	1.50 [0.79;2.93]	0.222	
Zoox	16 (13.6%)	7 (5.11%)	0.34 [0.12;0.85]	0.021	

The odds ratio (OR) can be expressed as:

$$\mathrm{OR}(\bar{\beta}x_1x_2) = \frac{\mathrm{Odd}(\bar{\beta}x_1)}{\mathrm{Odd}(\bar{\beta}x_2)} = e^{(x_1'\bar{\beta} - x_2'\bar{\beta})}$$

The outcomes from Table 2 show that the vehicle manufacturers, number of vehicles involved, and crash location cities do not vary widely based on whether the prior mode was conventional or automated. Bigger companies like Cruise and Google/Waymo are dominantly presented in the crashes with automated driving mode, and the smaller companies are more conservative. Most of the AV crashes that involve vehicles from smaller companies are in conventional driving mode. This finding indicates that larger and more prominent companies have more confidence in their AVs. They are either more advanced in automated driving technologies or are more willing to take the risk of testing new technologies. The odds of weekday-related automated mode AV crashes are higher (Odds Ratio or OR: 1.78, 95% Confidence Interval: 1.06–2.99). If AV was stopped, the odds of automated mode AV crashes are higher. Out of the different weather categories, the odds of automated mode AV crashes during inclement weather such as rain are higher. In relation to collision types, the odds of automated mode AV crashes in rear-end crashes are higher.

3.3. Association rules mining (ARM)

Rule-based modelling is gaining popularity in transportation safety analysis (Das, Kong, & Tsapakis, 2019b; Md Mahmud Hossain et al., 2021; Kong, Das, Jha, & Zhang, 2020; Montella, Mauriello, Pernetti, & Rella Riccardi, 2021). Association rules are very effective for a large set of unsupervised data by providing key insights that can be used for data-driven decision-making. The current study used an 'a priori' algorithm to generate the rules. This algorithm has the applicability of a 'bottom-up' approach in which frequent subsets are expanded one item at a time by using a breadth-first search following a Hash tree structure (Das et al., 2019b; Kong et al., 2020). Consider $I = \{i_1, i_2, i_3, \ldots i_n, \}$ as a set of N distinctive items. Let D be a set of transactions where each transaction T consists of a set of items, such that $T \in I$. Each transaction is associated with only an identifier. An association rule is denoted by a form of Antecedent → Consequent or A → B, where $A \in I$ and $B \in I$. In this unsupervised learning framework, the *a priori* algorithm provides a semi-supervised approach adopted by this study. This semi-supervised structure allows researchers to define the right-hand side of the mined rules. Thus, a group of rules that have the same consequent can be extracted and analyzed. In this study, two candidate consequents are: 'prior mode as conventional' and 'prior mode as automated'. To set these two consequents as the right-hand side of the mined rules, two groups of rules will be mined. Therefore, the AV crash patterns for two prior modes can be compared, and more insights can be extracted. The conventional measures for association rules are support (S), confidence (C), and lift (L). Many studies proposed newer interest measures for rule mining in the data mining research studies to generate interesting and insightful rules. Lift, the most common performance measure in association rules, measures how often A and B collectively occur compared to the expected value if they were statistically independent. A high lift value (greater than one) indicates dependence between A and B. If the value of the lift is greater than 1, it indicates that A and B appear more frequently together in the data and are said to be positively dependent on each other (Das et al., 2019b). The equations of ARM are specified below:

Support of A, $S(A) = \frac{n(A)}{n}$

Support of B, $S(B) = \frac{n(B)}{n}$

Support of rule A→ B, $S(A \to B) = \frac{n(AB)}{n} = S(B \to A)$

Confidence of rule A→ B, $C(A \to B) = \frac{S(A \to B)}{S(A)}$

Confidence of rule B→ A, $C(B \to A) = \frac{S(A \to B)}{S(B)}$

Lift of A→ B, $L(A \to B) = \frac{C(A \to B)}{S(B)} = \frac{C(B \to A)}{S(A)}$.

where $n()$ means count for each of individual (e.g. A or B) or combinations (e.g. AB).

3.4. Text network analysis (TNA)

The text network indicates the co-occurrence of keywords in the unstructured text data. The AV crash data reports provide crash narratives of crash occurrences. Exploration of these crash narratives can provide additional information that is often missing from structured crash data. In order to develop text networks from the crash narratives, there is a need to perform the key steps of text cleaning, such as the removal of stop words or redundant words. Two R-packages, 'quanteda' and 'igraph' (Benoit et al., 2018; Csardi, 2014), were used to map the text and extract the performance parameters. Since the analysis produced a large number of keywords, only the top 40 keywords were used for text network mapping.

To create a text network, the unstructured data is first converted to structured data by the normalization process. 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 keywords with high frequencies are illustrated by larger nodes. Similarly, the keywords with high co-occurrence frequencies result in a thick edge. The closer that the keywords are on the network, the closer they appear in the text. In addition, two network measures (degree centrality and betweenness centrality) were used to explore the text network. The difference between degree and betweenness centrality is that degree centrality quantifies the extent of the connection between nodes that fall within the same community, while betweenness centrality measures the extent of the connection one node outside of its community.

$$\text{Degree centrality}(m) = \sum_{n=1}^{l} c_{mn}$$

whereby c_{mn} takes a value of 1 if nodes m and n are connected, and a value of 0 otherwise (Kutela et al., 2021).

