REVIEW

Connecting tradition with modernity: Safety literature review

Daiquan Xiao[1], Bo Zhang[1], Zexi Chen[1], Xuecai Xu[1*] and Bo Du[2*]

[1] *School of Civil and Hydraulic Engineering Huazhong University of Science and Technology Wuhan 430074, China*
[2] *SMART Infrastructure Facility University of Wollongong, Wollongong, NSW 2522, Australia*
* Corresponding authors, E-mail: xuecai_xu@hust.edu.cn; bdu@uow.edu.au

Abstract

Road safety has long been considered as one of the most important issues. Numerous studies have been conducted to investigate crashes with significant progress, whereas most of the work concentrates on the lifespan period of roadways and safety influencing factors. This paper undertakes a systematic literature review from the crash procedure to identify the state-of-the-art knowledge, advantages and disadvantages of crash risk, crash prediction, crash prevention and safety of connected and autonomous vehicles (CAVs). As a result of this literature review, substantive issues in general, data source and modeling selection are discussed, and the outcome of this study aims to provide the summary of crash knowledge with potential insight into both traditional and emerging aspects, and guide the future research direction in safety.

Keywords: Road safety, Crash risk, Crash prediction, Crash prevention, Connected and autonomous vehicles

Introduction

Since its first emergence, transportation had experienced tremendous changes from tradition to modernity, meanwhile from traditional human-driven vehicles to modern connected and autonomous vehicles (CAVs). Due to the extreme significance, road safety has always been considered as one of the most important and significant topics in transportation engineering through the centuries, with the ultimate objective of reducing injuries and fatalities[1].

During recent decades, road safety evaluation has been explored from various perspectives[2−5], and crash has been widely employed to represent safety level. Currently, two main directions of safety research are the main focus: the lifespan period of roadways, i.e. planning (proactive safety), construction (work zone safety), operation (reactive safety), and management (behavioral and improvement safety); and the safety influencing factors, such as human (pedestrian, bicyclist, motorcyclist and driver), vehicle (motorcycle, car, bus/truck and heavy trucks), roadway (geometric design, classification, intersection type, etc.), environment (weather conditions, lighting, etc.). However, either direction may not concentrate on the true attribute of crash itself, thus the crash procedure still requires investigation, and the corresponding safety evaluation from crash procedure, i.e. crash risk, crash prediction, and crash prevention, which is worth acknowledging the challenges and opportunities in this study.

Crash risk, qualitatively, represents the state of unknown or potential crash within a traffic system, which may be developed into accidents under certain conditions, while quantitatively it denotes the probability of the danger converting into an accident, which is commonly employed to describe the safety situation before accident. Current studies mainly concentrate on crash risk analysis/evaluation and crash risk prediction, which can identify the potential influencing factors, analyze the possible consequences or results derived from, and evaluate the risk degree and impact range under traditional and CAVs conditions.

Crash prediction, in accordance with historic data of crashes, follows certain prediction theories and models, which may investigate the crash variation regulation, infer and estimate the developing trend and possible results. Currently, crash prediction studies include frequency, injury severity or crash rate from traditional statistical methods and econometric models to machine learning and deep learning approaches, which may improve the accuracy of online real-time crash prediction, and benefit the CAVs significantly.

Crash prevention refers to different measures pre-employed to avoid or reduce all types of fatalities or injury crashes. Nowadays, with the application of advanced and emerging technologies, especially the rapid progress of artificial intelligence (AI) and big data, intelligent transportation systems (ITS) have been widely applied to prevent crashes, real time with a variety of modeling and techniques.

With the rapid progress of the Internet of Things (IoTs) and the Internet of Vehicles (IoVs), safety modeling in the current literature is mainly concentrated on real-time crash prediction with artificial neural networks, deep learning or reinforcement learning models to manage the safety proactively. However, many studies related to CAVs are mainly concerned with hardware performance, such as sensor reflection speed or braking speed when crash occurs, which may not help increase the predictive performance. Therefore, in order to shed lights on the crash itself under traditional and modern situations, it is necessary to review previous studies systematically, summarize the current findings comprehensively, find out the gaps and connections, and decide where future direction is oriented.

Figure 1 gives the structure of the paper. The remainder of this paper is organized as follows. Section 2 provides details of

the crash-related literature review, and Section 3 discusses the issues generated from traditional to modern safety modeling and gives future direction. Section 4 reaches conclusions by summarizing the main contributions and findings.

Literature review

In this section, a review of related papers is provided to categorize crashes into crash risk, crash prediction and crash prevention. The literature search employs the core database of Web of Science, and the keywords cover crash risk analysis/ evaluation, crash risk prediction, crash frequency, crash injury severity, real-time crash prediction, crash prevention modeling, and crash prevention measures. In order to find out the existing issues and future gaps, the literature are explained in detail and the strengths and weaknesses of different methods are summarized in Table 1.

Crash risk

After reviewing the literature, we find that there are two main types of crash risk research, crash risk analysis/evaluation, and crash risk prediction. The former concentrates on the past influencing factors of crash risk while the latter focuses on the future possible factors of crash risk.

Crash risk analysis/evaluation

Some studies were conducted from the discrete models for crash risk analysis. Chen et al.[6] analyzed the risk factors that significantly influenced the severity of intersection crashes. Logistic regression was applied and seven risk factors obtained were found to be significantly associated with the severity of intersection crashes, including driver age and gender, speed zone, traffic control type, time of day, crash type, and seat belt usage. Lao et al.[7] established a highway rear-end crash risk estimation model using a generalized nonlinear model (GNM). The analysis concluded that the effect of truck percentage and slope on accident risk was parabolic: they increased crash risks initially, but decreased after the certain thresholds. Yu et al.[8] established disaggregate crash risk analysis models based on loop detector data and historical crash data for urban expressways. Bayesian semi-parametric inference technique was introduced to crash risk analysis to capture unobserved heterogeneity. However, due to the small sample size, weekend rush hour crashes were not considered. Cunto & Ferreira[9]
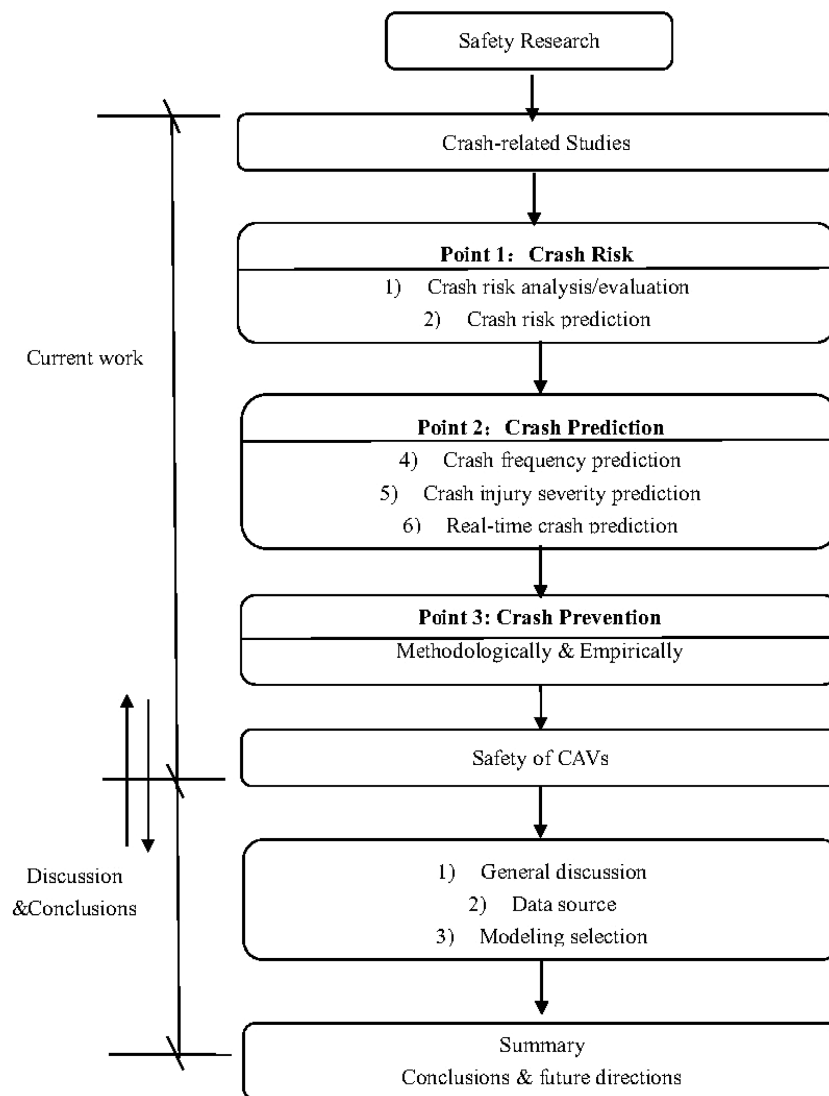

