




# A matched case-control analysis of autonomous vs human-driven vehicle accidents

Received: 30 September 2023

Accepted: 2 May 2024

Published online: 18 June 2024

 Check for updatesMohamed Abdel-Aty<sup>1</sup> & Shengxuan Ding<sup>1</sup>  

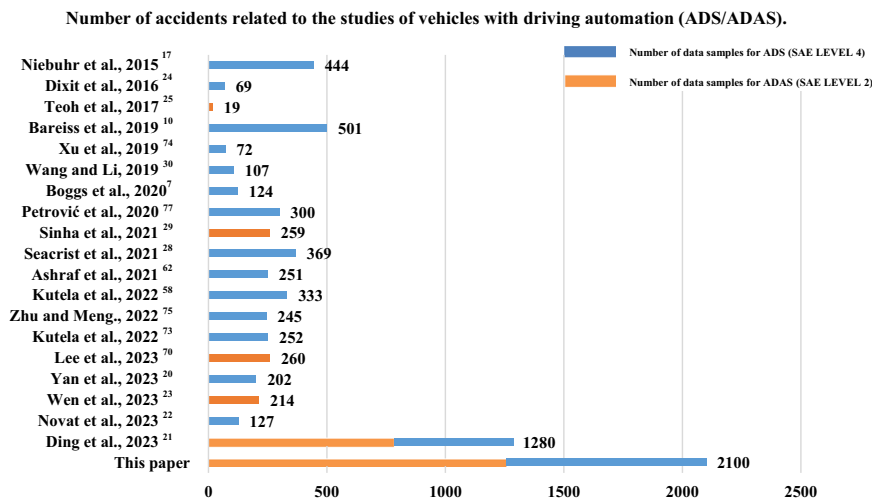
Despite the recent advancements that Autonomous Vehicles have shown in their potential to improve safety and operation, considering differences between Autonomous Vehicles and Human-Driven Vehicles in accidents remain unidentified due to the scarcity of real-world Autonomous Vehicles accident data. We investigated the difference in accident occurrence between Autonomous Vehicles' levels and Human-Driven Vehicles by utilizing 2100 Advanced Driving Systems and Advanced Driver Assistance Systems and 35,113 Human-Driven Vehicles accident data. A matched case-control design was conducted to investigate the differential characteristics involving Autonomous' versus Human-Driven Vehicles' accidents. The analysis suggests that accidents of vehicles equipped with Advanced Driving Systems generally have a lower chance of occurring than Human-Driven Vehicles in most of the similar accident scenarios. However, accidents involving Advanced Driving Systems occur more frequently than Human-Driven Vehicle accidents under dawn/dusk or turning conditions, which is 5.25 and 1.98 times higher, respectively. Our research reveals the accident risk disparities between Autonomous Vehicles and Human-Driven Vehicles, informing future development in Autonomous technology and safety enhancements.

Automation of systems has been experiencing rapid development and has brought about a revolution in the transportation industry. The introduction of Autonomous Vehicles (AV) technology has made the vision of a safe transportation system with effortless driving seem attainable. It is anticipated that the automation of systems will significantly reduce the number of accidents, as human errors contribute up to 90% of accidents<sup>1</sup>. While smart transportation has showcased several benefits, these emerging technologies have also exhibited drawbacks, particularly regarding safety risks. Accidents of on-road testing have already been documented in limited testing data<sup>2</sup>.

According to findings from the RAND corporation<sup>3</sup>, a mere advancement in safety features in the initial release of AVs could yield significant life-saving results. The research suggests that if AVs were to be introduced with an average safety level ten percent higher than that

of the typical human driver, approximately 600,000 fatalities could be averted in the United States over a span of 35 years. Nevertheless, it is crucial to acknowledge that between 2015 and 2022, there was a yearly rise in both the annual miles traveled by autonomous vehicles (AVMT) and the number of AV accidents on public roads in California, with the exception of a decline in 2020 attributable probably to the COVID-19 pandemic<sup>4</sup>. AV testing on California public roads is permitted by the California Department of Motor Vehicles (CADMV)<sup>5</sup>. Up until June 2023, we have identified and enriched with 598 Advanced Driving Systems (ADS (SAE Level 4)) accidents in California (AVOID dataset)<sup>6</sup>. These reports provide details about AV accidents and disengagements, which happen when the vehicle switches from autonomous mode due to technological issues or when the test driver or operator assumes manual control for safety purposes<sup>7</sup>. The detailed information included in these reports revealed various factors related to AV accidents.

<sup>1</sup>Smart and Safe Transportation Lab (SST), Department of Civil, Environmental and Construction Engineering, University of Central Florida, 12800 Pegasus Dr, Orlando, FL 32816, USA. ✉e-mail: [shengxuan.ding@ucf.edu](mailto:shengxuan.ding@ucf.edu)



**Fig. 1 | Number of accidents related to the studies of vehicles with driving automation (ADS/ADAS).** The blue line shows the number of ADS data samples, and the orange line shows the number of ADAS data samples used in related studies.

We further collected the corresponding reports of Human-Driven Vehicles (HDV) accidents to contrast the differences between AV and matched HDV accidents.

There are many potential benefits of AVs on traffic safety, such as a reduction in human error, reduced fatigue, and distraction. ADAS functions including Electronic Stability Control (ESC), Anti-Lock Braking (ABS), and Information and Communication Technology aids in ongoing driving tasks to prevent accidents<sup>8–10</sup>. For example, Tesla Autopilot consistently maintains a closer distance to the lane center than human drivers<sup>11</sup>. Additionally, Level 3 or higher levels of automation can further enhance traffic by reducing accidents<sup>12,13</sup>, improving mobility for the disabled and elderly, and minimizing traffic collisions through efficient driving and the reduction of human errors<sup>14</sup>. Koopman and Wagner highlighted the importance of understanding how AVs will interact with human drivers, pedestrians, and other road users<sup>15</sup>. Compared with HDV, accident avoidance features of AVs can reduce accidents and fatalities caused by distracted driving or human error by helping control the vehicle and alerting drivers to potential dangers<sup>16,17</sup>.

However, there are also some possible safety challenges that need to be addressed in AVs. Penmetsa et al.<sup>18</sup> identified several safety challenges that need to be addressed in the development of AV, such as the need for reliable sensing and perception, robust decision-making algorithms, and fail-safe mechanisms<sup>18</sup>. Kalra and Paddock presented a statistical methodology with 382 accidents per 100 million miles based on binomial and Poisson distributions for estimating the required mileage to establish the reliability of AVs, it shows that AVs would have to be driven hundreds of millions of miles and sometimes hundreds of billions of miles to demonstrate their reliability in terms of fatalities and injuries<sup>19</sup>. Some studies also have concentrated on analyzing the accident severity involving different types of AV<sup>20</sup>. Ding et al.<sup>6</sup> collected 1280 cases to compare factors related to injury severity between the ADAS and ADS accidents by random parameter multinomial logit models. Subject vehicle's contact area, road and environment, and pre-accident movement, significantly impact accidents injury severity. ADS-equipped vehicles in work zones have a higher probability of being involved in minor and moderate/severe injury accidents<sup>21</sup>.

The comparison of safety performance between AV and HDV is a topic of debate, with conflicting viewpoints<sup>22</sup>. On one side, numerous studies support the view that AVs are generally safer than HDVs<sup>23</sup>. For instance, Dixit et al.<sup>24</sup> analyzed the statistical distribution of reports on 69 manual disengagements and accidents, they compared the

accident rates of Google self-driving cars to those of human drivers and found that fatal accidents involving Google cars were lower than that of HDV, no fatalities occurred as compared to 1 death for every 108 million miles in California<sup>24</sup>. Additionally, from 2009 to 2015 in Mountain View, California, Google cars demonstrated a significantly lower rate of police-reportable accidents per million vehicle miles traveled (VMT) compared to human drivers (2.19 versus 6.06)<sup>25</sup>. On the other hand, some research challenges this view, suggesting that the safety of AVs may not always exceed that of HDVs. Schoettle and Sivak uncovered that AVs have a higher rate of accidents per million miles traveled compared to HDVs in limited (and generally less demanding) conditions (e.g., avoiding snowy areas)<sup>26</sup>. Favarò et al.<sup>27</sup> found that rear-end accidents were the most frequent type of AV accidents, with AVs being hit from behind by conventional vehicles at a rate twice that of rear-end accidents than rear-end “fender-benders” for conventional vehicles in California 2013<sup>27</sup>.

