Robotics and Autonomous Systems 182 (2024) 104831

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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

Navigating the blame game: Investigating automated vehicle fault in collisions under mixed traffic conditions

Boniphace Kutela [a,*], Jimoku Hinda Salum [b], Seif Rashidi Seif [c], Subasish Das [d], Emmanuel Kidando [e]

[a] Texas A&M Transportation Institute, 701N Post Oak Ln # 430, Houston, TX 77024, United States
[b] SRF Consulting Group, 2807 Remington Green Circle, Suite 205, Tallahassee, FL 32308, United States
[c] University of Dar es Salaam, P. O. Box 35091, Tanzania
[d] Civil Engineering Program, Texas State University, 601 University Drive, TX 78666, United States
[e] Department of Civil and Environmental Engineering, Cleveland State University, 2121 Euclid Avenue, Cleveland, OH 44115, United States

ARTICLE INFO

Keywords:
Automated vehicle
At-fault vehicle
Reliability
Mixed-effects logistic regression

ABSTRACT

Vehicle being at fault in a crash has extensively been associated with its driver's behaviors and other human errors for human-driven vehicles (HDV). The introduction of automated vehicles (AVs) is expected to eliminate such human errors due to the ability of AVs to communicate with the external environment. However, various reports have documented AVs being at fault in collisions. This study applied text mining and mixed-effects logistic regression (MELR) on crash data involving AVs collected between 2017 and 2022 in California to explore the likelihood of an AV being at fault during a collision. It was found that among 497 crashes, a relatively small percentage (14.29 %) involved AVs being at fault. The text network results revealed patterns of keywords associated with the AVs being at fault. Such patterns include conventional mode of operation, area of impact, and resulting injuries. Furthermore, with about a 93 % prediction accuracy and an 83 % sensitivity score, the MELR results revealed that the likelihood of AVs being at fault increases when they are operated in conventional mode or when disengagement is involved. Moreover, turning, merging, or changing lane movements, unclear weather conditions, and operating on roadways with four or more lanes significantly increased the odds of an AV being at fault during a crash. Conversely, AVs were less likely to be at fault in commercial land use than residential land use, at intersection locations, and when the crash involved a truck. The practical implications of the findings are presented to improve AV operations.

1. Background

Automated vehicles (AVs) represent a transformative advance in automotive technology, promising to redefine road safety, enhance mobility, and revolutionize the transportation landscape. The concept of autonomous vehicles has a surprisingly long history dating back to the 1930s when General Motors and Radio Corporation of America Sarnoff Laboratory jointly funded the research and development of AVs [41]. However, it was not until recently that this technology came into reality when companies such as General Motors, Google, Mercedes Benz, Toyota etc., started operating AVs on public roadways [13].

The introduction of automated vehicle (AV) technology has the potential to improve safety for all road users [5,6,12,34,39]. AV technology is expected to improve safety by overcoming human errors such as speeding, aggressive driving, slow reaction times, inexperience, inattention, and many other shortcomings [9].

Early studies evaluated the AV's safety using survey questionnaires, simulations, or virtual reality [11,15,33] due to lack of crash data. The survey-based studies identified that technology-savvy people's inclination toward AV as a safer mode [12,37]. The simulation-based studies revealed that safety benefits of autonomous vehicles can only be realized with widespread adoption [35]. As the number of autonomous vehicles on the road increases, complex interaction scenarios with vulnerable road users will emerge, potentially leading to a series of crashes involving autonomous vehicles.

Recently, numerous studies have used real-world crash data

* Corresponding author.
E-mail addresses: b-kutela@tti.tamu.edu (B. Kutela), jsalum@srfconsulting.com (J.H. Salum), seifrashid19@gmail.com (S.R. Seif), subasish@txstate.edu (S. Das), e.kidando@csuohio.edu (E. Kidando).

https://doi.org/10.1016/j.robot.2024.104831
Received 28 April 2023; Received in revised form 7 October 2024; Accepted 9 October 2024
Available online 10 October 2024
0921-8890/© 2024 Elsevier B.V. All rights reserved, including those for text and data mining, AI training, and similar technologies.

collected from California to analyze crashes involving AVs ([1,6,12]; F. M. [19,22]). A significant number of studies have focused on identifying the severity, manner of collision, and crash rates of AV crashes [6,40,44,47]. The findings from these studies vary significantly. While some studies reported that AVs performed worse than human-driven vehicles or HDVs [24], other studies praised AVs for being better [9,16]. These studies reported that rear-end collisions were the most common type of collision covering, making up between 61 % and 69 % of all crashes. Intersections are considered as the safety critical areas for AV crashes [9]. Moreover, crashes involving AVs occurred mostly on arterial roadways. Although AV exposure on different functional classifications is not readily available, it is generally assumed that AV trips are limited to urban areas rather than the rural locations. Other factors which determined the impact of a crash were weather conditions, time of the day, the direction of movement of the second party with respect to the AV, number of lanes, lane width, pavement markings, and the driving (i.e., operating) mode of the AV [44].

One of the promising safety aspects of AVs is the reduction of human errors due to their ability to communicate with the external environment by using sensors and reduce the likelihood of being at fault in any collision. By definition, an at-fault driver is one that greatly contributed to causing that crash to happen [26], in this case, the AVs are designed to avoid this scenario. However, previous studies have reported a few AVs being at fault in a crash [4,17,21,28]. A significant number of researchers have quantified the magnitude of AVs being at fault ([16]; F. M. [19,42,50]). The extent of AVs being at fault varies considerably per the number of observations used in the studies. One of the earliest studies [16] analyzed data between September 2014 and November 2015 and reported one crash involving the AV being at fault. Favarò et al. [19] studied 26 crashes involving AVs in California between 2014 and 2017 and found that AVs were at fault for about 15 % (4 out of 26). The most recent study [42] used crash data between 2014 and 2018 and found that the AVs were at fault in 6 % of the crashes. Among the analyzed crashes, 63 % occurred when AVs were in autonomous mode. Furthermore, Wang and Li [50] reported that crash severity significantly increased if the AV was at fault. However, the study found that the AVs were at fault in only 13 % of 133 crashes.

As AVs gradually integrate into public roadways, understanding the dynamics of AV-involved crashes becomes critically important. At-fault determination in AV crashes is crucial for developing effective safety regulations and standards. It provides regulatory bodies with the necessary data to formulate and refine traffic laws that accommodate the unique capabilities and limitations of automated driving systems. Moreover, liability in crashes involving AVs must be clearly defined to manage compensation and insurance claims effectively. There is also a significant research need to analyze AV crash data thoroughly to understand under what conditions AVs are most likely to be at fault. This research can drive technological improvements, enhancing AV systems to prevent crashes.

2. Study objective and contribution

To accurately determine faults in AV involved collisions, it is essential to conduct a detailed investigation that considers all of the relevant factors, including the actions of the human driver and the behavior of the AV system. This requires a thorough understanding of the capabilities and limitations of AV systems, as well as an understanding of how human drivers interact with these systems. The current study intends to explore AV crash data to understand the factors associated with AVs being at fault in a crash. To researchers and operators, the factors associated with the AV being at fault would help better design the infrastructure and the vehicle. On the other hand, the general public's trust in AVs greatly deteriorates when they know that the AV was at fault in a collision [46], and such trust rarely returns to its original level. Therefore, this study evaluated the likelihood of the AVs being at fault in a crash. Understanding the factors associated with the likelihood of the AVs being at fault would help improve the operations in mixed vehicle composition; thus, developers would be able to provide the necessary remedy for the technology. In addition, previous studies have shown that the public is likely to blame the machine in a crash that involves an AV regardless of who is at fault [4,14,38]. Thus, the improved safety performance of AVs would eventually elevate users' perception of AVs' safety. Many existing studies lack comprehensive data on interactions between AVs and HDVs. Furthermore, there is limited analysis of how AVs perform under different road conditions and traffic scenarios, particularly during crashes. This study addresses these gaps by utilizing the California AV crash dataset, which mostly includes collisions between AVs and HDVs. Notably, our research employs a more recent and larger sample size (497 crashes between 2017 and 2022) compared to previous studies using this dataset. By examining the specific circumstances and outcomes of these crashes, our analysis provides new insights into the factors contributing to AVs being at fault.

3. Data description

This study used AV crash data from California. The California Department of Motor Vehicles requires manufacturers who are testing AVs within the state to report any collision resulting in property damage, bodily injury, or death within ten days of the incident when an AV is involved in a crash. The data is archived on the California Department of Motor Vehicles website and publicly available [8]. The collision reports contain several variables, including the day and time of a crash, location, collision type, vehicle damage, characteristics of the other vehicle/participant, and narratives of a crash, to mention a few.