The betweenness centrality can be expressed as the ratio of the total number of the shortest paths from cluster m directly to cluster n over the number of the shortest paths from cluster m to cluster n through cluster t.

$$\text{Betweenness centrality}(i) = \sum_{n=1}^{l} \frac{b_{mn(k)}}{b_{mn}}$$

4. Results and discussions

4.1. Comparison between autonomous and conventional mode

A simple Chi-square test was performed to examine the difference between the attributes by prior modes (conventional or automated). Statistical analyses were performed on 'R' software using the package 'compareGroups' (Salvador, 2019). This study defined statistical significance as a p-value <.05 (statistically significant values are shown with an asterisk mark in Table 2). Consider for a given data matrix with the response and predictor variables that the mean parameter vector is $\bar{\beta}$. Using the odds of the combination x,

$$\text{Odd}(\bar{\beta}x) = \frac{\pi(\bar{\beta}x)}{1 - \pi(\bar{\beta}x)} = e^{(x'\bar{\beta})}$$

The odds ratio (OR) can be expressed as:

$$\text{OR}(\bar{\beta}x_1x_2) = \frac{\text{Odd}(\bar{\beta}x_1)}{\text{Odd}(\bar{\beta}x_2)} = e^{(x_1'\bar{\beta} - x_2'\bar{\beta})}$$

The outcomes from Table 2 show that the vehicle manufacturers, number of vehicles involved, and crash location cities do not vary widely based on whether the prior mode was conventional or automated. Bigger companies like Cruise and Google/Waymo are dominantly presented in the crashes with automated driving mode, and the smaller companies are more conservative. Most of the AV crashes that involve vehicles from smaller companies are in conventional driving mode. This finding indicates that larger and more prominent companies have more confidence in their AVs. They are either more advanced in automated driving technologies or are more willing to take the risk of testing new technologies. The odds of weekday-related automated mode AV crashes are higher (Odds Ratio or OR: 1.78, 95% Confidence Interval: 1.06–2.99). If AV was stopped, the odds of automated mode AV crashes are higher. Out of the different weather categories, the odds of automated mode AV crashes during inclement weather such as rain are higher. In relation to collision types, the odds of automated mode AV crashes in rear-end crashes are higher.

4.2. Rules mining results

This study used the R package 'arules' to perform ARM analysis (Hahsler, Chelluboina, Hornik, & Buchta, 2011). The chi-square test showed that variables such as the number of vehicles involved, crash locations, and vehicle manufacturers are not statistically significant (Table 2). Therefore, these variables were removed from the databases to perform association rules mining. The final dataset contains 255 transactions with 28 items (or vehicle attributes). It is important to provide support and confidence thresholds to generate intuitive rules. This study conducted a supervised version of association rules by keeping the consequent fixed for prior mode (either conventional or automated). After several trials and errors, the following thresholds have been considered:

– Minimum support = 0.05.
– Minimum confidence = 0.20.
– Minimum number of items = 2 (including consequent).
– Maximum number of items = 4 (including consequent).

4.2.1. Rules by prior mode = conventional

After setting the threshold, 170 rules were generated by keeping prior mode = conventional as the fixed consequent. Table 3 shows the top 20 rules based on the 'lift' measure. The first rule (A#1) has a lift measure of 2.161 (S = 5.5%, Confidence = 100%, Count = 14). It indicates that the proportion of AV collisions (in conventional mode) during weekdays in clear weather is 2.161 times the proportion of all AV collisions (during weekdays in clear weather) in the complete dataset. Business = Others is present in the top seven rules (A#1–A#7). Only 9% of the AV vehicles involved in crashes are manufactured by other companies such as Toyota, Nissan, Lyft. The frequent presence of this attribute indicates that small companies (in respect to the number of AVs on the road) are most likely conservative while testing the vehicles in real-world scenarios to mitigate risk. In five rules (A#12, A#13, A#14, A#15, A#18), the vehicle was in the moving condition. In relation to collision types, sideswipe crashes are also dominant in several rules (A#12, A#14, A#15, A#16, A#19, A#20). The explanation of rule A#12 (S = 7.1%, C = 85.7%, L = 1.85) is (a) 7.1% of AV crashes in manual driving mode is side-swipe when the testing vehicle is moving in clear weather condition, (b) Out of all incidents in the similar scenario, 85.7% produces collisions, and (c) the proportion between

Table 3. Top 20 rules for Prior Mode = Conventional as consequent.

Rule	LHS	S (%)	C (%)	L	Count
A#1	DOW = MTWT + Weather = Clear + Business = Others	5.5	100.0	2.16	14
A#2	DOW = MTWT + Business = Others	7.5	95.0	2.05	19
A#3	Weather = Clear + Business = Others	6.3	94.1	2.03	16
A#4	Weather = Clear + Lighting = Daylight + Business = Others	6.3	94.1	2.03	16
A##5	DOW = MTWT + Lighting = Daylight + Business = Others	5.9	93.8	2.02	15
A#6	DOW = MTWT + Vehicle was OtherParty = Moving + Business = Others	5.9	93.8	2.02	15
A#7	Vehicle was OtherParty = Moving + Weather = Clear + Business = Others	5.9	93.8	2.02	15
A#8	Business = Others	8.2	91.3	1.97	21
A#9	Lighting = Daylight + Business = Others	6.7	89.5	1.93	17
A#10	Vehicle was OtherParty = Moving + Business = Others	6.7	89.5	1.93	17
A#11	Vehicle was OtherParty = Moving + Lighting = Daylight + Business = Others	5.9	88.2	1.91	15
A#12	Vehicle Was = Moving + Weather = Clear + Collision Type = Side Swipe	7.1	85.7	1.85	18
A#13	Vehicle Was = Moving + Vehicle was OtherParty = Not Reported	6.3	80.0	1.72	16
A#14	Vehicle Was = Moving + Collision Type = Side Swipe	7.8	80.0	1.72	20
A#15	Vehicle Was = Moving + Vehicle was OtherParty = Moving + Collision Type = Side Swipe	6.3	80.0	1.72	16
A#16	Weather = Clear + Collision Type = Side Swipe + Business = Cruise	5.9	78.9	1.71	15
A#17	Vehicle was OtherParty = Not Reported	6.7	77.3	1.67	17
A#18	DOW = FSS + Vehicle Was = Moving + Lighting = Daylight	10.2	76.5	1.65	26
A#19	Weather = Clear + Collision Type = Side Swipe	9.0	74.2	1.60	23
A#20	Vehicle was OtherParty = Moving + Weather = Clear + Collision Type = Side Swipe	7.8	74.1	1.60	20