Fig. 1   Flowchart of this review.

Table 1. Summary of safety literature.

Crash procedure		Representative studies	Methods	Strengths and weaknesses
Crash Risk	Crash risk analysis/evaluation	Chen et al. (2012)[6], Lao et al. (2014)[7], Yu et al. (2016)[8], Cunto & Ferreira (2017)[9], Wu et al. (2018)[10], Gu et al. (2019)[11]	Discrete models (logistic regression, generalized nonlinear model, mixed ordered response, random parameter logistic regression)	Significant influencing factors can be clearly revealed while the cause-and-effect relations need to be explained by operators.
		Theofilatos & Yannis (2014)[12], Weng et al. (2014)[13], Weng et al. (2015)[14], Dingus et al. (2016)[15], Papadimitriou et al. (2019)[16], Wang et al. (2021)[17], Adeyemi et al. (2021)[18], Mahajan et al. (2022)[19]	Empirical perspectives (e.g. rear-end collision, drivers merging behavior, naturalistic driving data)	Results can be obtained from empirical testing or experiment, whereas the transferability needs to be confirmed.
		Roshandel et al. (2015)[1], Papadimitriou & Theofilatos (2017)[20]	Meta analysis (e.g. random-effects meta-analysis)	Comprehensive but complicated
	Crash risk prediction	Yu & Abdel-Aty (2013)[23], Yuan & Abdel-Aty (2018)[24], Yasmin et al. (2018)[25], Wang et al. (2019)[26], Guo et al. (2021)[27]	Real-time crash risk prediction (SVM, Bayesian approach, random forest)	Good results can be obtained by combing the machine learning or data mining with traditional methods, but the prediction accuracy needs to be improved.
		Bao et al. (2019)[28], Li et al. (2020)[29], Wang et al. (2021)[30]	Deep neural network (STCL-Net, LSTM-CNN)	The prediction accuracy is better whereas the large data and complicated modeling procedure are required.
Crash prediction	Crash frequency prediction	Qin et al. (2004)[31], Caliendo et al. (2007)[32], Ma et al. (2008)[33], Hou et al. (2022)[34]	Discrete models (ZIP model, negative binomial, multivariate Poisson-lognormal, random parameter logit model)	Significant influencing factors can be clearly revealed while the cause-and-effect relations need to be explained by operators.
		Hossain & Muromachi (2012)[35], Sun & Sun (2015)[36], Dong et al. (2015)[37], Huang et al (2016)[38], Tang et al. (2021)[39]	Bayesian approach (random multinomial logit, spatial model, hierarchical random parameter Tobit model)	The prediction accuracy is improved while the modeling is becoming complicated.
		Dong et al. (2015)[37], Huang et al. (2016)[38], Ambros et al. (2018)[40], Wu & Tsu (2021)[41]	Regional level (SVM with spatial weight, Bayesian spatial model, CNN-GRU)	The prediction accuracy is better while the modeling procedure is complicated.
	Crash injury severity prediction	Delen et al. (2017)[42], Iranitalab & Khattak (2017)[43], Huang et al. (2018)[44], Santos et al. (2022)[45],	Machine learning methods (SVM, NNC, CART, random forest)	The prediction is accuracy is increased whereas the data requirement is large.
		Li et al. (2019)[46], Hou et al. (2022)[34]	Unobserved heterogeneity (mixed logit model, random parameters logit model)	Heterogeneity issue can be addressed while temporal instability is still neglected.
	Real-time crash prediction	Basso et al. (2021)[47], Thapa et al. (2022)[48], Man et al. (2022)[49], Ma et al. (2022)[50], Li & Abdel-Aty (2022)[51], Hu et al. (2022)[52]	Deep neural network (generative adversarial network, TA-LSTM, FC-LSTM, ConvLSTM)	The prediction accuracy is better but the data requirement is improved.
		Ahmed & Abdel-Aty (2011)[53], Basso et al. (2021)[47], Li & Abdel-Aty (2022)[51]	Real-time data (speed data, trajectory fusion data)	Multisource data increase the prediction accuracy but data processing is complicated.
Crash prevention	Modeling perspective	Lee et al. (2003)[54], Mirzaei et al. (2014)[55]	Probabilistic model and logistic regression model	Traditional methods can identify the impact factors clearly but the accuracy needs to be improved.
	Empirical perspective	Ker et al. (2005)[56], El Khoury & Hobeika (2006)[57], Chen & Qin (2019)[58], Yue et al. (2020)[59], Hinnant & Stavrinos (2020)[60], Gidion et al. (2021)[61], Peng & Xu (2021)[62]	Test or simulation	Real scenarios benefit the realization of crash prevention, while the generality needs to be demonstrated.
Safety of CAVs	Crash risk	Jang et al. (2020)[63]	Data from CVs	The results were effective in reducing crash potential, but the transferability needs to be examined.
	Crash prediction	Xu et al. (2019)[64], Sinha et al. (2020)[65]	Road testing or simulation	The prediction accuracy is better, but the result didn't achieve the expected safety benefits.
	Crash prevention	Wang et al. (2020)[66], Wang et al. (2021)[30],	Meta-analysis or surrogate safety measures	The number of crashes could be reduced whereas the transferability still needs to be demonstrated.

investigated factors that influence the severity of motorcycle accidents in the urban streets of Fortaleza. The mixed ordered response models were employed and the results suggested that motorcyclists using helmets reduced their chances by 9% of suffering severe and fatal injuries after the crash. Accidents during the daylight, as well as on weekdays, presented lower risk of resulting in fatal injuries. Wu et al.[10] proposed the crash risk increase indicator to investigate the differences of crash risk between foggy and clear conditions. The binary logistic regression model was employed and the results found that the crash risk was about the increase at ramp vicinities in fog conditions. In the study by Gu et al.[11], a multilevel random parameters logistic regression model was presented to investigate driver's merging behavior in the acceleration lane with unmaned aerial vehicle (UAV) videos. The results showed that the merging speed, driving ability and the merging location affected the crash risk at interchange merging areas.

