



Toward safer highway work zones: An empirical analysis of crash risks using improved safety potential field and machine learning techniques

Bo Wang^{a,b,c,1}, Tianyi Chen^{b,d,1}, Chi Zhang^{a,c,*}, Yiik Diew Wong^b, Hong Zhang^e, Yunhao Zhou^b

^a School of Highway, Chang'An University, Xi'an 710064, China

^b School of Civil and Environmental Engineering, Nanyang Technological University, Singapore 639798, Singapore

^c Engineering Research Center of Highway Infrastructure Digitalization, Ministry of Education, Xi'an 710000, China

^d Department of Civil and Environmental Engineering, National University of Singapore, Singapore 117576, Singapore

^e Transportation Institute, Inner Mongolia University, Hohhot 010021, China

ARTICLE INFO

Keywords:

Work zone safety
Real-time trajectory data
Improved crash risk field
Machine learning
Key risk-contributing features

ABSTRACT

Due to complex traffic conditions, transition areas in highway work zones are associated with a higher crash risk than other highway areas. Understanding risk-contributing features in transition areas is essential for ensuring traffic safety on highways. However, conventional surrogate safety measures (SSMs) are quite limited in identifying the crash risk in transition areas due to the complex traffic environment. To this end, this study proposes an improved safety potential field, named the Work-Zone Crash Risk Field (WCRF). The WCRF force can be used to measure the crash risk of individual vehicles that enter a work zone considering the influence of multiple features, upon which the overall crash risk of the road segment in a specific time window can be estimated. With the overall crash risk used as a label, the time-window-based traffic data are used to train and validate an eXtreme Gradient Boosting (XGBoost) classifier, and the Shapley Additive Explanations (SHAP) method is integrated with the XGBoost classifier to identify the key risk-contributing traffic features. To assess the proposed approach, a case study is conducted using real-time vehicle trajectory data collected in two work zones along a highway in China. The results demonstrate that the WCRF-based SSM outperforms conventional SSMs in identifying crash risks in work zone transition areas on highways. In addition, we perform lane-based analysis regarding the impact of setting up work zones on highway safety and investigate the heterogeneity in risk-contributing features across different work zones. Several interesting findings from the analysis are reported in this paper. Compared to existing SSMs, the WCRF-based SSM offers a more practical and comprehensive way to describe the crash risk in work zones. The approach using the developed WCRF technique offers improved capabilities in identifying key risk-contributing features, which is expected to facilitate the development of safety management strategies for work zones.

1. Introduction

Road maintenance is necessary to ensure the service performance of highways, which requires the closure of certain lane(s) for maintenance work (Zhang et al., 2020) to avoid passing vehicles crashing into workers or equipment. Although various safety management strategies for work zones are required in many countries, such as Singapore (Land Transport Authority, 2019), China (Ministry of Transport, 2017), and the United States (AASHTO, 2018), many work zone crashes still occur

every year. According to statistical data published by the National Highway Traffic Safety Administration (NHTSA) and the Federal Highway Administration (FHWA) in the United States, the number of crash-related fatalities in work zones in 2019 reached 842 (NHTSA, 2019), of which approximately 85 % were drivers or vehicle passengers (FHWA, 2020). Of note, according to the "Code of Practice: Traffic Control at Work Zone" (Land Transport Authority, 2019), a standard work zone on a highway comprises five areas, namely, advance warning area, transition area, buffer space, activity area, and termination area, the

* Corresponding author at: School of Highway, Chang'An University, Xi'an 710064, China.

E-mail address: zhangchi@chd.edu.cn (C. Zhang).

¹ Bo Wang and Tianyi Chen contributed equally.

<https://doi.org/10.1016/j.aap.2023.107361>

Received 18 July 2023; Received in revised form 7 October 2023; Accepted 18 October 2023

Available online 25 October 2023

0001-4575/© 2023 Elsevier Ltd. All rights reserved.

descriptions of which are summarized in Fig. 1. Traffic accidents predominantly occur within the transition area, primarily attributed to frequent lane changing and braking in these zones (Idewu and Wolshon, 2010; Weng et al., 2016). Therefore, a comprehensive analysis of the crash risks and their contributing features within the transition area of the work zone is essential for enhancing the effectiveness of safety management strategies.

Due to the limited quality and quantity of crash reports pertaining to work zones, researchers are turning to either simulations or surrogate measures for investigating traffic safety in work zones (Hou and Chen, 2020; Bidkar et al., 2023). Compared to simulation methods, conflict analysis based on real-world vehicle trajectory data can better reflect the actual risky traffic situations in work zones (Weng et al., 2014). Time to collision (TTC) and deceleration rate to avoid a crash (DRAC) are the two SSMs that have been widely used for work zone crash risk analysis (J. Wang et al., 2022). The traffic environment, including the number of lanes, speed limit, etc., can be greatly changed when a work zone is established, which may result in frequent lane changing and braking by vehicles (Kummetha et al., 2020). However, most conventional SSMs have the following two limitations when measuring crash risks in work zones: they cannot a) fully consider the impact of several factors (e.g., vehicle type or road geometry) on the crash risk (Kordani et al., 2018), and b) sufficiently describe the crash risk in some scenarios involving vehicle interaction, such as vehicle following with similar speeds (Mahmud et al., 2017) or lane changing (Arun et al., 2021).

Given the inherent constraints of conventional SSMs, some scholars have developed SSMs based on potential field theory (Wolf and Burdick, 2008; Wang et al., 2016; Li et al., 2019; Arun et al., 2023). These methods creatively used virtual fields to describe the impact of surrounding objects, such as vehicles and road facilities, on the driving safety of a target vehicle (Ma et al., 2023). Since the SSMs based on the safety potential field consider the interaction between all road users, they provide a suitable approach for measuring crash risk in complex traffic environments. Currently, safety potential field models have been widely used in motion control and collision warning scenarios (Li et al., 2020a; Li et al., 2020b). However, applying safety potential field theory to measure crash risks in highway work zones faces the following challenges: a) it is difficult to determine the risk-related thresholds of safety potential fields when multiple objects are involved; b) the influence of road geometry on crash risk is often ignored; and c) most previous models require many parameters, which leads to a heavy workload in parameter calibration.

In recent years, the rapid development of artificial intelligence (AI) technologies has opened new avenues for in-depth analysis of traffic risks. On the one hand, compared to conventional radar-based techniques (C. Zhang et al., 2022), AI-powered computer vision technologies can more accurately capture the real-time trajectories and basic information of vehicles (St-Aubin et al., 2015), providing a solid foundation for crash risk measurement within safety potential field theory. On the other hand, machine learning techniques have shown enormous

potential for investigating the correlations between crash risk and risk-contributing features (Hu et al., 2022). Compared to conventional statistical methods, machine learning techniques possess superior capability in solving high-dimensional data manipulation and nonlinear problems (Yuan et al., 2022). Therefore, this study employs machine vision technologies to obtain real-time trajectory data and adopts machine learning techniques to identify the key risk-contributing features.

To summarize, in this study, a comprehensive approach is proposed to identify key risk-contributing features and investigate crash risks in highway work zones. To resolve the abovementioned gaps in previous studies, the proposed approach adopts an improved safety potential field model, named the Work-Zone Crash Risk Field (WCRF), to measure the crash risks of vehicles. Compared to conventional field-based SSMs, the WCRF-based SSM (WCRF force) is improved from multiple perspectives, which makes it more practical to implement and better suited for measuring crash risks in work zones. By employing the WCRF force, the overall crash risk of a road segment with a work zone during a specific time window can be estimated, and the key risk-contributing features can be further identified using machine learning techniques. As a case study, real-time vehicle trajectory data, which are collected from surveillance videos along a highway in China, are used to validate and assess the proposed approach. This study makes the following two major contributions. First, at the methodological level, we develop an improved field-based SSM that is able to describe crash risk in work zones in a practical and comprehensive way. Second, at the application level, the findings related to key risk-contributing features reveal the impact of setting up work zones as well as the heterogeneity in crash risk across different work zones on highways. The proposed approach would facilitate highway management authorities in identifying the key risk-contributing features of work zones, leading to the development of more effective safety management strategies.

The remainder of the paper is structured as follows. Section 2 provides a literature review. Section 3 introduces the methodology. Section 4 introduces the data preparation steps. The results and discussion are presented in Section 5. Finally, Section 6 provides the conclusions of our work.

2. Literature review

2.1. Features influencing crash risk in work zones

Previous research has explored the features that might affect the crash risk in work zones from four perspectives: work zone layouts (Mashhadi et al., 2021), traffic flow features (Meng and Weng, 2011), weather conditions (Hou and Chen, 2020), and road conditions (Zhang et al., 2020). Scientific and rational planning of work zones can effectively improve their safety. Within the work zone layouts, key features affecting crash risk include the length of different areas within the work zone (Yang et al., 2013), the lanes occupied by the work zone (Zhang et al., 2020), and the placement and type of traffic signs (Bella, 2005).

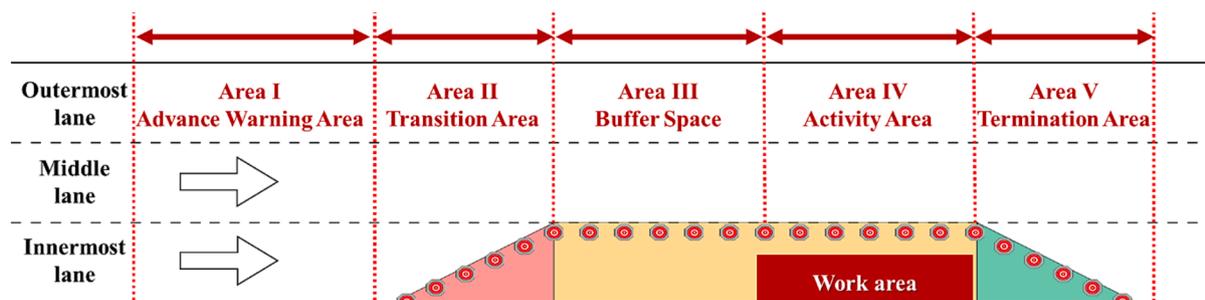


Fig. 1. Components of a work zone on a one-way three-lane road: advance warning area, which alerts drivers of upcoming changes; transition area, which redirects traffic and indicates driver requirements; buffer space, providing worker protection and separating traffic from the work area; activity area, where work takes place with workers and equipment; and termination area, allowing traffic to resume normal conditions.

Compared to static work zone layouts, the influence of dynamic traffic flow features on traffic safety is more complex (Bidkar et al., 2023). In terms of traffic flow, key features include the proportion of trucks (Meng and Weng, 2011), the capacity values of work zones (Chatterjee et al., 2009), traffic volume and speed across lanes (Weng et al., 2016), and vehicle-following patterns (Weng et al., 2014). Weather conditions and road geometry also play a significant role in crash risk. For instance, work zones face increased crash risks in adverse weather conditions such as rain and fog (Adomah et al., 2022), and work zones situated on steep slopes or tight curves are more hazardous than in other segments (Li and Bai, 2009).

Due to time and space constraints, it is difficult to avoid placing work zones during periods of poor weather or in areas with suboptimal road geometry. However, work zone layouts and traffic flow features can be optimized through measures to enhance safety (Almallah et al., 2021; Huang and Bai, 2014). Several countries have released guidelines specifically for work zone layouts (NHTSA, 2019; Land Transport Authority, 2019). Consequently, this study primarily explores the impact of traffic flow features on crash risk in work zones, aiming to provide a basis for developing safety management strategies.

2.2. Surrogate safety measures in work zones

Traffic conflict technology (TCT) is widely considered effective for conducting microlevel traffic safety research (Guo et al., 2020a; Pinnow et al., 2021; Guo et al., 2020b; Ha et al., 2012). Currently, surrogate safety measures (SSMs) (also called conflict measures) can be broadly classified into three types: time-based, distance-based, and deceleration-based measures (Mahmud et al., 2017). To avoid the risk of false positives for distance-based measures (Kuang et al., 2015; Park et al., 2018), time-based and deceleration-based measures are more commonly used in work zone assessment. Currently, the SSMs commonly used in work zone traffic safety research include TTC (Osman et al., 2018; Adomah et al., 2022), work zone time to collision (WTTTC) (Ge et al., 2019), DRAC (Meng and Weng, 2011; Weng et al., 2018; Yang et al., 2020), and time headway (TWH) (Chatterjee et al., 2009; Zhang et al., 2018). TWH may not be suitable for accurate crash risk assessments since the length of vehicles is ignored. TTC and DRAC, which are constructed based on the relative speed $v_B - v_F$ and relative distance $S_F - S_B$ of the preceding (F) and following (B) vehicles, are two widely used measures. $TTC = (S_F - S_B) / (v_B - v_F)$, and the conflict threshold can be determined based on the driver's braking response time (Chen et al., 2019). Additionally, $DRAC = (v_B - v_F)^2 / (S_F - S_B)$, and in this case, the conflict threshold can be determined based on the maximum rate of deceleration of the vehicle (Shi et al., 2018).