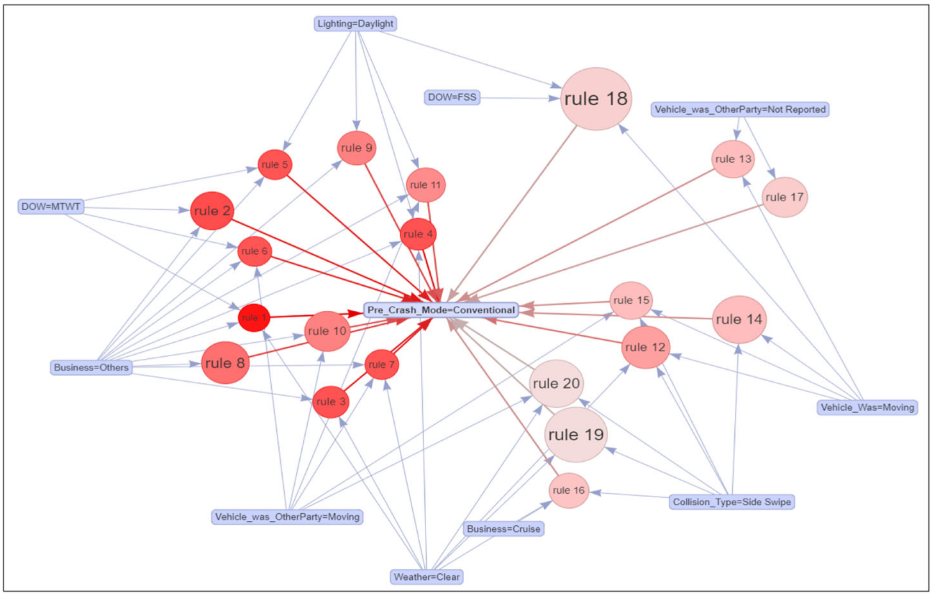

Figure 3. Top 20 rules (Prior Mode as Conventional)

moving AV sideswipe crashes in clear weather is 1.85 times the same proportion in the complete dataset.

In Figure 3, the first 20 rules are categorized based on all the frequent itemsets in these rules. The figure provides a clear picture of how these frequent itemsets are distributed in these rules. The size of the circle indicates the support value of that rule, and the colour indicates the lift value. A larger size means a higher support value, and a darker red colour means a higher lift value. On the left-hand side, a group of rules (rules A#1–A#11) associates factors like daylight, weekdays, moving vehicle, clear weather, and non-major AV manufacturers with the consequent prior mode = conventional. On the right-hand side, another group of rules (rules A#12–A#20) associates the conventional driving mode with the weekend (including Friday), moving vehicle, sideswipe collision type, and Cruise as manufacturer. The number of links between the itemset and the rules indicates the number of times that the itemset is present in the top 20 rules. More links indicate that the factor is more dominant among all other factors in the top 20 rules. Itemsets like clear weather, sideswipe collision type, and moving vehicles are presented in more than 5 out of 20 rules.

4.2.2. Rules by prior mode = automated

By considering prior mode = automated, a set of 213 rules was produced. Table 4 lists top 20 rules based on the lift measures. The business company ‘Waymo’ is present in nine rules, and ‘Cruise’ is present in seven rules. The first rule (B#1) exhibits a lift measure of 1.712. This indicates that the proportion of AV Cruise vehicle collisions (in automated mode) in clear weather with unknown lighting conditions is 1.712 greater than the proportion of all AV Cruise vehicle collisions (in clear weather with unknown lighting conditions) in the complete dataset. When the light condition is daylight, the likelihood of

Table 4. Top 20 rules for Prior Mode = Automated as consequent.

