Investigating the Interrelationships among Factors Associated with Automated Vehicle Crashes Using Additive Bayesian Network

Chunxi Huang
Robotics and Autonomous Systems, Division of Emerging Interdisciplinary Areas (EMIA) under Interdisciplinary Programs Office (IPO), The Hong Kong University of Science and Technology, Hong Kong SAR, China, 999077
Email: tracy.huang@connect.ust.hk

Ange Wang
Thrust of Intelligent Transportation, Systems Hub, The Hong Kong University of Science and Technology (Guangzhou), Guangdong, China, 511400
Email: awang324@connect.hkust-gz.edu.cn

Song Yan
Thrust of Intelligent Transportation, Systems Hub, The Hong Kong University of Science and Technology (Guangzhou), Guangdong, China, 511400
Email: syan931@connect.hkust-gz.edu.cn

Dengbo He, Corresponding Author
Thrust of Intelligent Transportation, Systems Hub, The Hong Kong University of Science and Technology (Guangzhou), Guangdong, China, 511400
Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong SAR, China, 999077
HKUST Shenzhen-Hong Kong Collaborative Innovation Research Institute, Futian, Shenzhen
Email: dengbohe@hkust-gz.edu.cn

ABSTRACT

Although automated vehicles (AV) were considered a promising solution to enhance traffic safety by eliminating human errors, AV crashes still happen in mixed traffic consisting of human-driven vehicles and AVs. Thus, to reduce AV-involved crashes, it is necessary to understand the factors leading to AV crashes. However, traditional regression-based methods may not reveal a structured relationship among leading factors of AV crashes, which hinders the exploration of countermeasures to AV crashes. Based on the 246 AV crash records collected by the National Highway Traffic Safety Administration, this study investigated the factors associated with AV crashes. An additive Bayesian network (ABN) approach was utilized to construct the topological relationship among potential influential factors of AV crashes, followed by post-ABN regression analyses. Results show that, though AV technologies have developed rapidly in the past few years, rear-end crashes are still dominating among AV-involved crashes, potentially because of the discrepancy in the driving behaviors between AV and human-driven vehicles. Compared to the pre-crash movements of AV, the crash type of AV-involved crashes is more related to the pre-crash movements of crash partners; while crash outcomes (e.g., injury severity) are associated with the environmental factors (e.g., operating entities) and crash-procedure-related factors (e.g., crash type). Findings from this study aid in understanding AV crash patterns, which can inform targeted interventions and technology advancements to improve safety outcomes for all road users.

Keywords: AV crashes, mixed traffic, traffic safety, additive Bayesian network, regression analysis

INTRODUCTION

Over the past few decades, with the advancements of both hardware (e.g., sensors, chips) and software (e.g., algorithms), automated vehicles (AVs) have evolved from rudimentary prototypes to sophisticated machines that are capable of cruising through complex traffic situations. Numerous companies are conducting extensive testing of AVs and even deployed them on public open roads, such as Baidu Apollo (*1*) in China and Waymo (*2*) in the U.S. This evolution has been shifting our transportation system from human-driven traffic only to mixed traffic (i.e., traffic consisted of human-driven vehicles and AVs), which is promised to enhance mobility (*3*), reduce congestion (*4*), and improve environmental sustainability (*5*) and traffic safety (*6*).

However, though over 90% of crashes are attributed to human error (*7*) and automation is expected to relieve humans from vehicle-controlling tasks (*8*), crashes still happen between AVs and human-driven vehicles (HDV) (*9*), given the complex interactions among road agents in mixed traffic. Thus, understanding the scenarios where conflicts between AV and HDV are most likely to occur and revealing the AV crash characteristics can provide insights into improving the safety of mixed traffic, for example, by refining AV control algorithms and enacting laws and regulations (*10*). The factors associated with AV-HDV crashes have been widely investigated in previous studies (*11–16*). However, constrained by traditional statistical models, previous research was not able to reveal the topological relationships among crash-related factors. For example, if one factor affects the crash rate by affecting another factor; one of the factors has to be abandoned or they need to be aggregated. Thus, the information regarding the association among multiple factors leading to AV crashes might be lost. Revealing the topological relationships among crash-related factors can help quantify the influential factors of AV-involved crashes and thus provide insights into AV crash prevention and intervention.

On the other hand, to the best of our knowledge, so far, all published studies that investigated the AV crash characteristics and explored associated factors were based on the crash report data collected by the California Department of Motor Vehicles (CA-DMV) (*11–16*). All these studies used the crash data before 2021 only. However, given the fast revolution of AV technologies, the perception and planning capabilities of AVs have improved significantly in the past few years (*17*). At the same time, with the prevalence of AVs being tested and deployed on public open roads in recent years, human drivers' attitudes towards AVs may also vary over time (*18*). Thus, insights from previous studies based on the CA-DMV dataset might be limited and may not guide the optimization of AV technology development and regulation design for mixed traffic. With more recent AV crash data available, it is imperative to reveal the most up-to-date high-risk scenarios in mixed traffic, which can provide up-to-date insights guiding the optimization of mixed traffic rules and design of AV testing scenarios.

Therefore, the objective of this study is to reveal the interrelationships among factors associated with recent AV-involved crashes. Specifically, based on an up-to-date AV crash dataset collected by the National Highway Traffic Safety Administration (NHTSA), 246 AV crashes that happened from July 2021 to April 15, 2023, were analysed using a graphical network-based mixed approach combining the additive Bayesian network (ABN) (*19*) and regression analysis. Findings from this study could help better understand the characteristics of AV-involved crashes and support the design and optimization of relevant technologies to enhance mixed traffic safety.

LITERATURE REVIEW

Previous studies have investigated AV crashes mainly from three perspectives: prevalence of AV crashes, characteristics of AV crashes, and factors influencing AV crash outcomes. Table 1 summarizes the main findings from these three perspectives.