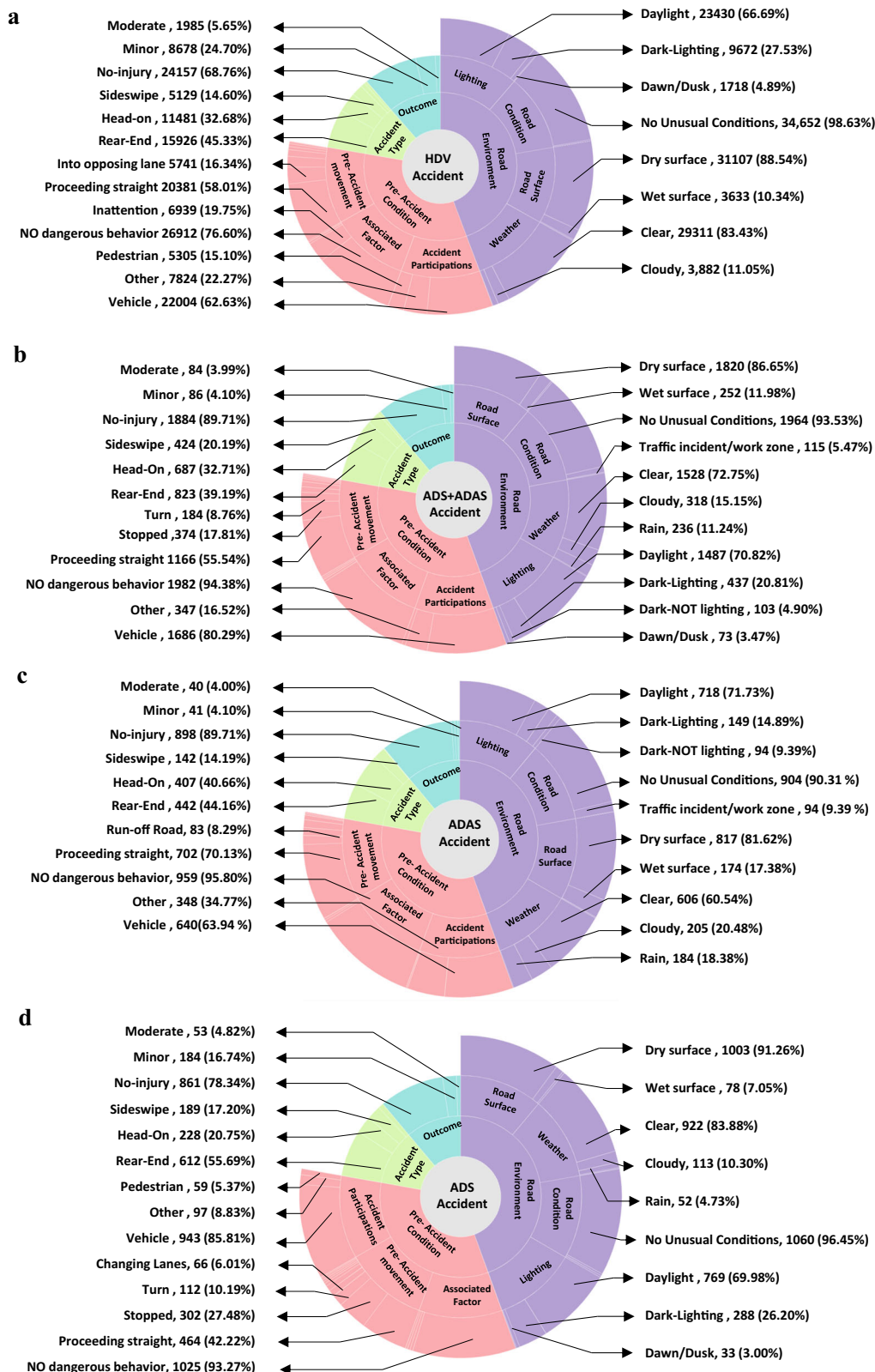
These studies offered significant insight into the factors that contribute to AV and HDV accidents, while earlier studies may not consider sufficient factors of road environment, accident outcome, pre-accident condition, and accident type for accident analysis due to a lack of matched AV and HDV data for identifying the characteristics of AV accidents and how they differ from regular HDV accidents<sup>28–30</sup>. The number of accidents of the relevant studies and their automation level are shown in Fig. 1.

In this work, we examined a dataset comprising both AV and HDV accidents. The dataset utilized in this study was compiled from various sources and encompassed information on accident types, road and environmental conditions, pre-accident vehicle movements, and accident outcomes. We use a matched case-control logistic regression model. We further add the National Highway Traffic Safety Administration (NHTSA) AV database comprising an additional 495 ADS and 1001 Advanced driver assistance systems (ADAS (SAE Level 2)) accidents<sup>31</sup>. These findings illuminate the factors that contribute to accidents involving AVs vs. HDVs.

## Results

### General trends in the full accident data

We first present general trends and comparisons between AV and HDV accidents of the full dataset. Figure 2a–d displays the distribution of factors affecting AV (2100), including SAE level 4 ADS (1099), level 2 ADAS (1001) and HDV accidents (35,133), respectively. Notably, vehicles make up 80% of participants in AV accidents, with pedestrians accounting for 3%. In contrast, for HDVs, pedestrians constitute 15%

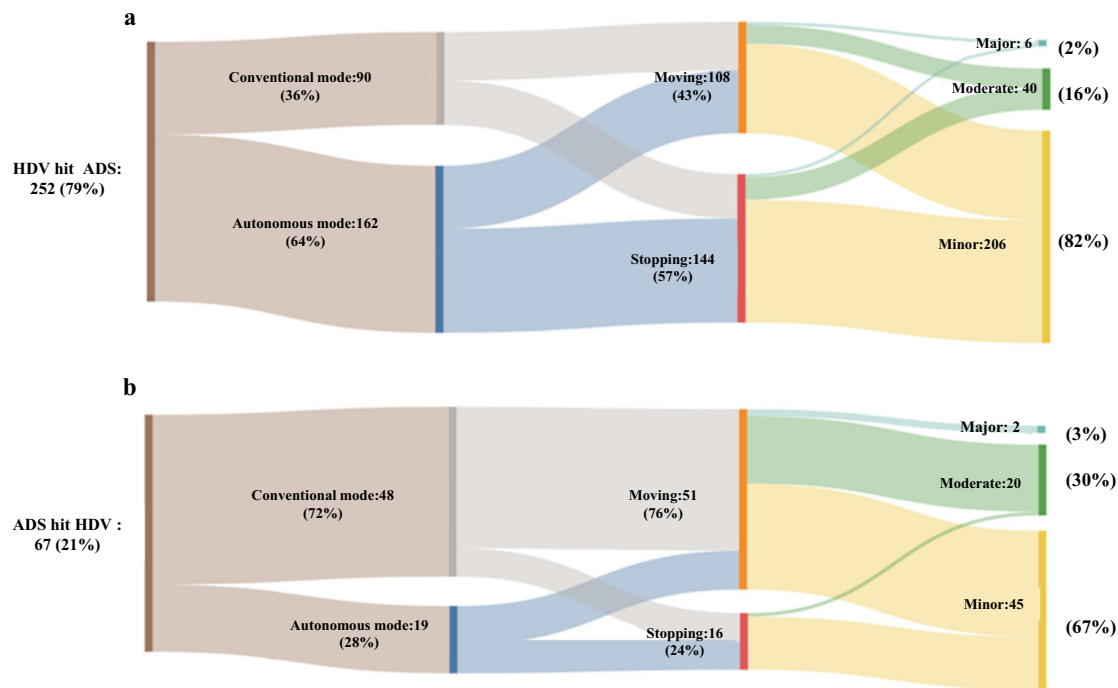


**Fig. 2 | Distribution of the factors influencing accidents of various vehicle types. a** HDV accidents with a sample of 35,133. **b** SAE level 4 ADS + SAE level 2 ADAS accidents with a sample of 2100. **c** SAE level 2 ADAS accidents with a sample of 1001. **d** SAE level 4 ADS accidents with a sample of 1099.

and vehicles 63% of accident participants, as depicted in Fig. 2a, b. When examining the outcomes of accidents, both AVs and HDVs lead to either no injuries or minor injuries occurring in 94%.

Significant disparities between AV and HDV accidents can be seen in work zones, traffic events, and pre-accident movements such as

slowing down, proceeding straight, and moving into opposing lanes, with AVs exhibiting higher accident rates. For both AVs and HDVs, the most frequent pre-accident movement is proceeding straight. It is observed that 56% of AV accidents and 58% of HDV accidents occur under this specific condition. About 5% of AV accidents take place in



**Fig. 3 | Rear-end accident conditions between ADS and HDV. a** Rear-end accidents that HDV hit an ADS from behind with a sample of 252. **b** Rear-end accidents that ADS hit an HDV from behind with a sample of 67.

locations impacted by previous traffic events or work zones, where normal traffic is disrupted by earlier incidents like disabled vehicles or spilled cargo. In comparison, just 1.3% of HDV accidents occur in similar settings. Analyzing the pre-accident scenario, a distinctive observation is that only 1.8% of AV accidents are attributed to inattention or poor driving behavior, in contrast to a much higher 19.8% for HDVs.

Evaluating environmental factors, the majority of accidents involving both AVs and HDVs tend to happen under clear weather conditions. Notably, accidents involving HDVs occur slightly more frequently under these conditions, at a rate of 83%, compared to 73% for AVs. However, AVs are more commonly involved in accidents during rainy conditions, accounting for 11% of such accidents, compared to HDVs which experience these conditions in only 5% of accidents. Dawn or dusk conditions experience 3.5% of AV accidents, which is lower than the 4.9% rate for HDVs.