Despite advantages such as simple computations and high interpretability, most conventional SSMs still exhibit the following drawbacks when used to measure crash risks in work zones. a) The effects of vehicle and road attributes on crash risk are not considered. For example, vehicle type and road grade can directly affect braking performance (Kordani et al., 2018). b) Due to the neglect of the required lateral safety distance of vehicles, many risks associated with complicated driving behaviors (e.g., lane changing) cannot be fully described (Arun et al., 2021). c) Simply measuring the crash risk of vehicles on highways based on relative speeds is insufficient. For instance, in two different car-following scenarios, even if the following vehicles have the same relative velocity and distance to the preceding car (same TTC and DRAC), it does not mean that their crash risks are the same. In fact, the car-following scenario with a higher absolute velocity presents a higher crash risk compared to the other scenario (Mahmud et al., 2017).

2.3. Safety potential field

In physics, a field refers to the spatial distribution of a physical quantity. The central goal of safety potential field theory is to use a

unified index to describe the spatial distribution of driving risks under "human, vehicle, road, and environmental effects". Currently, the safety potential field model is mainly applied in autonomous driving scenarios. Early on, Wolf and Burdick (2008) applied potential functions to vehicle safety control, constructing the lane potential, road potential, car potential, and velocity potential. Based on this approach, Li et al. (2020a) further incorporated acceleration and steering angle into the driving safety field model and proposed the potential field indicator (PFI) to characterize the comprehensive crash risk of vehicles. Furthermore, they have conducted extensive research in the domain of autonomous driving, such as car-following model (Li et al., 2019), lane-changing mechanisms (Li et al., 2020c), and the establishment of warning strategies (Li et al., 2020b). Some scholars have attempted to construct safety potential field models for human-driven scenarios (Mullakkal-Babu et al., 2020), among which Wang et al. (2015); Wang et al. (2016) proposed a unified driving safety field model that included static potential fields, kinetic potential fields, and behavioral fields considering the interaction among humans, vehicles, and roads and enhanced the definition of virtual mass. Arun et al. (2023) proposed a physics-informed road user safety field theory, upon which they developed an SSM for rear-end crash risk.

However, applying safety potential field theory to measure crash risks in work zones is associated with the following problems. a) Compared to conventional SSM-based models, safety potential field models often have many built-in parameters, and a large amount of data is required for parameter calibration when the models are applied to work zone scenarios. b) The overlapping of the safety potential fields of various objects necessitates the weight assigned to each object, and the determination of the weights can be highly difficult due to the multitude of objects in work zones. c) In typical human-driven scenarios, safety potential field models mostly consider driver factors but fail to fully consider the risk heterogeneity associated with road factors, such as the road grade. d) In previous studies, the risk charge of objects was often related to the weight of the target vehicle. However, it is challenging to obtain accurate vehicle weights in practical research.

3. Methodology

In this study, we propose a key risk-contributing feature identification for work zones approach integrating improved safety potential field and machine learning techniques. Fig. 2 illustrates the overall approach framework which consists of two main parts, namely, crash risk measurement and risk-contributing feature identification. The first part aims to apply the safety potential field to measure the crash risk in work zones. The second part focuses on using machine learning techniques to identify risk-contributing features. This part contains two phases, namely, dataset preprocessing and risk-contributing feature learning. The goal of dataset preprocessing is to construct a dataset suited for supervised machine learning, which includes the steps of sample splitting, labeling, and feature extraction. In phase of risk-contributing feature learning, we combined machine learning techniques with resampling algorithms and model interpretation methods to identify key risk-contributing features.

3.1. Crash risk measurement

3.1.1. Work zone crash risk field

To comprehensively identify crash risks within work zones, an improved safety potential field model, termed the Work-Zone Crash Risk Field (WCRF), is proposed. In work zone scenarios, objects are categorized into two types: target vehicles and interactive objects. The target vehicle is defined as the observed vehicle for which the crash risk is being measured. To reduce computational efforts, this study focuses solely on the crash risks between each target vehicle and objects in its front or side. Thus, interactive objects are limited to those positioned either in front of or to the side of the target vehicle, which includes

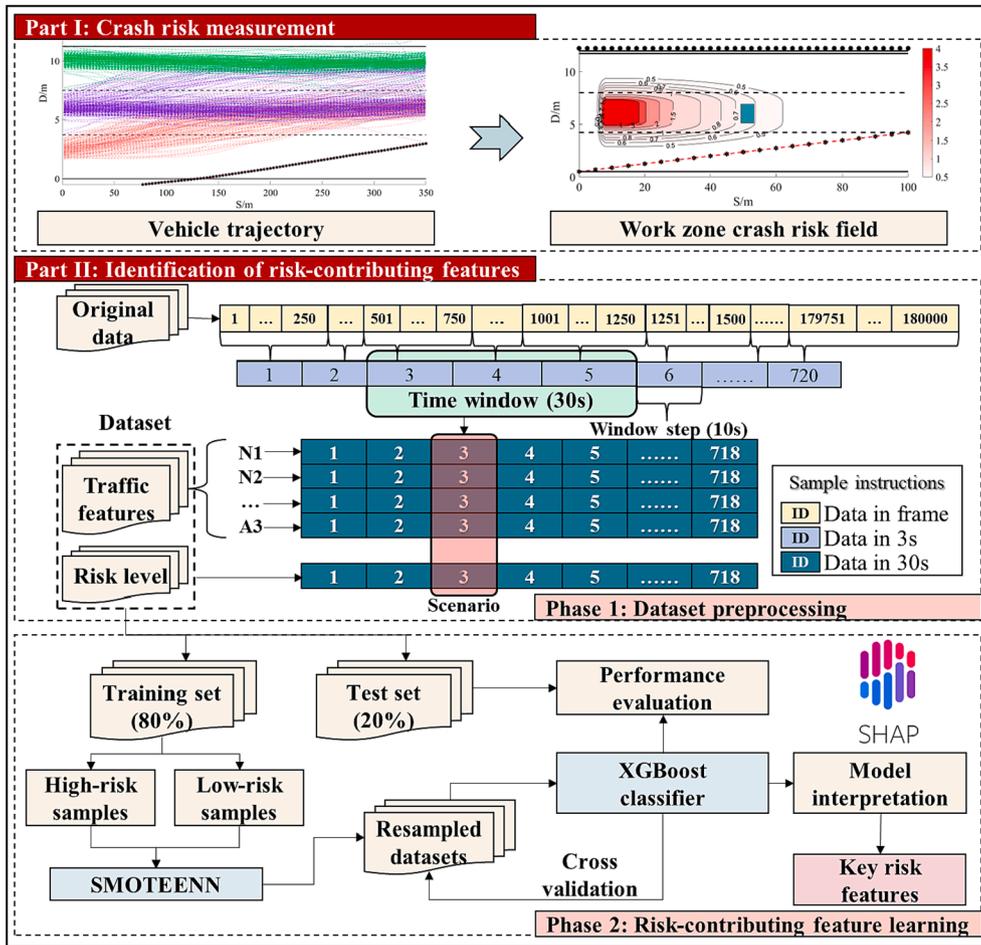


Fig. 2. Framework of key risk-contributing feature identification for work zones.

surrounding vehicles, guardrails, and traffic cones. In addition, each interactive object is not treated as a point, but an entity with a unique shape (as illustrated in Fig. 3). We define the virtual crash point as the point on the contour of the interactive object with the maximum field intensity, which is most likely the position for a potential crash to occur.

The target vehicle serves as the field source for the WCRF, and time is used as the primary indicator for field intensity, describing the potential distribution of crash risk in the space surrounding the vehicle. The field intensity E is solely determined by factors such as the distance from the vehicle, the attributes of the target vehicle (such as velocity and type), and the attributes of the road (such as grade). The field force F_i represents the crash risk between the target vehicle and the interactive object i , and it is related to the field intensity E and the risk charge Q_i of interactive object i . The maximum field force F_V represents the driving

risk of the vehicle. The basic relationship between the WCRF field force F , field intensity E , and risk charge Q is expressed in Eq. (1).

$$F_V = \max\{F_i\} = \max\{Q_i E\}$$

The core of the potential field model is the field intensity formula, which depends on the distribution characteristics of the crash risk. When modeling the field intensity, three issues need to be considered: the influence of distance on the field intensity, the anisotropy of the field intensity, and the key factors influencing the field intensity. Furthermore, to more accurately describe the interaction relationship between the target vehicle and other objects, the target vehicle is conceptualized as a rectangle. Considering the anisotropy of crash risk and focusing only on the crash risk of the target vehicle with front and side interactive objects,

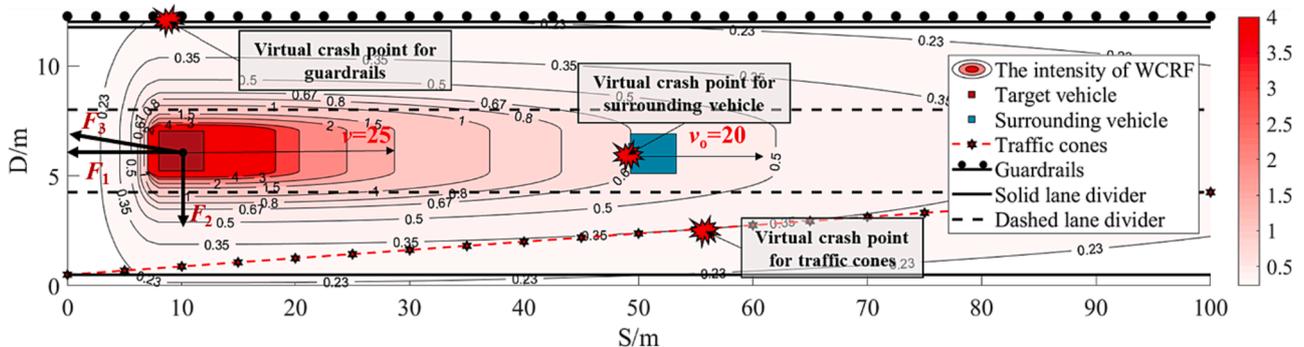


Fig. 3. Example work-zone crash risk field.

the distribution of the potential field proposed in this study is wedge-shaped. This shape is similar to that of the vehicle potential field discussed in previous studies (Wolf and Burdick, 2008). The improved field intensity formula proposed in this study is shown in Eq. (2), where the field intensity model E consists of two parts: the equivalent time-based distance T^* and the vehicle-road combination function η .

$$E = \frac{\eta}{T^*} \quad (2)$$

In addition, it is necessary to determine the risk threshold of the maximum field force F_v . Compared to WCRF, TTC and DRAC are highly interpretable SSMS, and their conflict thresholds can be theoretically determined. The threshold of TTC is related to the driver's reaction time (Mahmud et al., 2017). Additionally, the conflict threshold of DRAC can be determined based on the maximum possible deceleration (AASHTO, 2018). In this study, TTC and DRAC are employed to initially identify vehicles that are involved in conflicts (in this study, these vehicles are referred to as high-risk vehicles). Subsequently, the minimum F_v among these conflicting vehicles is computed as the risk threshold. To obtain an adequate number of high-risk samples, this study, based on previous research, selects a conflict threshold of 5 s for TTC (Xin et al., 2021) and 1.4 m/s² for DRAC (Chen et al., 2019). Furthermore, to facilitate subsequent data analysis, $1/\text{TTC}$ is used in place of TTC, meaning that the threshold for $1/\text{TTC}$ is set at 0.2 s^{-1} .

3.1.2. Model structure

In this section, the WCRF model and the formulas for the main model parameters are introduced. The final model for the WCRF force F is shown in Eq. (3), where the three main parameters are the equivalent time-based distance T^* , the vehicle-road combination function η , and the risk charge of interactive object Q . The methods for calculating the three main parameters (T^* , η and Q) will be introduced in detail later.

$$F = Q \frac{\eta}{T^*} = e^{0.125(v-v_0)\cos\varphi} \frac{\delta v M I}{\sqrt{y^2 + \delta^2 x^2}} \quad (3)$$

where (x, y) are the position of the interactive object's virtual crash point in the target vehicle's coordinate system. δ is the equivalent parameter. M is the vehicle attribute parameter, and I is the road attribute parameter. v is the vehicle velocity (m/s). v_0 represents the velocity of the interactive object, and φ is the equivalent angle between the interactive object and the direction of vehicle travel.