TABLE 1 Summary of the previous studies on AV crashes

Perspectives	Main Findings	Study
Prevalence of AV crashes	The crash rate of AVs was higher than that of HDVs.	(*20*, *21*)
	AV fleet by Google had lower crash rates compared to that of HDVs.	(*22*)
Characteristics of AV crashes	The exitance of crosswalks, intersections, and traffic signals was positively associated with AV crashes involving vulnerable road users.	(*14*)
	The AV being rear-ended was the most frequent type of AV-involved crash.	(*10*, *12*, *13*, *15*, *23–25*)
	The chances of AV being rear-ended were lower in the parking lot when the traffic density was low, and when nonvehicular participants were present.	(*12*, *13*)
	Among AV-involved crashes, AVs were more likely to be rear-ended when AVs stopped at intersections.	(*11*)
Factors influencing AV crash outcomes	The AV at fault was positively associated with the chance of road agents (e.g., occupants of involved vehicles or pedestrians) getting injured.	(*13*, *16*)
	The likelihood of having an injured person in a crash was higher when the AV was traveling through road segments with higher speed limits.	(*13*, *16*)
	Crashes were more likely to lead to injuries on road segments with roadside parking, intersections, and arterial roads, and when the AV was rear-ended.	(*12*, *26*)
	The likelihood of having an injured person in a crash was higher when the AV was traveling through wet roads and with dark lighting conditions.	(*13*, *27*)

Prevalence of AV crashes

The crash rate was one of the most straightforward metrics of driving safety. Given the increasing prevalence of AVs in recent years, the comparison of crash rates between AV and HDV has attracted attention from researchers. For example, using the CA-DMV dataset and the safety record of HDVs in the U.S. in 2013, Schoettle and Sivak (*20*) first compared the crash rates between AVs tested on the road and HDVs. They found that the crash rate of AVs (9.1 crashes per million miles) was higher than that of HDVs (4.1 crashes per million miles), though the difference was not significant. Blanco et al. (*21*) did a more comprehensive comparison based on the internal crash records of Google's Self-Driving Car Program, in which the minor AV crashes that did not meet police-reporting thresholds were accounted for. The HDV crash rate from the Second Strategic Highway Research Program Naturalistic Driving Study (SHRP 2 NDS) was used as the baseline. As a result, the estimated AV crash rate was 20.2 crashes per million miles, which was significantly higher compared to the estimated national-level HDV crash rate of 2.75-9.87 crashes per million miles. However, it should be noted that the crash rate of AV is highly data-source-dependent and time-variant. For example, one year after Blanco et al. (*21*), Teoh and Kidd (*22*) compared the crash rates of Google's AV fleet to the crash rates of HDV in the SHRP 2 NDS dataset again. This time, they found that Google's AV fleet had significantly lower crash rates (9.7 crashes per million miles) compared to HDVs in the SHRP 2 NDS dataset (27.6 crashes per million miles), based on their proposed data processing method.

Characteristics of AV crashes

The analysis of crash rates alone may not reveal factors leading to crashes and thus may not guide the countermeasures that can improve the safety of mixed traffic. Thus, researchers also tried to explore

the characteristics of AV-involved crashes. For example, by using the narratives of 252 AV crashes from 2017 to 2020 in the CA-DMV dataset, Kutela, Das, et al. (*14*) found that the exitance of crosswalks, intersections, and traffic signals was positively associated with AV crashes with vulnerable road users. Further, by investigating the crash reports in the CA-DMV dataset, many studies identified the AV being rear-ended as the most frequent type of AV-involved crashes (*10*, *12*, *13*, *15*, *23–25*). For example, Teoh and Kidd (*22*) found that Google's AV was significantly more likely to be rear-ended (7.1 cases per million miles) than that of HDVs (2.7 cases per million miles). By analysing the crash reports from Oct. 2014 to Dec. 2019 collected by CA-DMV, a more recent study also found that AVs were 4.8 times more likely to be rear-ended compared to HDVs in the SHRP 2 NDS dataset (*23*).

Given the high prevalence of rear-end collisions in AV-involved crashes, the scenarios leading to rear-end collisions in mixed traffic were analysed. Specifically, the AV collision types could be affected by whether the AV was in automated driving mode, whether the crashes involved pedestrians or cyclists, and the roadway environment (*16*). For example, by using the AV crash reports collected by CA-DMV, it was found that, compared to when the driver disengaged the AV system before a crash, the likelihood of an AV being rear-ended was substantially higher when the driving automation was controlling the AV (*12*). Kutela, Avelar, et al. (*13*) found that the chances of AV being rear-ended were lower in the parking lot when the traffic density was low, and when nonvehicular participants were present. Besides, Ashraf et al. (*11*) found that, among AV-involved crashes, AVs were more likely to be rear-ended when AVs stopped at intersections.

Factors influencing AV crash outcomes

The injury severity level of road agents is a commonly adopted metric of crash outcomes (*28–30*). Thus, previous research also analysed factors associated with the injury severity level in AV-involved crashes. For example, by analysing the CA-DMV AV crash reports from 2017 to 2020, Kutela, Avelar, et al. (*13*) found that irrespective of the crash type, the AV at fault was associated with a higher chance of road agents (e.g., occupants in crash partners or pedestrians) getting injured. Such a trend has also been observed using the same CA-DMV AV crash reports from 2014 to 2018, in which more severe injuries were observed when the AV was at fault compared to when the HDV was at fault (*16*).

At the same time, the traveling speed was associated with the crash outcome. For example, both Kutela, Avelar, et al. (*13*) and Wang and Li (*16*) found that the likelihood of having an injured person in a crash was higher when the AV was traveling through road segments with higher speed limit (e.g., on highways as compared to that in urban areas). Similarly, Wang and Li (*16*) also found that severe injuries were more likely to happen on highways compared to that in urban areas.

Further, the road type and crash types may also influence crash severity. For example, Boggs et al. (*12*) found that crashes with injuries were more likely to happen on roads with roadside parking, intersections, and arterial roads, and when the AV was rear-ended. Similarly, using 72 CA-DMV AV crash reports from Jan. 2015 to Jun. 2018, Xu et al. (*26*) found that the collision location, existence of roadside parking, whether it was a rear-end collision or not, and road type (one-way vs. two-direction) were all factors associated with the injury severity of AV crashes.

Lastly, the environmental conditions may also affect the crash outcomes. For example, Kutela, Avelar, et al. (*13*) found that the likelihood of having an injured person in a crash was higher when the AV was traveling through wet roads. Based on the CA-DMV AV crash reports from Sep. 2014 to May. 2019, Das et al. (*27*) found that dark lighting conditions were associated with a higher likelihood of having a severely injured person in a crash.

Scope of the current study

Although existing studies have investigated factors associated with AV crashes with statistical (e.g., (*26*, *31*, *32*)) and machine-learning models (e.g., (*11*, *24*, *33*)), they mostly neglected the relationships among the underlying factors. Thus, they cannot fully reveal the interdependency among the crash-related factors and hence can provide limited guidance on the choice of countermeasures to moderate crash outcomes. To overcome these problems, recent studies have explored the topological

relationships among multiple contributing factors of AV crashes using the Bayesian Network (BN) (*13*, *25*). However, traditional BN requires arbitrary categorization of continuous variables, which may lead to a loss of information. Further, the BN can only inform dependency but not correlation. Thus, to overcome the above-mentioned problems, we adopted a graphical network-based mixed approach combining the additive Bayesian network (ABN) (*19*) and regression analysis. The ABN can also handle continuous variables and reduce information loss in the traditional BN approach. Further, being different from traditional regression-based analysis, the ABN can help identify potential interrelationships among candidate contributing factors of AV crashes and guide the choice of regression models, which can reduce the chance of type I errors when building multiple statistical models exhaustively.