Rule	LHS	S (%)	C (%)	L	Count
B#1	DOW = MTWT + Lighting = Not Reported + Business = Cruise	9.0	92.0	1.71	23
B#2	DOW = MTWT + Weather = Not Reported + Business = Cruise	9.0	92.0	1.71	23
B#3	Vehicle was OtherParty = Moving + Lighting = Not Reported + Business = Cruise	10.2	86.7	1.61	26
B#4	Vehicle was OtherParty = Moving + Weather = Not Reported + Business = Cruise	10.2	86.7	1.61	26
B#5	DOW = MTWT + Vehicle Was = Stopped in Traffic + Business = Waymo	9.4	85.7	1.59	24
B#6	DOW = MTWT + Collision Type = Others + Business = Cruise	9.0	85.2	1.58	23
B#7	DOW = MTWT + Vehicle was OtherParty = Moving + Business = Waymo	10.6	84.4	1.57	27
B#8	Vehicle Was = Stopped in Traffic + Vehicle was OtherParty = Moving + Lighting = Not Reported	6.3	84.2	1.56	16
B#9	Vehicle Was = Stopped in Traffic + Vehicle was OtherParty = Moving + Weather = Not Reported	6.3	84.2	1.56	16
B#10	Vehicle Was = Stopped in Traffic + Business = Waymo	12.9	82.5	1.53	33
B#11	Vehicle Was = Stopped in Traffic + Vehicle was OtherParty = Moving + Business = Waymo	12.9	82.5	1.53	33
B#12	Vehicle Was = Moving + Lighting = Not Reported + Business = Cruise	9.0	82.1	1.52	23
B#13	Vehicle Was = Moving + Weather = Not Reported + Business = Cruise	9.0	82.1	1.52	23
B#14	Lighting = Not Reported + Collision Type = Others + Business = Cruise	10.6	81.8	1.52	27
B#15	Weather = Not Reported + Collision Type = Others + Business = Cruise	10.6	81.8	1.52	27
B#16	Vehicle Was = Stopped in Traffic + Lighting = Daylight + Business = Waymo	10.2	81.3	1.51	26
B#17	DOW = MTWT + Vehicle Was = Stopped in Traffic + Lighting = Not Reported	5.1	81.3	1.51	13
B#18	DOW = MTWT + Vehicle Was = Stopped in Traffic + Weather = Not Reported	5.1	81.3	1.51	13
B#19	Vehicle was OtherParty = Moving + Collision Type = Others + Business = Cruise	10.2	81.3	1.51	26
B#20	Vehicle Was = Stopped in Traffic + Collision Type = Rear End + Business = Cruise	5.1	81.3	1.51	13

more severe crashes drops, accounting for 16% of all AV crashes. The findings are consistent with previous research (Tarmizi & Abd Aziz, 2018; Wang et al., 2019; Wiseman, Adler-Golden, Ientilucci, & Perkins, 2021), which suggested that AVs could detect road users more successfully during the day than at night or in poor lighting conditions. Furthermore, approximately 50% of the top 20 rules are associated with a vehicle of the other party moving. Similar finding is found in Song et al. (2021). This may indicate that AV technologies have difficulties adjusting their movement track when the other vehicles are moving nearby. The majority of collision types are either rear-end or considered as 'other'. Other studies (Boggs et al., 2020a; Das et al., 2020) confirmed that rear-end crashes were the most common for both types of vehicles. However, over 50% of AV crashes were caused by conventional vehicles rear-ending AVs, which happened 1.6 times more often than rear-ending conventional vehicles (Liu, Wang, Wu, Glase, & He, 2021). Zhang and Xu (2021) used association rule analysis to identify six other crash patterns, including broadside collision, sideswipe collision, and four different rear-end collision scenarios.

In Figure 4, the first 20 rules with consequent prior mode = automated are categorized based on all the frequent itemsets shown in these rules. The figure provides a clear picture of how these frequent itemsets are distributed in these rules. Multiple itemsets have a higher frequency in the top 20 rules compared to other itemsets. These itemsets, like vehicles stopped in traffic, AVs with Cruise company, and weekdays, are dominant in these top rules.

4.3. Results of text network analysis (TNA)

The word counts in California AV narrative data are limited (min = 2 words, max = 264 words, and mean = 93 words). For variable selection process, this study selected variables

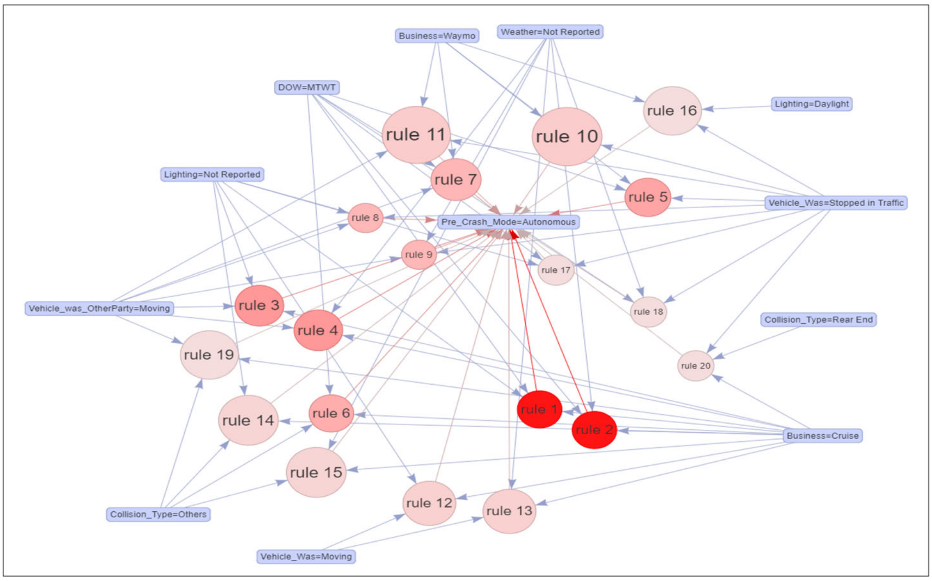

Figure 4. Top 20 rules (Prior Mode as Automated)

with less missing values. Some variables such as 'Other party involved' is a critical variable. However, very few incidents are associated with non-motorists and inclusion of these variables will not generate interesting results. Table 5 lists examples of crash narratives with different involved parties.