Since 2014, the State of California has been compiling AV collision reports. By December 2022, 522 crash records were available in the database. However, the documents for the 25 crashes occurred between 2014 and 2016 are of low quality, with several missing information and a different format from the documents of the crashes after 2016. Therefore, the crashes that occurred between 2014 and 2016 crash data were not used in this study. This study analyzed 497 crashes involving AVs collected between January 2017 through December 2022 in California

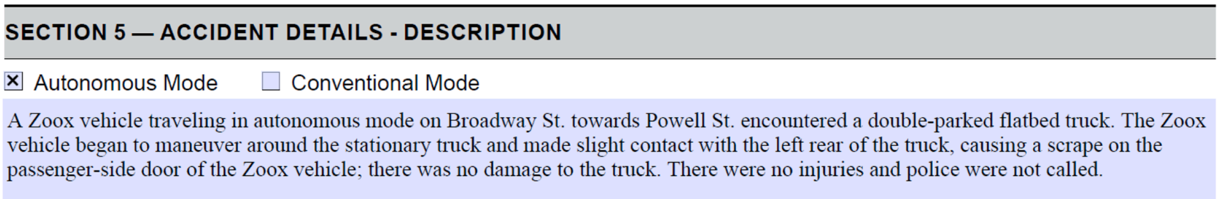
The research team reviewed individual crash report for suitability of this study. The review involved checking the available information in the crash report against the identified variables described and presented in the descriptive analysis section. It was found that all 497 crash reports had all needed information for this study. In addition, the narratives were analyzed to understand the key themes associated with the AVs being at fault via text networks analysis (TNA) as detailed in the methodology section. Some data that could not be obtained in crash reports, such as the number of lanes, were collected via Google Maps or Google Earth. Land use information was collected from the California Census data.

3.1. Identifying vehicle at fault in a crash

To identify the vehicle at fault in a crash, the narrative section of the crash report was used. The narratives clearly state the vehicle at fault, collision type, and sequence of events before the collision. The vehicle that "made contact" with the other was considered to be at fault. For instance, the excerpt in Fig. 1 states that "*The Zoox vehicle began to maneuver around the stationary truck and made slight contact with the left rear of the truck,*" which means the Zoox AV was at fault.

3.2. Descriptive analysis of the variables

Table 1 presents the descriptive statistics of 12 categorical variables of interest, including the response variable (i.e., the at-fault vehicle). These variables were selected based on the review of previous literature ([7,16]; F. M. [19,42,50]) and engineering judgments. The engineering judgment was applied to include variables that have not been



SECTION 5 — ACCIDENT DETAILS - DESCRIPTION

☒ Autonomous Mode ☐ Conventional Mode

A Zoox vehicle traveling in autonomous mode on Broadway St. towards Powell St. encountered a double-parked flatbed truck. The Zoox vehicle began to maneuver around the stationary truck and made slight contact with the left rear of the truck, causing a scrape on the passenger-side door of the Zoox vehicle; there was no damage to the truck. There were no injuries and police were not called.

Fig. 1. Excerpt from crash reports showing automated vehicle (AV) at fault in a collision.

Table 1
Descriptive statistics of the variables ($N = 497$).

Variable	Description	Categories	Count	Percent	Source
At fault vehicle	A vehicle at fault during a collision	AV at fault	71	14.29 %	Crash report - narrative
		HDV at fault	426	85.71 %	
Mode of operation	Mode of operation during a collision	Conventional	173	34.81 %	Crash report - narrative
		Autonomous	324	65.19 %	
Disengagement	Whether the autonomous technology was disengaged before a crash	Yes	83	16.70 %	Crash report - narrative
		No	414	83.30 %	
AV movement	Movement of AV during a collision	Straight	156	31.39 %	Crash report
		Stopping slowing down	243	48.89 %	
		Turning left/right	48	9.66 %	
		Merging/change lane	16	3.22 %	
		Parking related	34	6.84 %	
HDV movement	Movement of HDV during a collision	Straight	221	44.47 %	Crash report
		Stopping slowing down	81	16.30 %	
		Turning left/right	79	15.90 %	
		Merging/change lane	66	13.28 %	
		Others	50	10.06 %	
Land use	Land use where a crash occurred	Residential land use	180	36.22 %	Census data
		Commercial land use	126	25.35 %	
		Mixed land use	191	38.43 %	
Route class	Roadway class where a crash occurred	Local	168	33.80 %	HPMS
		Major Collector	50	10.06 %	
		Minor Arterial	110	22.13 %	
		Principal Arterial/ Interstates	169	34.00 %	
Weather condition	Weather conditions during a crash	Clear	449	90.34 %	Crash report
		Unclear	48	9.66 %	
VRU involvement	Whether a vulnerable road user (VRU), e.g., a pedestrian, was involved in a crash	Yes	72	14.49 %	Crash report - narrative
		No	425	85.51 %	
Number of lanes	The number of lanes at a crash location (includes the turning lanes)	One	102	20.52 %	Google maps
		Two	158	31.79 %	
		Three	87	17.51 %	
		Four or more	150	30.18 %	
Truck involvement	Whether a truck was involved in a crash	Yes	68	13.68 %	Crash report - narrative
		No	429	86.32 %	
Intersection crash	Whether a crash occurred at an intersection	Yes	377	75.86 %	Google maps
		No	120	24.14 %	
AV Company	The manufacturer or licensed AV testing company	GM Cruise	197	39.64 %	Crash report
		Waymo	179	36.02 %	
		Zoox	65	13.08 %	
		Other	56	11.27 %	

Note: HPMS = Highway performance management system.

extensively covered in previous studies. For example, disengagement, the transition from autonomous to human control, was included to explore its potential impact on crash dynamics, as abrupt transitions may introduce human error or delayed responses. Disengagement also means that the vehicle is no longer operating as an AV. Conversely, truck involvement was considered because, due to their larger size, trucks can be more easily detected by AV sensors (from object detection point of view), which may tend to reduce the likelihood of the AV being at fault. It can be observed that the AVs were at fault with a relatively small percentage (14.29 %) compared to human-driven vehicles (HDVs). Previous studies have reported a similar distribution but fewer observations ([16]; F. M. [19,42,50]). Further, more than half (65.19 %) of the crashes involved AVs operating in autonomous mode. Approximately three-quarters of crashes occurred at intersections (75.86 %), highlighting a critical location for crashes involving AVs. Stopping or slowing down movements of AVs tallied the highest crash count of about 243 out of 497 (48.89 %). Interestingly, HDVs were most involved in a crash when making a straight movement (44.47 %).

Besides, the turning movements of AVs registered a significant crash count as well, accounting for 48 of 497 (9.66 %). As for land use, mixed land use recorded the highest number of crashes, with about 191 out of 497 (38.43 %), followed by residential land use, which was in 180 of 497 (36.22 %). The remaining 126 (25.35 %) crashes occurred in commercial land-use areas. According to the roadway class where a crash occurred, principal arterials or interstates had the most crashes with 169 of 497 (34.00 %), followed by local roads with 168 of 497 (33.80 %) crashes. Major collectors had the lowest crash count, with 50 of 497 (10.06 %). The distributions of other categorical variables, such as weather conditions, VRU involvement, number of lanes, and truck involvement, are presented in Table 1.

The AV company variable represents the manufacturer or licensed AV testing company in California. The distribution of crashes by the AV companies shows that most crashes involved GM Cruise, Waymo, and Zoox. The three companies account for almost 88 % of all crashes, with large proportion from GM Cruise. This variable will later be used as the grouping variable for Mixed Effect Logistic Regression (MELR) since it represents unobserved company related protocols, including the reporting style, which may vary considerably.

3.3. Crash trend by vehicle at fault

Fig. 2 shows the trend of crashes involving AVs and HDVs over six years. It also presents the proportion of vehicles at fault during collisions. It can be observed that the crashes involving AVs increased from 29 in 2017 to 105 in 2019. There was a decline in the number of crashes in 2020, most probably due to COVID-19 restrictions, which resulted in fewer vehicles operating on the roadways during the lockdown.

Since 2020, the number of crashes has been on the rise, with 2022 registering the highest number of crashes (147) over the six years. Fig. 2 also shows that the proportion of AVs being at fault during a crash relatively increased over the six years despite being only 12.9 % in 2022. In fact, for the first three years, the proportion of AVs being at fault was less than 10 % even though 2019 had a significant number of crashes involving AVs. On the other hand, for the two most recent years (2020 and 2021), the proportion of AVs being at fault increased about three times. This shows that operators should explore the scenarios that increased the proportion of AVs being at fault.