METHODS

The overall methodological framework of this study is depicted in Figure 1. We first extracted three types of variables (i.e., environmental variables, crash-procedure-related variables, and crash outcome variables) from the NHTSA crash reports and aggregated some categorical variables to facilitate the following analyses. Then, we performed the ABN modeling to explore the topological interrelationships among the selected variables. Finally, based on the results of ABN, we further constructed post-ABN regression models to interpret the interrelationships quantitatively.

<Insert FIGURE 1 Here>

FIGURE 1 The overall methodological framework of this study.

Data Source

To obtain timely and transparent real-world crash data associated with automated driving systems and advanced driver assistance systems (ADAS), the National Highway Traffic Safety Administration (NHTSA) has issued a Standing General Order (SGO) in June 2021, which asked identified entities (e.g., manufactures and operators) to report crashes involving ADAS vehicles (i.e., vehicles with SAE level-2 driving automation) and AV (*34*). It should be noted that the AVs in the NHTSA dataset refer to vehicles with SAE Level 3, Level 4, or Level 5 systems that can perform all dynamic driving tasks (e.g., lane keeping and speed control), regardless of the designed operational region and the presence of a safety operator. According to the definition in the SGO, the AV-involved crashes included crashes in which the automated driving system was in use at any time 30 sec before the crash and the crash should have resulted in property damage or injury.

As of April 15, 2023, there were 281 AV-involved crashes reported by 25 reporting entities in the NHTSA dataset. Reporting entities submitted crash reports in electronic format using the web-based incident report form. For each crash, NHTSA collected 122 variables regarding six types of information: (1) AV Information (e.g., model, operating entity, and mileage of the involved AV); (2) Incident Information (e.g., incident date); (3) Incident Scene (e.g., location, roadway type, speed limit); (4) Crash Description (e.g., injury severity, crash with, contact area); (5) Post-Crash Information (e.g., data availability, whether investigated by a law enforcement agency or not); and (6) Narrative (i.e., written description of the pre-crash, crash, and post-crash details). The data used in this study was extracted from the AV crash report dataset collected by NHTSA, from July 2021 to April 15, 2023.

Data Pre-processing

In data pre-processing, we first filtered out repeated reports (due to multiple reporting entities or multiple versions of reports for the same crash) for each crash by using the vehicle identification number and same incident ID recorded in the crash report dataset, in which the latest version of the report (i.e., the

report with most recently updated information) was kept for analysis. Then, crash reports with missing information in five variables (i.e., the contact area of the crash partner (CP), the contact area of the AV, the pre-crash movement of CP, the injury severity, and the property damage) were excluded, as we could not determine variables of interest in this study, i.e., collision type, injury severity, and property damage of the crash. As a result, 246 AV-involved crashes were kept for analysis in our current study. All the data pre-processing was conducted using R 4.1.2 (*35*). It should be noted that, in the current study, we have analysed all AV-involved crashes, including the ones where the AVs were at fault or as victims.

Variable Extraction and Data Aggregation

Inspired by the classification method of NHTSA (*9*) and previous driving safety research (*10*, *25*, *26*), we extracted three types of factors from the NHTSA dataset, i.e., environmental factors (EF), crash-procedure-related factors (CPF), and crash-outcome factors (COF). To facilitate ABN analysis, all categorical variables were aggregated into two classes. Table 2 summarizes the extracted variables. The number of data points in each category before and after data aggregation is provided in the appendix.

TABLE 2 Summary of the extracted variables

[Type] Variable	Definition and Description	Distribution after Aggregation (Number of cases)	Ratio
[EF] Roadway type (RT)	The road type where the AV was operating at the time of the incident.	Intersection (113)	45.9%
		Other (133)	54.1%
[EF] Roadway surface (RS)	The roadway surface conditions at the time of the incident.	Dry (231)	93.9%
		Other (15)	6.1%
[EF] Weather	The weather conditions at the time and location of the incident.	Clear (211)	85.8%
		Other (35)	14.2%
[EF] Lighting	The lighting conditions at the time and location of the incident.	Daylight (155)	63.0%
		Other (91)	37.0%
[EF] State	The state or territory where the incident occurred.	California (162)	65.9%
		Other (84)	34.1%
[EF] Operating entity	The entity that was operating the AV at the time of the incident.	Waymo (116)	47.2%
		Other (130)	52.8%
[CPF] Crash type	The crash type of the AV-involved crash.	Rear-end (103)	41.9%
		Other (143)	58.1%
[CPF] Crash partner (CP)	Other road agents that involved in a crash with the AV.	Passenger car (154)	62.6%
		Other (92)	37.4%
[CPF] Pre-crash movement of CP (PMCP)	The pre-crash movement of the crash partner before the incident.	Proceeding straight (113)	45.9%
		Other (133)	54.1%
[CPF] Pre-crash movement of AV(PMAV)	The pre-crash movement of the AV before the incident.	Stopped (101)	62.6%
		Other (145)	37.4%
[CPF] Speed gap ratio (SGR)	$\mathrm{SGR} = \frac{\mathrm{PSL} - \mathrm{AVPCS}}{\mathrm{PSL}}$ (1) where PSL is the posted speed limit (in mph) of the roadway where the incident occurred; AVPCS, as defined in the NHTSA dataset, is the speed (in mph) of the AV at the time of the incident.	Mean = 0.77 Standard deviation = 0.32	/

[COF] Injury severity	The highest severity level of injury resulting from the incident.	No injuries (220)	89.4%
		Other (26)	10.6%
[COF] Property damage	Whether the incident resulted in any property damage or not.	Yes (219)	89.0%
		No (27)	11.0%
[COF] AV towed or not (AVToN)	Whether the AV was towed from the incident scene or not.	Yes (55)	22.4%
		No (191)	77.6%

Additive Bayesian Network (ABN)

Bayesian networks (BN) have been used in many studies to study observational data. Typically, for a set of random variables $X = \{X_1, \ldots, X_n\}$, a BN can be constructed as a directed acyclic graph (DAG) in which the nodes represent the random variables, and the directed links represent the statistical dependencies between the nodes. In a DAG, a node j can be regarded as the parent of a node i if there is an edge connecting j and i. A set of parents for node i can be denoted as Pa_i. The factorization of the joint probability in the BN was based on the Markov property, by which a node i is only dependent on its parent set Pa_i:

$$P(X) = P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid Pa_i) \qquad (2)$$

where n denoted the total number of nodes in a BN. As mentioned, the BN requires arbitrary categorization of continuous variables. To overcome this limitation, the ABN method was proposed (*19*) to provide flexibility in modelling complex systems and incorporate variable types of variables(e.g., countable, categorical, and continuous variables). ABN extends BN by incorporating multiple generalized linear models (GLMs) into the BN network (*36*). In the ABN, each node is viewed as the dependent variable in a GLM, and its parent variables are modelled as covariates. Unlike structural equation modelling which relies on expert knowledge to determine an optimal model structure (*37*), ABN modelling is data-driven (*36*), although expert knowledge can still be incorporated if necessary.