Figure 5 shows the text networks from the crash narratives of the AV crashes either in conventional or automated mode. Although the two networks and associated metrics show many similarities, several differences can be explored. For example, keywords such as police and intersection are present in the conventional mode network but not in the automated mode network. Similarly, rear-ended, stopped, minor, and information are present in the automated mode network but not in the conventional mode network. Additionally, the between centrality and degree centralities vary by the prior modes. For vehicles driving in the conventional mode, crashes are more likely to occur during moving status, and sideswipe is a dominant factor associated with this type of collision. For vehicles operating in automated mode, the crashes primarily occurred when the vehicles stopped for traffic and the other vehicle failed to stop. The cause of these crashes could be that the computer in automated vehicles has a shorter reaction time when there is a traffic stop ahead to ensure a successful stop. However, for any vehicle following an AV, the driver may have a longer reaction time than the AV and will then be more likely to collide with a suddenly stopped leading vehicle. The investigation on crash narrative analysis indicates that the CA DMV needs to make improvements to the required length of the crash narratives so that a complete picture can be drawn from the narrated details. The current narrative length is not adequate to identify key contributing factors. As AV crashes are associated with human–computer interactions, more real-time information on reaction time and associated movement-related variables are needed to understand the

Table 5. Examples of different involved parties and crash narratives.

Involved party	Crash narrative
Pedestrian	A Cruise AV, operating in autonomous mode, was involved in a collision while making a right hand turn from northbound Valencia Street onto 16th Street. The Cruise AV was stopped at a green light in between crosswalks of Valencia Street and 16th Street, waiting for pedestrians to cross over 16th Street. 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. There were no injuries, but the Cruise AV sustained some damage to its left rear light. The police were not called.
Bicyclist	A Waymo AV was travelling in autonomous mode on northbound View Street at California Street in Mountain View, approaching a four-way intersection with a traffic calming island. After coming to a complete stop at a two-way stop sign, the Waymo AV determined it was safe to proceed through the intersection and began to do so, when it detected a bicyclist approaching from the right. The Waymo AV then stopped for the bicyclist, whose front tire made contact with the passenger side of the stationary Waymo AV at approximately 3 MPH. The bicyclist remained upright and rode away without exchanging information. No injuries or damage were reported or observed.
Skateboarder	A Cruise AV was in autonomous mode and travelling northbound on De Haro Street between 23rd and 22nd streets. The AV braked in response to a pedestrian in the middle of the intersection on De Haro, running directly towards the AV. The AV, concerned for their safety, disengaged from autonomous mode to attempt to move the AV around the pedestrian who was still rapidly heading directly towards the AV in the centre of the street. The Cruise AV was then struck by a skateboarder that suddenly entered the crosswalk from the southeast sidewalk, coming downhill from 22nd street. Contact was made by the skateboarder to the Cruise AV. This caused damage to the front quarter panel area and the articulating radar of the Cruise AV. Because the skateboarder appeared injured, 911 was immediately called. The skateboarder left the scene shortly after the incident and before the police arrived. The Cruise remained on the scene and the driver gave a statement to the police.
Parked Vehicle	On 12 March 2019, at approximately 4:07 pm, a Lyft AV, operating at full manual mode, made contact at approximately 2 MPH with a parked 3rd party vehicle while reversing during a multi-point turn. There was no driver or passenger present in the 3rd party vehicle. The incident occurred on Pettis Avenue near El Camino Boulevard in Mountain View, CA. Minor damage is limited to the left rear bumper area of the Lyft AV and the left rear driver's side of the 3rd party, pending further inspection. The owner of the 3rd party vehicle came outside to exchange information with the operators. No injuries were reported, and the police were not notified.
Vehicle	A Hyundai rear-ended a stationary Zoox AV that was operating in Autonomous Mode when struck. The collision occurred while the Zoox AV was travelling westbound on Pacific A venue, intending to make a right turn at a green light onto Sansome St. The Zoox AV had slowly decelerated from a low speed and stopped, waiting for a pedestrian to clear the intersection, when the Hyundai's front, passenger-side bumper made contact with the Zoox AV's driver-side rear bumper and sensor. At the time of collision, the Hyundai was travelling at just under 4 MPH and was slightly accelerating towards the Zoox AV prior to impact.

trigger points of crash occurrence. CA DMV can provide more guidance on real-time crash data from the AV companies.

5. Conclusions

The technologies of AV are rapidly growing in recent years. This growth not only represents a paradigm shift on conventional vehicle movement concepts but also advantages for safe, secure, and sustainable transportation networks. However, safety concerns associated with AVs prevail. To understand the safety-related factors, it is essential to obtain enough data regarding the crash history and the contributing factors in AV crashes. In this study, a comprehensive analysis was conducted using the crash reports filed by various manufacturers from September 2014 to April 2020. The variables included collision type, lighting conditions, the number of vehicles involved, weather conditions, the event prior to the crash, prior mode, crash type, industry or business name, vehicle manufacturer, location of the crash, and whether the vehicle was