Some work was performed from the empirical perspective of crash risk. Theofilatos & Yannis[12] summarized the effect of traffic and weather characteristics on road safety. It was found that traffic flow had a non-linear relationship with crash rates, while speed limits had a positive relation with crash occurrence. On the other hand, the effect of precipitation increased crash frequency but didn't have a consistent effect on injury severity, and other weather parameters on safety were not significant. Weng et al.[13] used the deceleration rate to avoid the crash in the vehicle trajectory data to measure the rear-end collision risk under four different vehicle following modes: car-car, car-truck, truck-car and truck-truck in the construction area. The results showed that the car-truck follow mode had the highest risk of rear-end crash, followed by truck-truck, truck-car and car-car. Weng et al.[14] investigated the correlation between the drivers' merging behavior and the rear-end crash risk in work zone merging areas. The time to collision and the deceleration rate were employed to avoid the crash to calculate the rear-end crash risk between the merging vehicle and its adjacent vehicles. It was found that the rear-end crash risk increased when the merging vehicle or the adjacent vehicle was a heavy vehicle. Dingus et al.[15] evaluated risk factors with naturalistic driving data collected from multiple onboard video cameras and sensors. The results revealed that crash causation has shifted significantly in recent years, and distraction is detrimental to driver safety. Papadimitriou et al.[16] summarized the review of crash risk factors related to road infrastructure. Ten areas (alignment features, cross-section characteristics, road surface deficiencies, work zones, junction deficiencies, etc.) were structured and synthesis of results were made on individual risk factors. In view of the shortcomings of the single-dimensional risk source analysis method of crash risk in the past, Wang et al.[17] proposed a multi-dimensional risk source method, which assigned the weight of crash responsibility to risk factors, so as to incorporate crash responsibility into crash risk estimation, and under the combination of multiple risk factors quantify crash risk. The analysis concluded that the superposition effect of risk factors on crash was non-linear, and multi-dimensional risk factors had amplifying effect on the accumulation of crash risks. Adeyemi et al.[18] evaluated the association between the rush hour period and fatal and non-fatal crash injuries. Results of the meta-analysis revealed that the rush-hour period was associated with a 41% increased risk of fatal crash injury in the United States while the morning

rush hour period was related with increased crash injury risk compared to the afternoon rush hour period. Mahajan et al.[19] proposed a method for estimation of rear-end crash risk with a large naturalistic traffic dataset. The results showed that speed-drop was connected with increased crash risk as well as lane changing.

Meta analysis has been popular in recent years. Roshandel et al.[1] undertook a systematic literature review on the relationships between traffic characteristics and crash occurrence. Meta-analysis was conducted and the results showed that three summary estimates (speed variation, speed difference and average volume) had statistically significant negative impacts on crash occurrence. It then outlined the shortcomings and the common issues shared among the selected studies from five aspects, and described where future research should be directed. Papadimitriou & Theofilatos[20] meta-analyzed the crash-risk factors in freeway entrance and exit areas. A random-effects meta-analysis was conducted on the effect of ramp length on crash severity, and a nonsignificant overall effect was observed. And random-effects meta-analyses regarding deceleration lane length suggested a nonsignificant effect on road safety (both on frequency and severity) at a 95% level of confidence. It was found there was no indication of strong publication bias in any of the meta-analyses performed.

From the perspective of drivers, as for older drivers, Asbridge et al.[21] focused on the impact of restricted driver's licenses on crash risk. The results found that restricted driver licensing may be effective in reducing crash risk and decreasing traffic violations for older drivers. As for young drivers, Banz et al.[22] performed a systematic review of databases on crash-risk behaviors. Driving impairment mainly focused on drowsy/ fatigued driving or alcohol-impaired driving while distraction driving primarily concentrated on cognitive load, auditory and visual distractors. The findings showed that coupling neuroscience with driving simulation was feasible in examining driving behavior of contributing factors for fatal motor vehicle crashes.

Crash risk prediction

Some methods or approaches have been applied in real-time crash risk prediction under traditional conditions. Yu & Abdel-Aty[23] employed supported vector machine (SVM) to evaluate real-time crash risk. Model comparisons' results showed that the SVM model with RBF kernel provided the best goodness-of-fit. While the SVM models with linear kernel had similar results as the logistic regression models. Based on 23 signalized intersections in central Florida (USA), Yuan & Abdel-Aty[24] divided crashes into intersection crashes and intersection entrance crashes, and developed Bayesian conditional logistic models for the two types of crashes, respectively. It was found that the significant influencing factors differed in the real-time crash prediction of intersection crashes and intersection entrance crashes. Yasmin et al.[25] developed a joint reactive and proactive crash modeling framework by coupling the monthly crash risk and real-time crash risk in a unified econometric framework for a microscopic analysis unit. Among them, the monthly crash risk was evaluated by using static road attributes to establish a binary logit model, and the real-time crash risk is evaluated by using different real-time traffic attributes to establish multiple logit models. However, the traffic characteristics of the nearest downstream or upstream road segment were not considered in the real-time crash risk

prediction model. Wang et al.[26] established Bayesian logistic regression model and SVM model respectively by considering the geometric, socio-demographic, and trip generation prediction data to reflect drivers' characteristics and behaviors when analyzing the real-time crash risk of expressway ramps. The results showed that models taking into sociodemographic and trip generation prediction data outperformed models without considering these factors. Guo et al.[27] developed a crash risk model based on risky driving behavior and traffic flow. Random forest was considered to select variables with strong impacts on crashes and synthetic minority oversampling technique (SMOTE) was used to adjust the imbalanced dataset so that a logistic regression model was developed for predicting crash risk. The results indicated that the crash risk prediction model had high accuracy of 84.48% of the crashes.

With the introduction of deep neural network, crash risk prediction has been transmitted from tradition to CAVs era. Bao et al.[28] proposed a spatiotemporal convolutional long short-term memory network (STCL-Net) for predicting citywide short-term crash risk with multi-source data. It was found that the prediction performance decreased as the spatiotemporal resolution of prediction task increased. Li et al.[29] proposed a real-time crash risk prediction model with a long short-term memory convolutional neural network (LSTM-CNN), in which LSTM captured the long-term dependency while CNN extracted the time-in-variant features. Wang et al.[30] provided a comprehensive and systematic review of surrogate safety measures (SSM) under CAV environment. Simulation was considered as the most viable solution to evaluate CAV risk modeling, but road test was still the main approach.

Crash prediction

Crash frequency prediction

Discrete models have been widely applied in frequency prediction. Qin et al.[31] presented zero-inflated-Poisson (ZIP) model to predict crash counts for different types of crashes by considering the influencing factors, e.g. annual average daily traffic (AADT), segment length, speed limit and roadway width. It was found that the relationship between crashes and AADT was non-linear and varied by crash types. Caliendo et al.[32] predicted the crash frequency with Poisson, Negative Binomial and Negative Multinomial regression models for multi-lane roads in Italy. The results showed that for curves, length, curvature and AADT were significant while for tangents length, AADT and junctions were significant. Ma et al.[33] proposed a multivariate Poisson-lognormal (MVPLN) model to simultaneously model crash count predictions for different injury severity. This overcame the drawbacks of using univariate prediction models that ignored the effects of unobserved factors between crash rate of different injury severities on a particular road segment. Hou et al.[34] simulated four random parameter models and random parameter logit model with heterogeneity in the means and variances was found to provide the best accuracy. The temporal instability was evaluated and pairwise comparison provided potential insights into temporal variability.