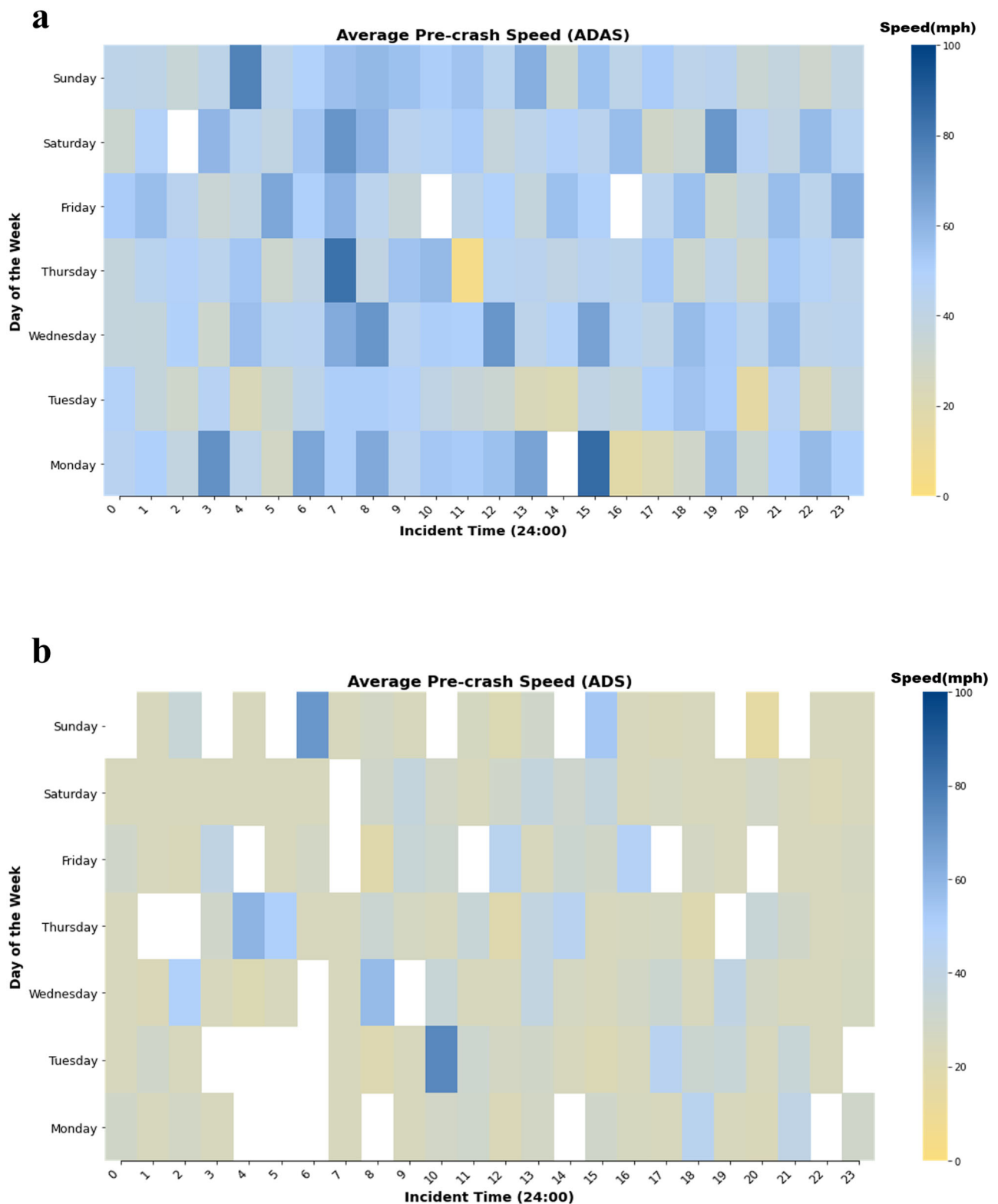
In terms of accident type, rear-end accidents constitute a majority for both AVs and HDVs. Furthermore, we determine the accident type associated with AV and HDV based on the accident angle. Our data indicates that for HDVs, the rear-end stands at 45% (other vehicles hit HDV), while head-on accidents (HDV hit other vehicles) occur at a rate of 33%. In contrast, AVs have a slightly lower rear-end accidents (other vehicles hit AV) rate of 39%, but a similar head-on accidents (AV hit other vehicles) rate of 33%. This suggests that while AVs have a marginally lower percentage of being rear-ended compared to HDVs, their percentage of head-on accidents is almost the same.

Figure 3 describes two conditions related to ADS rear-end accidents: a: HDV has hit an ADS from behind (252) and b: an ADS has hit an HDV from behind (67). The left side of the diagram starts with two conditions of HDV hit ADS or ADS hit HDV. The middle section shows the movement of vehicle: Moving or Stopping. The right side categorizes the severity of the accidents: Minor, Moderate and Major. The numbers indicated in each section of the diagram correspond to the total count of each specific category. The width of each link connecting the sections of the diagrams represents the proportion of scenarios that fall into the subsequent category. For example, In Fig. 3a, which details accidents where a HDV hit an ADS, we see that for the accidents where the ADS was moving (middle section), 62 cases (57%) involved

ADS in autonomous mode while 46 cases (43%) involved ADS in conventional mode (left section). Among the 108 accidents categorized under the 'moving condition', there were 2 cases (2%) that resulted in major injuries, 18 cases (17%) that led to moderate injuries, and the remaining 88 cases (81%) involved minor injuries.

The analysis reveals that 79% of rear-end accidents involve HDV hitting ADS, while 21% of rear-end accidents involve ADS hitting HDV. When the ADS hit HDV in conventional mode, most of the ADS are moving. We may conclude that compared with the autonomous mode, human drivers may not react as quickly or may not notice the object in time to take appropriate action. In terms of accident severity, 206 of accidents (82%) occur as minor injury when HDV hit ADS. This percentage is 67% when the ADS hit HDV. It is important to note that a majority of moderate and major accidents involving an ADS hitting an HDV occur when both vehicles are moving in the conventional mode. Notably, in cases where HDVs hit ADS, 64% of ADS are operating in autonomous mode. Conversely, when ADS are responsible for hitting HDVs, 72% of ADS are operating in the conventional mode. According to Dixit et al. in 2016, 56.1% of disengagements were attributed to system failures, 26.57% were initiated by the driver, and 9.89% were related to road infrastructure issues<sup>24</sup>. This observation suggests that conventional mode occurred more frequently than autonomous mode where ADS hit the HDV. This may be attributed to the advance autonomous mode of ADS. Autonomous mode uses advanced algorithms to detect and avoid obstacles and other vehicles in the path of the vehicle<sup>32</sup>.

We also contrast the ADS vs the ADAS, accidents related to ADAS and ADS display differences across various conditions, as shown in Fig. 2c, d. Regarding weather and road conditions, ADAS has 23.34% fewer accident number in clear skies but a 13.65% higher in rain compared to ADS. For road conditions, ADAS accidents experience a 7.48% higher accident number in traffic events or work zones and 10.33% higher accident number on wet roads than ADS. Analyzing pre-accident movements, ADAS indicate a 27.91% higher accident number for proceeding straight, while reporting 3.0% fewer turning accident numbers than ADS. In terms of accident types, ADAS is 3.0% higher accident number than ADS in broadside accidents and lags by 5.4% in sideswipe accidents. From an injury outcome perspective, ADAS



**Fig. 4 | Distribution of the pre-accidents Speed. a** ADAS Average Pre-accident Speed Heatmap. **b** ADS Average Pre-accident Speed Heatmap.

accidents present a 11.37% higher accident number of no-injury but a 2.1% lower number in fatal injuries against ADS. To enhance the understanding of pre-accident speeds, heatmaps that visually represent these speed patterns for ADS and ADAS vehicles are shown in Fig. 4. This chart offers a detailed comparison of how speeds vary across different days of the week and at various times of the day. This trend can be attributed to the fact that ADAS is primarily designed for

highway use, leading to a higher average pre-accident speed when compared to ADS vehicles, which are designed for a complex urban driving scenario.

We have also analyzed the full data to identify the influence of roadway elements and factors related to time using a random parameter logit model<sup>33</sup>. Only a single random parameter “Day of the week” demonstrated a significant effect. Upon analyzing the model, we found



that the dawn/dusk and turn conditions exhibit positive coefficients that are statistically significant at a 95% confidence level. This indicates a higher odd of an AV accident occurrence when these variables are significant in the random parameter logit model. Furthermore, we discovered that several variables demonstrate high significance and exhibit negative coefficients, suggesting a reduced likelihood of an accident when these factors are significant. These variables include the rain conditions, rear-end conditions, broadside conditions (a broadside condition is a car accident that occurs when the front of one vehicle slams into the side of another vehicle), moderate severity, proceeding straight, run-off road, backing, and entering traffic lane.