(1) The equivalent time-based distance

The time required for a vehicle to reach a specified location is known as the time-based distance, which can be used directly to measure crash risk (Shbeeb, 2000). The concept of equivalent time-distance T^* is introduced to address the anisotropy issue. When interactive object A is directly ahead in a vehicle's line of travel and interactive object B is not, assuming they both present equal crash risks, the time-distance from the vehicle to object A equates to the equivalent time-distance from the vehicle to object B. The coordinates of interactive objects A and B in the vehicle's coordinate system are $(x, 0)$ and $(0, y)$, respectively. Based on the vehicle's velocity v , the equivalent time-distance of interactive object B can be obtained as $T_B^* = T_A = x/v$. Here, we assume the existence of a virtual velocity $v_y = \delta v$ along the y -axis of the vehicle, such that $T_B^* = y/\delta v$. Any point (x, y) in the space around the vehicle satisfies Eq. (4), and T^* can be expressed by Eq. (5). To facilitate application in the Frenet coordinate system used for roads, coordinate transformation can be performed according to Eq. (6).

$$\begin{cases} x/v = T^* \cos\theta \\ y/\delta v = T^* \sin\theta \end{cases} \quad (4)$$

$$T^* = \frac{\sqrt{y^2 + \delta^2 x^2}}{\delta v} \quad (5)$$

$$[x, y, 1] = [s, d, 1] \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ s_0 & d_0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (6)$$

where (s, d) are the point coordinates of the Frenet coordinate system and (s_0, d_0) are the vehicle's center coordinates. θ is the clockwise angle between the vehicle's driving direction and the S -axis.

The equivalent parameter δ depends on the ratio of the lateral and longitudinal safety distances. According to previous research (Gunay, 2007; Budhkar and Maurya, 2017), the relationship between the lateral safety distance y_s and vehicle speed v can be expressed as $y_s = 0.02214v + 1.2$. Combined with the research on the threshold of time to accident (TA), the relationship between the longitudinal safety distance x_s and vehicle speed v can be expressed as $x_s = 1.5v$ (Shbeeb, 2000). Therefore, the equivalent parameter δ can be calculated as shown in Eq. (7).

$$\delta = 0.01476 + 0.8v^{-1} \quad (7)$$

(2) The vehicle-road combination function

Braking performance is one of the important factors that affects crash risk and is closely related to vehicle and road attributes. Previous research on the stopping sight distance (SSD) has extensively explored the effects of vehicle and road attributes on vehicle braking performance and has achieved valuable results (Samson et al., 2022). Therefore, in this study, the impact of different attributes (vehicle attributes M and road attributes I) on the SSD is assessed, and a vehicle-road combination function η is established to determine the intensity of the WCRF, as shown in Eq. (8). The standard value of η is 1 when a car is traveling on a flat road.

$$\eta = M I \quad (8)$$

Vehicle type is the most prominent vehicle attribute, and previous SSD research categorized vehicles into two types: trucks and cars (AASHTO, 2018). The Virginia Code - Tables of Speed and Stopping Distances (46.2–880) specifies the SSDs for these two types of vehicles (Code of Virginia, 2021). The vehicle parameter for cars is $M_{car} = 1$, and that for trucks M_{truck} is the ratio of the truck's SSD to the car's SSD. A further regression analysis was performed on the M_{truck} parameter, resulting in Eq. (9), with a coefficient of determination R^2 of 0.9952 and a root mean square error (RMSE) of 0.0029.

$$M_{truck} = 0.935v^{0.07269} \quad (9)$$

Road grade is one of the important factors that influences vehicle braking. Based on the SSD calculation method for different grades provided by the American Association of State Highway and Transportation Officials (AASHTO, 2018) and assuming a driver reaction time of 2.5 s and a vehicle braking deceleration rate of 3.4 m/s^2 , the SSD formula shown in Eq. (10) is established. The road attribute parameter I is the ratio of the SSD at different road grades G to the SSD on a flat road ($G = 0$), as shown in Eq. (11).

$$SSD = 2.502v + \frac{v^2}{(6.7926 + 19.6 * G)} \quad (10)$$

$$I = \frac{SSD}{SSD_{G=0}} \quad (11)$$

(3) The risk charge of an interactive object

Relative velocity is an important parameter for conventional SSMS

such as TTC and DRAC, as it can better measure crash risk. Therefore, a method for calculating the interactive object risk charge Q based on the relative velocity between an interactive object and a vehicle is proposed, as shown in Eqs. (12) and (13). Notably, this study mainly focuses on the likelihood of a crash, hence the attributes of the interactive objects only include their spatial position and speed. Both the guardrail and the traffic cone have zero speed. The speed of the surrounding vehicles is based on the scalar value of their actual speed, while the speeds of both the guardrails and traffic cones are zero. Meanwhile, the position of the virtual crash point is used as the spatial position of the interactive object. Moreover, the greater the positive speed difference between the target vehicle and the interactive object, the higher the likelihood of a crash, and the higher the risk charge as well. When the speed difference between a vehicle and an interactive object is 0, $Q = 1$, and when the speed difference is 5.56 m/s (20 km/h), $Q = 2$.

$$Q = e^{0.125(v-v_0)\cos\varphi} \quad (12)$$

$$\varphi = \arctan \frac{y}{\delta x} \quad (13)$$

To better illustrate how to utilize the WCRF force F to measure the crash risk of a target vehicle with different interactive objects in the work zone, we exemplify using a typical scenario illustrated in Fig. 3. In this scenario, two cars are driving on a flat road with a guardrail on the left and a traffic cone on the right. We select the following vehicle as the target vehicle. There are three interactive objects: the preceding vehicle (surrounding vehicles), traffic cones, and guardrails. Firstly, using Eqs. (2), (4)-(11), we can determine the potential field distribution formed by the target vehicle. Subsequently, the virtual crash points of the three interactive objects can be identified based on field intensity E . Building on this, the risk charge Q for the three interactive objects can be determined using Eqs. (12) and (13). Finally, using Eq. (3), the WCRF force F exerted on the target vehicle by the three interactive objects are deduced. In this example, the field forces of the surrounding vehicle, guardrails, and traffic cones are $F_1 = 1.25$, $F_2 = 1.21$, and $F_3 = 0.23$, respectively. According to Eq. (1), the F_V of the vehicle is determined to be 1.25.

3.1.3. Improvements

Compared with the field-based models proposed in previous studies (Wang et al., 2016; Li et al., 2020b), the improvements of the WCRF model are made as follows:

- 1) **Field source:** Previous scholars have proposed the use of interactive objects as field sources in safety potential fields, as shown in Fig. 4 (a). While this method effectively illustrates the distribution of collision risks on the road, applying it directly to work zone scenarios faces the following challenges: a) difficulty in determining weights

for fields generated by different interactive objects, b) numerous undetermined parameters require calibration, and c) potential fields for each interactive object must be computed separately. Drawing inspiration from the concept that forces are mutual, the WCRF model designates the target vehicle as the exclusive field source, as presented in Fig. 4(b). In contrast to previous model, this adjustment in the field source negates the need to determine weights and substantially reduces model parameters and computational demands.

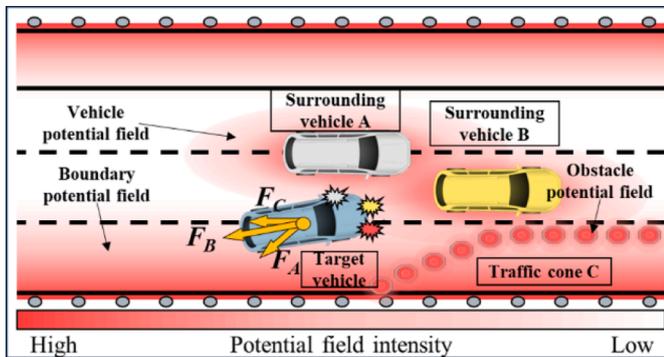
- 2) **Risk quantification:** In previous studies, the risk field forces of multiple interactive objects were often superimposed into a comprehensive field force to represent the risk state of vehicles. This approach might be highly susceptible to traffic flow influences. For instance, in multi-vehicle scenarios, the comprehensive field force might be greater even if each vehicle poses a low risk, compared to scenarios with fewer vehicles but higher risks. However, the maximum safety field force F_V can more effectively represent the vehicle's extreme crash risks. Therefore, in this study, only the F_V is used to represent the crash risk of the target vehicle. Although this approach might result in overlooking some high-risk vehicles when considered as interactive objects, it can accurately identify vehicles that have a high risk of crash with objects in front or to the side of them. This provides an effective SSM for measuring crash risk in the work zone.

3.2. Risk-contributing feature identification

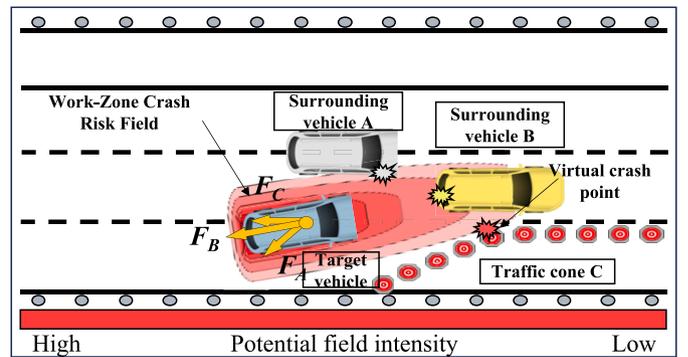
3.2.1. Dataset preprocessing

(1) Scenario splitting

To explore the impact of traffic features on crash risk, it is necessary to split the original data extracted frame-by-frame from a video. Utilizing a fixed time interval for data splitting can potentially cause the loss of correlation between features and sample labels, especially for infrequent high-risk samples (Polyzotis et al., 2017). However, employing a sliding window approach can help the model identify the associations between neighboring statistical units, thereby improving its performance (Gertz et al., 2020). To a certain degree, integrating the sliding window with a machine learning algorithm can alleviate issues related to extreme sample imbalance (Chiong et al., 2022). Relevant research has confirmed that traffic datasets collected at 30-second intervals are superior to other related datasets in machine learning (Hu et al., 2022). Moreover, a window step that is too small can lead to overfitting, while a step that is too large can cause high-risk samples to be incorrectly split. Considering that the passing time for most vehicles is approximately 10 s, the window step size should ideally be less than 10 s. Therefore, a sliding window method is applied to split the original data into scenario datasets. The size of the sliding window is 30 s, with a



(a) Safety potential field



(b) Work-zone crash risk field

Fig. 4. Conceptual diagram of improvements to the safety potential field.

step of 3 s. That is, each scenario contains all the information observed within 30 s.

(2) Scenario labeling

Whether there are high-risk interactions within the time window of the scenarios is used as the criterion for assessing their risk level. Based on the WCRF force F_V and its risk thresholds, interactions between vehicles and surrounding objects can be classified into two groups, namely, high-risk and low-risk interactions. Accordingly, the scenarios containing high-risk interactions are labeled with high-risk, while the remainder are labeled with low-risk.

In this study, TTC and DRAC are employed to determine the risk threshold for the WCRF field force F_V . First, high-risk vehicles are initially identified using the known thresholds of SSMS. Then, the F_V of each vehicle is calculated, and the smallest F_V is selected as the risk threshold for the WCRF field force. Since TTC and DRAC are highly interpretable SSMS, they are chosen as the SSMS for the initial identification of high-risk vehicles. The threshold of TTC is related to the driver's reaction time (Mahmud et al., 2017), while the conflict threshold of DRAC is dependent on the maximum possible deceleration (AASHTO, 2018). Based on previous research, we select a conflict threshold of 5 s for TTC (Xin et al., 2021) and 1.4 m/s² for DRAC (Chen et al., 2019). Furthermore, to facilitate subsequent data analysis, 1/TTC is used in place of TTC, meaning that the threshold for 1/TTC is set at 0.2 s⁻¹.

(3) Scenario features

In response to the requirements for developing traffic management strategies, the traffic features of scenario samples are determined, as shown in Table 1, based on common traffic control strategies in work zones. These strategies include speed limits (Nnaji et al., 2020), early lane merging (Almallah et al., 2021), and the opening of emergency lanes (La Torre et al., 2017). The speed limit can effectively regulate the speed of travel in different lanes. Installing speed-recording cameras can improve the effectiveness of speed limit control and reduce the number of speeding vehicles. Early lane merging can effectively reduce the number of vehicles in closed lanes. Additionally, opening emergency lanes when necessary can reduce the number of vehicles in other lanes.

3.2.2. Risk-contributing feature learning

XGBoost (eXtreme Gradient Boosting) is an ensemble learning algorithm based on decision trees that provides excellent performance and flexibility in solving supervised learning problems (Chen and Guestrin, 2016). Previous studies have confirmed that XGBoost usually performs well in traffic feature data mining and crash risk prediction (Shi et al., 2019). Furthermore, considering the imbalance between high-risk and low-risk samples in a dataset can enhance XGBoost model training. A resampling algorithm can effectively improve the problem of sample imbalance. SMOTEENN involves a combination of oversampling and

Table 1
Introduction to the traffic features of scenarios.