In our study, we utilized a Bayesian approach for both ABN structure discovery and parameter learning. Specifically, a uniform structural prior was applied, weighting all eligible DAG structures equally in the absence of data. At the same time, uninformative priors were used for all parameters of each edge, assuming that all values were equally probable in the absence of data. The ABN modelling in this study was conducted using R (*35*) and JAGS (*38*), which involved four consecutive steps as shown in Figure 2.

<Insert FIGURE 2 Here>

FIGURE 2 The procedures of ABN modelling. In the figure, BN is the abbreviation for Bayesian networks.

Step 1: Initialize the input variables

In the ABN model, all variables were modelled using specific probability distributions based on their types: binomial distribution for binary variables and Gaussian distribution for quantitative variables (*19*). Thus, in the current study, the speed gap ratio (SGR) was modelled using the Gaussian distribution while all other variables were modelled using the binomial distribution. To avoid potential numerical accuracy issues, following previous research (*39*), we standardized each SGR using the mean and standard deviation of all samples. Besides, we incorporated prior knowledge about the data structure to

guide the optimal model searching. Specifically, all edges directing to the environmental variables were banned with the assumption that these variables were not influenced by any other variables.

Step 2: Identify the optimal model

To identify the optimal model from the data, we adopted an exact order-based exhaustive search method (*40*), in which we varied the maximum number of nodes allowed to be directed to each node from one to seven following the previous study (*19*). After that, the model was evaluated using the network score (i.e., the log marginal likelihood), in which the model with the highest goodness-of-fit (i.e., the largest network score) was selected as the optimal model (*19*). The R package '*abn*' (version 2.7-3) was used in this step to construct and identify the optimal model.

Step 3: Adjust to overcome potential overfitting

To validate the identified optimal model and address the potential issue of overfitting in ABN (*41*), we utilized a parametric bootstrapping method based on Markov Chain Monte Carlo (MCMC) simulations (*42*) in our study. As illustrated in Fig. 1, we first estimated the posterior parameters from the optimal model structure using Laplace approximations. Subsequently, we employed MCMC samplers (i.e., JAGS) to generate 10,000 bootstrap datasets based on the optimal model structure identified from Step 2 and the estimated posterior parameters. Each generated bootstrap dataset had the same size as the original dataset. Next, we conducted an exact model search (i.e., repeating Step 2) on all 10,000 generated bootstrap datasets, resulting in a total of 10,000 bootstrap ABN structures (i.e., 10,000 DAGs). By calculating the frequency of occurrence for each edge in the bootstrap ABN structures, we were able to identify the edges in the DAG that were recovered in over 50% of the bootstrap ABN structures, which were retained in the final pruned ABN model as they could be considered as robust.

Step 4: Calculate the link strength

As the p-value in frequentist statistics is not appropriate for ABN, the link strength was adopted, which can be regarded as a proxy for edge uncertainty. Following a previous study (*19*), we adopted the percentage link strength (PLS) as the link strength metric. Specifically, the PLS between variable X and Y is defined as the percentage reduction in uncertainty of variable Y given the state of variable X, when the states of all other variables directed to Y are known. The PLS can be expressed as follows:

$$PLS\ (X \rightarrow Y)\ = \frac{H(Y|Z)\ -\ H(Y|X,Z)}{H\ (Y|Z)} \qquad (3)$$

where Z represents all other variables directed to Y (except for X) and H is the entropy computed using the empirical distribution of the random variable. When there is no arc connecting X and Y, the PLS is set to zero.

Post-ABN Regression Analysis

To provide more detailed interpretations of the DAG constructed by the ABN method, we conducted post-ABN regression analyses for all edges in the DAG. For each child variable (i.e., variable being directed by other variables), a regression model was fitted by viewing the child variable itself as the dependent variable and all variables directing to it as independent variables. More specifically, when the dependent variable was continuous, we fitted a linear regression model; while when the dependent variable was binomial distributed, we fitted a binomial logistic regression model. Because of the additivity assumption in ABN, the potential interaction effects between variables were not modeled in ABN (*19*). To account for this, in the post-ABN regression analysis, for models with two or more independent variables, we included all potential two-way interaction effects between independent variables in the initial full model. For example, for a model with Y (e.g., binary variable with possible values 1 and 0) as the dependent variable and X_1, X_2, X_3 as independent variables, the equation for the full models is as follows:

$$ln\left(\frac{p(Y=1)}{p(Y=0)}\right) = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_3 + \beta_{1|2} * (X_1|X_2) + \beta_{1|3} * (X_1|X_3) + \beta_{2|3} * (X_2|X_3) \quad (4)$$

where, $(X_1|X_2)$, $(X_1|X_3)$, and $(X_2|X_3)$ stands for the two-way interaction effects between all independent variables. For all the models that we fitted in the post-ABN regression analysis, we applied the backward stepwise selection method based on Bayesian Information Criterion (BIC) (*43*) for model selection. Only significant interaction effects ($p < .05$) were reported. All the regression analyses were conducted using R (*35*).

RESULTS

ABN Modelling Analysis

The extracted optimal DAG is shown in Figure 2, in which rectangles represent binary variables, and ovals represent continuous variables. In Figure 2, the percentage values next to the arcs are the percentage link strength (PLS) of each link and the widths of the arcs are proportional to the PLS. The ABN modelling analysis performed structure discovery based on the observed AV crash data and its outcome reveals the topological interrelationships among all variables of interest, which were used to guide the post-ABN regression analyses. The larger PLS means that more uncertainty can be explained by the parent variable. For example, based on the DAG, the speed gap ratio (PLS=20.7%), crash partner (PLS=11.1%), and pre-crash movement of crash partner (PMCP, PLS=28.5%) may be associated with the crash type, among which, the PMCP accounted for most uncertainty in the crash type. However, the PLS cannot inform if a linear association between the parent variables and the child variable exists. Therefore, we further fit a regression model by taking the crash type as the dependent variable and speed gap ratio, crash partner, and PMCP as independent variables.

<Insert FIGURE 3 Here>

FIGURE 3 The final tuned optimal DAG.

Post-ABN Regression Analysis

Based on the optimal DAG model constructed by the ABN method, in total, we fitted 7 regression models, with roadway type, speed gap ratio, PMAV, crash type, injury severity, property damage, and roadway surface as dependent variables, respectively. Post-hoc contrasts were conducted if the independent variable was significant in the model. Table 3 summarizes the model results and the significant post-hoc contrasts.