### Findings of road, environment, and accident type

Based on the results of the matched case control logistic regression model, compared with HDV, the odds of an ADS accident occurring in rainy weather are 0.335 times. This indicates a lower likelihood of an ADS accident in rainy weather compared to an HDV accident. RADAR is capable of detecting objects at distances exceeding 150 m, even in adverse weather conditions such as fog or rain<sup>34</sup>. In contrast, human drivers may only be able to perceive objects up to approximately 10 meters away under similar circumstances<sup>35</sup>. Although adverse weather can increase the likelihood of potential failures or loss of sensors<sup>36–38</sup>, recent innovations in visual algorithms, coupled with the combined use of cameras, LIDAR, GNSS, and RADAR sensors<sup>39,40</sup>, are crafted to recognize pedestrians and vehicles under varying weather scenarios, such as cloudiness, snow, rain, and darkness<sup>41,42</sup>. This offers solutions to the challenges associated with driving in less-than-ideal conditions. In contrast, human drivers may have difficulties seeing through heavy rain or fog, leading to a delay in detecting potential hazards or reacting appropriately<sup>40</sup>.

Interestingly, the dawn/dusk odds ratio indicates a 5.250 higher probability of ADS accident than HDV accident. This could be attributed to the sensors and cameras used by AVs may not be able to quickly adapt to changes in lighting conditions, which could affect their ability to detect obstacles, pedestrians, and other vehicles<sup>39,43</sup>. At dawn and dusk, for instance, the sun's shadows and reflections may confuse sensors, making it hard for them to distinguish between objects and identify potential hazards. Furthermore, the fluctuating light conditions can impact the accuracy of object detection and recognition algorithms used by AVs, which can result in false positives or negatives<sup>35,44</sup>.

Accident types related findings for ADS and HDVs are worth noting. Compared to HDV accidents, AVs experience relatively lower risks in rear-end and broadside accidents (0.457 times and 0.171 times, respectively). This finding indicates that AVs may detect and react to potential rear-end and sideswipe accident situations much faster than humans can. This is because they are equipped with advanced sensors and software that can quickly analyze the surrounding environment and make decisions based on the data received<sup>45</sup>. In addition, the kinematic method used by ACC system keeps track of and regulates the distance between vehicles, alerting drivers if this space becomes smaller than the safe limit, especially at highways<sup>46</sup>. By ensuring that vehicles keep a consistent speed and distance between vehicles, thus effectively reduces the risk of rear-end accidents<sup>47</sup>. Compared with ADS, HDVs tend to display greater velocity differences at larger spacing ranges<sup>48</sup>, a factor that significantly contributes to a higher incidence of rear-end and sideswipe accidents<sup>49</sup>.

### Findings of pre-accident conditions and accident outcomes

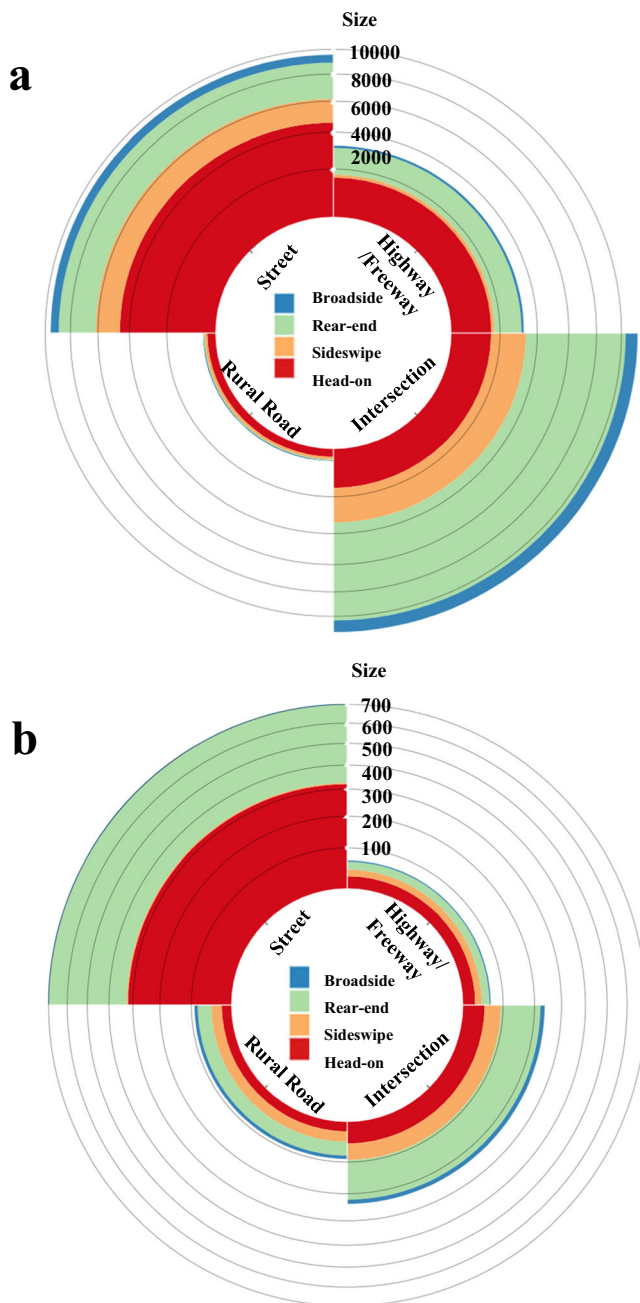
In terms of pre-accident conditions, most of the pre-accident movements made by ADS reduce the probability of accidents from the results of the matched case control logistic regression, except for turning, which increases the likelihood of an accident by 1.988 times compared to HDVs. One possible reason is a lack of situational awareness. Situational awareness of AVs can be defined as the ability of these vehicles to perceive essential elements in their surroundings,

understand the importance of these elements, and anticipate their future state or changes<sup>50</sup>. The complexity of turning in autonomous driving scenarios arises from three primary challenges: choosing the appropriate lane (target lane selection), devising and computing a safe and efficient path (trajectory planning and calculation), and executing the turn while adjusting to dynamic conditions (vehicle controlling and tracking)<sup>51</sup>. AVs rely on sensors and algorithms to perceive their surroundings and make driving decisions<sup>45</sup>. However, these systems may not detect all obstacles and hazards, particularly in complex and dynamic driving scenarios like turning at intersections<sup>52,53</sup>. It is a significant challenge to generate sufficient information and achieve comprehensive detection of the surrounding environment from a single independent source due to limited sensor ranges and limited coverage of the environment by sensors in AV<sup>45,54</sup>. Additionally, some AVs are programmed to follow predefined rules and scenarios, which may not encompass every possible driving situation<sup>55–57</sup>. The modifications of scenarios can present difficulty for AVs in perceiving and responding to them, thereby raising the risk of an accident<sup>58</sup>. Moreover, multiple oncoming HDVs and the complexity of such driving scenarios are a considerable challenge for AVs such as unprotected left turns at intersections<sup>59</sup>. These situations are complicated by factors like limited priority and variation in trajectories<sup>60</sup>. AVs tend to be overcautious (such as having a longer startup delay during the turning at intersections)<sup>61</sup>, which can lead to rear-end or sideswipe accidents with HDVs<sup>62</sup>. Furthermore, multi-interactions caused by mixed flows aggravate uncertainties in detection, such as the superposition of distance and angle measurement error<sup>63</sup>. Conversely, HDVs can adapt and modify their speed more seamlessly than AVs, highlighting the limitations of AVs in comparison to the adeptness of experienced drivers<sup>64</sup>. And AVs face difficulties with executing lane changes or turning in heavy traffic and lack psychological insight<sup>65,66</sup>. In addition, HDVs can predict pedestrian movements and exercise caution based on their driving experience, whereas AVs may struggle with recognizing pedestrians' intentions, potentially leading to emergency braking or accidents due to a lack of understanding of social cues and psychological reasoning<sup>65,67,68</sup>.

ADS accidents are less likely to occur than HDV accidents in situations such as proceeding straight, run-off road (a vehicle leaves the designated roadway and travels onto an area that is not intended for regular traffic) and entering traffic lane conditions (a vehicle transitioning from a stationary or parked position to enter a traffic lane and become physically present within the flow of traffic). When considering the proceeding straight condition, it was found that AV accident resulted in a 0.299 lower probability of an HDV accident. Remarkably, ADS accident risk is 0.021 times as high as that of an HDV accident in run-off road condition, which can be explained by the faster reaction time of AVs<sup>24</sup>. AVs can detect these situations and apply corrective actions, such as adjusting the speed or steering angle<sup>69,70</sup>, more quickly and accurately than a human driver<sup>71,72</sup>. The result of matched-case control model revealed a significant correlation between the entering traffic lane condition of ADS accident, the risk of which is 0.267 times as high as HDV accident. According to the results of the matched case-control logistic regression, the impact of backing is noteworthy, which shows that the ADS is less likely to be affected than the HDV. According to the analysis, the model using accidents of AVs showed a decreased probability of accidents for moderate and fatal severity in comparison to HDV.