4. Methodology

As presented in a flowchart in Fig. 3, following the data collection from the three sources, this study employed two approaches to explore the likelihood of the AVs being at fault. First, text networks were used to understand the key themes associated with AVs being at fault. Second, mixed-effects logistic regression was applied to quantify the impact of predictors on the likelihood of AV being at fault during a collision. Incorporating both qualitative and quantitative methods, specifically Text Network Analysis (TNA) and regression analysis, offers a comprehensive approach to investigating the complex dynamics of AV interactions and crash outcomes. The mixed-effect logistic regression is adept at handling the variability across multiple observations of the same entity, which is crucial in traffic safety studies where repeated measures on the same traffic segments or vehicles often occur. This statistical method allows for the exploration of fixed effects of observed variables while accounting for random variations, making it highly suitable for analyzing the influence of specific factors on crash severity across different conditions and contexts. On the other hand, TNA provides a qualitative depth that complements the quantitative breadth of logistic regression. It enables the extraction and interpretation of nuanced themes and patterns from textual data, such as drivers' descriptions of incidents or comments on road conditions, which are often overlooked in purely quantitative analyses. This dual approach is particularly effective in addressing complex systems like traffic environments where both human factors and technical elements interplay. The details for the two approaches are presented below.

4.1. Text networks

Text network analysis (TNA) is a branch of text mining that uses edges and nodes to provide insights from unstructured textual data [52]. The *nodes* represent the keywords, and the *edges* indicate the co-occurrence of the keywords. The thicker the edge connecting two nodes, the higher the frequency of the keywords' co-occurrence. Keywords with a similar pattern are connected to form a *community*.

Text normalization is the first step to creating a network, followed by transforming unstructured to structured data and generating nodes and links. Normalization involves the removal of connecting words and symbols and converting all words to lowercase. The transformation from structured to unstructured data is performed by creating a matrix of keywords from a corpus of words. Lastly, mapping nodes and links is performed by systemically assigning a new node for a new keyword and increasing a unit frequency for a node with a repeating keyword. At first, all the consecutive keywords are given an identification whose name is equal to that of the node. The algorithm then checks if the keywords reappear in other sentences. If the two keywords appear, the algorithm first checks if the pair exists. If the pair does not exist, the algorithm sets a new node with a connection, but if the pair exists, the algorithm adds a unit weight to the existing edge. The higher the co-occurrence frequency of keywords is, the higher the connecting edge's weight [30,31,36]. The last step is providing key insights from the networks [52]. In general, several network measures, including degree centrality, betweenness centrality, collocations, and co-occurrences, among others, exist. However, this study used the network's topology and co-occurrence of the keywords to deduce insights from the networks.

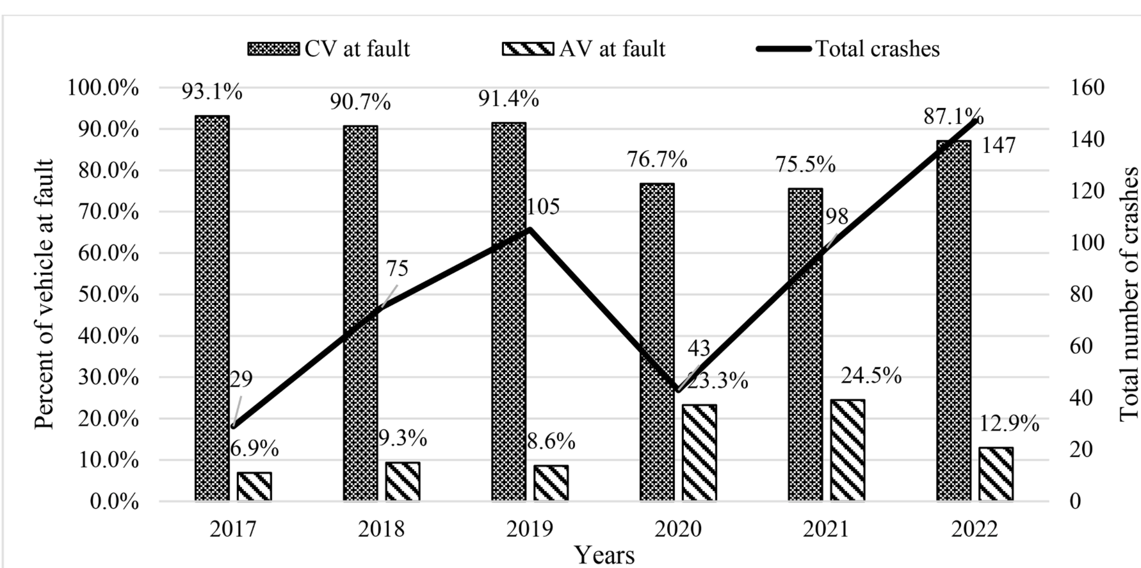


Fig. 2. The trend of crashes involving automated vehicles (AVs) and the proportion of vehicles at fault.

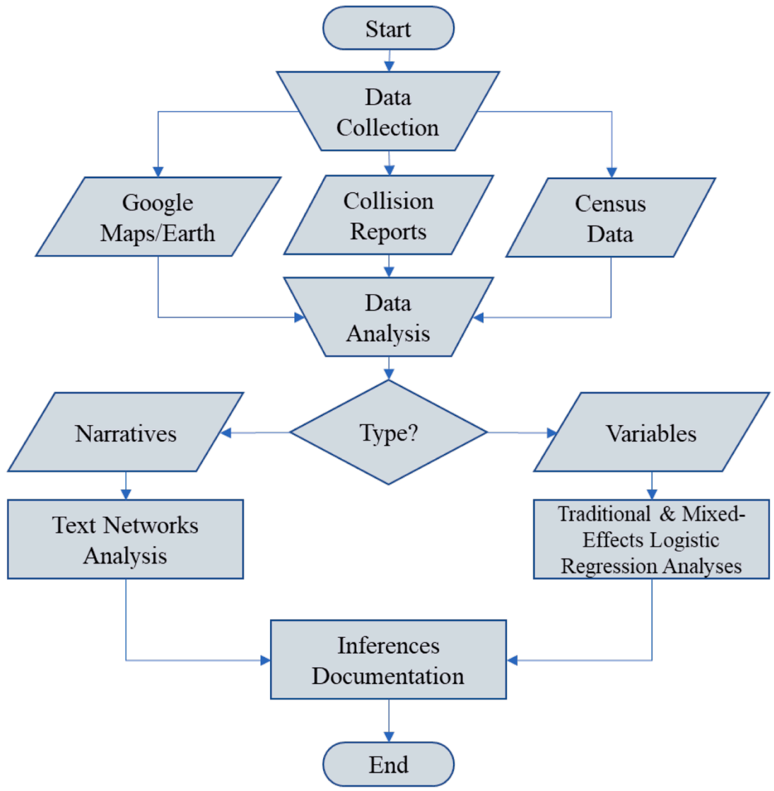


Fig. 3. Methodology flowchart.

4.2. Mixed-effects logistic regression

The AV being at fault in a collision can be expressed as a binary variable (yes/no). Thus, binary regression models, such as logit and probit, are the best candidate for this data type. Previous studies have favored logit over the probit model due to the underlying latent assumptions and the ease of interpretation of the parameters using the odds ratios [51].

The logistic regression follows a Bernoulli probability function with the dependent variable Y_i having two outcomes (either 1 if a success, or 0 if a failure). The θ_i, which is the probability that the event is successful, can be expressed as an inverse logistic function of a vector X_i of explanatory variables as:

$$\theta_i = \frac{1}{1+ e^{-X_i\beta}} \tag{1}$$

Upon linearizing the θ_i, it can be written as

$$logit(\theta_i) = ln\left(\frac{\theta_i}{1-\theta_i}\right) = \widehat{\beta}_0 + \widehat{\beta}_1 X_1 + \ .... \ + \widehat{\beta}_n X_n + \varepsilon_j \tag{2}$$

Whereby $\widehat{\beta}_1$ up to $\widehat{\beta}_n$ are the variable coefficients to be estimated, the coefficient $\widehat{\beta}_0$ is a constant term, and ε_j is the random effect term due to different AV companies.

As expressed in Eq. (2), the logistic regression model incorporated the mixed-effects term to account for random effects due to cluster-level variation, different AV companies in this case. Being a generalized linear mixed model (GLMM) type, a mixed-effects logistic regression estimates both fixed and random effects simultaneously, allowing for more comprehensive data analysis [3]. In the current study, the fixed effects represent the relationship between the predictor variables and the likelihood that an AV is at fault, while the random effects account for the variation that may exist from AV company-level factors that cannot be easily measured or controlled.

Both traditional and mixed-effects logistic regression models were developed to compare the estimated results. The comparison was based on the prediction accuracy and sensitivity of the models, using in and out of samples. The model interpretation was based on the odds ratios (OR), which are the exponents of the coefficients. A variable with an OR greater than one is associated with an increased probability of AV at fault occurrence, while the vice versa is true for an OR of less than one. On the other hand, a variable with an OR equal to one indicates that the variable has no significant impact on an AV being at fault [32,48].

5. Results and discussion

This section presents the results and discussions. It is divided into two parts: Text Network and Regression results and discussions. The text network results show the major themes associated with the at-fault AV during a crash. On the other hand, the regression results quantify the extent to which each predictor affects the likelihood of AVs being at fault. The detailed discussions are presented in the next sections.