Bayesian approach has been employed in crash prediction. Hossain & Muromachi[35] employed random multinomial logit model to identify the predictors and then Bayesian belief net was applied to establish the real-time crash prediction model. The results reflected that at an average threshold value the

accuracy reached 66% of the future crashes. Sun & Sun[36] proposed a dynamic Bayesian network model of time sequence traffic data to find out the relationship between crash occurrence and dynamic speed data. It was found that the proposed model with speed condition data and nine traffic state combinations can achieve 76.5% crash prediction accuracy. Dong et al.[37] proposed support vector machine (SVM) to assess multi-dimensional spatial data in crash prediction at the level of traffic analysis zones. Bayesian spatial model with conditional autoregressive prior was compared and the results revealed that SVM models outperformed the non-spatial model and addressed complex spatial data in regional crash prediction modeling. Huang et al.[38] developed a macro-level Bayesian spatial model with conditional autoregressive prior and a micro-level Bayesian spatial joint model to predict zonal crashes. It was found that the micro-level Bayesian spatial model revealed better performance, while the macro-level crash analysis required less detailed data. Tang et al.[39] proposed a conditional quantile-based Bayesian hierarchical random parameter Tobit model investigate the regional varying effects of road-related factors on crash rate at different quantiles of the crash rate distribution. This was used to explore crash rate in areas with extremely high crash rate.

Some scholars have established crash prediction models for regional crash rate. Dong et al.[37] considered the spatial correlation between adjacent regions when establishing a regional crash prediction model, and established a SVM model with spatial weight characteristics. Through comparison, it was found that the model was better than the non-spatial model in terms of model fitting and prediction performance. Huang et al.[38] compared the predictive performance of a macro method and micro method for regional crash prediction models. The macro method employed a macro-Bayesian space model and the micro-method employed the summation of expected crashes across all road entities within a sub-area to estimate the frequency of sub-area crashes, where each subregion adopted a micro-Bayesian spatial model. The results showed that the micro-level model has better overall fitting and prediction performance, and can better understand the micro-factors closely related to the crash, which was easy to obtain more direct countermeasures. The advantage of crash analysis at the macro level is that it requires less detailed data and is an essential means of incorporating traffic safety considerations into long-term transportation planning. Ambros et al.[40] summarized the crash prediction models (CPMs) from state-of-the-art and state-of-the-practice, specifically including data collection, road network segmentation, variable selection, functional form, validation models and how to use them in practice for current applications to help practitioners rationally use crash prediction models in the context of lag theory. Wu & Tsu[41] developed a fusion deep learning approach combining a convolution neural network (CNN) and gated recurrent units (GRU) to predict at-fault crash driver frequency with city-level traffic enforcement predictors. The CNN-GRU prediction accuracy outperformed other methods and the findings can facilitate the development of traffic safety measures.

Crash injury severity prediction

Machine learning and related methods have been applied in injury severity prediction. Delen et al.[42] identified significant influencing factors affecting injury severity through SVM and applied sensitivity analysis to the predictive model, determi-

ning the relative importance of these factors. The results showed that the use of seat belts and manner of collision were the primary factors affecting the severity of the crash, but the study only made a dichotomous classification of injury severity. Iranitalab & Khattak[43] compared multinomial logit (MNL), nearest neighbor classification (NNC), SVM and random forests (RF) in predicting crash severity, and investigated the effects of data clustering methods on the performance of crash severity prediction models. The results showed that NNC had the best performance in overall and more severe crashes, and data clustering didn't affect the prediction results of SVM. Huang et al.[44] used a classification and regression tree (CART) model to examine the interactive effects of various influencing factors on injury severity in mountain highway crashes. It was found that a combination of the following factors had a significant impact on the occurrence of serious crashes: coach drivers involved in improper lane changing and other improper actions, drivers involved in speeding during afternoon or evening, drivers involved in speeding along large curves and straight segments during morning, noon or night, and drivers experiencing fatigue while passing along the downgrade. However, in this literature, injury severity measures were only divided into two categories due to data limitations. Santos et al.[45] summarized the crash injury severity modeling methods with 20 different statistical or machine learning techniques. Random forest showed the best performance, followed by support vector machine and decision tree. Casualty issues, unobserved heterogeneity and temporal instability need to be considered.

In order to capture the unobserved heterogeneity in the influencing factors of single-vehicle injury severity, Li et al.[46] divided the entire dataset into seven sub-data sets by latent class analysis, and then built a mixed logit model on each sub-data set. This study only assumed the widely used normal distribution as the assumption of randomly distributed variables in the mixed logit model, which may not be realistic. Hou et al.[34] compared the performance of different random parameters logit models for injury severity prediction. The comparison found that the random parameters logit model with heterogeneity in the means and variances outperformed other models in terms of predictive performance.

Real-time crash prediction

Deep neural network has provided alternatives for real-time crash prediction. Based on convolutional neural networks, Basso et al.[47] built an accident prediction model. It was found that deep convolutional generative adversarial networks technique with random undersampling performed better for real-time crash prediction using vehicle-by-vehicle data. Thapa et al.[48] developed a duration-based, real-time crash prediction model by considering time-varying covariates, and equal time intervals of crashes were modeled as alternative with multinomial logit models with large data. Different datasets were compared and resulted in reasonable accuracy. In order to improve the spatiotemporal transferability of real-time crash prediction model, Man et al.[49] developed Deep Neural Network (DNN) as a baseline model with imbalanced dataset and incorporated Generative Adversarial Network (GAN) to generate synthetic crash data. The results revealed that the predictability of the transferred models outperformed the existing ones with 95% accuracy. Ma et al.[50] presented am improved genetic programming (GP) for real-time crash prediction. Logistic regression and backward-propagation

neural network were considered as baseline methods to examine the interpretability and accuracy of GP, and the results displayed that GP prediction model can solve the trade-off between interpretability and accuracy. Li & Abdel-Aty[51] developed a deep learning model to predict real-time crash likelihood with trajectory data. A temporal attention-based long short-term memory (TA-LSTM) was cooperated to capture temporal correlation between time-series data and a convolutional neural network (CNN) were combined to predict the crash likelihood. The findings showed that the proposed model performed well and trajectory fusion improved the prediction accuracy. Hu et al.[52] proposed to improve the defect of fully connected long short-term memory (FC-LSTM) network model of ignoring the spatial features of crash by adopting Convolutional Long Short-Term Memory (ConvLSTM) network, which can effectively capture the spatiotemporal characteristics of crashes within the road network. By comparison, it was found that ConvLSTM has better accuracy, lower loss value and higher computational efficiency.