### Discussion

A comprehensive examination was performed using a dataset of AV and HDV accidents in this study. A total of 2100 AV accidents and 35,133 HDV accident records were collected, which accurately reflected the accident details. The analysis first dealt the whole available data that includes both ADS and ADAS (SAE levels 4 and 2, respectively) using general descriptive statistics, percentages, and a random



**Fig. 5 | Distribution of the accident's types and scenarios. a** HDV accidents with a sample of 35,113. **b** ADS accidents with a sample of 1099. The distributions of accident types (head-on, sideswipe, rear-end, broadside) for vehicle categories (HDV and ADS) illustrate the frequency and proportion of each accident type in the respective locations by vehicle categories.

parameter logit model (not shown in the paper for brevity and since results are almost consistent with the other matched model). The analysis considered four categories of variables, including accident type, road and environment, pre-accident movement, and accident outcomes.

The accident data of AV and HDV were compared using the matched case-control logistic regression. The impact of different variables on the potential of accidents in AV vs HDV was conducted using a matched case-control logistic regression model. Based on the model estimation results, it can be concluded that ADS in general are safer than HDVs in most accident scenarios for their object detection and avoidance, precision control, and better decision-making.

However, the odds ratio of an ADS accident happening under dawn/dusk or turning conditions is 5.250 and 1.988 times higher, respectively, than the probability of an HDV accident occurring under the same conditions. The possible reasons might be a lack of situational awareness in complex driving scenarios and limited driving experience of AVs<sup>21</sup>. Improving the safety of ADS under dawn/dusk or turning conditions necessitates a holistic approach that involves advanced sensors, robust algorithms, and smart design considerations. Key strategies include enhancing weather and lighting sensors, implementing redundancy measures, and integrating sensor data effectively. By focusing on these aspects, the safety of ADS can be significantly enhanced in challenging scenarios.

Compared with current studies that only focus on AV accidents<sup>73–75</sup> or analyze AV and HDV with limited samples<sup>4,76,77</sup>, this paper has analyzed the factors that contribute to AV in comparison to HDV accidents through the analysis of real-world accidents and multi-source data. Furthermore, this research encompassed both AV and HDV accidents, instead of solely concentrating on different levels of AV accidents without considering a comparison with HDV accidents. Moreover, this study addresses the issue of unbalanced data between AV and HDV accidents by employing a matched case-control study design. One of the constraints of this study is analyzing the detailed levels of AV and the specific activated ADAS or AV system in an accident. Understanding and modeling different classifications of AV versus HDV accidents can be challenging and may require more data. It would also be crucial in the future to incorporate data about right-of-way at intersections, encompassing yield signs, stop signs, priority signals, and traffic lights, to enhance the comparative analysis between AV and HDV. Future research could also benefit from consulting a group of AV experts to identify and report on the factors contributing to safety differences between HDVs and AVs. Reporting their responses could provide qualitative depth to the research findings.

## Methods

### Data preparation

The full AV data set includes 2100 (ADS + ADAS) accidents based on AVOID<sup>6</sup> (CADMV and NHTSA's AV accident databases<sup>31</sup>). Supplementary Table 3 presents the general descriptive statistics of the databases, and the description and explanation for the variables are given in Supplementary Table 4. It provides a summary of the characteristics of the available variables that are classified into four major categories in the final data. The categories of variables include road and environmental characteristics (such as weather, road condition, road surface, and lighting conditions), pre-accident conditions (including vehicle manufacturers, AV driving modes, and pre-accident vehicle movement status), accident type, and accident outcome (accident severity), as shown in Supplementary Table 3. Among this information, the day of week, time of day and road location are typical confounders relevant to the traffic accident risk<sup>78</sup>. To be specific, the risk of traffic accidents can vary by the day of the week and time of day due to differences in traffic volume and driver behavior (e.g., commuters mostly in peak periods). In addition, the road location can affect the risk of traffic accidents by influencing traffic volume, speed limits, and the presence of other risk factors such as road design factors, intersections, pedestrians, and bicyclists.

The second group of accidents data comprised information of HDV, which was gathered from the Statewide Integrated Traffic Records System (SWITRS)<sup>79</sup>. This format and structure of data can be matched with the AV data, we collected 35,113 cases of HDV accidents according to the year of AV accidents as the first step. The distributions of accident types (head-on, sideswipe, rear-end, broadside) for vehicle categories (HDV and ADS) are shown in Fig. 5, which visually illustrates the frequency and proportion of each accident type in the respective locations by vehicle categories. For HDV accidents, intersections are the primary locations (significantly higher than other HDV accident

location types with  $F=5.1043$  and  $p=0.0166$ ), where 61.5% of HDV intersection accidents are rear-end, making it the most common type. Urban streets are the second most common scenario, with head-on accounting for 48.0% of HDV street accidents. Conversely, ADS accidents occur more frequently on urban streets (significantly higher than other ADS accident location types with  $F=10.4982$  and  $p=0.0011$ ), where 45.6% of ADS street accidents were head-on. Accidents at intersections are the second most common for ADS, with rear-end making up 53.8% of these ADS intersection accidents.

### A matched case-control logistic regression model

A matched case-control study is an observational study that involves comparing individuals who have a specific health outcome or disease (the cases) with individuals who do not have the health outcome or disease (the controls)<sup>80</sup>. The study design involves selecting cases and controls based on their exposure to a particular risk factor or characteristic, and then comparing the frequency of that exposure between the two groups. In the context of this paper, a matched case-control study could be used to investigate the relationship between accident-related risk factors<sup>81,82</sup>. Cases would be AVs involved in accidents, and controls would be HDVs involved in accidents.

A matched control study has been designed to investigate the impact of various factors on the likelihood of accidents in two specific scenarios: AV and HDV.

Conditional logistic regression is a variant of logistic regression that specifically tackles the issue of stratification within matched case-control studies<sup>83</sup>. In this research, there are  $N$  strata denoted by  $i=1,2,\dots,N$ . Each stratum has one AV accident case sample and  $k$  HDV accident control samples denoted by  $j=1,2,\dots,k$ . The conditional likelihood for the  $i$ th strata depends on the probability of the total number of cases (AV accident case samples) and controls ( $k$  HDV accident control samples) recorded in the stratum<sup>84</sup>.  $P_{ri}(x_{ji})$  refers to the probability of the  $j$ th samples in the  $i$ th stratum is a AV accident where  $x_{ji}=(x_{1ji}, x_{2ji}, \dots, x_{pji})$  is determined by a vector of variables ( $x_1, x_2, \dots, x_p$ ). A logistic regression model with linear parameters is employed to estimate the likelihood of an accident, as described by Abdel-Aty et al.<sup>85</sup>:

$$\text{logit}[P_{ri}(x_{ji})] = (a_i, b_1x_{1ji}, \dots, b_kx_{pji}) \quad (1)$$

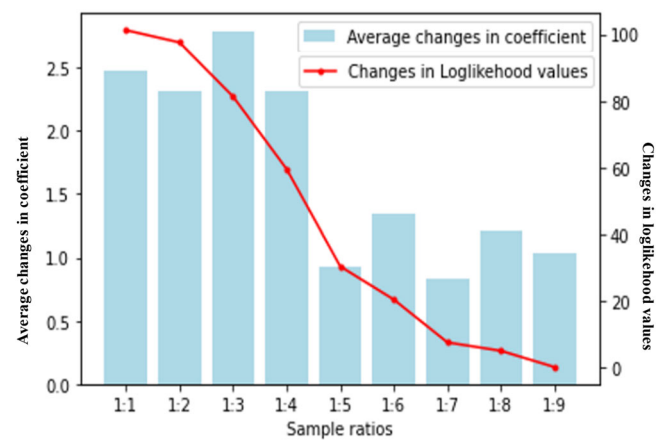
The controlled variables used to create strata are reflected in the intercept term. To incorporate the impact of stratification in the analysis, it is possible to construct a conditional log-likelihood. This log-likelihood function comprises multiple terms, each representing the conditional probability of an accident occurring within a specific stratum<sup>86</sup>. The following equation presents the formula for the conditional likelihood function, as stated by Abdel-Aty et al.<sup>85</sup>:

$$\mathcal{L}(\beta) = \prod_{i=1}^N \left[ \left( 1 + \sum_{j=1}^k \exp \left\{ \sum_{u=1}^p \beta_u (x_{uji} - x_{u0i}) \right\} \right) \right]^{-1} \quad (2)$$

The coefficients' estimates in Eq. (1) are identical to the maximum likelihood function values in Eq. (2). These estimates are log-odds ratios that can provide an approximation of the relative risk of an accident and are also referred to as hazard ratios (i.e., the ratio of odds for accident occurrence versus non-occurrence). The hazard ratio is calculated by raising the exponential value to the coefficient's power. For a dummy variable, the odds ratio is a statistic defined as the ratio of the odds of the case. The odds ratio can be written as

$$OR(x_k) = \frac{\Pr(y_{i0}=1, x_k=1, Z)/[1 - \Pr(y_{i0}=1, x_k=1, Z)]}{\Pr(y_{i0}=1, x_k=0, Z)/[1 - \Pr(y_{i0}=1, x_k=0, Z)]} = \exp(\beta_k) \quad (3)$$

where,  $Z$  represents the vector of explanatory variables excluding  $x_k$ .  $\beta_k$  is estimated coefficient for  $x_k$ .