Risk Feature	Unit	Lane	Code
Total number of vehicles	vehicle	Innermost lane	N1
		Middle lane	N2
		Outermost lane	N3
Total number of speeding vehicles	vehicle	Innermost lane	S1
		Middle lane	S2
		Outermost lane	S3
Maximum speed	m/s ⁻¹	Innermost lane	X1
		Middle lane	X2
		Outermost lane	X3
Mean speed	m/s ⁻¹	Innermost lane	A1
		Middle lane	A2
		Outermost lane	A3

undersampling and is based on the synthetic minority oversampling technique (SMOTE) and edited nearest neighbor (ENN) technique (Chawla et al., 2002). SMOTEENN can reduce the number of majority-class samples and increase the number of minority-class samples so that the model classification ability can be improved while maintaining a data balance (B. Wang et al., 2022).

In this study, a traffic risk classification model that combines the XGBoost classifier and the SMOTEENN resampling algorithm is developed. First, the dataset is divided into training and test sets at an 8:2 ratio based on previous research experience (Qi et al., 2022). Second, the training set is resampled using SMOTEENN. Subsequently, the XGBoost model is trained using a balanced training set with 5-fold cross-validation (Wang et al., 2020). The model hyperparameters are optimized using the grid search method. Finally, the trained model is evaluated using the test set.

In addition, the SHAP (SHapley Additive exPlanations) method is applied to interpret the trained model. SHAP is a popular method for explaining feature importance; it simulates the contribution of each feature to model predictions and calculates the Shapley value of each feature (Lundberg and Lee, 2017). The Shapley value is used in game theory to measure each player's contribution to the payoff of a cooperative game, ensuring that each feature's contribution to the model prediction is unique. SHAP can explain the contribution of each feature to the prediction results and the interactions among features.

4. Data collection

4.1. Data sources

The data for this study are from the S4 highway in China, which has 3 lanes in each direction, and the designed speed of travel is 120 km/h (Fig. 5). The data collected include surveillance video and highway design drawings. Four surveillance videos are selected as case studies: Case AW (road segment A with a work zone), Case AN (road segment A without a work zone), Case BW (road segment B with a work zone), and Case BN (road segment B without a work zone). In the two cases of AW and BW, the work zones are located in the innermost lane. The traffic flows in the four cases are generally similar. Each video, with a resolution of 1920 pixels × 1080 pixels and a frame rate of 24 frames/second, is 2 h long. The statistical descriptions of the four case studies are shown in Table 2. In the table, "car" refers to vehicles that have no more than two axles or no more than four wheels (e.g., private cars and vans), while "truck" refers to vehicles that have more than two axles or more than four wheels (e.g., trucks and buses).

4.2. Real-time trajectory extraction

In this study, a framework was developed using machine vision algorithms to extract vehicle trajectories from surveillance videos for crash risk analysis, as shown in Fig. 6. The framework consists of three main components: 1) vehicle detection and tracking, 2) camera calibration and coordinate transformation, and 3) the extraction of vehicle geometry information. The WCRF force can be calculated using Eq. (3) based on the extracted trajectory and vehicle geometry information.

1) **Vehicle detection and tracking:** The R-CNN (Region-based Convolutional Neural Network) and YOLO (You Only Look Once) algorithms are popular in vehicle detection (Kim et al., 2020). An R-CNN is a "two-stage" algorithm with high accuracy but a slow speed, and YOLO uses a single network for fast detection and easy deployment (Maity et al., 2021). To track vehicle trajectories, object detection needs to be combined with multiobject tracking (MOT). DeepSORT is a classic MOT algorithm, and StrongSORT combines it with Gaussian-smoothed interpolation for enhanced accuracy and speed (Du et al., 2023). Therefore, YOLOv5 and StrongSORT were used for real-time trajectory data extraction.



Fig. 5. Study case and area.

Table 2
Descriptive statistics of data collected from the case studies.

Indicator	Case AW	Case AN	Case BW	Case BN
Video recording date	2022-11-30	2022-12-05	2022-12-05	2022-12-08
Time period	09:00–11:00	09:00–11:00	14:00–16:00	14:00–16:00
Work zone lane	Innermost lane	/	Innermost lane	/
Road grade	−2.64 %	−2.64 %	0.9 %	0.9 %
Total number of vehicles	534	501	604	557
Total number of cars	337	317	371	414
Total number of trucks	197	184	233	143
Innermost lane vehicle count	134	144	139	101
Middle lane vehicle count	244	216	203	172
Outermost lane vehicle count	156	141	262	284

Note: A vehicle's lane is determined based on its position upon entering the transition area.

- 2) **Camera calibration and coordinate system transformation:** Accurately estimating surveillance camera parameters is crucial for coordinate transformation. Due to the lack of reference objects on curved roads, the 'VWL' (one vanishing point, known width and length) method is used for camera parameter calibration (Kanhere and Birchfield, 2010; Wang et al., 2007); in this approach, a vanishing point is determined based on road markings, and camera parameters are obtained based on the known length of road markings and a set of parallel markings with known spacing (Eqs. (S1)-(S5) in Supplementary Material). Upon obtaining the camera parameters, the trajectory coordinates in the world coordinate system are obtained using Eq. (S6) in Supplementary Material. Finally, based on the road markings as a reference, the trajectory coordinates in the Frenet coordinate system of the road are obtained.
- 3) **Extraction of vehicle geometry information:** Vehicle geometry information, such as length and width, is crucial for traffic risk measurement. In this study, the method proposed by Dubská et al. (2014) is improved, and a method that utilizes background subtraction and YOLO vehicle detection to automatically fit the 3D shapes of vehicles is introduced. The program identifies the outer contour of vehicles using a Gaussian mixed-background subtraction algorithm and extracts the vehicle geometry information using a coordinate transformation matrix obtained from camera calibration. The final value

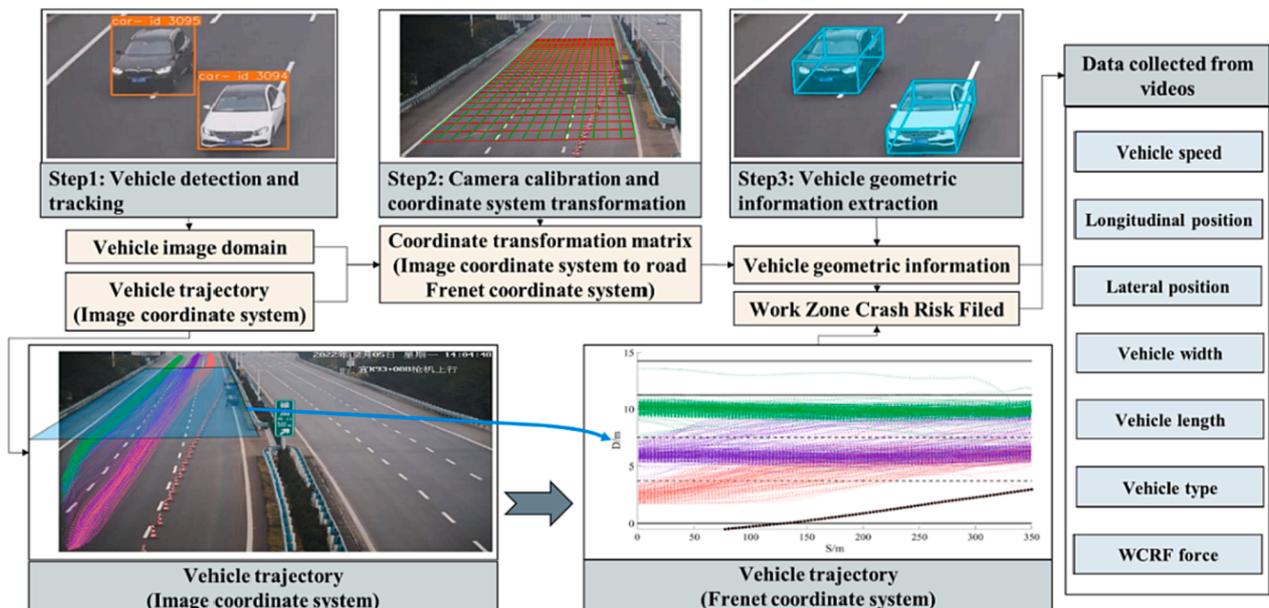


Fig. 6. A framework for trajectory data extraction and crash risk measurement based on machine vision.

is the mean of the geometry information for the 20 frames before the vehicle leaves the monitoring area.

In addition, data filtering is conducted to remove noise and enhance the accuracy of trajectory extraction. As an effective and efficient data filtering method which has recommended in previous studies (Ven-thuruthiyil and Chunchu, 2018), Savitzky–Golay filter is used to clean the vehicle trajectory data.

5. Results and discussion

5.1. Models

5.1.1. Threshold determination

The risk threshold serves as an important basis for risk identification and sample labeling. According to the method described in Section 3.1.1, a total of 27 high-risk vehicles are identified using 1/TTC and DRAC. Based on the maximum WCRF force F_V of these vehicles, the threshold of F_V is determined to be 1.9. Fig. 7 illustrates the results of identifying high-risk vehicles using DRAC, 1/TTC and F_V , with each point representing a vehicle. Using the WCRF force F_V , a total of 57 high-risk vehicles are identified, and the numbers of high-risk vehicles identified using 1/TTC and DRAC are 21 and 26, respectively. Compared to 1/TTC and DRAC, WCRF displays a higher sensitivity for identifying high-risk vehicles.

5.1.2. Measure comparison

To explain the differences among different SSMS in identifying high-risk vehicles, we further analyze the key indicators at the timestamp of maximum risk for all vehicles. The SSMS adopted as baselines for comparison include the potential field indicator (PFI) (Li et al., 2019; Li et al., 2020a), 1/TTC, and DRAC. Herein, we assume that the mass of cars is 2 tons and that of trucks is 7 tons when using PFI. The key indicators include the vehicle type, speed, distance, and motion directions of the target vehicle and its interactive object. Fig. 8 displays six parallel coordinate plots, each illustrating results from different SSMS used to identify high-risk vehicles. Within these plots, each line represents a state in which a target vehicle interacts most dangerously with surrounding vehicles. Additionally, we have highlighted some typical examples for clarity.

From Fig. 8, we find that:

- a) Fig. 8(a) illustrates the result of high-risk vehicle identification using the PFI. Two types of low-risk scenarios are observed to be often misidentified as high-risk. The typical sample A shows one type

where the speed of the following vehicle is much slower than that of the preceding vehicle. On the other hand, the typical sample B represents the other type, where vehicles are spaced significantly apart laterally. The misidentifications can be attributed to certain aspects of PFI: (i) it does not fully account for the speed difference between vehicles; (ii) it might not give enough weight to lateral safety distances; and (iii) it tends to emphasize mass attributes, resulting in scenarios involving trucks having a generally higher PFI than those without trucks. In essence, while the PFI has been proven effective in comparing the crash risk of a single target vehicle with different interactive object, it is less suitable for multiple different target vehicles.