The data used by real-time crash prediction models was also changing. Ahmed & Abdel-Aty[53] used real-time speed data collected by a tag reader on a toll road called an automatic vehicle identification (AVI) system to build a RF model for real-time crash prediction, which showed a 70% prediction accuracy rate. Basso et al.[47] proposed a new image-inspired data architecture for most past crash real-time prediction models using data aggregated every five or ten minutes, which used random undersampling algorithm to rebalance the data and established the Deep Convolutional Generative Adversarial Networks model. It was found that the model outperformed other traditional forecasting methods in terms of AUC and sensitivity values to a range of false positives. Li & Abdel-Aty[51] applied trajectory fusion data to real-time crash prediction. The features extracted from the data were used to predict the real-time crash probability, and the temporal attention mechanism was adopted to improve the prediction accuracy of the deep learning crash probability prediction model.

Crash prevention

Some works were performed from modeling perspective to prevent the crashes. Lee et al.[54] predicted the likelihood of crashes on freeways on the basis of traffic flow conditions, and suggested the risk-based evaluation framework for real-time traffic control. A probabilistic model was adopted, and the test showed that this model overcame the limitations of many existing static crash prediction models. Crash potential estimated by this model was sensitive to short-term variation of traffic flow. Mirzaei et al.[55] evaluated the relation between drivers' knowledge, attitude, and practice (KAP) regarding traffic regulations, and their deterministic effect on road traffic crashes (RTCs). After a sampling survey, logistic regression was used to analyze the questionnaire results and evaluated the relationship between RTCs and KAP variables. The results showed that safer attitude, and safer practice were associated with a decreased number of RTC, but only attitude was significantly concerned with a decrease of RTC.

A large amount of prevention measures have been conducted empirically. Ker et al.[56] investigated the effectiveness of post-license driver education for preventing road traffic crashes. Through a systematic review and meta-analyses of random controlled trials, the results provided no evidence that post-license driver education was effective in preventing road

injuries or crashes. El Khoury & Hobeika[57] developed a new simulation in vertical curve on a two-lane two-way highway. This system detected and warned the violating vehicle in real time, and also warned the opposite vehicles in the same lane as the violating vehicles were being warned. The results showed that the system would reduce the possible crashes from the base case by a mean of 26.3% in the eastbound and 33.3% in the westbound. Chen & Qin[58] proposed a crash prediction and prevention method based on simulated traffic data to detect imminent crash risk and help recommend traffic control strategies (TCS) to prevent crashes. The proposed method was tested in a case study with variable speed limit (VSL) strategies for demonstration, and results showed that the method could effectively detect crash-prone conditions and evaluate the safety and mobility impacts of various TCS alternatives before their deployment. Yue et al.[59] conducted an in-depth investigation of pedestrian crashes and identified crash causation patterns and its implications for pedestrian crash prevention. The results showed that the pattern concerned with distracted driving and unexpected change of pedestrian trajectory accounted for a large number of the crashes. and the findings presented the implications for roadway facility design as well as roadway safety education and pedestrian prevention system development. Hinnant & Stavrinos[60] evaluated how rewards favoring safe choices affected decision making while teens played a driving game with and without peer observation and whether rewards were more effective for adolescents with the riskiest driving styles. It was found that rewards for safe driving can be an effective mechanism for reducing motor vehicle crashes, especially for the most at-risk drivers, if they can be made appetizing to adolescents. Gidion et al.[61] analyzed a sample of injured motorcycle riders from the German In-depth Accident Study (GIDAS) to identify priorities for injury assessment and prevention. The results indicated that the priorities for rider safety interventions were: fracture of the rib cage, femur fracture, tibia fracture, etc., which needed to be considered before using and developing procedures and test tools. Peng & Xu[62] developed a combined VSL and lane change guidance (LCG) controller to prevent secondary crashes (SCs). The combined controller was based on distributed deep reinforcement learning (RL). Simulation experiments indicated that the developed combined controller achieved higher performance in general than any single sub-controller, and was able to accurately capture the spatial and temporal impact areas caused by prior crashes and generate proper interventions of traffic flow proactively.

Safety of CAVs

As for the crash risk, Jang et al.[63] analyzed crash risks according to the data obtained from coonected vehicles (CVs) equipped with in-vehicle forward collision warning systems, and estimated the safety benefits of the forward hazardous situation warning (FHSW) information presented by a C-ITS pre-deployment project for Korean freeways. The results suggested that providing FHSW based on V2X in a CV environment was effective in reducing the crash potential.

As for crash prediction, Xu et al.[64] investigated the characteristics and patterns of CAVs involved crashes. The descriptive statistics analysis was employed to investigate the characteristics of CAVs involved crashes and a bootstrap based binary logistic regressions were then developed to investigate the factors contributing to the collision type and severity. The

results suggested that the CAV driving mode, collision location, etc., were the main factors contributing to the severity level of CAV involved crashes. The CAV driving mode, CAV stopped or not, CAV turning or not, etc, were the factors affecting the collision type of CAV involved crashes. Sinha et al.[65] investigated the effect of the introduction of CAVs on both injury severity and frequency through a microsimulation modelling exercise. The results indicated that the introduction of CAVs did not achieve the expected decrease in crash severity and rates involving manual vehicles, despite the network performance has been improved. And the safety benefits of CAVs were not proportional to CAV penetration, full-scale benefits of CAVs can only be achieved at 100% CAV penetration.

From the prevention perspective, Wang et al.[66] evaluated the safety effectiveness of nine common and important CV or AV technologies, and tested the safety effectiveness of these technologies for six countries. Meta-analysis was conducted and the results displayed that if all of technologies were implemented in the six countries, the average number of crashes could be reduced by 3.40 million. Wang et al.[17] made a comprehensive and critical review of SSM (Surrogate Safety Measures) and discussed their various applications, especially in CAV related safety studies. It was found that when modeling safety in mixed autonomy traffic or fully automated traffic, whether the SSM validated in traditional traffic environments can still be applicable was a critical issue, and the transferability of SSM, using real-world automated driving data for deriving SSM, would be interesting areas for future research.

Discussion

During recent decades, a number of researchers have made considerable progress in investigating roadway safety, especially the relationship between crashes and the influencing factors. Due to the big data and emerging AI technologies, data-driven crash related studies have been the common understanding nowadays. Although much progress has been made in this area, challenging issues are still available from traditional to modernity. Consequently, the current state of crash related studies is valuable so as to identify the future orientation.

General discussion

As is known, the causation of crashes is a complicated and instant procedure, which may involve the interactions of human beings (drivers, motorcyclists, cyclists and pedestrians), vehicles (motorized and non-motorized), roadways (classification, geometric design and roadside facilities), and environmental factors (lighting or weather, or facilities). Generally speaking, during model processes, the more influencing factors included, the more accurate the crash estimation/prediction is. However, there are some issues when selecting the variables to include. First, the co-linearity between influencing variables should be examined before the final model is determined. When the co-linearity is involved, the model may incorrectly reflect the actual relation, which may lead to modelling mistakes. There are some alternatives to be considered to remove the co-linearity. For example, the more significant one is selected while the other is eliminated between two influencing variables, or some interaction form, plus/subtraction, multiplying/dividing, even Log, can be chosen to address the co-linearity, which generates the second point, the

interactions between variables. Crashes may happen due to more than one influencing factor, and the interactions among human beings, vehicles, roadways and environment accounts for over 30% of crashes[1] , thus the crash prediction with only one type of factor may omit some important information and may cause error rates or false positives.