TABLE 3 Wald statistics of type 3 analysis for all models

Model	Dependent Variable	Independent Variable	Statistics	*p*
1	Roadway Type	State	$\chi^2(1) = 15.60$	< .0001
		Operating Entity	$\chi^2(1) = 30.63$	< .0001
2	Speed Gap Ratio	Roadway Type	$F(1, 244) = 18.99$	< .0001
3	PMAV	Speed Gap Ratio	$\chi^2(1) = 35.61$	< .0001
4	Crash Type	Speed Gap Ratio	$\chi^2(1) = 10.87$	.001
		Crash Partner	$\chi^2(1) = 10.62$	.001
		PMCP	$\chi^2(1) = 52.46$	< .0001
5	Injury Severity	Crash Type	$\chi^2(1) = 10.78$	.0009
		Operating Entity	$\chi^2(1) = 16.02$	< .0001

		AVToN	$\chi^2(1) = 14.13$	.0002
6	Property Damage	Operating Entity	$\chi^2(1) = 10.33$	.001
7	Roadway Surface	Weather	$\chi^2(1) = 27.69$	< .0001

Model 1: Roadway type. It was found that both state and operating entity were significantly associated with the roadway type. Compared to other states, the crashes that happened in California were more likely to happen at intersections compared to those on other types of roads, with an odds ratio (OR) of 3.45 and a 95% confidence interval (CI) of [1.87, 6.39]. Besides, compared to other operating entities, AV vehicles operated by Waymo were more likely to be involved in crashes at intersections compared to those happening at other locations (OR = 4.91, 95%CI: [2.80, 8.64]).

Model 2: Speed Gap Ratio (SGR). The roadway type was significantly associated with SGR. Specifically, the intersection was associated with higher SGR compared to other roadway types (mean difference (Δ) = 0.17, 95%CI: [0.09, 0.25], t (244) = 4.36, $p < .0001$).

Model 3: PMAV. The SGR was associated with the PMAV. Specifically, every 0.1-unit increase in speed gap ratio (SGR) led to a 3.05 multiplicative increase in the odds of PMAV being stopped, with a 95%CI: [2.11, 4.39].

Model 4: Crash type. It was found that the speed gap ratio (SGR), Crash Partner, and PMCP were all significantly associated with the crash type. Every 0.1-unit increase in speed gap ratio (SGR) led to a 1.23 multiplicative increase in the odds of the AV being rear-ended, with a 95%CI: [1.09, 1.39]. Compared to other crash partners, the passenger car had a higher probability of being involved in rear-end crashes with AV (OR = 3.23, 95%CI: [1.60, 6.54]). When the PMCP was *Proceeding straight*, the AV crashes had a higher probability of being rear-end crashes compared to when the PMCP was *Others*, with an OR of 11.58, 95%CI: [5.97, 22.46].

Model 5: Injury severity. The crash type, operating entity, and AVToN were all significantly associated with the injury severity. Specifically, when a rear-end crash happened, the AV-involved crashes had a higher probability of having an injured person compared to when the crash type was *Others*, with an odds ratio (OR) of 3.46, 95%CI: [1.58, 6.64]. Compared to AV operated by other entities, we found that AV operated by Waymo had lower probabilities of having injured persons in crashes (OR = 0.17, 95%CI: [0.07, 0.41]). Further, if the AV was towed, the AV-involved crashes had a higher probability of having an injured person (OR = 4.75, 95%CI: [2.11, 10.71]).

Model 6: Property damage. The operating entity was also associated with property damage. Specifically, compared to other operating entities, AV operated by Waymo had a higher probability of having property damage in crashes, with an odds ratio (OR) of 6.02, 95%CI: [2.01, 17.98].

Model 7: Roadway surface. It was found that the weather was associated with the roadway surface. Compared to other weather, the roadway surface had a higher probability of being dry when the weather was clear (OR = 36.17, 95%CI: [9.51, 137.66]).

DISCUSSION

This study explored the interrelationships among crash scenario factors, crash-procedure-related factors, and crash outcome factors among 246 AV crashes using an ABN-regression mixed approach. The ABN approach facilitated structure discovery among all selected variables and obtained a graphical model, while the regression analysis further provided a more detailed interpretation of the interrelationships. To our knowledge, this is the first attempt to use the ABN-based approach to explore factors related to AV crashes.

Interrelationships among factors related to AV crashes

Environmental factors

For AV crashes, we found that compared to crashes that happened in other states, the crashes that happened in California were more likely to be at intersections. Similarly, compared to other companies, the AVs operated by Waymo were more likely to be at intersections. However, we can hardly comment

on this phenomenon from an AV technology perspective of view, as we have no information regarding the exposure of AVs in different states and by different companies. Potentially, the driving culture and traffic regulations may differ between states and technologies may differ between different companies, which may explain the differences in crash characteristics. Alternatively, it is also possible that California opened more urban areas for AV tests compared to other states, or Waymo had a higher ratio of vehicles being tested in urban areas compared to other companies. However, controlling for exposure is not easy for AV crash studies, as this piece of information might be a commercial secret. If possible, future research may better answer this research question with exposure considered.

Crash-procedure-related factors

Compared to crashes that happened on other types of roadways, AV-involved crashes were associated with higher SGR at intersections. In other words, AVs tended to drive at lower speeds relative to the posted speed limit when navigating through intersections. The larger SGR at interactions indicates that the AVs were more cautious at interactions compared to those on other types of roadway, potentially because the interactions are complex and had a higher level of uncertainty (*10*, *24*, *44*). The conservativeness of driving automation in mixed traffic has been repeatedly discovered in previous studies based on different datasets (e.g., Huang et al. (*45*) based on the Waymo dataset, Song et al. (*24*) based on the CA-DMV dataset).

With a larger SGR, the AV had a higher chance of stopping and a higher chance of being rear-ended by other vehicles. The association between the SGR and the PMAV is straightforward – stopping is an extreme scenario of reducing speed. However, we should be cautious when interpreting the relationship between the SGR and the chance of being rear-ended. Intuitively, driving slower than the posted speed should not increase the crash risk if the AV is following the speed of the traffic flow. However, if the AV is driving slower than it should be, then the risk of being rear-ended by others might increase, which was in line with previous studies (*11–13*). This is especially the case when human drivers are not able to anticipate the sudden speed drops of AVs (*44*), particularly on roadways with a high-speed limit, even human drivers have kept a larger distance from AVs (*45*, *46*).

The lack of anticipation may also explain why the chance of an AV being rear-ended was associated with pre-crash movement (PMCP) of proceeding straight – the human driver's sight of view might be blocked by the leading AVs on the straight section of a road and thus drivers' capability to anticipate the unexpected stop of the leading AV may be impaired. This finding further echoes another association discovered in our study. We found that among AV-involved crashes, human-driven passenger cars had higher probabilities of rear-ending a leading AV. Intuitively, passenger cars generally have a shorter stopping distance compared to larger vehicles due to their smaller size and smaller mass (*47*, *48*) and thus should have a lower chance of rear-ending a lead vehicle. Thus, the high chance of passenger cars involving rear-end crashes might be explained as the occlusion of the front view by AVs. Specifically, given the relatively lower sitting position of passenger car drivers compared to that of other types of vehicles (e.g., SUVs), it is likely that the driver of passenger cars is more likely to be occluded and less able to anticipate unexpected behaviours of the leading vehicle.