**Fig. 6 | Selection for the optimal number of controls.** The blue column line indicates the average coefficient changes (left y-axis), and the red line shows the changes in loglikelihood values (right y-axis).

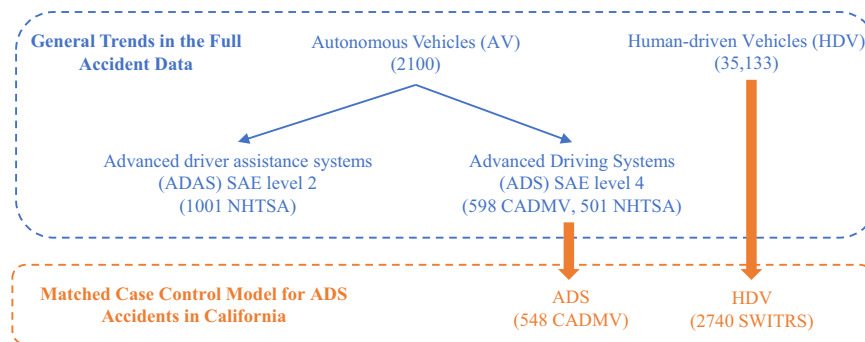
### Matched case control study for ADS accidents

Our aim is to explore the differential characteristics of accidents involving AVs and HDVs, rather than comparing accidents and non-accidents. Direct comparisons between AV and HDV accidents are still not viable as the difference in exposure and number of vehicles of both types is extremely unbalanced. We incorporated Annual Average Daily Traffic (AADT) data of various road types from the California Traffic Census Program<sup>87</sup>, which is provided in the supplementary methods. HDVs show a higher incidence of accidents on highways, intersections, and streets, particularly on highways. For rural roads, HDVs and AVs exhibit almost similar accident rates. Across all road types, HDVs consistently record significantly higher accident figures than AVs. To examine the impact of exogenous variables on accident risk for different vehicle types, we conducted a matched case-control logistic regression model for AV (ADS) and HDV accidents. The coordinates of accidents are extracted by Google Map API, and then the type of road is identified to conduct matched case-control logistic regression. The distribution of AV accidents over various situations differs from the distribution of HDV accidents is concluded from the matched case-control study.

To overcome this challenge of variables that confound the relationship between risk factors and traffic accident outcomes, the first principle is to match cases and controls at the same location. In the case of a location that does not have enough controls, similar locations within a radius of 5 miles for intersections and urban segments were used, and the day of the week and time of day were controlled. We assumed that cases and controls were under similar traffic patterns based on the controlled time and space. Aside from intersections and streets, the location of each stratum for AV and HDV accidents is on the same highways and expressways. In addition, the same road type for each stratum is controlled to ensure similar geometric design. As the manual override and conventional modes of ADAS closely resemble HDV, we only focus on ADS cases from CA for the matched case-control study. Furthermore, some cases were removed due to difficulties in obtaining or imputing precise road types for matched case-control logistic regression.

We organized the data into  $N$  strata according to the occurrence of AV accidents. Each stratum consisted of one case and  $k$  corresponding controls. To ensure consistency across strata, we employed a matched case-control logistic regression by adjusting the number of control samples and assessed the resulting estimates for each model. Samples generally refer to the groups of accidents selected for comparison within each stratum. Case samples are specific accidents who have the outcome that is the focus of the





**Fig. 7 | Sample size of data.** The blue fonts indicate general accident trends, while the orange fonts represent data for the matched case control model of ADS accidents in California. The HDV data is sourced from SWITRS<sup>79</sup>. The ADS (SAE

Level 4) data is sourced from CADMV<sup>5</sup> and NHTSA<sup>31</sup>, while the ADAS (SAE Level 2) data is sourced from NHTSA<sup>31</sup>.

**Table 1 | Matched case-control logistic regression model**

Variable	Estimated parameter	Odds Ratio	t-statistic	P value	95% Confidence Interval of OR
<i>Road and environment</i>					
Rain condition (1 if weather is rain, 0 otherwise)	-1.093	0.335	-3.948	<0.001	[0.195,0.577]
Dawn/dusk condition (1 if it is dawn/dusk, 0 otherwise)	1.658	5.25	4.508	<0.001	[2.552,10.788]
<i>Accident type</i>					
Rear-end (1 if accident type is rear-end, 0 otherwise)	-0.783	0.457	-4.186	<0.001	[0.317,0.659]
Broadside (1 if accident type is broadside, 0 otherwise)	-1.766	0.171	-5.221	<0.001	[0.088,0.332]
<i>Accident outcomes</i>					
Moderate (1 if injury severity outcome is moderate, 0 otherwise)	-0.455	0.634	-1.998	0.046	[0.406,0.991]
Fatal (1 if injury severity outcome is fatal, 0 otherwise)	-2.335	0.097	-8.733	<0.001	[0.057,0.164]
<i>Pre-accident conditions</i>					
Proceeding straight (1 if pre-accident movement is proceeding straight, 0 otherwise)	-1.210	0.299	-4.544	<0.001	[0.178,0.504]
Run-off Road (1 if pre-accident movement is Run-off Road, 0 otherwise)	-3.88	0.021	-9.545	<0.001	[0.009,0.046]
Entering traffic lane (1 if pre-accident movement is entering traffic lane, 0 otherwise)	-1.320	0.267	-10.954	<0.001	[0.211,0.338]
Turn (1 if pre-accident movement is turn, 0 otherwise)	0.687	1.988	4.012	<0.001	[1.421,2.781]
Backing (1 if pre-accident movement is backing, 0 otherwise)	-0.712	0.491	-4.803	<0.001	[0.367,0.656]
Number of strata	548				
McFadden pseudo-R-squared	0.810				
Likelihood ratio test	628.9				
Wald test	344.2				
Score (log rank) test	590.7				

1 if accident is AV, 0 HDV; Sampling ratio is 1:5.  
Sample size: 548 ADS accidents and HDV strata.

study. Control samples are accidents who do not have the specific outcome being studied. In this study, 'AV accident case sample' consists of AV accidents within each stratum, and the 'HDV accident control samples' consist of the HDV accidents within the same stratum. The method begins by utilizing an initial equal proportion of AV accident case samples to HDV accident control samples (1:1) and progressively increasing the ratio (1:3, 1:5, 1:7, 1:9...) until the coefficients between consecutive models exhibit no significant change. From Fig. 6, it is evident that there are no notable disparities between the models employing sample ratios of 1:5 and 1:6. As a result, we opt for the 1:5 ratio for our analysis. To further support our decision, we evaluate the improvement in log-likelihood across the models, aligning with our hypothesis of selecting an AV accident case sample to HDV-accident control sample ratio of 1:5.

As a result, 548 ADS accident accidents were applied for the matched case-control design and are discussed in this paper. The

sample of the data is shown in Fig. 7. The estimation results and 95% confidence intervals of the odds ratio are presented in Table 1, which was generated using the survival package in R programming<sup>88</sup>. A total of 11 significant variables were identified by combining road and environment, accident type, accident outcomes, and pre-accident conditions during the estimation process.

### Data availability

The Human Driven-Vehicle (HDV) accidents dataset that we used to is publicly available at <https://www.chp.ca.gov/programs-services/services-information/switrs-internet-statewide-integrated-traffic-records-system>. The Autonomous Vehicle (AV) accidents dataset is available at <https://github.com/UCF-SST-Lab/AVOID-Autonomous-Vehicle-Operation-Incident-Dataset/tree/main/asset>. The Annual Average Daily Traffic (AADT) data of various road types from the California Traffic Census Program is available at <https://dot.ca.gov/>

[programs/traffic-operations/census](#). Source data for figures are provided with this paper. All other data used in this study are available from the corresponding authors upon request. Source data are provided with this paper.

## Code availability

The codes for data validation and processing are available on Zenodo with a (<https://doi.org/10.5281/zenodo.11081206>). The quick tutorial and README file are also included in the repository for reference. Python scripts for geospatial data processing are prepared with the OSMnx package and offered in the repository, which can be referred to in the file Address2OSM.ipynb under the folder of code. All other code used in this study are available from the corresponding authors upon request.