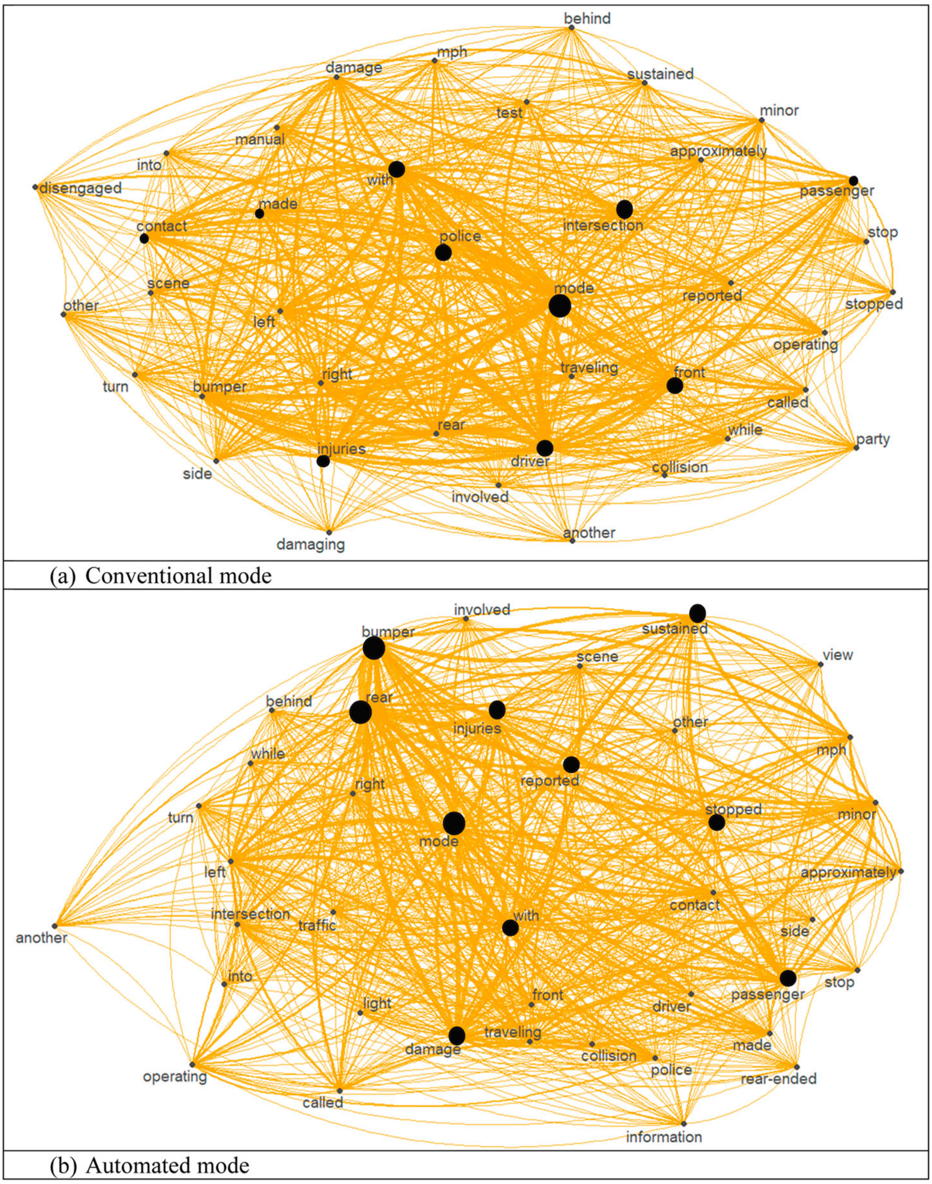

Figure 5. Text Networks from Crash Narratives. (a) Conventional mode. (b) Automated mode.

moving or stopped. The common perception is that AV-involved crashes depend on built-environment dynamics as the infrastructure components are usually incorporated in the AV routing design.

A crash event sequence depicts the progression of scenes involving interactions between an AV and its moving surroundings (Ulbrich, Menzel, Reschka, Schuldt, & Maurer, 2015). In AV collision data, multiple variables such as weather, lighting, road surface, road geometries, traffic control, and traffic conditions influence the crash occurrence and associated sequences based on the prior mode (Nitsche, Thomas, Stuetz, &

Welsh, 2017; Sander & Lubbe, 2018). Liu et al. (2021) identified 15 types of pre-crash scenarios for AV modes and 26 scenarios for conventional modes and the results showed that there were big differences between how AVs and other vehicles crashed and what happened before. It is found that the prior mode (conventional vs. automated) differs significantly for several factors. The current study examined the prior modes in depth by using two different 'rules-based modeling' approaches. The association rules for crashes involving AVs with conventional or automated prior modes were mined. The rules for each of the prior modes provide additional context for AV-related crash occurrence scenarios. Besides, two text networks were developed from crash narrative reports based on the prior modes.

This study shows that bigger companies, such as Google/Waymo, have more confidence and thus grant more power to automated technology. AV-related crashes with automated driving modes prior to the collisions were found to be dominantly associated with these bigger AV companies. Moreover, AV-related crashes are more likely to be the sideswipe collision type when the prior driving mode is conventional, and they are more likely to be the rear-end collision type when the prior driving mode is automated. These rear-end collisions are found to frequently occur to these AVs under the automated driving mode, stopped in traffic, and hit by another moving vehicle from behind. These findings suggest that the AV manufactory could design the vehicle to be more alert towards the rear vehicles' trajectories and provide more comprehensive evasive movement when allowed by the environment. For example, when an AV reacts to traffic with a sudden stop, the following vehicle driving by a human driver may not act as quickly as the AV. Thus, a collision may occur. In that case, with prior knowledge of an incoming event, the AV can perform an evasive manoeuver to avoid the collision, such as moving forward several feet or changing lanes if there is some available room on the adjacent lanes.

This study highlights the complexity and challenges of identifying key risk factors associated with AV crashes. It also extended the analysis by performing text network analysis to identify intuitive knowledge. The crash narratives are very brief. These narratives need more contextual information by providing both parties experience. The study is not without limitations. The conventional reactive approach of traffic safety analysis, using historical crash data, has several limitations, such as data reliability, omission of contributing factors, and exclusion of human errors in generalized safety performance functions. As the current paper depends on the precision of the limited information provided by the California DMV AV crash database, the limitations prevail. More states are required to make AV-related crash details open to the public so that the contexts can be understood in a more detailed fashion. Future studies can use a larger dataset and collect additional geometric data (such as curvatures, superelevation) from Google Earth to extend this analysis.