Crash-outcome factors

First, it is straightforward that when a vehicle was towed away, there was a higher chance that someone got injured in the accident, as the vehicle being towed away indicates a higher severity of a collision (*49*). Similarly, the association between pre-crash movement of moving straight of crash partner (PMCP) and the chance of having someone injured in a collision is also easy to understand. When a vehicle is proceeding straight, it is likely to be traveling at a higher speed compared to that of other pre-crash movements, such as turning or merging. As a result, crashes involving straight-moving vehicles may involve higher impact forces and energy transfer, leading to a higher likelihood of injury (*50*).

At the same time, it is interesting to find that the AVs operated by Waymo had a lower probability of resulting in injuries while a higher probability of incurring property damage compared to AVs operated by other entities. This might be related to different operational strategies or testing

environments of AVs operated by different companies. Specifically, AVs by Waymo have been found to implement more conservative driving strategies (e.g., maintaining longer following distances and adhering to lower speeds) compared to HDVs (*51*). While these strategies may reduce injury severity effectively in crashes, they may also increase the likelihood of property damage, as human-driven vehicles may not anticipate or adapt to the over-conservative behavior of AVs, leading to more frequent minor collisions with AVs. Furthermore, the testing environments of AVs by Waymo might also play a role. Compared to other AVs included in the NHTSA dataset, Waymo is more frequently tested and operated in urban areas with higher traffic complexity but lower posted speed limits, which increases the chance of minor collisions without injury but with property damage.

Limitations and future research

Despite the interesting insights, the findings from this study were restricted to the current data and should be interpreted carefully. For example, the exposure of AVs should be considered given that different states in the U.S. may have different policies toward AVs. Thus, the AVs have different exposures in different states (e.g., many companies tested and deployed their AVs in California). At the same time, due to the limited data collection duration (i.e., < 2 years) in the dataset, we were not able to consider the temporal factor. Given the fast development of AV technologies and the increasing exposure of AVs, it will be interesting to investigate the effect of temporal factors on AV crash characteristics. Therefore, future research may consider including spatial factors (e.g., the exposure of AVs) and temporal factors (e.g., years) in the modeling process, which may support a better understanding of the characteristics of AV-involved crashes and thus enhance mixed traffic safety. Next, although we have considered the roadway type in our analysis, the road type is multi-dimensional. For example, the lane width and traffic volume can be different across types of roads and some road-type-related dimensions have been found to be related to crash outcomes. For example, the crash has been found to be more likely to happen on roads with higher traffic volumes (*52–54*). Unfortunately, neither the NHTSA dataset nor the CA-DMV dataset provides information regarding these subdimensions of road type. Thus, future research may also consider modeling how the subdimensions of road type can affect AV crashes when the data becomes available. Besides, it should be noted that, due to the limit of the ABN, we aggregated the categorical data into binary variables and there might be information loss in this process. Future research may explore better ways to model the topological data. Finally, as limited by the NHTSA dataset, we were not able to differentiate the crashes with AVs at fault and as victims. Future categorization of crash responsibilities may provide higher resolution on factors leading to AV-involved crashes.

At the same time, the limitation of the high computational cost of ABN should also be noted here. For example, currently (implementation in R following (*40*)), it is difficult to conduct exact structure discovery with bootstrapping to fit over 25 variables, as the computation cost grows exponentially as a function of the variable number. Although we could adopt the less ideal heuristic searches (*55*) and inexact order-based searches (*56*) as alternatives, the allowable maximum number of variables was still around 40 (*36*). Thus, we could only analyse limited variables using ABN before future advancements in technology that make intensive computation feasible. Future optimization is needed to increase the applicability of the ABN approach.

CONCLUSIONS

Based on 246 AV crashes extracted from the crash reports in the NHTSA dataset, we investigated the interrelationships among factors related to AV crashes using a mixed approach combining additive Bayesian network and regression analysis. The key findings and conclusions of this study are summarized as follows:

- Although driving automation technologies have been evolving rapidly in the past few years, rear-end crashes are still dominating among AV-involved crashes, potentially because of the discrepancy in the driving behaviors between AV and human-driven vehicles.
- Compared to the pre-crash movements of AV, the crash type of AV-involved crashes is more related to the pre-crash movements of crash partners; while crash outcomes (e.g.,

injury severity) are associated with the environmental factors (e.g., operating entities) and crash-procedure-related factors (e.g., crash type).

- The mixed approach combining ABN and post-ABN regression can model the interrelationships among the variables of interest and better reveal how the crash outcomes and crash characteristics are associated with different types of factors compared to the traditional regression-only approach.

In summary, the findings from this study have revealed the characteristics of AV that may lead to collisions and thus may guide the optimization of AV. For example, we found that the environment-related factors are not directly associated with the crash outcomes, indicating that the sensors so far may not be the bottleneck of AV safety at this stage. In contrast, the traffic- and motion-related factors were directly and indirectly associated with the crash outcomes. For example, over-conservation is likely not always good – it reduces the likelihood of having an injured person in a crash but increases the chance of crashes with property damages. This calls for more human-like AV control strategies or new external human-machine interfaces (eHMI) for AV to enhance HDV drivers' understanding of AV intentions so that mixed traffic safety can be improved.

ACKNOWLEDGMENTS

This research was supported by the National Natural Science Foundation of China (Grant No. 52202425) and partially by the Guangzhou Municipal Science and Technology Project (No. 2023A03J0011), and Project of Hetao Shenzhen-Hong Kong Science and Technology Innovation Cooperation Zone (HZQB-KCZYB-2020083).

The authors would like to thank the Transportation Behavioral Psychology and Safety Lab at the Hong Kong University of Science and Technology (Guangzhou). The authors would also express their most sincere thanks to the National Highway Traffic Safety Administration (NHTSA) for their efforts in collecting and releasing the raw data used in this study.

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: C.H., D.H.; data processing: C.H., A.W.; analysis and interpretation of results: C.H., A.W., S. Y., D.H.; draft manuscript preparation: C.H., A.W., S. Y., D.H. All authors reviewed the results and approved the final version of the manuscript.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

REFERENCES

1. Baidu Apollo. Autonomous Driving Solution from Apollo. https://apollo.auto. Accessed Feb. 17, 2022.
2. Waymo. We Are Building the World's Most Experienced Driver. waymo.com. Accessed Feb. 17, 2022.
3. Zhao, L., and A. A. Malikopoulos. Enhanced Mobility with Connectivity and Automation: A Review of Shared Autonomous Vehicle Systems. *IEEE Intelligent Transportation Systems Magazine*, Vol. 14, No. 1, 2020, pp. 87–102.
4. Cohen, T., and C. Cavoli. Automated Vehicles: Exploring Possible Consequences of Government (Non) Intervention for Congestion and Accessibility. *Transport reviews*, Vol. 39, No. 1, 2019, pp. 129–151.
5. Taiebat, M., A. L. Brown, H. R. Safford, S. Qu, and M. Xu. A Review on Energy, Environmental, and Sustainability Implications of Connected and Automated Vehicles. *Environmental science & technology*, Vol. 52, No. 20, 2018, pp. 11449–11465.