- b) In Fig. 8(b), the typical sample C can be identified based on TTC and the WCRF force but not with DRAC. This indicates that when the relative speed of travel is low, DRAC demonstrates lower sensitivity to high-risk interactions with vehicles in close proximity.
- c) In Fig. 8(c), the typical sample D can be identified with DRAC and the WCRF force but not based on TTC and PFI. This indicates that when the distance is large, both TTC and PFI demonstrate lower sensitivity to high-risk interactions involving vehicles traveling at high relative speeds.
- d) In the process of measuring crash risk, TTC and DRAC do not consider factors such as road grade and vehicle type. Therefore, in Eq. (3), we set the vehicle attribute parameter M and the road attribute parameter I to 1.00 to more effectively compare these SSMS, as shown in Fig. 8(d). As for the typical sample E, although the distance to the interaction object is short, the vehicle is not identified as a high-risk vehicle by DRAC and TTC due to its low Δv . As for the typical sample F, despite the high speed of travel and close distance between the vehicle and object, the vehicle is not identified as a high-risk vehicle due to the insensitivity of DRAC and TTC to the lateral safety distance. However, the WCRF is constructed based on the equivalent time-based distance, and it successfully overcomes these limitations. As a result, it proves to be more suitable for measuring crash risks in complex traffic environments, such as work zones.
- e) Fig. 8(e) displays the high-risk vehicle identification results by WCRF force, considering the vehicle attribute parameter M . In comparison to Fig. 8(d), there are more high-risk vehicles identified by the WCRF force, such as trucks with insufficient following safety distance represented by the typical sample G. Fig. 8(f) shows the final high-risk vehicle identification results using the WCRF force. Since the effect of road grade is considered, more high-risk vehicles are identified, such as vehicles with insufficient following safety distance in the downhill road segments represented by typical sample H. By

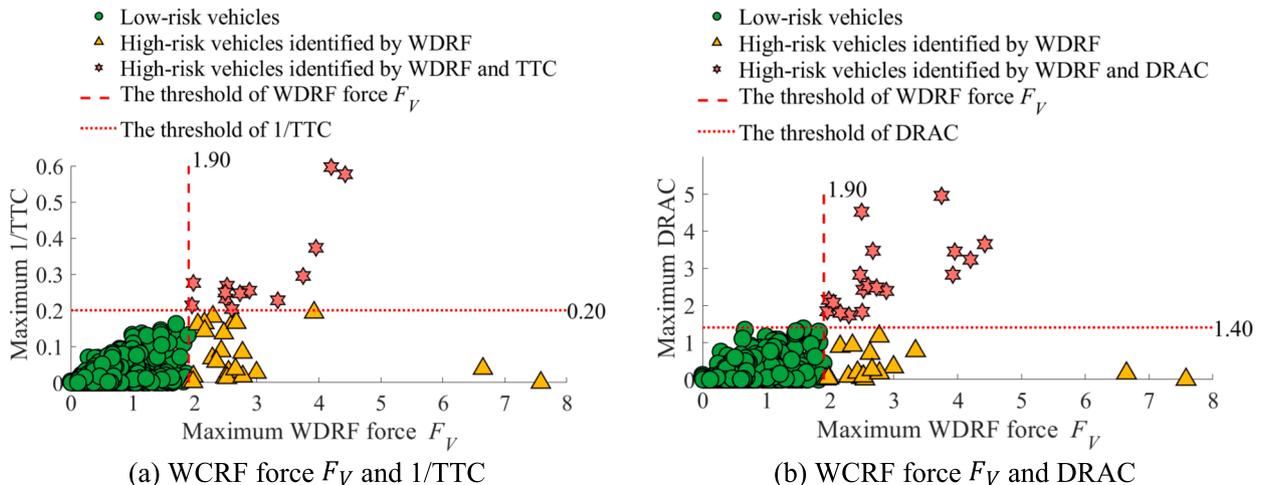


Fig. 7. Comparison of high-risk vehicle identification results based on 1/TTC, DRAC, and F_V .

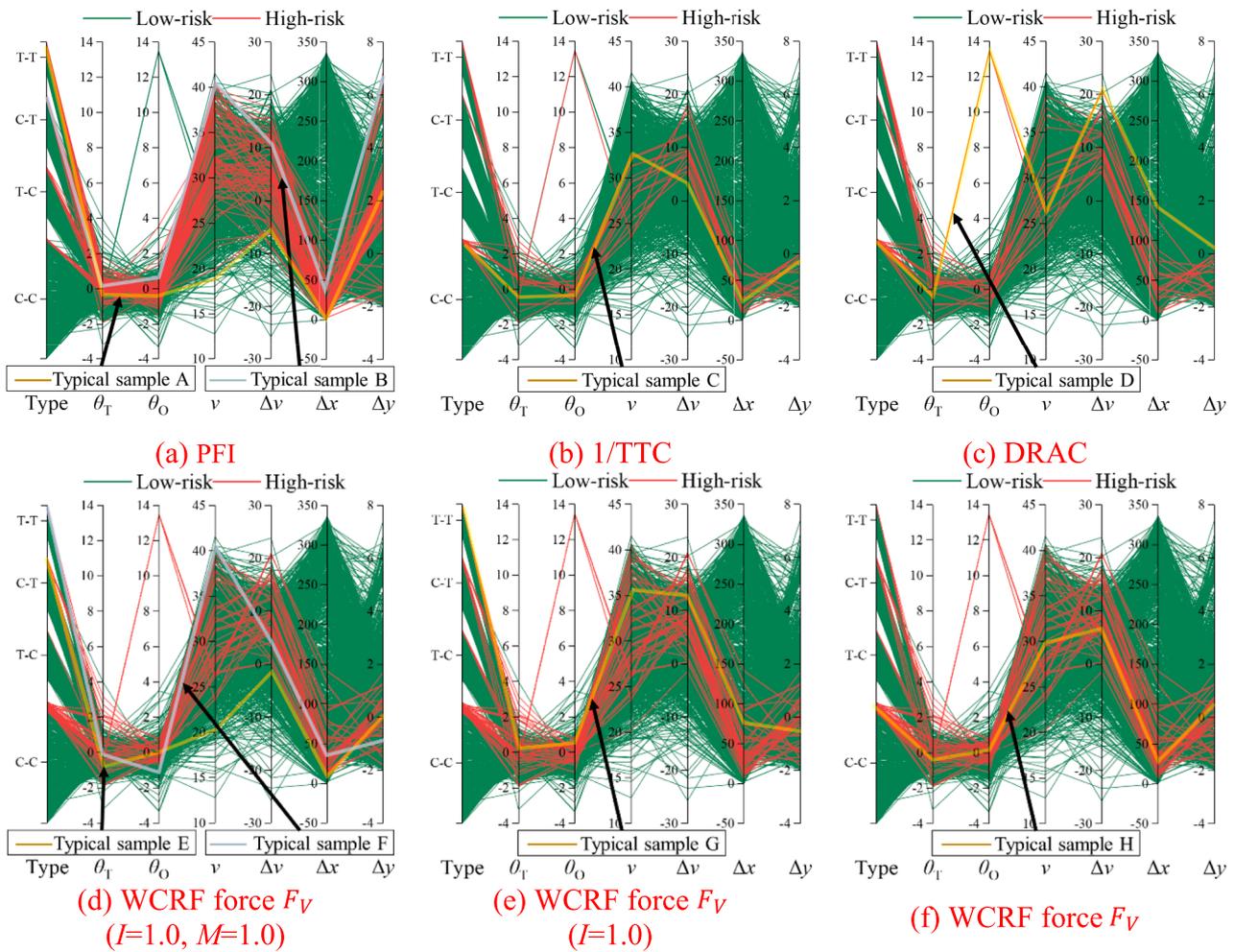


Fig. 8. Parallel coordinates plots of high-risk vehicle identification results. The horizontal axis represents vehicle status indicators, while the vertical axis shows the corresponding values for each indicator. Herein, the “Type” refers to vehicle-pairs between the target vehicle and its interactive vehicle. For example, “C-T” means a pair between a target vehicle which is a “Car” and its interactive vehicle which is a “Truck”. θ_T and θ_O are the movement directions of the target vehicle and the interactive vehicle, respectively. Δ_v , Δ_x and Δ_y represent the relative speed and the lateral and longitudinal distances between the target vehicle and the interactive vehicle, respectively.

considering vehicle types and road grade, the WCRF force can more accurately reflect the crash risk of vehicles than can other SSMS.

5.2. Risk analysis

5.2.1. Exploratory data analysis

To understand the impact of setting up a work zone on the spatial

distribution of crash risks, the WDRF F_V of each vehicle in every frame of the collected videos are calculated. In case studies, both work zones are located in the innermost lane, where the crash risk between the vehicles and the guardrail is extremely low. Therefore, in this section, the analysis is focused on two parts: (a) the distribution of virtual crash points on the traffic cones, and (b) the distribution of crash risks in each lane.

Fig. 9 shows the distribution of virtual crash points between each

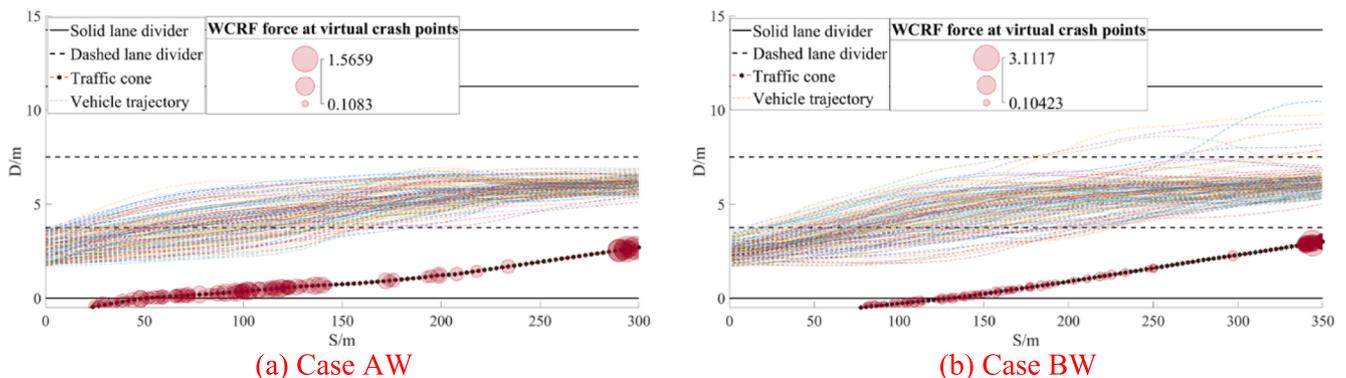


Fig. 9. Spatial distribution of virtual crash points for traffic cones.

vehicle (from innermost lane) and the traffic cones. Each bubble indicates a virtual crash point with the highest crash risk as a vehicle passes the transition area, and the size of bubble represents the strength of the WCRF force. We can observe that these virtual crash points are mainly distributed at the head and tail of the transition area and the crash risks at the tail are always higher. This means traffic cones might not be detected by some drivers earlier, making it less likely for those vehicles to decelerate or change lanes earlier, and those vehicles will inevitably have a higher crash risk with traffic cones when changing lane at the tail of the transition area.

Furthermore, the road area was partitioned into grids, where each grid unit measured 3.0 m along the S-axis and 0.5 m along the D-axis. Fig. 10 illustrates the heatmap depicting the maximum F_V within each grid. The color in the figure represents the maximum WCRF force of all vehicles at the selected location, with lighter colors indicating a higher crash risk. It intuitively reflects the spatial distribution of crash risk for four cases. By comparing the four cases, it can be observed that the presence of work zones noticeably increases the crash risk in road segments. However, there is an obvious difference in the spatial distribution of crash risks in the transition areas for Cases AW and BW. Specifically, most crash risks for Case AW plot in the middle lane, and in Case BW, the risks are distributed in the middle and outermost lanes more evenly.

Furthermore, to investigate the reasons for the differences in the spatial distribution of crash risks, we perform a statistical analysis based on the lanes each vehicle occupies before entering the road segment. This analysis includes the WCRF force F_V , the number of vehicles changing lanes, and the number of high-risk vehicles. Table 3 presents the statistical results for the four cases, revealing the following key findings.

- For the cases in which there are no work zones (Cases AN and BN), according to the mean values of F_V , the overall crash risk of the vehicles in the three lanes is ranked in the following order: innermost > middle > outermost. This trend may be linked to the speed limits for each lane: innermost lane (120 ~ 100 km/h), middle lane (100 ~ 80 km/h), and outermost lane (100 ~ 60 km/h).
- In the cases in which work zones are established (Cases AW and BW), according to the mean values of F_V , the overall crash risk of the vehicles in the middle lane is the highest. This suggests that establishing work zones in the innermost lane has the greatest impact on the safety status of the middle lane. This finding is consistent with previous research results (Meng and Weng, 2011). Notably, an important difference between Cases BW and AW is that in BW, the vehicles in the outermost lane also experience a substantial increase in crash risk.
- By comparing the four cases, we find that the increase in high-risk vehicles may be related to vehicles changing lanes. In terms of the total numbers of lane-changing vehicles and high-risk vehicles, the

ranking is the same for all four cases: BW > AW > BN > AN. This suggests that lane changing might complicate traffic and heighten the crash risk. Furthermore, combined with video analysis, of the high-risk vehicles, 47 vehicles (82.45 %) participate in or are affected by lane changing; notably, either they are lane-changing vehicles or they are involved in high-risk interactions with lane-changing vehicles. This further indicates that vehicle lane changes significantly increase the collision risk in work zones.

- By comparing road segments A and B, we found that for road segment B, many vehicles in the middle lane generally move to the outermost lane. This behavior is likely influenced by the existence of the nearby intersection exit and the service area ahead on road B. Due to the presence of a work zone, many drivers planning to exit the highway choose to change lanes at this location, moving to the outermost lane. This also explains why in Case BW, the crash risk for vehicles in the outermost lane markedly increased.

5.2.2. Analysis of key risk-contributing features

(1) Model performance evaluation

The XGBoost model is trained using the resampled datasets for Cases AW and BW, and the model hyperparameters are optimized as shown in Supplementary Material Figs. S1 and S2 and Table S1. The performance of the trained models is evaluated based on the test set, and the results are summarized in Table 4. The performance of the models developed based on the data from Cases AW and BW is excellent, with accuracy, area under the curve (AUC), and recall values all exceeding 0.90. This indicates that the models are capable of accurately distinguishing among different levels of risk based on the traffic features of the scenario samples. Additionally, the precision of the two models is comparatively low compared to the recall values at 0.870 and 0.774, respectively. It is possible that the lower precision of models is a result of high-risk traffic features being present in some low-risk scenario samples. However, due to the cautious driving behavior of the drivers in those scenario samples, the WCRF force may not exceed the risk threshold.