Acknowledgements

The authors like to thank two anonymous reviewers for their excellent feedback.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Subasish Das http://orcid.org/0000-0002-1671-2753
Xiaoqiang "Jack" Kong http://orcid.org/0000-0002-8120-0754
Md Mahmud Hossain http://orcid.org/0000-0002-2737-6951

References

Ashraf, M. T., Dey, K., Mishra, S., & Rahman, M. T. (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. 036119812110184.

Banerjee, S. S., Jha, S., Cyriac, J., Kalbarczyk, Z. T., & Iyer, R. K. (2018). Hands off the wheel in autonomous vehicles? A systems perspective on over a million miles of field data. In: *2018 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*. IEEE (pp. 586–597).

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. *Journal of Open Source Software*, *3*(30), 774. doi:10.21105/joss.00774

Boggs, A. M., Arvin, R., & Khattak, A. J. (2020a). Exploring the who, what, when, where, and why of automated vehicle disengagements. *Accident Analysis & Prevention*, *136*, 105406. doi:10.1016/j.aap.2019.105406

Brandenburg, S., & Skottke, E. M. (2014). Switching from manual to automated driving and reverse: Are drivers behaving more risky after highly automated driving? In *2014 17th IEEE Int. Conf. Intell. Transp. Syst. ITSC 2014* (pp. 2978–2983). doi:10.1109/ITSC.2014.6958168

Campbell, M., Egerstedt, M., How, J. P., & Murray, R. M. (2010). Autonomous driving in urban environments: Approaches, lessons and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *368*(1928), 4649–4672. doi:10.1098/rsta.2010.0110

Cayeux, E., Mihai, R., Carlsen, L., Ørevik, M., Birgisson, K., & Bergerud, R. (2021). A technical approach to safe mode management for a smooth transition from automatic to manual drilling. *SPE/IADC International Drilling Conference and Exhibition, Virtual*. doi:10.2118/204114-ms

Chang, J., Healy, T., & Wood, J. (2012). The potential regulatory challenges of increasingly autonomous motor vehicles. *Santa Clara Law Review*, *52*(4), 1423.

Csardi, G. (2014). igraph: Network analysis and visualization (pp. 1–322).

Das, S., Dutta, A., & Tsapakis, I. (2020). Automated vehicle collisions in California: Applying Bayesian latent class model. *IATSS Research*, *44*(4), 300–308. doi:10.1016/j.iatssr.2020.03.001

Das, S., Hossain, M. M., Rahman, M. A., Kong, X., Sun, X., & Mamun, G. M. (2022). Understanding patterns of moped and seated motor scooter (50 cc or less) involved fatal crashes using cluster correspondence analysis. *Transportmetrica A: Transport Science*. doi:10.1080/23249935.2022.2029613

Das, S., Jha, K., Fitzpatrick, K., Brewer, M., & Shimu, T. H. (2019a). Pattern identification from older bicyclist fatal crashes. *Transportation Research Record*, *2673*(6), 638–649. doi:10.1177/0361198119841863

Das, S., Kong, X., & Tsapakis, I. (2019b). Hit and run crash analysis using association rules mining. *Journal of Transportation Safety & Security*. doi:10.1080/19439962.2019.1611682

Eriksson, A., & Stanton, N. A. (2017). Takeover time in highly automated vehicles: Noncritical transitions to and from manual control. *Human Factors*, *59*(4), 689–705. doi:10.1177/0018720816685832

EUA. (2012). Preliminary statement of policy concerning automated vehicles. *The National Highway Traffic Safety Administration, Washington DC*, 1–14.

Fagnant, D. J., & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice*, *77*, 167–181. doi:10.1016/j.tra.2015.04.003

Favarò, F., Eurich, S., & Nader, N. (2018). Autonomous vehicles' disengagements: Trends, triggers, and regulatory limitations. *Accident Analysis & Prevention*, *110*, 136–148. doi:10.1016/j.aap.2017.11.001

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. doi:10.1371/journal.pone.0184952

Hahsler, M., Chelluboina, S., Hornik, K., & Buchta, C. (2011). The arules R-package ecosystem: Analyzing interesting patterns from large transaction data sets. *Journal of Machine Learning Research*, *12*, 2021–2025.

Hossain, M. M., Rahman, M. A., Sun, X., & Mitran, E. (2021). Investigating underage alcohol-intoxicated driver crash patterns in Louisiana. *Transportation Research Record*, doi:10.1177/03611981211019742

Khattak, Z. H., Fontaine, M. D., Smith, B. L., & Board, T. R. (2019). An exploratory investigation of disengagements and crashes in autonomous vehicles 7p-7p.

Kong, X., Das, S., Jha, K., & Zhang, Y. (2020). Understanding speeding behavior from naturalistic driving data: Applying classification based association rule mining. *Accident Analysis & Prevention*, *144*. doi:10.1016/j.aap.2020.105620

Kutela, B., Das, S., & Dadadhova, B. (2021). Mining patterns of autonomous vehicle crashes involving vulnerable road users to understand the associated factors. *Accident Analysis & Prevention*, *165*.

Leanne, M., Yang, S., & Kuo, J. (2018). Cutting through the hype: measuring driver attention and actual driver take-over times in automated driving. In *Australas. Road Saf. Conf.* (p. 2).