6. Tafidis, P., H. Farah, T. Brijs, and A. Pirdavani. Safety Implications of Higher Levels of Automated Vehicles: A Scoping Review. *Transport reviews*, Vol. 42, No. 2, 2022, pp. 245–267.
7. Singh, S. *Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey (DOT HS 812 115)*. 2015.
8. Sam, D., C. Velanganni, and T. E. Evangelin. A Vehicle Control System Using a Time Synchronized Hybrid VANET to Reduce Road Accidents Caused by Human Error. *Vehicular communications*, Vol. 6, 2016, pp. 17–28.
9. National Highway Traffic Safety Administration. *Summary Report: Standing General Order on Crash Reporting for Automated Driving Systems*. 2022.
10. Liu, Q., X. Wang, X. Wu, Y. Glaser, and L. He. Crash Comparison of Autonomous and Conventional Vehicles Using Pre-Crash Scenario Typology. *Accident Analysis & Prevention*, Vol. 159, 2021, p. 106281. https://doi.org/10.1016/j.aap.2021.106281.
11. Ashraf, M. T., K. Dey, S. Mishra, and M. T. Rahman. Extracting Rules from Autonomous-Vehicle-Involved Crashes by Applying Decision Tree and Association Rule Methods. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2675, No. 11, 2021, pp. 522–533. https://doi.org/10.1177/03611981211018461.
12. Boggs, A. M., B. Wali, and A. J. Khattak. Exploratory Analysis of Automated Vehicle Crashes in California: A Text Analytics & Hierarchical Bayesian Heterogeneity-Based Approach. *Accident Analysis & Prevention*, Vol. 135, 2020, p. 105354. https://doi.org/10.1016/j.aap.2019.105354.
13. Kutela, B., R. E. Avelar, and P. Bansal. Modeling Automated Vehicle Crashes with a Focus on Vehicle At-Fault, Collision Type, and Injury Outcome. *Journal of Transportation Engineering, Part A: Systems*, Vol. 148, No. 6, 2022. https://doi.org/10.1061/JTEPBS.0000680.
14. Kutela, B., S. Das, and B. Dadashova. Mining Patterns of Autonomous Vehicle Crashes Involving Vulnerable Road Users to Understand the Associated Factors. *Accident Analysis & Prevention*, Vol. 165, 2022, p. 106473. https://doi.org/10.1016/j.aap.2021.106473.
15. Petrović, Đ., R. Mijailović, and D. Pešić. Traffic Accidents with Autonomous Vehicles: Type of Collisions, Manoeuvres and Errors of Conventional Vehicles' Drivers. *Transportation Research Procedia*, Vol. 45, 2020, pp. 161–168. https://doi.org/10.1016/j.trpro.2020.03.003.
16. Wang, S., and Z. Li. Exploring the Mechanism of Crashes with Automated Vehicles Using Statistical Modeling Approaches. *PLOS ONE*, Vol. 14, No. 3, 2019, p. e0214550. https://doi.org/10.1371/journal.pone.0214550.
17. Kontar, W., T. Li, A. Srivastava, Y. Zhou, D. Chen, and S. Ahn. On Multi-Class Automated Vehicles: Car-Following Behavior and Its Implications for Traffic Dynamics. *Transportation Research Part C: Emerging Technologies*, Vol. 128, 2021, p. 103166. https://doi.org/10.1016/j.trc.2021.103166.
18. Kaye, S.-A., K. Somoray, D. Rodwell, and I. Lewis. Users' Acceptance of Private Automated Vehicles: A Systematic Review and Meta-Analysis. *Journal of safety research*, Vol. 79, 2021, pp. 352–367.
19. Kratzer, G., F. Lewis, A. Comin, M. Pittavino, and R. Furrer. Additive Bayesian Network Modeling with the R Package Abn. *Journal of Statistical Software*, Vol. 105, No. 8, 2023. https://doi.org/10.18637/jss.v105.i08.
20. Schoettle, B., and M. Sivak. A Preliminary Analysis of Real-World Crashes Involving Self-Driving Vehicles. *University of Michigan Transportation Research Institute*, 2015.
21. Blanco, M., J. Atwood, S. M. Russell, T. E. Trimble, J. A. McClafferty, and M. A. Perez. *Automated Vehicle Crash Rate Comparison Using Naturalistic Data*. Virginia Tech Transportation Institute, 2016.
22. Teoh, E. R., and D. G. Kidd. Rage against the Machine? Google's Self-Driving Cars versus Human Drivers. *Journal of Safety Research*, Vol. 63, 2017, pp. 57–60. https://doi.org/10.1016/j.jsr.2017.08.008.