5.1. Text networks results and discussion

In the text networks, the arrangement of the co-occurrence explains the sequence of the events. The key patterns revealed in the text networks were validated by the actual sentences from the raw data to make sure that the authors' assumptions/insights aligned with the raw data. This was done by taking the key phrases and searching them in the raw data. A portion of the raw statement from the data was then presented in the results and discussion to support the argument.

Fig. 4 presents the text networks for at fault AVs during a crash. The figure shows that the keyword *"vehicle"* is the center of the community formed by the other five major keywords- *front, damage, passenger, side, and driver.*

The co-occurrences of the keywords *"front vehicle", "front bumper", "front damage"* and *"passenger vehicle"* indicate that the front bumper of the AV at fault was damaged most of the time when the AV was involved

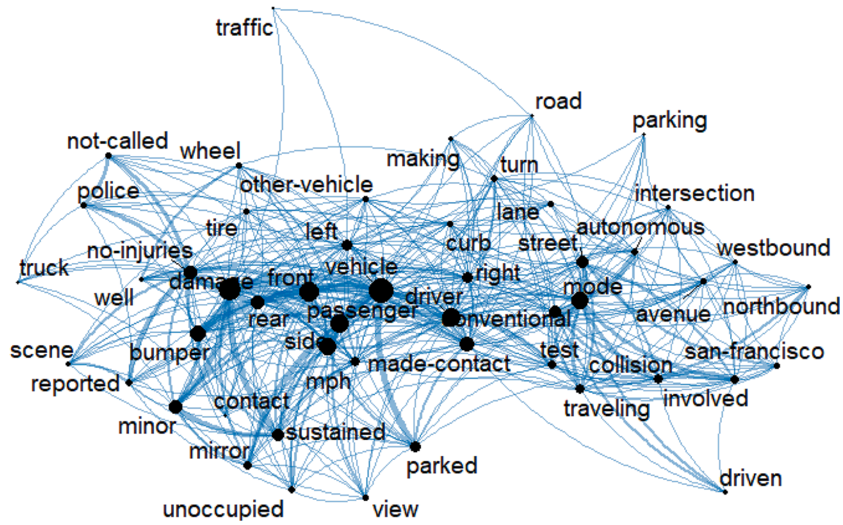


Fig. 4. Text networks for at-fault AV during a crash.

in a crash. It also indicates that the at-fault AV mostly impacted the front vehicle, with the front vehicle being a passenger car in most cases. For instance, a part of one of the narratives states that *"While making maneuver to avoid parallel parking vehicle, driver Cruise-AV made-contact with other-vehicle that was pulling out the driveway on the opposite side street, scratching front center bumper other-vehicle and denting front left bumper Cruise-AV."* The word "*vehicle"* is heavily linked to "*parked",* indicating that the AV is likely to be at fault when the crash involves an AV and a parked vehicle. The following narrative, "*While pulling out parking spot on Jackson, driver Cruise-AV made-contact with other-vehicle parked in front it, damage Cruise-AV front bumper assembly, left headlamp assembly, and left front radar",* justifies that AVs were at fault when collided with a parked vehicle. The word "*damage"* is heavily linked to the words "*minor"* and "*sustained",* indicating that there is a high possibility of getting minor damage when the AV is at fault in a crash. This observation suggests that crash severity increases when AV is in autonomous mode; therefore, if most of the crashes had minor damage, it can be concluded that AVs were likely to be in conventional mode when they were at fault in collisions. This can be justified by the text networks, which show a heavy link between the words "*conventional*" and "*mode"*.

The link between the words "*no-injuries*", "*police*", and "*not-called"* indicate that the police were not called most of the time when the crash did not involve injuries. The link between the keywords "*damage*", "*reported*", and "*bumper*" indicate that most damages reported were for bumper damage. The keyword "*street"* is heavily linked to "*autonomous"* and "*mode"* while the word "*avenue*" is heavily linked to "*conventional"* and "*mode"*. This shows that the AV is likely to be in autonomous mode in crashes occurring on streets, while for crashes occurring in avenues, the AV is likely to be in conventional mode.

The text-mining results facilitated an understanding of the key patterns for crashes that the AVs are either at fault or not at fault. Further, the text mining facilitated identifying other key variables such as truck involvement, disengagement, and VRU involvement, which would not be identified using the structured data. To quantify the associated factors for the likelihood of the AV being at fault, a regression was necessary for this case. The next section presents the mixed-effects logistic regression model results and discussions.

5.2. Mixed-effects logistic regression results and discussion

This study used several variables that are assumed to be associated with the likelihood of an AV being at fault. Prior to developing statistical models, the correlation analysis between independent variables was performed. The rationale of this step was to avoid highly correlated independent variables in the model [27]. It can be observed that all variables have low correlation coefficients except weather conditions (Fig. 5). Therefore, all variables were used in the fitted model and used for inference.

To demonstrate the benefit of using the mixed-effect model over the traditional logistic regression, performance of these two models were evaluated. The performance metrics adopted in this study are prediction accuracy, sensitivity score, and the Bayesian Information Criterion (BIC). The prediction accuracy estimates the proportion of correctly predicted at-fault and not-at-fault events, while the sensitivity score represents the ability of the model to estimate the at-fault vehicle group correctly. On the other hand, BIC is the metric that provides a tradeoff between prediction accuracy and model complexity [49]. To determine the performance of the developed models, the in-sample and out-of-sample prediction was performed. The in-sample prediction was performed using an entire dataset used to develop the models. On the other hand, the data was partitioned into 60 % for training and 40 % for testing for out-of-sample prediction.

Fig. 6 presents the models' performance based on accuracy and

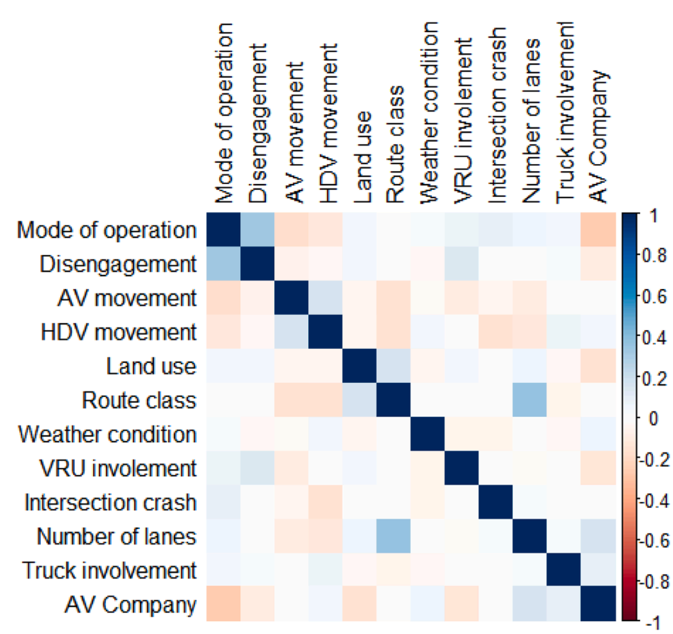


Fig. 5. Correlation matrix of independent variables.

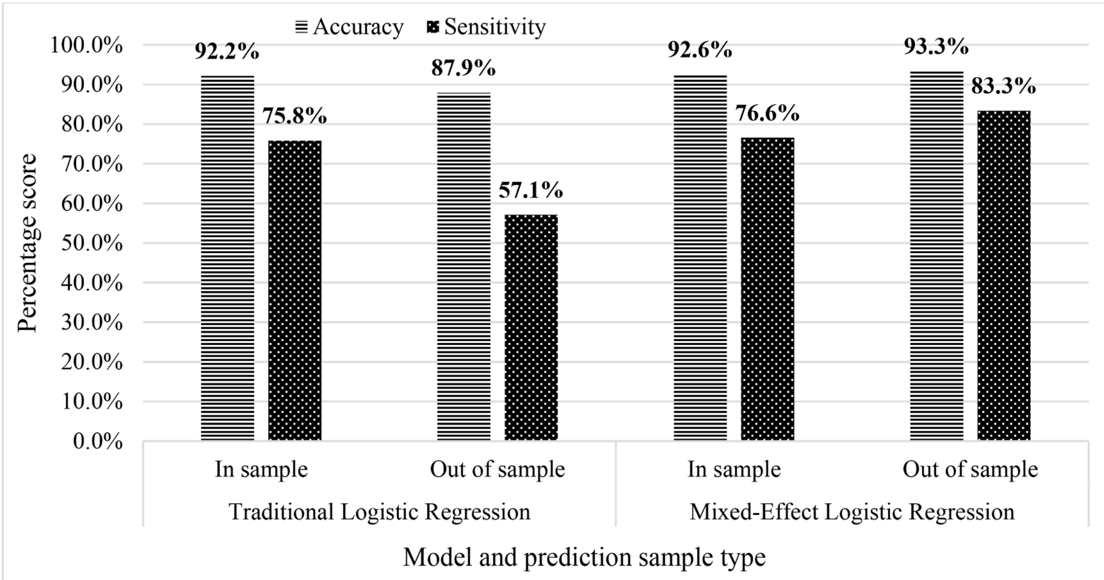


Fig. 6. Models' performance assessment parameters.