Liu, Q., Wang, X., Wu, X., Glase, Y., & He, L. (2021). Crash comparison of autonomous and conventional vehicles using pre-crash scenario typology. *Accident Analysis & Prevention*, *159*, 106281. doi:10.1016/j.aap.2021.106281

Montella, A., Mauriello, F., Pernetti, M., & Rella Riccardi, M. (2021). Rule discovery to identify patterns contributing to overrepresentation and severity of run-off-the-road crashes. *Accident Analysis & Prevention*, *155*. doi:10.1016/j.aap.2021.106119

Montoro, L., Useche, S. A., Alonso, F., Lijarcio, I., Bosó-Seguí, P., & Martí-Belda, A. (2019). Perceived safety and attributed value as predictors of the intention to use autonomous vehicles: A national study with Spanish drivers. *Safety Science*, *120*, 865–876. doi:10.1016/j.ssci.2019.07.041

Morales-Alvarez, W., Sipele, O., Léberon, R., Tadjine, H. H., & Olaverri-Monreal, C. (2020). Automated driving: A literature review of the take over request in conditional automation. *Electron*, *9*(12), 1–34. doi:10.3390/electronics9122087

NHTSA. (2020). Automated vehicles for safety [WWW Document]. Retrieved from: https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety

Nitsche, P., Thomas, P., Stuetz, R., & Welsh, R. (2017). Pre-crash scenarios at road junctions: A clustering method for car crash data. *Accident Analysis & Prevention*, *107*, 137–151. doi:10.1016/j.aap.2017.07.011

NTSB. (2018). Preliminary report highway HWY18MH010 [WWW Document]. Reprieved from: https://www.urbanismnext.org/resources/preliminary-report-highway-hwy18mh010

Russell, H. E. B., Harbott, L. K., Nisky, I., Pan, S., Okamura, A. M., & Gerdes, J. C. (2016). Motor learning affects car-to-driver handover in automated vehicles. *Science Robotics*, *1*, 1. doi:10.1126/scirobotics.aah5682

SAE. (2019). Society of automotive engineers SAE. SAE Standards News: J3016 automated-driving graphic update [WWW Document].

Salvador, I. S. J. (2019). Package 'compareGroups'. [WWW Document]. Retrieved from: https://cran.r-project.org/web/packages/compareGroups/compareGroups.pdf

Sander, U., & Lubbe, N. (2018). The potential of clustering methods to define intersection test scenarios: Assessing real-life performance of AEB. *Accident Analysis & Prevention*, *113*, 1–11. doi:10.1016/j.aap.2018.01.010

Sinha, A., Vu, V., Chand, S., Wijayaratna, K., & Dixit, V. (2021). A crash injury model involving autonomous vehicle: Investigating of crash and disengagement reports. *Sustain*, *13*, 14. doi:10.3390/su13147938

Song, Y., Chitturi, M., & Noyce, D. (2021). Automated vehicle crash sequences: Patterns and potential uses in safety testing. *Accident Analysis & Prevention*, *153*, 106017. doi:10.1016/j.aap.2021.106017

Taeihagh, A., & Lim, H. S. M. (2019). Governing autonomous vehicles: Emerging responses for safety, liability, privacy, cybersecurity, and industry risks. *Transport Reviews*, *39*(1), 103–128. doi:10.1080/01441647.2018.1494640

Tarmizi, I. A., & Abd Aziz, A. (2018). Vehicle detection using convolutional neural network for autonomous vehicles. In: *2018 International Conference on Intelligent and Advanced System (ICIAS)*. IEEE (pp. 1–5).

Tscharaktschiew, S., & Evangelinos, C. (2019). Pigouvian road congestion pricing under autonomous driving mode choice. *Transportation Research Part C: Emerging Technologies*, *101*, 79–95. doi:10.1016/j.trc.2019.02.004

Ulbrich, S., Menzel, T., Reschka, A., Schuldt, F., & Maurer, M. (2015). Defining and substantiating the terms scene, situation, and scenario for automated driving. In *2015 IEEE 18th International Conference on Intelligent Transportation Systems. Presented at the 2015 IEEE 18th International Conference on Intelligent Transportation Systems* (pp. 982–988).

Wang, H., Yu, Y., Cai, Y., Chen, X., Chen, L., & Liu, Q. (2019). A comparative study of state-of-the-art deep learning algorithms for vehicle detection. *IEEE Intelligent Transportation Systems Magazine*, *11*(2), 82–95. doi:10.1109/MITS.2019.2903518

Wang, J., Zhang, L., Huang, Y., & Zhao, J. (2020). Safety of autonomous vehicles. *Journal of Advanced Transportation*, *2020*. doi:10.1155/2020/8867757

Wang, S., & Li, Z. (2019a). Exploring causes and effects of automated vehicle disengagement using statistical modeling and classification tree based on field test data. *Accident Analysis & Prevention*, *129*, 44–54. doi:10.1016/j.aap.2019.04.015

Wang, S., & Li, Z. (2019b). Exploring the mechanism of crashes with automated vehicles using statistical modeling approaches. *PLoS One*, *14*(3), e0214550.

Wiseman, S., Adler-Golden, S., Ientilucci, E., & Perkins, T. (2021). Enhanced target detection under poorly illuminated conditions. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. IEEE (pp. 1425–1428).

Xu, C., Ding, Z., Wang, C., & Li, Z. (2019). Statistical analysis of the patterns and characteristics of connected and autonomous vehicle involved crashes. *Journal of Safety Research*, *71*, 41–47. doi:10.1016/j.jsr.2019.09.001

Zhang, J., & Xu, C. (2021). Investigating the typical scenarios and contributory factors to crash severity of autonomous vehicle involved collisions using association rule analysis. In *Proceedings of the 100th Annual Meeting of the Transportation Research Board*.