More importantly, two model specification issues are often discussed during modeling. On one hand, when data are collected, some important factors may be unobserved or omitted, thus the heterogeneity issue occurs, so the specification results of crashes are probably biased or the model assessment may be incorrectly estimated. On the other hand, there may exist intrinsic relations between crashes and impact factors (e.g. crash rate vs travel speed)[67], and vice versa, which may generate endogeneity issue. Similarly, without taking into account of the endogenous variables, the model specification may be biased or the resulting impact may be postulated.

Therefore, because of these reasons above, the performance of the current crash analysis/evaluation and prediction models are less accurate, which may need comprehensive and diverse datasets to increase the preciseness and consistency.

Data source

Traditionally, the crash data were collected by official transportation departments, specifically from police reports to reflect the time, location and related characteristics of the crash. However, due to different reasons, not all the crashes were documented in the police reports since some of them were not reported to the police, so the data may not cover all the cases, thus the modeling accuracy may be biased. Consequently, the cause-and-effect relationship may not be precisely derived from the partial datasets, hence more advanced data collection technologies have been applied to improve the data quality.

Currently, video surveillance has been considered as the most direct and precise method, which can not only 'see' the crash occurrence through the video footage, but render image processing techniques to extract, identify and track the trajectories of vehicles so that the crash can be predicted and detected. For instance, YOLO (You Only Look Once) series can be used to detect the vehicles from the videos while SORT (Simple Online and Real-time Tracking) algorithms can be employed to track the vehicle trajectories so that the crash can be forecasted in advance, which may help improve the data accuracy. Another merit of video cameras is to validate the information from the police report through crosschecking, and more neglected or unreported crashes can be captured or retrieved[1].

One of the widely used devices of data collection is unmanned aerial vehicle or drones, which has been paid more attention by researchers due to direct, cheap and convenient advantages. Similar to video surveillance, drones can be adopted to sense the traffic scenarios, detect the vehicles with advanced techniques and pre-estimate the moving conditions so that crashes can be predicted and managed in advance. On the other hand, the drones can be manipulated for certain area with aerial photographs, and the statistics of traffic flow can be obtained so that the traffic conditions can be analyzed and congestion reasons can be deduced from continuous monitoring within certain periods, which may provide a foundation for real-time dispersion of traffic flow.

The emerging technique around traffic parameters is real-time online web crawler based on Python, which is one type of automatic data collection methods. Through this crawler technique, the traffic variables (e.g. volume, speed and density) can be collected directly every 5 or 10 min, which is an empirically superior option, compared to the conventional loop detectors for traffic variables. Furthermore, for some specific segments within certain periods the spatial and temporal features can be obtained from such data, which may benefit the vehicle trajectories tracking, crash detection and prediction. This method belongs to smart transportation, which is convenient and efficient, satisfying the accurate requirements of real-time traffic conditions, and worthy of promotion.

As for CAVs, a variety of sensors embedded in the vehicles can detect all the vehicles and objects around, and make the decisions as soon as possible if something abnormal is about to happen. Identical to the video or image processing approach, the sensors can detect, identify and track the moving objects or images, and then artificial intelligence algorithms (e.g. deep learning, reinforcement learning) are employed to process them immediately. Meanwhile, the CAVs need to communicate with other vehicles (V2V), infrastructures (V2I), and roadside facilities and devices (V2X) so that the vehicles-roadway synchronization and real-time traffic conditions can be realized within seconds through the cloud and big data, in this way the crash prediction tends to be more accurate so as to avoid the conflict in advance. Although a large number of high-tech corporations and motor companies are investing in huge finances to develop the CAVs, the testing mileage has been increasing day by day, so far no company can guarantee that their CAVs are 100% safe since crashes continue to occur. Meanwhile, as stated by Li et al.[68], accompanied with CAVs, there are many issues (e.g. ethics, reliability, law and enforcement) to be dealt with, but CAVs are the future transport modes, and will be realized with the progress of science and technologies.

Modeling selection

After reviewing the literature as mentioned above, we generally categorize the models into three types: statistical and econometric models, machine learning and AI algorithms, and empirical experiments.

Conventionally, statistical and econometric models are widely employed by most studies of crashes, and the main reason lies in that these models can reflect certain principles about the crash analysis or estimation with some reasonable assumptions, and some results may reveal certain generality and transferability. However, with the increasing requirement of massive data, the conventional methods can't meet the demand of big data, thus machine learning and AI algorithms reveal strong potentiality for nonlinear, dynamic, real time and complex situations. Among them, deep neural networks has been widely applied in crash analysis, estimation and prediction, and convolutional neural network, LSTM, and hybrid models have been demonstrated by various studies[49,69−70].

Another critical approach of modeling is empirical experiments, i.e. through actual testing or real experiments, the safety level can be evaluated or predicted, especially for the CAVs. Currently, most of the CAVs are still testing the software and hardware, and with mileages of roadway testing increasing, different types of scenarios have been provided, and a variety of the risk evaluation schemes have been training and learning.

Finally, how to select the modeling depends on the problem description, dataset and objectives about crashes: if the problem belongs to the traditional statistical issue, econometric modeling may be a better option, while the massive data may turn to machine learning or AI algorithms, and if the modeling needs to be established through actual testing, empirical experiment and simulation may be the alternative.

Conclusions

This paper presents a literature review of safety from traditional to the CAVs era, focusing on the crash procedure with crash risk, crash prediction, crash prevention and safety of CAVs. Then substantive issues in general discussion, data source, and modeling selection are discussed, and the outcomes of this work tend to provide the summary of crash knowledge in the traditional aspect and emerging aspect, and guide the future direction in safety.

Although safety evaluation has been acknowledged from various perspectives, there is still interest in exploring crash procedures. It can be found from the literature review that:

1) Crash risk analysis/evaluation is mainly conducted with discrete models, empirical and meta analysis, while crash risk prediction relies on machine learning and AI algorithms.

2) As for crash frequency prediction, discrete models, Bayesian approach and machine learning methods have been employed whereas machine learning methods in crash injury severity prediction play an important role and real-time prediction relies on the deep neural network and datasets.

3) Crash prevention emphasizes modeling and countermeasures.

4) Safety of CAVs is mainly counting on the testing and simulation right now.

Furthermore, the discussion section reaches the following points:

1) Co-linearity and interactions between influencing factors may lead to errors during modeling, and two model specification issues heterogeneity and endogeneity may cause biased results, so these problems should be emphasized during crash modeling;

2) Video surveillance is a significant data source, not only for traditional data collection, but for advanced drones, web crawlers, and even CAVs.

3) Modeling selection depends on the problem description, but machine learning and AI algorithms may be the better option for crashes currently and in the future, while testing and simulation are suitable for CAVs in the current state.

By summarizing the status of current studies of safety, some guidance and recommendations are proposed for future direction:

1) For traditional crash-related studies, the estimation or prediction accuracy can't meet the requirement of complex modeling, so more advanced machine learning methods or AI algorithms (e.g. edge computing, deep neural network) can be integrated into the econometric models in order to satisfy the big data requirements and estimation or prediction accuracy;

2) As for CAVs, road testing or simulation is the main approach currently to demonstrate the safety of CAVs, while autonomous driving (AD) and vehicle-infrastructure cooperated autonomous driving (VICAD) may provide alternatives. AD safety is the critical reason of influencing the commercialization, and cooperation sensing, decision-making and control of VICAD can improve the AD safety significantly, which may boost the rapid development of CAVs.