Table 2
Traditional and mixed-effects logistic regression results.

	Traditional Logistic Regression				Mixed-Effects Logistic Regression			
	Estimate	OR	S. E	p-value	Estimate	OR	S. E	p-value
Intercept	−1.281	0.28	0.805	0.112	−1.074	0.34	0.870	0.217
Mode of operation								
Autonomous	***−2.321***	***0.10***	***0.544***	***<0.001***	***−2.308***	***0.10***	***0.551***	***<0.001***
Disengagement								
Yes	***1.859***	***6.42***	***0.614***	***0.002***	***1.882***	***6.57***	***0.624***	***0.003***
AV movement								
Stopping slowing down	***−2.831***	***0.06***	***0.711***	***<0.001***	***−2.965***	***0.05***	***0.735***	***<0.001***
Turning left/right	**1.086**	**2.96**	**0.631**	**0.085**	**1.075**	**2.93**	**0.640**	**0.093**
Merging/change lane	***2.181***	***8.86***	***0.803***	***0.007***	***2.113***	***8.27***	***0.812***	***0.009***
Parking related	−0.296	0.74	0.607	0.626	−0.282	0.75	0.616	0.647
HDV movement								
Stopping slowing down	0.054	1.06	0.692	0.938	0.052	1.05	0.696	0.941
Turning left/right	***2.951***	***19.13***	***0.663***	***<0.001***	***2.846***	***17.22***	***0.680***	***<0.001***
Merging/change lane	−0.907	0.40	0.927	0.328	−0.880	0.41	0.940	0.349
Other movements	***3.371***	***29.11***	***0.674***	***<0.001***	***3.294***	***26.96***	***0.682***	***<0.001***
Land use								
Commercial land use	***−1.904***	***0.15***	***0.681***	***0.005***	***−1.749***	***0.17***	***0.720***	***0.015***
Mixed land use	0.258	1.29	0.449	0.565	0.322	1.38	0.465	0.489
Route class								
Major Collector	−1.003	0.37	0.716	0.161	−0.834	0.43	0.750	0.266
Minor Arterial	−0.425	0.65	0.569	0.456	−0.365	0.69	0.582	0.530
Principal Arterials /Interstate	**−0.924**	**0.40**	**0.561**	**0.099**	−0.862	0.42	0.576	0.135
Weather condition								
Unclear	***1.358***	***3.89***	***0.654***	***0.038***	**1.276**	**3.58**	**0.662**	**0.054**
VRU involvement								
Yes	***−1.803***	***0.16***	***0.852***	***0.034***	***−1.707***	***0.18***	***0.861***	***0.047***
Intersection crash								
Yes	**−0.762**	**0.47**	**0.420**	**0.070**	**−0.850**	**0.43**	**0.442**	**0.054**
Number of lanes								
Two lanes	0.035	1.04	0.572	0.951	−0.149	0.86	0.625	0.812
Three lanes	0.765	2.15	0.761	0.315	0.603	1.83	0.787	0.444
Four or more lanes	***1.640***	***5.16***	***0.693***	***0.018***	**1.393**	**4.03**	**0.762**	**0.067**
Truck involvement								
Yes	***−1.408***	***0.24***	***0.676***	***0.037***	***−1.476***	***0.23***	***0.691***	***0.033***
Random effects (*AV company*)								
	σ^2				0.104		0.322	
Model Summary								
Number of observations	497							
Groups (AV company)	4							
BIC	−244.3				−331			

sensitivity. It can be observed that mixed-effect logistic regression performs relatively better than traditional logistic regression. While the in-sample prediction accuracies do not vary significantly across models (92.2 % for traditional and 92.6 % for Mixed-effect), the out-of-sample and sensitivity scores have improved for the mixed-effect logistic regression. The out-of-sample prediction accuracy for the mixed-effect model (93.3 %) is about 5 % higher than that of traditional regression (87.9 %), which is deemed significant. Further, the difference in the sensitivity scores for out-of-sample (83.3 % against 57.1 %) provides more insights into the performance of the two models. The results of BIC also align with those of the sensitivity score, suggesting that the mixed-effect logistic regression offers benefits over using traditional logistic regression. This finding could be attributed to the ability of the mixed-effect logistic regression to address the issue of unobserved heterogeneity of the dataset. Several previous studies have made a similar conclusion regarding the benefit of mixed-effect models over traditional regressions.

Table 2 provides the results of the traditional logistic regression and mixed-effects logistic regression analyses. This table presents the estimate, odds ratio (OR), standard error (S.E.), p-value of each variable, and the random parameter estimate. The OR measures the change in the odds of the outcome variable for a one-unit change in the independent variable. A value greater than 1 indicates an increase in the odds of the outcome, whereas a value less than 1 indicates a decrease in the odds of the outcome [29,39]. The mixed effect coefficient 0.104 of the random parameter term σ^2 signify the unobserved variability across the AV companies. Based on the results of model performance, the discussions of the model results in the next sections are based on the mixed-effects logistic model.

5.2.1. Mode of operation

Two modes of operation that is conventional, and autonomous modes were assessed. An AV is considered autonomous when operates or drives in autonomous mode. Such a technology performs the dynamic driving task without a human actively supervising it. On the other side, when the human driver operates the AV is considered as it operates in a conventional mode, disengaged autonomous technology [8]. Regardless of the operation mode, what the general public observes on the roadway is the "automated vehicle" and not the mode of operation. Thus, it is imperative to include manually operated automated vehicles in the analysis.

The model results in Table 2 show that the likelihood of an AV being at fault during a crash decline when it operates in autonomous mode compared to conventional mode. The mixed-effects logistic regression model shows that the mode of operation significantly affects the likelihood of an AV being at fault during a crash. The OR for the autonomous mode of operation is 0.10 ($p < 0.001$), which indicates that the odds of an AV being at fault during a crash are about 90 % lower when the AV is operating in autonomous mode compared to conventional mode after controlling for other variables in the model. This finding suggests that the AV is less likely to be at fault during a crash when autonomous technology is engaged. This is consistent with the assumption that autonomous technology can improve driving safety by reducing human errors by improving situational awareness. It also suggests that human errors may contribute to crashes involving AVs operated in conventional mode. The effectiveness of AVs in reducing crash fault rates can be attributed to advanced technologies in autonomous systems that enhance situational awareness, continuous monitoring, and rapid response capabilities, which are less susceptible to common human limitations like distraction, fatigue, and delayed reaction times. These systems are designed to maintain consistent vigilance, adhere strictly to traffic laws, and execute precise maneuvers, thereby enhancing road safety. The inclusion of random effects for AV companies in the mixed-effects model indicates that there is heterogeneity in the effect of mode of operation across different AV companies. This suggests that different AV companies may have different levels of technology development or safety culture, which may affect the performance of AVs in different modes of operation. Several previous studies found similar results. For example, Sinha et al. [42] revealed that only 1 in 25 sideswipe crashes would involve an AV at fault when operating in autonomous mode compared to 1 in 8 when operating in conventional mode. Furthermore, a similar conclusion has been reported by Dixit et al. [16] study, which indicates that the AV is less likely to be at fault in a crash when operating in autonomous mode.

5.2.2. AV's disengagement

According to the mixed-effects logistic regression results, the OR of an AV being at fault during a crash when its autonomous technology was disengaged is 6.57 (p-value=0.003), when all other variables are held constant. This indicates that the odds of an AV being at fault during a crash are 6.57 times higher when its autonomous technology is disengaged compared to when it is operating autonomously. This statistically significant result suggests that disengagement of the autonomous technology may increase the risk of an AV being at fault during a crash. When human drivers take control, they might introduce errors in judgment, reaction time, or situational awareness compared to the AV's programming. Further, drivers might disengage the autonomous technology in poor driving conditions or complex situations (F. [18]), increasing the risk of a crash and the likelihood of the AV being at fault.

5.2.3. AV's movement prior to a collision

The mixed-effects model was also used to investigate the effect of the AV's movement on the likelihood of the AV being at-fault during a crash. The base for comparison was an AV moving straight, and the model assessed the effect of turning left, turning right, and changing lanes. The results showed that turning and merging or changing lanes both significantly increased the odds of the AV being at-fault during a crash compared to moving straight. Specifically, the odds ratio for turning left/right was 2.94, indicating that an AV turning left/right was almost thrice as likely to be at-fault during a crash compared to an AV moving straight. Similarly, the odds ratio for merging or changing lanes was 8.27, indicating that an AV merging or changing lanes was also eight times as likely to be at fault during a crash compared to an AV moving straight. On the other hand, parking-related did not have a significant effect on the likelihood of the AV being at fault during a crash compared to moving straight. Overall, these results suggest that the AV's movement before a crash plays an important role in determining the likelihood of the AV being at fault. AVs that turn, merge, or change lanes are more likely to be at fault during a crash compared to AVs that move straight. The findings suggest that the AVs algorithms for merging and turning require improvement to be able to navigate these conditions. Such improvement may include the risk-based merging decisions, which causes less abrupt deceleration of AVs, provides safe gap, and lower the conflict risk in terms of Time Exposed Time-to-Collision (TET) [23]. These results can be used to inform the designers and car manufacturers of AVs, focusing on improving their ability to navigate turns and lane changes safely.