## References

1. Fleetwood, J. Public health, ethics, and autonomous vehicles. *Am. J. Public Health* **107**, 532–537 (2017).
2. Lee, D. & Hess, D. J. Regulations for on-road testing of connected and automated vehicles: Assessing the potential for global safety harmonization. *Transp. Res. Part A* **136**, 85–98 (2020).
3. Zhang, L. *Cruise's Safety Record Over 1 Million Driverless Miles*, <https://getcruise.com/news/blog/2023/cruises-safety-record-over-one-million-driverless-miles> (2023).
4. Liu, Q., Wang, X., Wu, X., Glaser, Y. & He, L. Crash comparison of autonomous and conventional vehicles using pre-crash scenario typology. *Accid. Anal. Prev.* **159**, 106281 (2021).
5. DMV, C. Autonomous vehicle collision reports, <https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/autonomous-vehicle-collision-reports/> (2023).
6. Zheng, O. et al. AVOID: Autonomous Vehicle Operation Incident Dataset Across the Globe. Preprint at <https://arxiv.org/abs/2303.12889> (2023).
7. Boggs, A. M., Wali, B. & Khattak, A. J. Exploratory analysis of automated vehicle crashes in California: A text analytics & hierarchical Bayesian heterogeneity-based approach. *Accid. Anal. Prev.* **135**, 105354 (2020).
8. Scanlon, J. M. et al. Waymo simulated driving behavior in reconstructed fatal crashes within an autonomous vehicle operating domain. *Accident. Anal. Prev.* **163**, 106454 (2021).
9. Ahangarnejad, A. H., Radmehr, A. & Ahmadian, M. A review of vehicle active safety control methods: From antilock brakes to semiautonomy. *J. Vib. Control* **27**, 1683–1712 (2021).
10. Bareiss, M., Scanlon, J., Sherony, R. & Gabler, H. C. Crash and injury prevention estimates for intersection driver assistance systems in left turn across path/opposite direction crashes in the United States. *Traffic Inj. Prev.* **20**, S133–S138 (2019).
11. Gordon, T. J. & Lidberg, M. Automated driving and autonomous functions on road vehicles. *Vehicle System Dynamics* **53**, 958–994 (2015).
12. Milakis, D., Van Arem, B. & Van Wee, B. Policy and society related implications of automated driving: A review of literature and directions for future research. *J. Intell. Transp. Syst.* **21**, 324–348 (2017).
13. Yue, L., Abdel-Aty, M., Wu, Y. & Wang, L. Assessment of the safety benefits of vehicles' advanced driver assistance, connectivity and low level automation systems. *Accident. Anal. Prev.* **117**, 55–64 (2018).
14. Chan, C.-Y. Advancements, prospects, and impacts of automated driving systems. *Int. J. Transp. Sci. Technol.* **6**, 208–216 (2017).
15. Koopman, P. & Wagner, M. Autonomous vehicle safety: An interdisciplinary challenge. *IEEE Intell. Transp. Syst. Mag.* **9**, 90–96 (2017).
16. Harper, C. D., Hendrickson, C. T. & Samaras, C. Cost and benefit estimates of partially-automated vehicle collision avoidance technologies. *Accid. Anal. Prev.* **95**, 104–115 (2016).
17. Niebuhr, T., Junge, M. & Achmus, S. Expanding pedestrian injury risk to the body region level: how to model passive safety systems in pedestrian injury risk functions. *Traffic Inj. Prev.* **16**, 519–531 (2015).
18. Penmetsa, P., Sheinidashtegol, P., Musaev, A., Adanu, E. K. & Huddall, M. Effects of the autonomous vehicle crashes on public perception of the technology. *IATSS Res.* **45**, 485–492 (2021).
19. Kalra, N. & Paddock, S. M. Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability? *Transp. Res. Part A* **94**, 182–193 (2016).
20. Yan, S., Huang, C. & He, D. A comparison of patterns and contributing factors of ADAS and ADS involved crashes. *J. Transp. Saf. Sec.* **15**, 1–28 (2023).
21. Ding, S. et al. Exploratory Analysis of the Crash Severity between Vehicular Automation (SAE L2-5) with Multi-Source Data. Preprint at <https://arxiv.org/abs/2303.17788> (2023).
22. Norris, N., Emmanuel, K., Boniphace, K. & Angela, E. K. A comparative study of collision types between automated and conventional vehicles using Bayesian probabilistic inferences. *J. Saf. Res.* **84**, 251–260 (2023).
23. Wen, X., Huang, C., Jian, S. & He, D. Analysis of discretionary lane-changing behaviours of autonomous vehicles based on real-world data. *Transportmetrica A: Transport Science* **19**, 1–24 (2023).
24. Dixit, V. V., Chand, S. & Nair, D. J. Autonomous vehicles: disengagements, accidents and reaction times. *PLoS One* **11**, e0168054 (2016).
25. Teoh, E. R. & Kidd, D. G. Rage against the machine? Google's self-driving cars versus human drivers. *J. Saf. Res.* **63**, 57–60 (2017).
26. Schoettle, B. & Sivak, M. *A preliminary analysis of real-world crashes involving self-driving vehicles* (University of Michigan Transportation Research Institute, 2015).
27. Favarò, F. M., Nader, N., Eurich, S. O., Tripp, M. & Varadaraju, N. Examining accident reports involving autonomous vehicles in California. *PLoS One* **12**, e0184952 (2017).
28. Seacrist, T. et al. In-depth analysis of crash contributing factors and potential ADAS interventions among at-risk drivers using the SHRP 2 naturalistic driving study. *Traffic Inj. Prev.* **22**, S68–S73 (2021).
29. Sinha, A., Vu, V., Chand, S., Wijayaratra, K. & Dixit, V. A crash injury model involving autonomous vehicle: Investigating of crash and disengagement reports. *Sustainability* **13**, 7938 (2021).
30. Wang, S. & Li, Z. Exploring the mechanism of crashes with automated vehicles using statistical modeling approaches. *PLoS One* **14**, e0214550 (2019).
31. NHTSA. *Standing General Order on Crash Reporting*, <https://www.nhtsa.gov/laws-regulations/standing-general-order-crash-reporting#data> (2023).
32. Ahangar, M. N., Ahmed, Q. Z., Khan, F. A. & Hafeez, M. A survey of autonomous vehicles: Enabling communication technologies and challenges. *Sensors* **21**, 706 (2021).
33. Yuan, R., Ding, S., Fang, Z., Gu, X., & Xiang, Q. Investigating the spatial heterogeneity of factors influencing speeding-related crash severities using correlated random parameter order models with heterogeneity-in-means. *Transp. Lett.* **15**, 1–13 (2023).
34. Sun, Z., Bebis, G. & Miller, R. On-road vehicle detection: A review. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**, 694–711 (2006).
35. Zang, S. et al. The impact of adverse weather conditions on autonomous vehicles: How rain, snow, fog, and hail affect the performance of a self-driving car. *IEEE Vehicular Technol. Mag.* **14**, 103–111 (2019).
36. Gehrig, S., Reznitskii, M., Schneider, N., Franke, U. & Weickert, J. Priors for stereo vision under adverse weather conditions. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 238–245 (2013).
37. Cui, Z., Yang, S.-W. & Tsai, H.-M. A vision-based hierarchical framework for autonomous frontvehicle taillights detection and signal