3) As for researchers who are interested in safety, the first thing to do is to find out whether the safety problem belongs to traditional or emerging issue, and then determine which methods to conduct the research as listed above.

Due to the limitation of articles reviewed, some issues of crashes may be neglected, which doesn't mean that they are not important, but for the aspects of crashes mentioned in this study they are not highly related. If possible, crash procedure may be extended to a broader area in the future to reflect the safety comprehensively and systematically.

Acknowledgments

Conflict of interest

Xuecai Xu and Bo Du are the Editorial Board members of the Journal *Digital Transportation and Safety*. They were blinded from reviewing or making decisions on the manuscript. The article was subject to the journal's standard procedures, with peer-review handled independently of these Editorial Board members and their research groups.

Dates

REFERENCES

1. Roshandel S, Zheng Z, Washington S. 2015. Impact of real-time traffic characteristics on freeway crash occurrence: systematic review and meta-analysis. *Accident Analysis & Prevention* 79:198−211

2. Savolainen PT, Mannering FL, Lord D, Quddus MA. 2011. The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. *Accident Analysis & Prevention* 43(5):1666−76

3. Mannering FL, Bhat CR. 2014. Analytic methods in accident research: Methodological frontier and future directions. *Analytic Methods in Accident Research* 1:1−22

4. Mannering FL, Shankar V, Bhat CR. 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. *Analytic Methods in Accident Research* 11:1−16

5. Mannering FL. 2018. Temporal instability and the analysis of highway accident data. *Analytic Methods in Accident Research* 17:1−13

6. Chen H, Cao L, Logan DB. 2012. Analysis of risk factors affecting the severity of intersection crashes by logistic regression. *Traffic Injury Prevention* 13(3):300−7

7. Lao Y, Zhang, G, Wang Y, Milton J. 2014. Generalized nonlinear models for rear-end crash risk analysis. *Accident Analysis & Prevention* 62:9−16

8. Yu R, Wang X, Yang K, Abdel-Aty, M. 2016. Crash risk analysis for Shanghai urban expressways: a Bayesian semi-parametric modeling approach. *Accident Analysis & Prevention* 95:495−502

9. Cunto FJC, Ferreira S. 2017. An analysis of the injury severity of motorcycle crashes in Brazil using mixed ordered response models. *Journal of Transportation Safety & Security* 9:33−46

10. Wu Y, Abdel-Aty M, Lee J. 2018. Crash risk analysis during fog conditions using real-time traffic data. *Accident Analysis & Prevention* 114:4−11

11. Gu X, Abdel-Aty M, Xiang Q, Cai Q, Yuan J. 2019. Utilizing UAV video data for in-depth analysis of drivers' crash risk at interchange merging areas. *Accident Analysis & Prevention* 123:159−69

12. Theofilatos A, Yannis G. 2014. A review of the effect of traffic and weather characteristics on road safety. *Accident Analysis & Prevention* 72:244−56

13. Weng J, Meng Q, Yan X. 2014. Analysis of work zone rear-end crash risk for different vehicle-following patterns. *Accident Analysis & Prevention* 72:449−57

14. Weng J, Xue S, Yang Y, Yan X, Qu X. 2015. In-depth analysis of drivers' merging behavior and rear-end crash risks in work zone merging areas. *Accident Analysis & Prevention* 77:51−61

15. Dingus TA, Guo F, Lee S, Antin JF, Perez M, et al. 2016. Driver crash risk factors and prevalence evaluation using naturalistic driving data. *PNAS,* 113(10):2636−41

16. Papadimitriou E, Filtness A, Theofilatos A, Ziakopoulos A, Quigley C, et al. 2019. Review and ranking of crash risk factors related to the road infrastructure. *Accident Analysis & Prevention* 125:85−97

17. Wang X, Qu Z, Song X, Bai Q, Pan Z, et al. 2021. Incorporating accident liability into crash risk analysis: A multidimensional risk source approach. *Accident Analysis & Prevention* 153:106035

18. Adeyemi OJ, Arif AA, Paul R. 2021. Exploring the relationship of rush hour period and fatal and non-fatal crash injuries in the US: a systematic review and meta-analysis. *Accident Analysis & Prevention* 163:106462

19. Mahajan V, Katrakazas C, Antoniou C. 2022. Crash risk estimation due to lane changing: A data-driven approach using naturalistic data. *IEEE Transactions on Intelligent Transportation Systems* 23(4):3756−65

20. Papadimitriou E, Theofilatos A. 2017. Meta-analysis of crash-risk factors in freeway entrance and exit areas. *Journal of Transportation Engineering, Part A: Systems* 143(10):04017050

21. Asbridge M, Desapriya E, Ogilvie R, Cartwright J, Mehrnoush V, et al. 2017. The impact of restricted driver's licenses on crash risk for older drivers: a systematic review. *Transportation Research Part A: Policy and Practice* 97:137−45

22. Banz BC, Hersey D, Vaca FE. 2021. Coupling neuroscience and driving simulation: A systematic review of studies on crash-risk behaviors in young drivers. *Traffic Injury Prevention* 22(1):90−95

23. Yu R, Abdel-Aty M. 2013. Utilizing support vector machine in real-time crash risk evaluation. *Accident Analysis & Prevention* 51:252−59

24. Yuan J, Abdel-Aty M. 2018. Approach-level real-time crash risk analysis for signalized intersections. *Accident Analysis & Prevention* 119:274−89

25. Yasmin S, Eluru N, Wang L, Abdel-Aty MA. 2018. A joint framework for static and real-time crash risk analysis. *Analytic Methods in Accident Research* 18:45−66

26. Wang L, Abdel-Aty M, Lee J, Shi Q. 2019. Analysis of real-time crash risk for expressway ramps using traffic, geometric, trip generation, and socio-demographic predictors. *Accident Analysis & Prevention* 122:378−84

27. Guo M, Zhao X, Yao Y, Yan P, Su Y, et al. 2021. A study of freeway crash risk prediction and interpretation based on risky driving behavior and traffic flow data. *Accident Analysis & Prevention* 160:106328

28. Bao J, Liu P, Ukkusuri SV. 2019. A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source data. *Accident Analysis & Prevention* 122:239−54

29. Li P, Abdel-Aty M, Yuan J. 2020. Real-time crash risk prediction on arterials based on LSTM-CNN. *Accident Analysis & Prevention* 135:105371

30. Wang C, Xie Y, Huang H, Liu P. 2021. A review of surrogate safety measures and their applications in connected and automated vehicles safety modeling. *Accident Analysis & Prevention* 157:106157

31. Qin X, Ivan JN, Ravishanker N. 2004. Selecting exposure measures in crash rate prediction for two-lane highway segments. *Accident Analysis & Prevention* 36(2):183−91

32. Caliendo C, Guida M, Parisi A. 2007. A crash-prediction model for multilane roads. *Accident Analysis & Prevention* 39(4):657−70

33. Ma J, Kockelman KM, Damien P. 2008. A multivariate Poisson-lognormal regression model for prediction of crash counts by severity using Bayesian methods. *Accident Analysis & Prevention* 40(3):964−75

34. Hou Q, Huo X, Leng J, Mannering F. 2022. A note on out-of-sample prediction, marginal effects computations, and temporal testing with random parameters crash-injury severity models. *Analytic Methods in Accident Research* 33:100191

35. Hossain M, Muromachi Y. 2012. A Bayesian network based framework for real-time crash prediction on the basic freeway segments of urban expressways. *Accident Analysis & Prevention* 45:373−81