5.2.4. HDV's movement prior to a collision

Based on the mixed-effects model results on the HDV's movement, it was found that when the HDV was making a turn, the odds of the AV being at fault were significantly higher compared to when the HDV was moving straight (OR=17.22, $p < 0.001$). These results suggest that the AV is more likely to be at fault during a crash when the HDV is making a turn, regardless of the direction of the turn. This could be because turns involve more complex maneuvers and may require more attention and decision-making from both the AV and the human driver of the HDV. Turns may also involve blind spots, which could increase the risk of a collision. This is, when a vehicle is shifting lanes or changing direction, there's a higher chance of another vehicle being in the driver's blind spot [10].

5.2.5. Land use

The model results on the land use variable suggest that the type of land use where a crash occurred significantly predicted the likelihood of an AV being at fault. Residential land use was considered as the base for comparison. The odds ratio (OR) of an AV being at fault during a crash in commercial land use was 0.17, meaning that crashes occurring in commercial areas are less likely to be the AV's fault than crashes in residential areas. On the contrary, the OR of an AV being at fault during a crash in mixed land uses was 1.38, meaning that crashes occurring in mixed land use areas were 1.38 times more likely to be the AV's fault than crashes in crashes in residential areas. However, the OR of an AV being at fault during a crash in mixed land uses was not statistically significant (OR=1.38, $p = 0.489$). Overall, these results suggest that the type of land use where a crash occurs is an important predictor of the likelihood of an AV being at fault. This could be attributed to the level of interaction between AVs and other road users in different land use areas. Other road users violate rules more often in residential areas and expect an AV to act as a conventional vehicle, i.e., to forgive their mistakes; unfortunately, AVs operate the way they are programmed [40]. For this reason, manufacturers of AVs could think of adding an indicator to alert other road users that the vehicle is operating in autonomous mode. Moreover, it is expected that residential areas have more parking facilities. This finding further supports the previous findings.

5.2.6. Route class

Based on the mixed-effects logistic regression model results, the "Route class" variable has no statistically significant effect on the likelihood of an AV being at fault during a crash. The variable showed statistical significance at a 90 % level for traditional logistic regression. The difference in the results between traditional and mixed-effect logistic regression can be due to the unexplained heterogeneity present in the traditional regression. However, the odds are less than 1 in all roadway categories, meaning that crashes on these types of roadways are associated with a lower likelihood of AV fault compared to crashes on local routes. These findings may be restricted to California only, as the AV testing is performed in a dedicated area [8]. Thus, most crashes that involve AVs at fault are likely to be observed in that area.

5.2.7. Weather condition

The odds of AVs being at fault in a crash were more than three times (OR = 3.58) more significant under unclear weather conditions than under clear weather conditions. It is worth noting that the unclear (adverse) weather conditions in this study included fog, rain, snow, storm, and cloudy. The underlying cause could be attributed to visibility issue, AV sensors having difficulty detecting pavement markings, traffic signs, signalizations, other vehicles, or any other obstacle when it is operated under unclear weather conditions. With limitations in sensor data due to unclear weather, AVs might struggle to make safe driving decisions [45]. They might misinterpret the environment [53], leading to errors in judgment and potentially being at fault in a crash. The finding should be considered as an opportunity for AV operators to collect more driving data in adverse weather conditions to train their AVs. The data collection during these conditions should be performed in the presence of the vehicle's AV operator to avoid crashes.

5.2.8. VRU involvement

The results presented in Table 2, the mixed-effects logistic regression model results suggest that the VRU involvement factor is statistically significant on the likelihood of an AV being at fault. The odds of an AV being at fault during a crash is 0.18 compared to when there is no VRU involvement. This result suggests that the odds of an AV being at fault during a crash are about 82 % lower when a VRU is involved compared to when there is no VRU involvement. AVs are likely programmed to be extra cautious around pedestrians, cyclists, and other VRUs. This might involve slower speeds, increased following distances, and more attention to potential VRU movements compared to interacting with other vehicles. However, the findings should be interpreted with caution as the data is reported by the operators, who might not want to position the AVs at fault, especially when the collision involves VRUs due to the pushback and riots that have recently emerged from crashes involving pedestrians [25].

5.2.9. Intersection crash

The mixed-effects logistic regression model results suggest that the occurrence of a crash at an intersection has a marginally significant effect on the likelihood of an AV being at fault during a crash. The estimated log odds ratio for an intersection crash is -0.850, with a standard error of 0.442 and a p-value of 0.054. This means that the odds of an AV being at fault during a crash are 0.427 times lower when a crash occurs at an intersection compared to when a crash occurs outside of an intersection, after accounting for other predictor variables in the model. The situation can be attributed by several possible reasons. First, intersections are more likely to have traffic signals and stop signs, which can provide clearer guidance to AVs compared to situations with less structured traffic flow. Also, traffic tends to slow down at intersections, reducing the severity of potential crashes and giving AVs more time to react. Further, drivers might be more attentive at intersections, anticipating potential hazards from other vehicles, especially in the presence of warning alerts [2], which could influence how they interact with AVs.

5.2.10. Number of lanes

The results suggest that crashes on roadways with three lanes and four or more lanes are associated with a higher likelihood of the AV being at fault compared to crashes on roadways with one lane. The OR for three lanes is 1.83 with a p-value of 0.444, which means that the odds of an AV being at fault in a crash on a roadway with three lanes are 1.83 times higher than the odds of the AV being at fault on a roadway with one lane, but this difference is not statistically significant. The OR for four or more lanes is 4.03 with a p-value of 0.067. This means that the odds of an AV being at fault in a crash on a roadway with four or more lanes are four times higher than that of an AV being at fault on a roadway with one lane. However, the p-value is greater than 0.05, indicating that the difference is not statistically significant at the 95 % confidence level. Possibly, the increased likelihood of the AVs being at fault may be explained by the increased traffic complexity in multilane compared to single-lane roadways. Multi-lane roads present a more complex environment for AVs to navigate. They need to track and react to more vehicles simultaneously, including those changing lanes, entering/exiting the road, and potentially exhibiting unpredictable behavior. Further, AVs rely on clear lane markings and signage to understand traffic flow. On complex multi-lane roads, exits, merges, and specific lane designations can create confusion for AVs if the markings or signage are unclear or misinterpreted [20].

5.2.11. Truck involvement

The mixed-effects logistic regression model results suggest that the involvement of a truck in a crash has a significant effect on the likelihood of an AV being at fault during a crash. The estimated log odds ratio for truck involvement is -1.476, with a standard error of 0.691 and a p-value of 0.033. This means that the odds of an AV being at fault during a crash are 0.229 times lower when a crash involves a truck compared to when a crash does not involve a truck, after accounting for other predictor variables in the model. In conclusion, the results suggest that the involvement of a truck in a crash is associated with a lower likelihood of an AV being at fault during a crash. This relationship is statistically significant and relatively strong. The findings may have important implications for AV development and safety efforts, as they suggest that AVs may be less likely to be at fault during a crash if they interact with trucks on the road. This can be linked to the fact that AVs are equipped with advanced sensor systems (LiDAR, radar, and cameras), which provide a 360° view around the vehicle. Such systems enable them to detect large objects, such as trucks, with high reliability and from greater distances than human drivers typically can, particularly when road geometry is adequate [43]. These systems are designed to continuously monitor the surrounding environment and are not affected by human factors like fatigue, distraction, or limited reaction time. The ability to identify large objects more quickly and accurately enables them to maintain safer following distances and execute maneuvers with precision. Additionally, these sensors work in conjunction with advanced algorithms such as Vehicle-to-Everything (V2X) communication to predict the patterns of surrounding vehicle movements, enhancing the AV's ability to respond proactively to changes in speed, direction, or road conditions. This capability allows AVs to make more informed decisions in complex traffic scenarios, such as navigating around large trucks during lane changes or responding to unexpected stops.

6. Conclusions and future works

AV technology holds promise for enhancing road safety and mobility. However, despite ongoing testing and deployment in states like California, crashes involving AVs and HDVs have been reported. This study investigated the likelihood of an AV being at fault in a collision, analyzing 497 crashes involving AVs reported in California between 2017 and 2022. By applying text mining and mixed-effects logistic regression, this research provides an in-depth analysis of factors contributing to at-fault AV crashes, which is less explored in earlier studies.