- recognition. In *2015 IEEE 18th International Conference on Intelligent Transportation Systems*. 931–937 (IEEE, 2015).
38. Hnawa, M. & Radha, H. Object detection under rainy conditions for autonomous vehicles: A review of state-of-the-art and emerging techniques. *IEEE Signal Process. Mag.* **38**, 53–67 (2020).
  39. Vargas, J., Alswiss, S., Toker, O., Razdan, R. & Santos, J. An Overview of Autonomous Vehicles Sensors and Their Vulnerability to Weather Conditions. *Sensors* **21**, 5397 (2021).
  40. Van Brummelen, J., O'Brien, M., Gruyer, D. & Najjaran, H. Autonomous vehicle perception: The technology of today and tomorrow. *Transp. Res. Part C* **89**, 384–406 (2018).
  41. Radecki, P., Campbell, M. & Matzen, K. All weather perception: Joint data association, tracking, and classification for autonomous ground vehicles. Preprint at <https://arxiv.org/abs/1605.02196> (2016).
  42. Filgueira, A., González-Jorge, H., Lagüela, S., Díaz-Vilariño, L. & Arias, P. Quantifying the influence of rain in LiDAR performance. *Measurement* **95**, 143–148 (2017).
  43. Parekh, D. et al. A review on autonomous vehicles: Progress, methods and challenges. *Electronics* **11**, 2162 (2022).
  44. Khatab, E., Onsy, A., Varley, M. & Abouelfarag, A. Vulnerable objects detection for autonomous driving: A review. *Integration* **78**, 36–48 (2021).
  45. Yeong, D. J., Velasco-Hernandez, G., Barry, J. & Walsh, J. Sensor and sensor fusion technology in autonomous vehicles: A review. *Sensors* **21**, 2140 (2021).
  46. Alotibi, F. & Abdelhakim, M. Anomaly detection for cooperative adaptive cruise control in autonomous vehicles using statistical learning and kinematic model. *IEEE Trans. Intell. Transp. Syst.* **22**, 3468–3478 (2020).
  47. Li, Y. et al. Evaluation of the impacts of cooperative adaptive cruise control on reducing rear-end collision risks on freeways. *Accid. Anal. Prev.* **98**, 87–95 (2017).
  48. Adewale, A. & Lee, C. Prediction of car-following behavior of autonomous vehicle and human-driven vehicle based on drivers' memory and cooperation with lead vehicle. *Transp. Res. Record.*, <https://doi.org/10.1177/03611981231195051> (2023).
  49. Li, Y., Wu, D., Lee, J., Yang, M. & Shi, Y. Analysis of the transition condition of rear-end collisions using time-to-collision index and vehicle trajectory data. *Accid. Anal. Prev.* **144**, 105676 (2020).
  50. Endsley, M. R. Toward a theory of situation awareness in dynamic systems. *Hum. factors* **37**, 32–64 (1995).
  51. Ding, Z., Sun, C., Zhou, M., Liu, Z. & Wu, C. Intersection vehicle turning control for fully autonomous driving scenarios. *Sensors* **21**, 3995 (2021).
  52. Levin, M. W. & Boyles, S. D. Intersection auctions and reservation-based control in dynamic traffic assignment. *Transp. Res. Rec.* **2497**, 35–44 (2015).
  53. Haris, M. & Hou, J. Obstacle detection and safely navigate the autonomous vehicle from unexpected obstacles on the driving lane. *Sensors* **20**, 4719 (2020).
  54. Bhavsar, P., Das, P., Paugh, M., Dey, K. & Chowdhury, M. Risk analysis of autonomous vehicles in mixed traffic streams. *Transp. Res. Rec.* **2625**, 51–61 (2017).
  55. Fagnant, D. J. & Kockelman, K. Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. *Transp. Res. Part A* **77**, 167–181 (2015).
  56. Zhang, Q. et al. A systematic framework to identify violations of scenario-dependent driving rules in autonomous vehicle software. *Proc. ACM Meas. Anal. Comput. Syst.* **5**, 1–25 (2021).
  57. Riedmaier, S., Ponn, T., Ludwig, D., Schick, B. & Diermeyer, F. Survey on scenario-based safety assessment of automated vehicles. *IEEE Access* **8**, 87456–87477 (2020).
  58. Kutela, B., Avelar, R. E. & Bansal, P. Modeling automated vehicle crashes with a focus on vehicle at-fault, collision type, and injury outcome. *J. Transp. Eng. Part A* **148**, 04022024 (2022).
  59. Zhou, D., Ma, Z., Zhang, X. & Sun, J. Autonomous vehicles' intended cooperative motion planning for unprotected turning at intersections. *IET Intell. Transp. Syst.* **16**, 1058–1073 (2022).
  60. Wael, K. M. A., Miho, A., Hideki, N. & Dang Minh, T. Stochastic approach for modeling the effects of intersection geometry on turning vehicle paths. *Transp. Res. Part C* **32**, 179–192 (2013).
  61. Noh, S. Decision-making framework for autonomous driving at road intersections: Safeguarding against collision, overly conservative behavior, and violation vehicles. *IEEE Trans. Ind. Electron.* **66**, 3275–3286 (2018).
  62. Ashraf, M. T., Dey, K., Mishra, S. & Rahman, M. T. Extracting rules from autonomous-vehicle-involved crashes by applying decision tree and association rule methods. *Transp. Res. Rec.* **2675**, 522–533 (2021).
  63. Zhou, D., Ma, Z. & Sun, J. Autonomous Vehicles' Turning Motion Planning for Conflict Areas at Mixed-Flow Intersections. *IEEE Trans. Intell. Veh.* **5**, 204–216 (2020).
  64. Grahm, H., Kujala, T., Silvennoinen, J., Leppänen, A. & Saariluoma, P. Expert drivers' prospective thinking-aloud to enhance automated driving technologies—Investigating uncertainty and anticipation in traffic. *Accid. Anal. Prev.* **146**, 105717 (2020).
  65. Lake, B. M., Ullman, T. D., Tenenbaum, J. B. & Gershman, S. J. Building machines that learn and think like people. *Behav. Brain Sci.* **40**, e253 (2017).
  66. Zhang, Y., Wang, W., Zhou, X., Wang, Q. & Sun, X. Tactical-level explanation is not enough: Effect of explaining AV's lane-changing decisions on drivers' decision-making, trust, and emotional experience. *Int. J. Hum. Comput. Interact.* **39**, 1438–1454 (2023).
  67. Rasouli, A. & Tsotsos, J. K. Autonomous vehicles that interact with pedestrians: A survey of theory and practice. *IEEE Trans. Intell. Transp. Syst.* **21**, 900–918 (2019).
  68. Schwarting, W., Pierson, A., Alonso-Mora, J., Karaman, S. & Rus, D. Social behavior for autonomous vehicles. *Proc. Natl Acad. Sci.* **116**, 24972–24978 (2019).
  69. Raouf, I. et al. Sensor-based prognostic health management of advanced driver assistance system for autonomous vehicles: A recent survey. *Mathematics* **10**, 3233 (2022).
  70. Lee, S., Arvin, R. & Khattak, A. J. Advancing investigation of automated vehicle crashes using text analytics of crash narratives and Bayesian analysis. *Accid. Anal. Prev.* **181**, 106932 (2023).
  71. Cui, J., Sabaliauskaite, G., Liew, L. S., Zhou, F. & Zhang, B. Collaborative analysis framework of safety and security for autonomous vehicles. *IEEE Access* **7**, 148672–148683 (2019).
  72. Sun, X., Cao, S. & Tang, P. Shaping driver-vehicle interaction in autonomous vehicles: How the new in-vehicle systems match the human needs. *Appl. Ergonomics* **90**, 103238 (2021).
  73. Kutela, B., Das, S. & Dadashova, B. Mining patterns of autonomous vehicle crashes involving vulnerable road users to understand the associated factors. *Accid. Anal. Prev.* **165**, 106473 (2022).
  74. Xu, C., Ding, Z., Wang, C. & Li, Z. Statistical analysis of the patterns and characteristics of connected and autonomous vehicle involved crashes. *J. Saf. Res.* **71**, 41–47 (2019).
  75. Zhu, S. & Meng, Q. What can we learn from autonomous vehicle collision data on crash severity? A cost-sensitive CART approach. *Accid. Anal. Prev.* **174**, 106769 (2022).
  76. Cascetta, E., Carteni, A. & Di Francesco, L. Do autonomous vehicles drive like humans? A Turing approach and an application to SAE automation Level 2 cars. *Transp. Res. Part C* **134**, 103499 (2022).

77. Petrović, Đ., Mijailović, R. & Pešić, D. Traffic accidents with autonomous vehicles: type of collisions, manoeuvres and errors of conventional vehicles' drivers. *Transp. Res. Proc.* **45**, 161–168 (2020).
78. Gross, F. & Jovanis, P. P. Estimation of the safety effectiveness of lane and shoulder width: Case-control approach. *J. Transp. Eng.* **133**, 362–369 (2007).
79. SWITRS. *Statewide Integrated Traffic Records System*, <https://iswitr.chp.ca.gov/Reports/jsp/index.jsp> (2023).
80. Ingram, D., Sanders, K., Kolybaba, M. & Lopez, D. Case-control study of phyto-oestrogens and breast cancer. *Lancet* **350**, 990–994 (1997).
81. Abdel-Aty, M. A., Hassan, H. M., Ahmed, M. & Al-Ghamdi, A. S. Real-time prediction of visibility related crashes. *Transp. Res. part C* **24**, 288–298 (2012).
82. Ahmed, M. M., Abdel-Aty, M. & Yu, R. Bayesian updating approach for real-time safety evaluation with automatic vehicle identification data. *Transp. Res. Rec.* **2280**, 60–67 (2012).
83. Rahman, M. M. & Lamsal, B. P. Ultrasound-assisted extraction and modification of plant-based proteins: Impact on physicochemical, functional, and nutritional properties. *Compr. Rev. Food Sci. Food Saf.* **20**, 1457–1480 (2021).
84. Rahman, R., Bhowmik, T., Eluru, N. & Hasan, S. Assessing the crash risks of evacuation: A matched case-control approach applied over data collected during Hurricane Irma. *Accid. Anal. Prev.* **159**, 106260 (2021).
85. Abdel-Aty, M., Uddin, N., Pande, A., Abdalla, M. F. & Hsia, L. Predicting freeway crashes from loop detector data by matched case-control logistic regression. *Transp. Res. Rec.* **1897**, 88–95 (2004).
86. Peck, R. C., Gebers, M. A., Voas, R. B. & Romano, E. The relationship between blood alcohol concentration (BAC), age, and crash risk. *J. Saf. Res.* **39**, 311–319 (2008).
87. Program, T. C. *Traffic Census Program*, <https://dot.ca.gov/programs/traffic-operations/census> (2024).
88. Therneau, T. M., Lunley, T., Atkinson, E. & Crowson, C. survival: Survival analysis. <https://cran.r-project.org/web/packages/survival/index.html> (2024).

## Acknowledgements

The authors wish to thank Dr. Ou Zheng for his role in creating the AVOID data used in this study.

## Author contributions

M.A.A. and S.D. conceived the study. M.A.A. and S.D. wrote the manuscript. M.A.A. and S.D. estimated the models and conducted the analysis. M.A.A. supervised the analysis and edited the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-48526-4>.

**Correspondence** and requests for materials should be addressed to Shengxuan Ding.

**Peer review information** *Nature Communications* thanks Yee Mun Lee and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024