23. Goodall, N. J. Comparison of Automated Vehicle Struck-from-behind Crash Rates with National Rates Using Naturalistic Data. *Accident Analysis & Prevention*, Vol. 154, 2021, p. 106056. https://doi.org/10.1016/j.aap.2021.106056.
24. Song, Y., M. V. Chitturi, and D. A. Noyce. Automated Vehicle Crash Sequences: Patterns and Potential Uses in Safety Testing. *Accident Analysis & Prevention*, Vol. 153, 2021, p. 106017. https://doi.org/10.1016/j.aap.2021.106017.
25. Lee, S., R. Arvin, and A. J. Khattak. Advancing Investigation of Automated Vehicle Crashes Using Text Analytics of Crash Narratives and Bayesian Analysis. *Accident Analysis & Prevention*, Vol. 181, 2023, p. 106932. https://doi.org/10.1016/j.aap.2022.106932.
26. Xu, C., Z. Ding, C. Wang, and Z. Li. Statistical Analysis of the Patterns and Characteristics of Connected and Autonomous Vehicle Involved Crashes. *Journal of Safety Research*, Vol. 71, 2019, pp. 41–47. https://doi.org/10.1016/j.jsr.2019.09.001.
27. Das, S., A. Dutta, and I. Tsapakis. Automated Vehicle Collisions in California: Applying Bayesian Latent Class Model. *IATSS Research*, Vol. 44, No. 4, 2020, pp. 300–308. https://doi.org/10.1016/j.iatssr.2020.03.001.
28. Abdel-Aty, M., and J. Keller. Exploring the Overall and Specific Crash Severity Levels at Signalized Intersections. *Accident Analysis & Prevention*, Vol. 37, No. 3, 2005, pp. 417–425.
29. Kockelman, K. M., and Y.-J. Kweon. Driver Injury Severity: An Application of Ordered Probit Models. *Accident Analysis & Prevention*, Vol. 34, No. 3, 2002, pp. 313–321.
30. Savolainen, P. T., F. L. Mannering, D. Lord, and M. A. Quddus. The Statistical Analysis of Highway Crash-Injury Severities: A Review and Assessment of Methodological Alternatives. *Accident Analysis & Prevention*, Vol. 43, No. 5, 2011, pp. 1666–1676.
31. Huang, C., X. Wen, and D. He. Characteristics of Rear-End Collisions: A Comparison between Automated Driving System-Involved Crashes and Advanced Driving Assistance System-Involved Crashes. *Transportation Research Record: Journal of the Transportation Research Board*, 2023. https://doi.org/10.1177/03611981231209319.
32. Yan, S., C. Huang, and D. He. A Comparison of Patterns and Contributing Factors of ADAS and ADS Involved Crashes. *Journal of Transportation Safety & Security*, 2023, pp. 1–28. https://doi.org/10.1080/19439962.2023.2284175.
33. Chen, H., H. Chen, Z. Liu, X. Sun, and R. Zhou. Analysis of Factors Affecting the Severity of Automated Vehicle Crashes Using XGBoost Model Combining POI Data. *Journal of Advanced Transportation*, Vol. 2020, 2020, pp. 1–12. https://doi.org/10.1155/2020/8881545.
34. National Highway Traffic Safety Administration. Standing General Order on Crash Reporting for Incidents Involving ADS and Level 2 ADAS. https://www.nhtsa.gov/laws-regulations/standing-general-order-crash-reporting. Accessed Jul. 21, 2022.
35. R Development Core Team. R Core Team (2020). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL Https://Www.R-Project.Org/. *R Foundation for Statistical Computing*.
36. Lewis, F. I., and M. P. Ward. Improving Epidemiologic Data Analyses through Multivariate Regression Modelling. *Emerging Themes in Epidemiology*, Vol. 10, No. 1, 2013, p. 4. https://doi.org/10.1186/1742-7622-10-4.
37. Hair, J. F. Multivariate Data Analysis. 2009.
38. Hornik, K., F. Leisch, A. Zeileis, and M. Plummer. JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling. No. 2, 2003.
39. Comin, A., A. Jeremiasson, G. Kratzer, and L. Keeling. Revealing the Structure of the Associations between Housing System, Facilities, Management and Welfare of Commercial Laying Hens Using Additive Bayesian Networks. *Preventive Veterinary Medicine*, Vol. 164, 2019, pp. 23–32. https://doi.org/10.1016/j.prevetmed.2019.01.004.
40. Koivisto, M., and K. Sood. Exact Bayesian Structure Discovery in Bayesian Networks. *The Journal of Machine Learning Research*, Vol. 5, 2004, pp. 549–573.

41. Babyak, M. A. What You See May Not Be What You Get: A Brief, Nontechnical Introduction to Overfitting in Regression-Type Models. *Psychosomatic medicine*, Vol. 66, No. 3, 2004, pp. 411–421.
42. Friedman, N., M. Goldszmidt, and A. Wyner. Data Analysis with Bayesian Networks: A Bootstrap Approach. 2013.
43. Burnham, K. P., and D. R. Anderson. Multimodel Inference. *Sociological Methods & Research*, Vol. 33, No. 2, 2004, pp. 261–304. https://doi.org/10.1177/0049124104268644.
44. McCarthy, R. L. Autonomous Vehicle Accident Data Analysis: California OL 316 Reports: 2015–2020. *ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg*, Vol. 8, No. 3, 2022. https://doi.org/10.1115/1.4051779.
45. Huang, C., X. Wen, D. He, and S. Jian. Sharing the Road: How Human Drivers Interact with Autonomous Vehicles on Highways. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 66, No. 1, 2022, pp. 1437–1441. https://doi.org/10.1177/1071181322661165.
46. Wen, X., Z. Cui, and S. Jian. Characterizing Car-Following Behaviors of Human Drivers When Following Automated Vehicles Using the Real-World Dataset. *Accident Analysis & Prevention*, Vol. 172, 2022, p. 106689. https://doi.org/10.1016/j.aap.2022.106689.
47. Bujanovic, P., and T. Lochrane. Capacity Predictions and Capacity Passenger Car Equivalents of Platooning Vehicles on Basic Segments. *Journal of Transportation Engineering, Part A: Systems*, Vol. 144, No. 10, 2018, p. 4018063.
48. Sarvi, M. Heavy Commercial Vehicles-following Behavior and Interactions with Different Vehicle Classes. *Journal of Advanced Transportation*, Vol. 47, No. 6, 2013, pp. 572–580.
49. Eluru, N., C. R. Bhat, and D. A. Hensher. A Mixed Generalized Ordered Response Model for Examining Pedestrian and Bicyclist Injury Severity Level in Traffic Crashes. *Accident Analysis & Prevention*, Vol. 40, No. 3, 2008, pp. 1033–1054.
50. Elvik, R. *The Power Model of the Relationship between Speed and Road Safety: Update and New Analyses*. 2009.
51. Bansal, P., and K. M. Kockelman. Forecasting Americans' Long-Term Adoption of Connected and Autonomous Vehicle Technologies. *Transportation Research Part A: Policy and Practice*, Vol. 95, 2017, pp. 49–63.
52. Martin, J.-L. Relationship between Crash Rate and Hourly Traffic Flow on Interurban Motorways. *Accident Analysis & Prevention*, Vol. 34, No. 5, 2002, pp. 619–629. https://doi.org/10.1016/S0001-4575(01)00061-6.
53. Høye, A. K., and I. S. Hesjevoll. Traffic Volume and Crashes and How Crash and Road Characteristics Affect Their Relationship – A Meta-Analysis. *Accident Analysis & Prevention*, Vol. 145, 2020, p. 105668. https://doi.org/10.1016/j.aap.2020.105668.
54. Roshandel, S., Z. Zheng, and S. Washington. Impact of Real-Time Traffic Characteristics on Freeway Crash Occurrence: Systematic Review and Meta-Analysis. *Accident Analysis & Prevention*, Vol. 79, 2015, pp. 198–211. https://doi.org/10.1016/j.aap.2015.03.013.
55. Heckerman, D., D. Geiger, and D. M. Chickering. Learning Bayesian Networks: The Combination of Knowledge and Statistical Data. *Machine learning*, Vol. 20, 1995, pp. 197–243.
56. Friedman, N., and D. Koller. Being Bayesian about Network Structure. A Bayesian Approach to Structure Discovery in Bayesian Networks. *Machine learning*, Vol. 50, 2003, pp. 95–125.