6.1. Key findings

The study's results revealed that only 14.29 % of the crashes involved AVs being at fault, predominantly when operating in conventional mode. The analysis identified several factors that significantly increase the likelihood of an AV being at fault, including merging or lane-changing movements, unclear weather conditions, and operation on roadways with four or more lanes. Conversely, AVs were less likely to be at fault at intersections, in commercial land use areas, and when crashes involved a truck. The study provides initial insights into the factors contributing to the likelihood of an AV being at fault in a crash. These new insights would inform the manufacturers of the areas that they need to improve. The findings may also challenge transportation engineers to design transport facilities that can optimally accommodate both HDVs and AVs.

6.2. Implications

The study's findings have direct implications for AV manufacturers and transportation engineers. For manufacturers, this research identifies critical areas, such as lane-changing, merging processes, and operation in adverse weather conditions, that require further refinement in AV technologies. Improved sensor capabilities and reprogramming of AVs to better handle these situations could mitigate the risk of AVs being at fault in crashes. For transportation engineers, the results suggest the need to design roadways and infrastructure that can optimally accommodate both HDVs and AVs, facilitating safer interactions between the two. The insights gained from this research serve as a foundation for developing safety protocols and operational guidelines to enhance AV integration into existing traffic systems.

6.3. Future work

While this study offers a focused analysis of AV fault in crashes, it also identifies areas for future research. Given the study's limitations regarding data availability, future work should explore the broader ecosystem of road use by explicitly investigating the interactions between AVs and other road users. Observational studies or simulations modeling various traffic scenarios can provide deeper insights into these interactions. Additionally, employing advanced modeling techniques to capture the nuances of human-AV interactions would further enhance our understanding of how AVs adapt to other road users' behaviors. Another avenue for future research involves examining the differences in AV crash statistics based on the source of crash reports, such as comparing data reported by operators with those collected by the police. Such investigations would help validate the reliability of AV crash data and provide a more comprehensive understanding of AV safety in real-world operations.

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: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation. **Jimoku Hinda Salum:** Writing – review & editing, Writing – original draft. **Seif Rashidi Seif:** Writing – review & editing, Writing – original draft, Data curation. **Subasish Das:** Writing – review & editing, Writing – original draft, Methodology. **Emmanuel Kidando:** Writing – review & editing, Writing – original draft, Methodology.

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.

Acknowledgments

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

Data availability

Data will be made available on request.

References

[1] H. Alambeigi, A.D. McDonald, S.R. Tankasala, Crash themes in automated vehicles: A topic modeling analysis of the California Department of Motor Vehicles automated vehicle crash database. arXiv preprint arXiv:2001.11087.
[2] S. Bakhtiari, T. Zhang, T. Zafian, S. Samuel, M. Knodler, C. Fitzpatrick, D.L. Fisher, Effect of visual and auditory alerts on older drivers' glances toward latent hazards while turning left at intersections, Transp. Res. Rec. 2673 (9) (2019) 117–126, https://doi.org/10.1177/0361198119844244/ASSET/IMAGES/LARGE/10.1177_0361198119844244-FIG6.JPEG.
[3] D. Bates, M. Mächler, B.M. Bolker, S.C. Walker, Fitting linear mixed-effects models using lme4, J. Stat. Softw. 67 (1) (2015), https://doi.org/10.18637/jss.v067.i01.
[4] J.M. Bennett, K.L. Challinor, O. Modesto, P. Prabhakharan, Attribution of blame of crash causation across varying levels of vehicle automation, Saf. Sci. 132 (2020) 104968, https://doi.org/10.1016/J.SSCI.2020.104968.
[5] P. Bhavsar, P. Das, M. Paugh, K. Dey, M. Chowdhury, Risk analysis of autonomous vehicles in mixed traffic streams, Transp. Res. Rec. J. Transp. Res. Board 2625 (1) (2017) 51–61, https://doi.org/10.3141/2625-06.
[6] A.M. Boggs, B. Wali, A.J. Khattak, Exploratory analysis of automated vehicle crashes in California: a text analytics & hierarchical Bayesian heterogeneity-based approach, Accid. Anal. Prev. 135 (2020) 105354, https://doi.org/10.1016/j.aap.2019.105354.
[7] A.M. Boggs, B. Wali, A.J. Khattak, Exploratory analysis of automated vehicle crashes in California: a text analytics & hierarchical Bayesian heterogeneity-based approach, Accid. Anal. Prev. 135 (2020) 105354, https://doi.org/10.1016/j.aap.2019.105354.
[8] California-DMV. (2023). Autonomous vehicle collision reports - California DMV. https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/autonomous-vehicle-collision-reports/. Accessed June 2023.
[9] H. Chen, H. Chen, Z. Liu, X. Sun, R. Zhou, Analysis of factors affecting the severity of automated vehicle crashes using XGBoost model combining POI data, J. Adv. Transp. (2020), https://doi.org/10.1155/2020/8881545.