36. Sun J, Sun J. 2015. A dynamic Bayesian network model for real-time crash prediction using traffic speed conditions data. *Transportation Research Part C: Emerging Technologies* 54:176−86

37. Dong N, Huang H, Zheng L. 2015. Support vector machine in crash prediction at the level of traffic analysis zones: assessing the spatial proximity effects. *Accident Analysis & Prevention* 82:192−98

38. Huang H, Song B, Xu P, Zeng Q, Lee J, et al. 2016. Macro and micro models for zonal crash prediction with application in hot zones identification. *Journal of Transport Geography* 54:248−56

39. Tang J, Yin W, Han C, Liu X, Huang H. 2021. A random parameters regional quantile analysis for the varying effect of road-level risk factors on crash rates. *Analytic Methods in Accident Research* 29:100153

40. Ambros J, Jurewicz C, Turner S, Kieć M. 2018. An international review of challenges and opportunities in development and use of crash prediction models. *European Transport Research Review* 10:35

41. Wu Y, Hsu TP. 2021. Mid-term prediction of at-fault crash driver frequency using fusion deep learning with city-level traffic violation data. *Accident Analysis & Prevention* 150:105910

42. Delen D, Tomak L, Topuz K, Eryarsoy E. 2017. Investigating injury severity risk factors in automobile crashes with predictive analytics and sensitivity analysis methods. *Journal of Transport & Health* 4:118−31

43. Iranitalab A, Khattak A. 2017. Comparison of four statistical and machine learning methods for crash severity prediction. *Accident Analysis and Prevention* 108:27−36

44. Huang H, Peng Y, Wang J, Luo Q, Li X. 2018. Interactive risk analysis on crash injury severity at a mountainous freeway with tunnel groups in China. *Accident Analysis and Prevention* 111:56−62

45. Santos K, Dias JP, Amado C. 2022. A literature review of machine learning algorithms for crash injury severity prediction. *Journal of Safety Research* 80:254−69

46. Li Z, Wu Q, Ci Y, Chen C, Chen X, et al. 2019. Using latent class analysis and mixed logit model to explore risk factors on driver injury severity in single-vehicle crashes. *Accident Analysis and Prevention* 129:230−40

47. Basso F, Pezoa R, Varas M, Villalobos M. 2021. A deep learning approach for real-time crash prediction using vehicle-by-vehicle data. *Accident Analysis and Prevention* 162:106409

48. Thapa D, Paleti R, Mishra S. 2022. Overcoming challenges in crash prediction modeling using discretized duration approach: An investigation of sampling approaches. *Accident Analysis and Prevention* 169:106639

49. Man CK, Quddus M, Theofilatos A. 2022. Transfer learning for spatio-temporal transferability of real-time crash prediction models. *Accident Analysis and Prevention* 165:106511

50. Ma X, Lu J, Liu X, Qu W. 2022. A genetic programming approach for real-time crash prediction to solve trade-off between interpretability and accuracy. *Journal of Transportation Safety & Security*

51. Li P, Abdel-Aty M. 2022. Real-time crash likelihood prediction using temporal attention–based deep learning and trajectory fusion. *Journal of Transportation Engineering, Part A: Systems* 148(7):04022043

52. Hu Z, Zhou J, Huang K, Zhang E. 2022. A data-driven approach for traffic crash prediction: A case study in Ningbo, China. *International Journal of Intelligent Transportation Systems Research* 20(2):508−18

53. Ahmed MM, Abdel-Aty MA. 2011. The viability of using automatic vehicle identification data for real-time crash prediction. *IEEE Transactions on Intelligent Transportation Systems* 13(2):459−68

54. Lee C, Hellinga B, Saccomanno F. 2003. Proactive freeway crash prevention using real-time traffic control. *Canadian Journal of Civil Engineering* 30(6):1034−41

55. Mirzaei R, Hafezi-Nejad N, Sadegh Sabagh M, Ansari Moghaddam A, Eslami V, et al. 2014. Dominant role of drivers' attitude in prevention of road traffic crashes: A study on knowledge, attitude, and practice of drivers in Iran. *Accident Analysis and Prevention* 66:36−42

56. Ker K, Roberts I, Collier T, Beyer F, Bunn F, et al. 2005. Post-licence driver education for the prevention of road traffic crashes: a systematic review of randomised controlled trials. *Accident Analysis and Prevention* 37(2):305−13

57. El Khoury J, Hobeika A. 2006. Simulation of an ITS crash prevention technology at a no-passing zone site. *Journal of Intelligent Transportation Systems* 10(2):75−87

58. Chen Z, Qin X. 2019. A novel method for imminent crash prediction and prevention. *Accident Analysis and Prevention* 125:320−29

59. Yue L, Abdel-Aty M, Wu Y, Zheng O, Yuan J. 2020. In-depth approach for identifying crash causation patterns and its implications for pedestrian crash prevention. *Journal of Safety Research* 73:119−32

60. Hinnant JB, Stavrinos D. 2020. Rewards decrease risky decisions for adolescent drivers: Implications for crash prevention. *Transportation Research Part F: Traffic Psychology and Behaviour* 74:272−79

61. Gidion F, Carroll J, Lubbe N. 2021. Motorcyclist injuries: Analysis of German in-depth crash data to identify priorities for injury assessment and prevention. *Accident Analysis and Prevention* 163:106463

62. Peng C, Xu C. 2021. Combined variable speed limit and lane change guidance for secondary crash prevention using distributed deep reinforcement learning. *Journal of Transportation Safety & Security* 14:2166−91

63. Jang J, Ko J, Park J, Oh C, Kim S. 2020. Identification of safety benefits by inter-vehicle crash risk analysis using connected vehicle systems data on Korean freeways. *Accident Analysis and Prevention* 144:105675

64. Xu C, Ding Z, Wang C, Li Z. 2019. Statistical analysis of the patterns and characteristics of connected and autonomous vehicle involved crashes. *Journal of Safety Research* 71:41−47

65. Sinha A, Chand S, Wijayaratna KP, Virdi N, Dixit V. 2020. Comprehensive safety assessment in mixed fleets with connected and automated vehicles: A crash severity and rate evaluation of conventional vehicles. *Accident Analysis and Prevention* 142:105567

66. Wang L, Zhong H, Ma W, Abdel-Aty M, Park J. 2020. How many crashes can connected vehicle and automated vehicle technologies prevent: a meta-analysis. *Accident Analysis and Prevention* 136:105299

67. Xu X, Kwigizile V, Teng H. 2013. Identifying access management factors associated with safety of urban arterials mid-blocks: A panel data simultaneous equation models approach. *Traffic Injury Prevention* 14(7):734−42

68. Li W, Huang Y, Wang S, Xu X. 2022. Safety criticism and ethical dilemma of autonomous vehicles. *AI and Ethics* 2:869−74

69. Cai Q, Abdel-Aty M, Yuan J, Lee J, Wu, Y. 2020. Real-time crash prediction on expressways using deep generative models. *Transportation Research Part C: Emerging Technologies* 117:102697

70. Kashifi MT, Al-Sghan IY, Rahman SM, Al-Ahmadi HM. 2022. Spatiotemporal grid-based crash prediction — application of a transparent deep hybrid modeling framework. *Neural Computing and Applications* 24:20655−69