(2) SHAP value analysis

Fig. 11 demonstrates the local explanations of the model using SHAP values. Traffic features are plotted on the vertical axis in order of their importance, with more important features placed higher. Each point on the horizontal axis represents a sample, and the color represents the feature value, where red indicates a large feature value and blue indicates a small feature value. When the SHAP value is positive, the likelihood of a sample being classified as high-risk increases, and a negative SHAP value indicates a decrease in the likelihood of the sample being classified as high-risk.

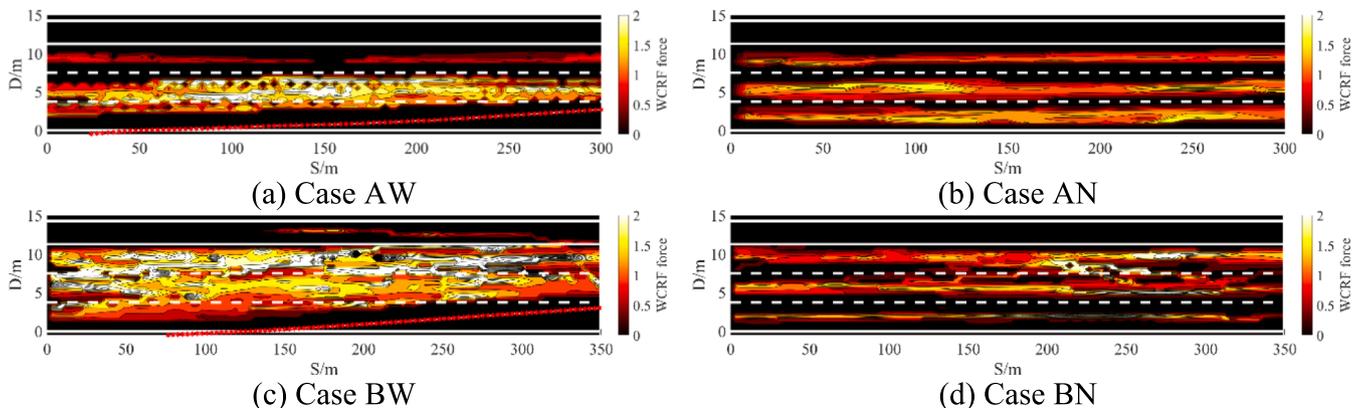


Fig. 10. Spatial distribution heatmap of crash risk.

Table 3
Statistical results for vehicles from different lanes.

Case	Lane	WCRF force F_V			Lane-changing vehicles		High-risk vehicles	
		Max	Mean	S.D.	Number	Proportion	Number	Proportion
AW	Innermost	2.49	0.58	0.31	134	100.00 %	3	2.24 %
	Middle	3.92	0.78	0.48	12	4.92 %	14	5.74 %
	Outermost	1.13	0.31	0.30	1	0.64 %	0	0.00 %
AN	Innermost	1.75	0.49	0.38	0	0.00 %	0	0.00 %
	Middle	1.83	0.40	0.35	0	0.00 %	0	0.00 %
	Outermost	1.73	0.35	0.41	0	0.00 %	0	0.00 %
BW	Innermost	2.73	0.61	0.33	139	100.00 %	2	1.44 %
	Middle	7.66	0.71	0.69	90	44.33 %	14	6.90 %
	Outermost	4.79	0.65	0.62	6	2.29 %	20	7.63 %
BN	Innermost	3.00	0.53	0.56	1	0.99 %	1	0.99 %
	Middle	1.97	0.47	0.22	33	19.19 %	1	0.58 %
	Outermost	2.52	0.44	0.39	2	0.70 %	2	0.70 %

Table 4
Results of classification performance evaluation.

Data source	Accuracy	Recall	Precision	F1 score	AUC
Case AW	0.992	0.952	0.870	0.909	0.973
Case BW	0.962	0.923	0.774	0.842	0.944

According to Fig. 11, for Case AW, the three most important features are N2 (number of vehicles in the middle lane), X2 (maximum speed of vehicles in the middle lane), and A2 (mean speed of vehicles in the middle lane), all associated with the middle lane. On the other hand, for Case BW, the three most important features are N3 (number of vehicles in the outermost lane), A3 (mean speed of vehicles in the outermost lane), and X2, which are associated with the outermost and middle

lanes. Notably, the three most important features in Cases AW and BW are associated with lanes with a high concentration of high-risk vehicles (Table 3). To a certain extent, this finding validates the proposed key risk feature identification method in this study.

In Cases AW and BW, N3 (number of vehicles in the outermost lane) makes a positive contribution to high crash risk. This indicates that many vehicles are in the outermost lane in both cases' high-risk scenarios. However, X3 (maximum speed of vehicles in the outermost lane) and A3 display noteworthy heterogeneity in their contribution to high-risk level scenarios. Specifically, in Case AW, X3 makes a negative contribution to high crash risk, while A3 has a positive impact, but in Case BW, the contributions of these two features are reversed. This inconsistency may be related to the vehicles changing from the middle lane to the outermost lane. When a vehicle changes lanes, it tends to accelerate to maintain a distance from its following vehicle in the target

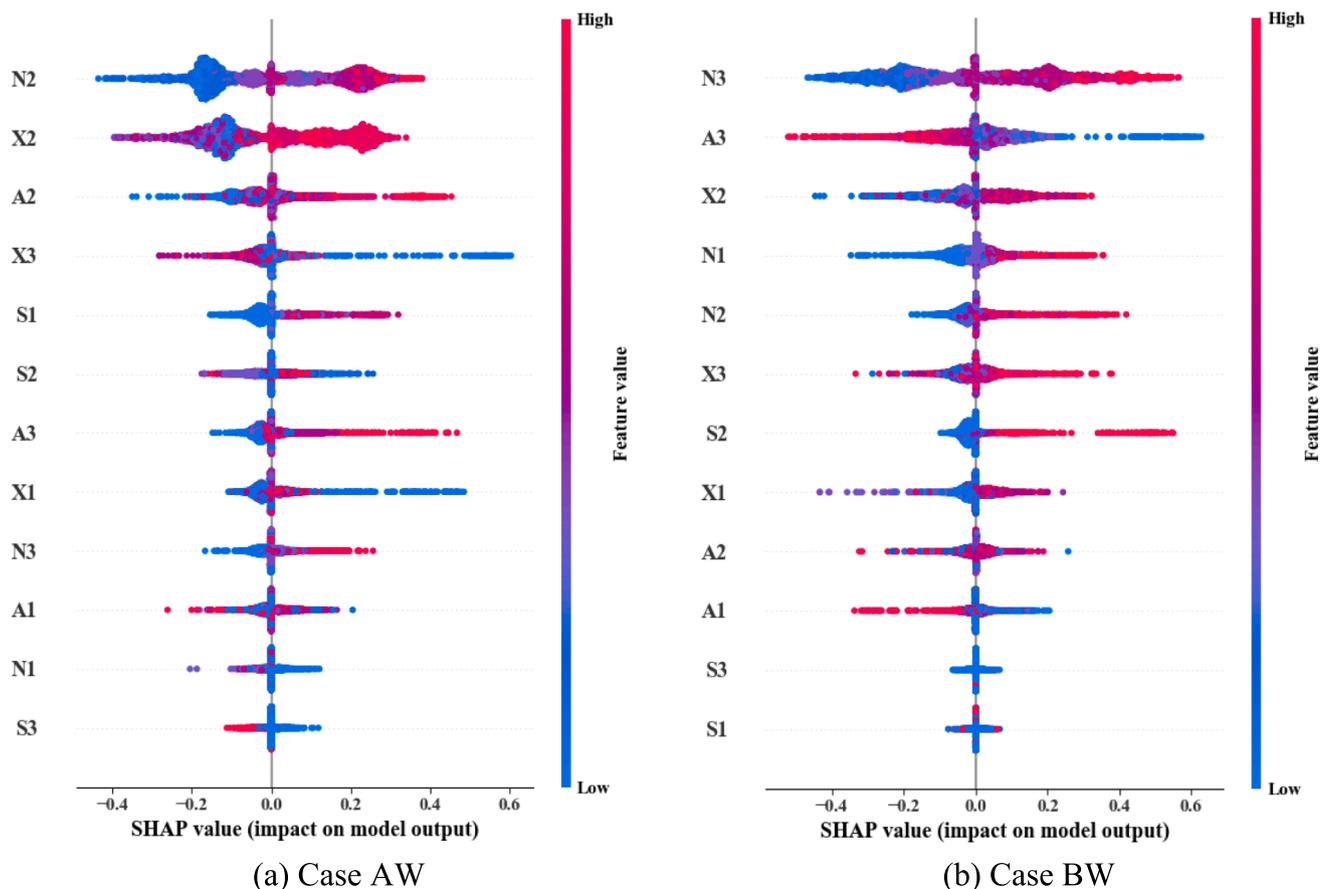


Fig. 11. SHAP summary plot.

lane. Meanwhile, the following vehicle in the target lane usually slows to avoid a crash with the lane-changing vehicle (Hidas, 2002). This means that when there is an increasing number of vehicles changing lane to the outermost lane, X3 tends to be higher, while A3 tends to be lower. As shown in Table 3, in Cases AW and BW, 4.92 % and 44.33 % of the vehicles in the middle lane change lane to the outermost lane. Thus, lane changing maneuvers might be a main cause of the heterogeneity in the contribution to high-risk level scenarios between X3 and A3. Additionally, from the collected video footage, we observed a high-risk situation where the vehicle in the middle lane was forced to change lanes to the outermost lane to avoid crashes with the vehicle changing lanes from the innermost lane. In Case BW, a total of 20 vehicles in the middle lane were involved in this high-risk situation, and Fig. 12 illustrates one of them as an example. This finding reveals how lane changing maneuvers might contribute to the crash risk of the outermost lane in a work zone.

In Fig. 11(a), we can determine that the key risk-contributing features for Case AW are the N2, X2, A2, X3, and S1 (number of speeding vehicles in the innermost lane), among which the N2, A2, and S1 make a positive contribution to high crash risk. The findings indicate that in scenario AW, high-risk scenarios exhibit the following characteristics when compared to low-risk scenarios: a) increased traffic flow in the middle lane; b) higher speeds observed in the middle lane; and c) more speeding vehicles entering the transition area from the innermost lane. Additionally, X2 only has a positive contribution to high crash risk when its value is high, but S2 (number of speeding vehicles in the middle lane) makes only a small contribution with no discernible regularity. Meanwhile, only three of Case AW's high-risk vehicles were speeding vehicles. This suggests that the current highest speed limit (i.e., 80 km/h) in the transition area may not be sufficient to ensure road safety. Therefore, when developing traffic safety management strategies for Case AW, the primary focus should be on lowering the highest operating speed limit in the middle lane.

Additionally, as shown in Fig. 11(b), the key risk-contributing features for Case BW are the N3, A3, X2, N1 (number of vehicles in the innermost lane), and N2. Among them, the N3, X2, N1, and N2 all make a positive contribution to high crash risk. Combining these results with those of the feature analysis of A3 above, in scenario AW, high-risk scenarios exhibit the following characteristics when compared to low-risk scenarios: a) increased traffic flow in road segment; b) higher speeds observed in the middle lane; and c) more vehicles changing lanes to the outermost lane. Additionally, there are 30 high-risk vehicles that were speeding in Case BW. The positive contribution of S2 (number of speeding vehicles in the middle lane) to high crash risk also indicates a serious speeding problem in the middle lane of the transition area. This might be related to the drivers not carefully observing the speed limit signs and not having enough time to slow in the transition area. Moreover, according to the analysis results in Section 5.2.1, many vehicles changing lanes to the outermost lane is a major reason for the higher crash risk in Case BW. Therefore, traffic safety management strategies for Case BW could be developed from the following three perspectives: a) optimize the speed limit sign placement (e.g., adequate distance from the transition area and ensuring that signs are visible enough); b) improve speeding vehicle management; c) reduce traffic flows in the outermost lane (e.g., opening emergency lanes); and d) reduce the

number of vehicles changing lanes to the outermost lane in the transition area (i.e., prompt vehicles to change lanes prior to entering the work zone).

6. Conclusion

In this study, a comprehensive approach is proposed to investigate crash risks and identify key risk-contributing features in transition areas of highway work zones. The primary contributions are summarized as follows. First, at the methodological level, we improve the safety potential field model and propose an SSM, that is WCRF force, for measuring crash risks. With the improvements made from multiple perspectives, the WCRF force is more practical to implement and is better suited for measuring crash risks in work zones than conventional SSMs including existing field-based models. Second, at the application level, we collected a large number of vehicle trajectories data from surveillance videos, sourced from two road segments under two conditions: with and without work zones. Based on this, we investigate key risk-contributing features and their heterogeneity in crash risk across different work zones on a highway. A case study is conducted to validate and examine the proposed approach, and we obtain the following primary conclusions from the case study.