[10] J.B. Cicchino, Effects of blind spot monitoring systems on police-reported lane-change crashes, Traffic Inj. Prev. 19 (6) (2018) 615–622, https://doi.org/10.1080/15389588.2018.1476973.
[11] T.S. Combs, L.S. Sandt, M.P. Clamann, N.C. McDonald, Automated vehicles and pedestrian safety: exploring the promise and limits of Pedestrian detection, Am. J. Prev. Med. 56 (1) (2019) 1–7, https://doi.org/10.1016/j.amepre.2018.06.024.
[12] S. Das, A. Dutta, I. Tsapakis, Automated vehicle collisions in California: applying Bayesian latent class model, IATSS Res. (2020), https://doi.org/10.1016/j.iatssr.2020.03.001.
[13] P. Davidson, A. Spinoulas, Autonomous vehicles: what could this mean for the future of transport, in: Australian Institute of Traffic Planning and Management (AITPM) National Conference, Brisbane, Queensland, 2015.
[14] J. De Freitas, X. Zhou, M. Atzei, S. Boardman, L. Di Lillo, Harvard Business School Research Paper Series Working Paper, 2023, pp. 23–036.
[15] S. Deb, L. Strawderman, D.W. Carruth, J. DuBien, B. Smith, T.M. Garrison, Development and validation of a questionnaire to assess pedestrian receptivity toward fully autonomous vehicles, Transp. Res. Part C Emerg. Technol. 84 (2017) 178–195, https://doi.org/10.1016/j.trc.2017.08.029.
[16] V.V. Dixit, S. Chand, D.J. Nair, Autonomous vehicles : disengagements, Accid. React. Times (2016) 1–14, https://doi.org/10.1371/journal.pone.0168054.
[17] V.R. Duddu, P. Penmetsa, S.S. Pulugurtha, Modeling and comparing injury severity of at-fault and not at-fault drivers in crashes, Accid. Anal. Prev. 120 (2018) 55–63, https://doi.org/10.1016/J.AAP.2018.07.036.
[18] F. Favarò, S. Eurich, N. Nader, Autonomous vehicles' disengagements: trends, triggers, and regulatory limitations, Accid. Anal. Prev. 110 (October 2017) (2018) 136–148, https://doi.org/10.1016/j.aap.2017.11.001.
[19] F.M. Favarò, N. Nader, S.O. Eurich, M. Tripp, N. Varadaraju, Examining accident reports involving autonomous vehicles in California, PLoS One 12 (9) (2017), https://doi.org/10.1371/journal.pone.0184952.
[20] N. Formosa, M. Quddus, C.K. Man, M.K. Singh, C. Morton, C.B. Masera, Evaluating the impact of lane marking quality on the operation of autonomous vehicles, J. Transp. Eng. Part A Syst. 150 (1) (2023) 04023126, https://doi.org/10.1061/JTEPBS.TEENG-7688.
[21] K. Goh, G. Currie, M. Sarvi, D. Logan, Factors affecting the probability of bus drivers being at-fault in bus-involved accidents, Accid. Anal. Prev. 66 (2014) 20–26, https://doi.org/10.1016/J.AAP.2013.12.022.
[22] N.J. Goodall, Comparison of automated vehicle struck-from-behind crash rates with national rates using naturalistic data, Accid. Anal. Prev. 154 (2021) 106056, https://doi.org/10.1016/j.aap.2021.106056.
[23] W. Jin, M. Islam, M. Chowdhury, Risk-based merging decisions for autonomous vehicles, J. Saf. Res. 83 (2022) 45–56, https://doi.org/10.1016/J.JSR.2022.08.003.
[24] M.J. Mclaughlin (2018). Analyzing autonomous car crashes in California. Masters Thesis. Stamford University. https://stacks.stanford.edu/file/druid:jb200qw0760/Thesis.pdf. Accessed April 2023.
[25] D. Kerr (2023). Protesters stop Waymo and cruise self-driving cars with only a traffic cone: NPR. https://www.npr.org/2023/08/26/1195695051/driverless-cars-san-francisco-waymo-cruise. Acessed September 2023.
[26] K. Kim, L. Li, J. Richardson, L. Nitz, Drivers at fault: influences of age, sex, and vehicle type, J. Saf. Res. 29 (3) (1998) 171–179, https://doi.org/10.1016/S0022-4375(98)00013-9.
[27] B. Kutela, The role of crosswalk-related features on drivers' spatial yielding compliance at signalized midblock crosswalks, J. Traffic Transp. Eng. (2022), https://doi.org/10.1016/J.JTTE.2021.11.001 (English Edition).
[28] B. Kutela, R.E. Avelar, P. Bansal, Modeling automated vehicle crashes with a focus on vehicle at-fault, collision type, and injury outcome, J. Transp. Eng. Part A Syst. 148 (6) (2022), https://doi.org/10.1061/JTEPBS.0000680.
[29] B. Kutela, R.E. Avelar, S.R. Geedipally, A. Jhamb, Prediction of occurrence and severity of run-off-roadway crashes on rural two-lane roadways using bayesian networks, Transp. Res. Rec. J. Transp. Res. 2676 (3) (2021), https://doi.org/10.1177/03611981211051351.
[30] B. Kutela, S. Das, B. Dadashova, Mining patterns of autonomous vehicle crashes involving vulnerable road users to understand the associated factors, Accid. Anal. Prev. (2021) 106473, https://doi.org/10.1016/J.AAP.2021.106473.
[31] B. Kutela, N. Langa, S. Mwende, E. Kidando, A.E. Kitali, P. Bansal, A text mining approach to elicit public perception of bike-sharing systems, Travel Behav. Soc. 24 (2021) 113–123, https://doi.org/10.1016/j.tbs.2021.03.002.
[32] B. Kutela, H.H. Teng, Assessment of methodological alternatives for modeling the spatiotemporal crossing compliance of pedestrians at signalized midblock crosswalks, J. Transp. Eng. A: Syst. 146 (2) (2020) 04019062.
[33] A. Millard-Ball, Pedestrians, autonomous vehicles, and cities, J. Plan. Educ. Res. 38 (1) (2018) 6–12, https://doi.org/10.1177/0739456X16675674.
[34] M.M. Morando, Q. Tian, L.T. Truong, H.L. Vu, Studying the safety impact of autonomous vehicles using simulation-based surrogate safety measures, J. Adv. Transp. 2018 (2018), https://doi.org/10.1155/2018/6135183.
[35] S.M. Mousavi, O.A. Osman, D. Lord, K.K. Dixon, B. Dadashova, Investigating the safety and operational benefits of mixed traffic environments with different automated vehicle market penetration rates in the proximity of a driveway on an urban arterial, Accid. Anal. Prev. 152 (2021) 105982, https://doi.org/10.1016/J.AAP.2021.105982.
[36] D. Paranyushkin (2011). Identifying the pathways for meaning circulation using text network analysis. Nodus Labs. https://noduslabs.com/research/pathways-meaning-circulation-text-network-analysis/. Accessed June 2023.
[37] P. Penmetsa, E. Ko, D. Wood, T. Wang, S.L. Jones, perceptions and expectations of autonomous vehicles – A snapshot of vulnerable road user opinion, Technol. Forecast. Soc. Change 143 (February) (2019) 9–13, https://doi.org/10.1016/j.techfore.2019.02.010.
[38] E. Pöllänen, G.J.M. Read, B.R. Lane, J. Thompson, P.M. Salmon, Who is to blame for crashes involving autonomous vehicles? Exploring blame attribution across the road transport system, Ergonomics 63 (5) (2020) 525–537, https://doi.org/10.1080/00140139.2020.1744064.
[39] J.H. Salum, B. Kutela, A.E. Kitali, E. Kidando, Stated preference analysis of automated vehicles among California residents using probabilistic inferences, Transp. Res. Rec. J. Transp. Res. 2676 (2) (2021) 036119812110398, https://doi.org/10.1177/03611981211039845.
[40] B. Schoettle, M. Sivak. A preliminary analysis of real-world crashes involving self-driving vehicles, University of Michigan Transportation Research Institute, 2015.
[41] S.E. Shladover, Connected and automated vehicle systems: introduction and overview, J. Intell. Transp. Syst. 22 (3) (2018) 190–200, https://doi.org/10.1080/15472450.2017.1336053.
[42] A. Sinha, V. Vu, S. Chand, K. Wijayaratna, V. Dixit, A crash injury model involving autonomous vehicle: Investigating of crash and disengagement reports, Sustainability 13 (14) (2021) 7938.
[43] J. So, J. Hwangbo, S.H. Kim, I. Yun, Analysis on autonomous vehicle detection performance according to various road geometry settings, J. Intell. Transp. Syst. 27 (3) (2023) 384–395, https://doi.org/10.1080/15472450.2022.2042280.
[44] Y. Song, M.V. Chitturi, D.A. Noyce, Automated vehicle crash sequences: patterns and potential uses in safety testing, Accid. Anal. Prev. 153 (2021) 1–24, https://doi.org/10.1016/j.aap.2021.106017.
[45] L. Tang, Y. Shi, Q. He, A.W. Sadek, C. Qiao, Performance test of autonomous vehicle lidar sensors under different weather conditions, Transp. Res. Record 2674 (1) (2020) 319–329.
[46] H. Tapiro, A. Wyman, A. Borowsky, T. Petzoldt, X. Wang, D.S. Hurwitz, Automated vehicle failure: The first pedestrian fatality and public perception, Transp. Res. Record 2676 (8) (2022) 198–208.
[47] E.R. Teoh, D.G. Kidd, Rage against the machine? Google TMs self-driving cars versus human drivers, J. Saf. Res. 63 (2016) 57–60.
[48] UCLA: Statistical Consulting Group. (2014). Introduction to generalized linear mixed models. https://stats.oarc.ucla.edu/other/mult-pkg/introduction-to-generalized-linear-mixed-models/. Accessed May 2022.
[49] S.I. Vrieze, Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), Psychol. Methods 17 (2) (2012) 228–243, https://doi.org/10.1037/A0027127.
[50] S. Wang, Z. Li, Exploring the mechanism of crashes with automated vehicles using statistical modeling approaches, PLoS One 14 (3) (2019), https://doi.org/10.1371/journal.pone.0214550.
[51] J.M. Woodridge, Introductory Economics a Modern Approach, 5th ed., South-Western, Cengage Learning, 2012.
[52] B. Yoon, Y. Park, A text-mining-based patent network: analytical tool for high-technology trend, J. High Technol. Manag. Res. 15 (1) (2004) 37–50, https://doi.org/10.1016/j.hitech.2003.09.003.
[53] S. Zang, M. Ding, D. Smith, P. Tyler, T. Rakotoarivelo, M.A. Kaafar, The impact of adverse weather conditions on autonomous vehicles: how rain, snow, fog, and hail affect the performance of a self-driving car, IEEE Veh. Technol. Mag. 14 (2) (2019) 103–111, https://doi.org/10.1109/MVT.2019.2892497.

Boniphace Kutela, Ph.D., P.E., is an Assistant Research Scientist at Texas A&M Transportation Institute. He has over ten years of experience in transportation safety and operations-related research projects. His major areas of expertise include roadway data collection and analysis methodologies, machine learning, and predictive modeling with a focus on traffic safety. He has authored/co-authored over thirty publications in reputable peer-reviewed journals and presented over twenty presentations at national and local conferences.

Jimoku Hinda Salum is a research associate and Ph.D. student at Florida International University. He obtained his master's degree in transportation engineering at Florida International University and his bachelor's degree in Civil Engineering from the University of Dar es Salaam. His research interests fall under intelligent transportation systems, traffic operations and safety, and connected and autonomous vehicles. He has authored/co-authored over ten publications.

Seif Rashidi Seif is a civile engineer with SMEC where he supervises the construction of Dar Es Salaam Bus Rapid Transit System (BRT) infrastructure Phase 3 - Lot 1. Prior to joining SMEC, he obtained his bachelor's degree in Civil Engineering from the University of Dar es Salaam. His research interests are in Transportation Engineering, Public Transportation, Traffic Operations, among others.

Subasish Das Ph.D: Dr. Subasish Das is an assistant professor at Texas State University (TSU). He has more than ten years of national and international experience associated with transportation safety and operations. His major areas of expertise include database management, statistical analysis, and machine learning with an emphasis on safety and transportation operations, machine vision, spatial analysis with modern web GIS tools, computer programming, interactive data visualization, and deep learning tools for autonomous cars. He is an active member of ITE, ASCE, and TRB Committee for Artificial Intelligence and Advanced Computing Applications (AED50).

Emmanuel Kidando, Ph.D., P.E., is an assistant professor in the Department of Civil & Environmental Engineering at Cleveland State University. His areas of expertise include intelligent transportation systems (ITS), connected vehicle systems, freeway operations, highway safety, incident management, and the application of data science in transportation engineering.