- 1) Compared to TTC and DRAC, the WCRF force developed in this study offers a more comprehensive and accurate measurement of crash risks in complex traffic environments. This is due to the advantages of the safety potential field in describing the crash risk of multifactor influence and multitarget interaction. The WCRF force not only identifies the vehicle's extreme risk state, but also accounts for the influence of lateral safety distance, vehicle type, and road grade on crash occurrences.
- 2) Setting up a work zone increases the crash risks on that road segment. The findings of the case study indicate that closing the innermost lane may not only increase the crash risk in the middle lane but also impact the crash risk in the outermost lane. Comparative analysis of crash risks in each lane indicates that the higher frequency of lane changes among vehicles might be the primary cause for these increased crash risks.
- 3) The influence of traffic features on crash risks exhibits great heterogeneity across different work zones. Not only are there variations in the key risk-contributing features of the transition area in different work zones, but the same feature might also make opposite contributions to high crash risk. These findings emphasize the importance of considering the key risk-contributing features specific to each work zone and implementing targeted improvement measures when developing traffic management strategies.

However, this study still has some limitations. First, from the perspective of application, this study solely focuses on the transition area and investigates only one type of work zone, that is, the closure of the innermost lane on the highway segment with three lanes in each direction. Second, from the perspective of methodology, method selection or optimization is not a focus of this study. We did not compare the performances of various classifiers, examine the effects of different time

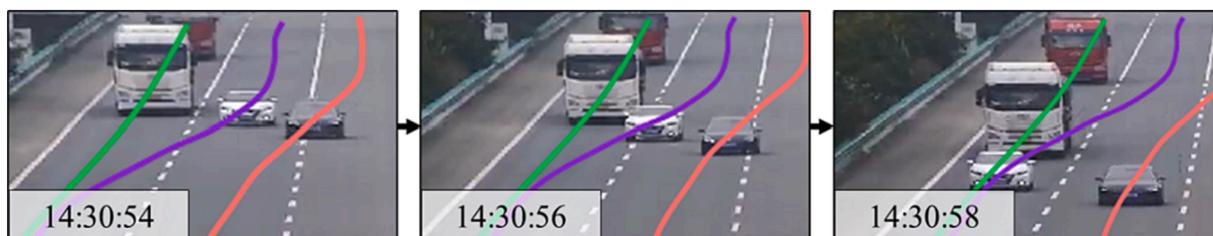


Fig. 12. Example of forced lane-changing maneuver. The white car is forced by the black car to change from the middle lane to the outermost lane.

windows on classification performance, or consider the pros and cons of various resampling algorithms. Hence, in future work, we will attempt to explore different types of work zones as well as the other areas apart from the transition area within the work zone, aiming to gain a more comprehensive understanding of the key risk-contributing factors in work-zone high-risk scenarios. Additionally, we will further optimize the proposed approach by incorporating the resampling algorithms and classifiers with more advanced performance (Ding et al., 2022; Yuan et al., 2019). Moreover, we will continue to improve the proposed approach for identifying risk-contributing features more effectively and efficiently, such as adopting more cutting-edge machine learning techniques and optimizing time window selection.

CRedit authorship contribution statement

Bo Wang: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Visualization. **Tianyi Chen:** Supervision, Writing – review & editing. **Chi Zhang:** Supervision, Project administration, Funding acquisition, Resources. **Yiik Diew Wong:** Supervision, Writing – review & editing. **Hong Zhang:** Writing – review & editing. **Yunhao Zhou:** Software.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

This work was supported by the Shaanxi Province Natural Science Basic Research Project General Program (No. 2023-JC-YB-391), the Key Research and Development Program of Sichuan Province (No. 2022YFG0048), the Key Research and Development Program of Shanxi Province (No. 202102020101014), and the Science and Technology Project of Sichuan Transportation Department (No. 2022-ZL-04). A fair amount of the research works are conducted in the Nanyang Technological University to which the first author was attached as a visiting PhD candidate sponsored by the China Scholarship Council.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.aap.2023.107361>.

References

- Federal Highway Administration, 2022. Work Zone Facts and Statistics, March 2022 Edition, FHWA, Washington, D.C., U. S. Available from: https://ops.fhwa.dot.gov/wz/resources/facts_stats.htm. Accessed February 26, 2023.
- Adomah, E., Bakshi, A.K., Ahmed, M.M., 2022. Safety impact of connected vehicles on driver behavior in rural work zones under foggy weather conditions. *Transp. Res. Rec.* 2676 (3), 88–107. <https://doi.org/10.1177/03611981211049147>.
- Almallah, M., Hussain, Q., Alhajyaseen, W.K.M., Pirdavani, A., Brijis, K., Dias, C., Brijis, T., 2021. Improved traffic safety at work zones through animation-based variable message signs. *Accid. Anal. Prev.* 159, 106284 <https://doi.org/10.1016/j.aap.2021.106284>.
- American Association of State Highway and Transportation Officials, 2018. A policy on geometric design of highways and streets, 7th ed. AASHTO, Washington, D.C., U. S.
- Arun, A., Haque, M.M., Bhaskar, A., Washington, S., Sayed, T., 2021. A systematic mapping review of surrogate safety assessment using traffic conflict techniques. *Accid. Anal. Prev.* 153, 106016 <https://doi.org/10.1016/j.aap.2021.106016>.
- Arun, A., Haque, M.M., Washington, S., Mannering, F., 2023. A physics-informed road user safety field theory for traffic safety assessments applying artificial intelligence-based video analytics. *Anal. Methods. Accid. Res.* 37, 100252 <https://doi.org/10.1016/j.amar.2022.100252>.
- Bella, F., 2005. Validation of a driving simulator for work zone design. *Transp. Res. Rec.* 1937 (1), 136–144. <https://doi.org/10.1177/0361198105193700119>.
- Bidkar, O., Arkatkar, S., Joshi, G., Easa, S.M., 2023. Effect of construction work zone on rear-end conflicts by vehicle type under heterogeneous traffic conditions. *J. Transp. Eng. a. Syst.* 149 (4), 05023001. <https://doi.org/10.1061/JTEPBS.TEENG-7275>.
- Budhkar, A.K., Maurya, A.K., 2017. Characteristics of lateral vehicular interactions in heterogeneous traffic with weak lane discipline. *J. Mod. Transp.* 25 (2), 74–89. <https://doi.org/10.1007/s40534-017-0130-1>.
- Chatterjee, I., Edara, P., Menneni, S., Sun, C., 2009. Replication of work zone capacity values in a simulation model. *Transp. Res. Rec.* 2130, 138–148. <https://doi.org/10.3141/2130-17>.
- Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* 16, 321–357. <https://doi.org/10.1613/jair.953>.
- Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, pp.785–794. [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785).
- Chen, T., Shi, X., Wong, Y.D., 2019. Key feature selection and risk prediction for lane-changing behaviors based on vehicles' trajectory data. *Accid. Anal. Prev.* 129, 156–169. <https://doi.org/10.1016/J.AAP.2019.05.017>.
- Chiong, R., Fan, Z., Hu, Z., Dhakal, S., 2022. A Novel Ensemble Learning Approach for Stock Market Prediction Based on Sentiment Analysis and the Sliding Window Method. *IEEE Trans. Comput. Soc. Syst.* <https://doi.org/10.1109/TCSS.2022.3182375>.
- Code of Virginia. Title 46.2. Motor Vehicles. Chapter 8. Regulation of Traffic. § 46.2-880 (2021). Available: <https://law.lis.virginia.gov/vacode/title46.2/chapter8/section46.2-880> Accessed February 26, 2023.
- Ding, H., Lu, Y., Sze, N.N., Chen, T., Guo, Y., Lin, Q., 2022. A deep generative approach for crash frequency model with heterogeneous imbalanced data. *Anal. Methods Accid. Res.* 34, 100212. <https://doi.org/10.1016/J.AMAR.2022.100212>.
- Du, Y., Zhao, Z., Song, Y., Zhao, Y., Su, F., Gong, T., Meng, H., 2023. StrongSORT: make deepsort great again. arXiv preprint arXiv, 2202, 13514. [10.48550/arXiv.2202.13514](https://arxiv.org/abs/2202.13514).
- Dubská, M., Herout, A., Sochor, J., 2014. Automatic camera calibration for traffic understanding. In: Nottingham, U.K. (Ed.), In: Proceedings of the British Machine Vision Conference 2014, pp. 1–12.
- Ge, H., Xia, R., Sun, H., Yang, Y., Huang, M., 2019. Construction and simulation of rear-end conflicts recognition model based on improved TTC algorithm. *IEEE Access* 7, 134763–134771. <https://doi.org/10.1109/ACCESS.2019.2937898>.
- Gertz, M., Grosse-Butenuth, K., Junge, W., Maassen-Francke, B., Renner, C., Sparenberg, H., Krieter, J., 2020. Using the XGBoost algorithm to classify neck and leg activity sensor data using on-farm health recordings for locomotor-associated diseases. *Comput. Electron. Agric.* 173, 105404 <https://doi.org/10.1016/J.COMPAG.2020.105404>.
- Gunay, B., 2007. Car following theory with lateral discomfort. *Transp. Res. Part B Methodol.* 41 (7), 722–735. <https://doi.org/10.1016/j.trb.2007.02.002>.
- Guo, Y., Sayed, T., Zheng, L., 2020b. A hierarchical bayesian peak over threshold approach for conflict-based before-after safety evaluation of leading pedestrian intervals. *Accid. Anal. Prev.* 147, 105772. [10.1016/J.AAP.2020.105772](https://doi.org/10.1016/J.AAP.2020.105772).
- Guo, Y., Sayed, T., Essa, M., 2020a. Real-time conflict-based Bayesian Tobit models for safety evaluation of signalized intersections. *Accid. Anal. Prev.* 144, 105660 <https://doi.org/10.1016/J.AAP.2020.105660>.
- Ha, D.H., Aron, M., Cohen, S., 2012. Time headway variable and probabilistic modeling. *Transp. Res. Part C Emerg. Technol.* 25, 181–201. [10.1016/j.trc.2012.06.002](https://doi.org/10.1016/j.trc.2012.06.002).
- Hidas, P., 2002. Modelling lane changing and merging in microscopic traffic simulation. *Transp. Res. Part C Emerg. Technol.* 10 (5–6), 351–371. [https://doi.org/10.1016/S0968-090X\(02\)00026-8](https://doi.org/10.1016/S0968-090X(02)00026-8).
- Hou, G., Chen, S., 2020. Study of work zone traffic safety under adverse driving conditions with a microscopic traffic simulation approach. *Accid. Anal. Prev.* 145, 105698 <https://doi.org/10.1016/j.aap.2020.105698>.
- Hu, Y., Li, Y., Huang, H., Lee, J., Yuan, C., Zou, G., 2022. A high-resolution trajectory data driven method for real-time evaluation of traffic safety. *Accid. Anal. Prev.* 165, 106503 <https://doi.org/10.1016/j.aap.2021.106503>.
- Huang, Y., Bai, Y., 2014. Effectiveness of graphic-aided portable changeable message signs in reducing vehicle speeds in highway work zones. *Transp. Res. Part C Emerg. Technol.* 48, 311–321. <https://doi.org/10.1016/j.trc.2014.09.007>.
- Idewu, W.I.A., Wolshon, B., 2010. Joint merge and its impact on merging speeds in lane reduction areas of construction zone. *Transp. Res. Rec.* 2169 (1), 31–39. <https://doi.org/10.3141/2169-04>.
- Kanhere, N.K., Birchfield, S.T., 2010. A taxonomy and analysis of camera calibration methods for traffic monitoring applications. *IEEE Trans. Intell. Transp. Syst.* 11 (2), 441–452. <https://doi.org/10.1109/TITS.2010.2045500>.
- Kim, J.A., Sung, J.Y., Park, S.H., 2020. Comparison of Faster-RCNN, YOLO, and SSD for real-time vehicle type recognition. In: Proceedings of 2020 IEEE International Conference on Consumer Electronics-Asia, Seoul, Korea (south), pp. 1–4. <https://doi.org/10.1109/ICCE-Asia49877.2020.9277040>.
- Kordani, A.A., Rahmani, O., Saman, A., Nasiri, A., Boroomandrad, S.M., 2018. Effect of adverse weather conditions on vehicle braking distance of highways. *Civil Eng. J.* 4 (1), 46–57. <https://doi.org/10.28991/cej-030967>.
- Kuang, Y., Qu, X., Wang, S., 2015. A tree-structured crash surrogate measure for freeways. *Accid. Anal. Prev.* 77, 137–148. <https://doi.org/10.1016/j.aap.2015.02.007>.
- Kummetha, V.C., Kondyli, A., Chrysikou, E.G., Schrock, S.D., 2020. Safety analysis of work zone complexity with respect to driver characteristics — A simulator study employing performance and gaze measures. *Accid. Anal. Prev.* 142, 105566 <https://doi.org/10.1016/j.aap.2020.105566>.

- La Torre, F., Domenichini, L., Nocentini, A., 2017. Effects of stationary work zones on motorway crashes. *Saf. Sci.* 92, 148–159. <https://doi.org/10.1016/j.ssci.2016.10.008>.
- Li, Y., Bai, Y., 2009. Highway Work Zone Risk Factors and Their Impact on Crash Severity. *J. Transp. Eng.* 135 (10), 694–701. [https://doi.org/10.1061/\(ASCE\)TE.1943-5436.00000055](https://doi.org/10.1061/(ASCE)TE.1943-5436.00000055).
- Li, L., Gan, J., Qu, X., Mao, P., Ran, B., 2019. Car-following model based on safety potential field theory under connected and automated vehicle environment. *China J. Highw. Transp.* 32 (12), 76–87. <https://doi.org/10.1017/CBO9781107415324.004>.
- Li, L., Gan, J., Ji, X., Qu, X., Ran, B., 2020a. Dynamic driving risk potential field model under the connected and automated vehicles environment and its application in car-following modeling. *IEEE Trans. Intell. Transp. Syst.* 23 (1), 122–141. <https://doi.org/10.1109/ITITS.2020.3008284>.
- Li, L., Gan, J., Yi, Z., Qu, X., Ran, B., 2020b. Risk perception and the warning strategy based on safety potential field theory. *Accid. Anal. Prev.* 148, 105805 <https://doi.org/10.1016/j.aap.2020.105805>.
- Li, L., Gan, J., Zhou, K., Qu, X., Ran, B., 2020c. A novel lane-changing model of connected and automated vehicles: Using the safety potential field theory. *Physica A* 559, 125039. <https://doi.org/10.1016/j.physa.2020.125039>.
- Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* 2017, 1–10.
- Ma, Y., Dong, F., Yin, B., Lou, Y., 2023. Real-time risk assessment model for multi-vehicle interaction of connected and autonomous vehicles in weaving area based on risk potential field. *Physica A* 620, 128725. <https://doi.org/10.1016/j.physa.2023.128725>.
- Mahmud, S.M.S., Ferreira, L., Hoque, M.S., Tavassoli, A., 2017. Application of proximal surrogate indicators for safety evaluation: A review of recent developments and research needs. *IATSS Res.* 41 (4), 153–163. <https://doi.org/10.1016/j.iatssr.2017.02.001>.
- Maity, M., Banerjee, S., Sinha Chaudhuri, S., 2021. Faster R-CNN and YOLO based vehicle detection: A survey. In: Proceedings of the 5th international conference on computing methodologies and communication, Erode, India. pp.1442-1447. <https://doi.org/10.1109/ICCMC51019.2021.9418274>.
- Mashhadi, A.H., Farhadmanesh, M., Rashidi, A., Marković, N., 2021. Review of Methods for Estimating Construction Work Zone Capacity. *Transp. Res. Rec.* 2675 (9), 382–397. <https://doi.org/10.1177/03611981211002202>.
- Meng, Q., Weng, J., 2011. Evaluation of rear-end crash risk at work zone using work zone traffic data. *Accid. Anal. Prev.* 43 (4), 1291–1300. <https://doi.org/10.1016/j.aap.2011.01.011>.
- Ministry of Transport of the People Republic of China, 2017. Road traffic signs and markings Part 4: Work zone (GB 5768.4-2017). China MOT, Beijing, China.
- Mullakkal-Babu, F.A., Wang, M., He, X., van Arem, B., Happee, R., 2020. Probabilistic field approach for motorway driving risk assessment. *Transp. Res. Part C Emerg. Technol.* 118, 102716 <https://doi.org/10.1016/j.trc.2020.102716>.
- National Highway Traffic Safety Administration, 2019. *Traffic Safety Facts 2019*. NHTSA, Washington, D.C., U. S.
- Nnaji, C., Gambate, J., Lee, H.W., Zhang, F., 2020. Improving construction work zone safety using technology: A systematic review of applicable technologies. *J. Traffic Transp. Eng.* 7 (1), 61–75. <https://doi.org/10.1016/j.jtte.2019.11.001>.
- Osman, M., Paleti, R., Mishra, S., 2018. Analysis of passenger-car crash injury severity in different work zone configurations. *Accid. Anal. Prev.* 111, 161–172. <https://doi.org/10.1016/j.aap.2017.11.026>.
- Park, H., Oh, C., Moon, J., Kim, S., 2018. Development of a lane change risk index using vehicle trajectory data. *Accid. Anal. Prev.* 110, 1–8. <https://doi.org/10.1016/j.aap.2017.10.015>.
- Pinnow, J., Masoud, M., Elhenawy, M., Glaser, S., 2021. A review of naturalistic driving study surrogates and surrogate indicator viability within the context of different road geometries. *Accid. Anal. Prev.* 157, 106185 <https://doi.org/10.1016/j.aap.2021.106185>.
- Polyzotis, N., Roy, S., Whang, S.E., Zinkevich, M., 2017. Data management challenges in production machine learning. In: Proceedings of the ACM SIGMOD International Conference on Management of Data Part F127746, pp.1723–1726. <https://doi.org/10.1145/3035918.3054782>.
- Qi, H., Yao, Y., Zhao, X., Guo, J., Zhang, Y., Bi, C., 2022. Applying an interpretable machine learning framework to the traffic safety order analysis of expressway exits based on aggregate driving behavior data. *Phys. A Stat. Mech. Appl.* 597, 127277 <https://doi.org/10.1016/j.physa.2022.127277>.
- Samson, C.J.R., Hussain, Q., Alhajyaseen, W.K.M., 2022. Analysis of Stopping Sight Distance (SSD) Parameters: A Review Study. *Procedia Comput. Sci.* 201, 126–133. <https://doi.org/10.1016/j.procs.2022.03.019>.
- Shbeeb, L., 2000. Development of traffic conflicts technique for different environments: A comparative study of pedestrian conflicts in Sweden and Jordan. Lund University, Lund.
- Shi, X., Wong, Y.D., Li, M.Z.F., Chai, C., 2018. Key risk indicators for accident assessment conditioned on pre-crash vehicle trajectory. *Accid. Anal. Prev.* 117, 346–356. <https://doi.org/10.1016/j.aap.2018.05.007>.
- Shi, X., Wong, Y.D., Li, M.Z.F., Palanisamy, C., Chai, C., 2019. A feature learning approach based on XGBoost for driving assessment and risk prediction. *Accid. Anal. Prev.* 129, 170–179. <https://doi.org/10.1016/j.aap.2019.05.005>.
- Singapore Land Transport Authority, 2019. Code of Practice: Traffic Control at Work Zone, Singapore LTA, 2019 Edition, Singapore.
- St-Aubin, P., Saunier, N., Miranda-Moreno, L., 2015. Large-scale automated proactive road safety analysis using video data. *Transp. Res. Part C Emerg. Technol.* 58, 363–379. <https://doi.org/10.1016/j.trc.2015.04.007>.
- Venthuruthiyil, S.P., Chunchu, M., 2018. Trajectory reconstruction using locally weighted regression: a new methodology to identify the optimum window size and polynomial order. *Transportmetrica A: Transport Science* 14 (10), 881–900. <https://doi.org/10.1080/23249935.2018.1449032>.
- Wang, K., Huang, H., Li, Y., Wang, F.Y., 2007. Research on lane-marking line based camera calibration. In: Proceedings of 2007 IEEE International Conference on Vehicular Electronics and Safety, Beijing, China, pp.1-6. <https://doi.org/10.1109/ICVES.2007.4456361>.
- Wang, K., Xue, Q., Xing, Y., Li, C., 2020. Improve Aggressive Driver Recognition Using Collision Surrogate Measurement and Imbalanced Class Boosting. *Int. J. Environ. Res. Public Health.* 2020, Vol. 17, Page 2375 17 7, 2375. <https://doi.org/10.3390/ijerph17072375>.
- Wang, B., Zhang, C., Wong, Y.D., Hou, L., Zhang, M., Xiang, Y., 2022. Comparing resampling algorithms and classifiers for modeling traffic risk prediction. *Int. J. Environ. Res. Public Health.* 19(20), 13693. <https://doi.org/10.3390/ijerph192013693>.
- Wang, J., Wu, J., Li, Y., 2015. The driving safety field based on driver-vehicle-road interactions. *IEEE Trans. Intell. Transp. Syst.* 16 (4), 2203–2214. <https://doi.org/10.1109/ITITS.2015.2401837>.
- Wang, J., Wu, J., Zheng, X., Ni, D., Li, K., 2016. Driving safety field theory modeling and its application in pre-collision warning system. *Transp. Res. Part C Emerg. Technol.* 72, 306–324. <https://doi.org/10.1016/j.trc.2016.10.003>.
- Wang, J., Song, H., Fu, T., Behan, M., Jie, L., He, Y., Shangquan, Q., 2022. Crash prediction for freeway work zones in real time: A comparison between Convolutional Neural Network and Binary Logistic Regression model. *Int. J. Transp. Sci. Technol.* 11 (3), 484–495. <https://doi.org/10.1016/j.ijst.2021.06.002>.
- Weng, J., Xue, S., Yan, X., 2016. Modeling vehicle merging behavior in work zone merging areas during the merging implementation period. *IEEE Trans. Intell. Transp. Syst.* 17 (4), 917–925. <https://doi.org/10.1109/ITITS.2015.2477335>.
- Weng, J., Du, G., Li, D., Yu, Y., 2018. Time-varying mixed logit model for vehicle merging behavior in work zone merging areas. *Accid. Anal. Prev.* 117, 328–339. <https://doi.org/10.1016/j.aap.2018.05.005>.
- Weng, J., M, Q., Yan, X., 2014. Analysis of work zone rear-end crash risk for different vehicle-following patterns. *Accid. Anal. Prev.* 72, 449–457. <https://doi.org/10.1016/j.aap.2014.08.003>.
- Wolf, M.T., Burdick, J.W., 2008. Artificial potential functions for highway driving with collision avoidance. In: Proceedings of 2008 IEEE International Conference on Robotics and Automation, Pasadena, pp.3731–3736. <https://doi.org/10.1109/ROBOT.2008.4543783>.
- Xin, F., Wang, X., Sun, C., 2021. Risk evaluation for conflicts between crossing pedestrians and right-turning vehicles at intersections. *Transp. Res. Rec.* 2675 (12), 1005–1014. <https://doi.org/10.1177/03611981211031885>.
- Yang, G., Ahmed, M., Adomah, E., 2020. An integrated microsimulation approach for safety performance assessment of the Wyoming connected vehicle pilot deployment program. *Accid. Anal. Prev.* 146, 105714 <https://doi.org/10.1016/j.aap.2020.105714>.
- Yang, H., Ozbay, K., Ozturk, O., Yildirimoglu, M., 2013. Modeling work zone crash frequency by quantifying measurement errors in work zone length. *Accid. Anal. Prev.* 55, 192–201. <https://doi.org/10.1016/j.aap.2013.02.031>.
- Yuan, J., Abdel-Aty, M., Gong, Y., Cai, Q., 2019. Real-Time Crash Risk Prediction using Long Short-Term Memory Recurrent Neural Network. *Transp. Res. Rec.* 2673 (4), 314–326. <https://doi.org/10.1177/0361198119840611>.
- Yuan, C., Li, Y., Huang, H., Wang, S., Sun, Z., Wang, H., 2022. Application of explainable machine learning for real-time safety analysis toward a connected vehicle environment. *Accid. Anal. Prev.* 171, 106681 <https://doi.org/10.1016/j.aap.2022.106681>.
- Zhang, C., Zhang, H., Ma, X., Zhang, M., Wang, S., 2018. Driving risk assessment in work zones using cloud model. *Math. Probl. Eng.* 2018 <https://doi.org/10.1155/2018/8759580>.
- Zhang, C., Wang, B., Yang, S., Zhang, M., Gong, Q., Zhang, H., 2020. The driving risk analysis and evaluation in rightward zone of expressway reconstruction and extension engineering. *J. Adv. Transp.* 2020 <https://doi.org/10.1155/2020/8943463>.
- Zhang, C., Wei, J., Hu, A.S., Fu, P., 2022. A novel method for calibration and verification of roadside millimetre-wave radar. *IET Intel. Transport Syst.* 16 (3), 408–419. <https://doi.org/10.1049/ITR2.12151>.