



Travel route safety estimation based on conflict simulation

Helai Huang^a, Yulu Wei^a, Chunyang Han^{b,*}, Jaeyoung Lee^a, Suyi Mao^a, Fan Gao^a

^a School of Traffic and Transportation Engineering, Smart Transport Key Laboratory of Hunan Province, Central South University, Changsha, China

^b Department of Automation, Tsinghua University, Beijing, China



ARTICLE INFO

Keywords:

Travel route safety estimation
Urban road network
Exposure measure
Conflict rate prediction
Route-based safety impedance

ABSTRACT

With the aim of providing travelers information about the safety levels of selectable routes, it is necessary to develop a method that can properly estimate the safety of alternative travel routes. This paper proposes a conflict-based approach for travel route safety estimation (TRSE). It is developed on the basis of the classical safety evaluation model where both the amount of exposure to safety risk and the risk under unit exposure are measured to estimate the route safety. A combination of a set of dynamic and static factors related to traffic flow characteristics and roadway features are selected to estimate conflict exposure and potential conflict risk. A route-based method is employed where two parallel estimations of conflict are conducted for both the component segments links and intersection turning links. Three machine learning models (i.e., random forest, k-nearest neighbor, and support vector machine) are tested in conflict risk estimation. A fuzzy reasoning process based on the fuzzy logic algorithm is employed to conduct the route safety estimation. The proposed TRSE is tested on a four-horizontal and six-vertical network extracted from a real road network in China. Conflict simulation results were obtained by Vissim and SSAM tools. The results illustrate the practicability and effectiveness of the proposed TRSE approach.

1. Introduction

In route navigation systems, the cost and efficiency-related criteria, such as shorter distance (Choi et al., 2014) and shorter travel time (George and Kim, 2013) are primary considerations. Recently, air pollution associated with road transportation has also been studied in route choice models (Zou et al., 2020). Undoubtedly, traffic safety is another equally important factor that potentially influences travelers' decision-making process (Näätänen and Summala, 1974). The advancement of real-time data collection and fast-processing technologies make it feasible to account the safer route that might reduce crash risk during the trip. The real-time estimation provides travelers a timely information about route safety, allowing them to have pre-trip or enroute information in making the route choice. To this end, the major challenge is to develop a method dealing with safety estimation of candidate travel routes between any available O-D pair, which is hereinafter referred to as travel route safety estimation (TRSE).

Unlike the travel efficiency that has standard time measurement, it is challenging to accurately measure the crash risk potential of a vehicle navigating through a series of segments and intersections by existing methods (Chandra, 2014). First, the amount of crash exposure for the

trip relates to the conditions of surrounding traffic and the distance of the selected route. Second, the risk status is associated with a large number of factors which are partly static and deterministic and partly stochastic and dynamic on a continual basis (Han et al., 2018; Huang and Abdel-Aty, 2010), leading to the temporal variation in the level of crash risk. Moreover, the component links of a route pass through a number of segments and intersections only cover a specific traffic approach of them (e.g., the short path occupying the left-turn bay of an intersection), which requires TRSE to account for the safety of the traffic approaches covered by the route, rather than that of the entire road entities.

Recently, several studies have dealt with TRSE modeling. The historic crash frequencies were deemed as the most direct way to represent road safety. However, the crash records only reflect safety that the road has long been performed in the past. It is insufficient to indicate the safety status at the period of traveling, and thus is hard to provide travelers dynamic safe-route navigation. Besides, other studies estimated the route safety by evaluating the possible effects of the stationary geometric features and traffic control characteristics of the route components, such as the baseline safety estimation using negative binomial model by Jiang et al. (2020) and the driving-challenge-based route risk

* Corresponding author.

E-mail address: sandiant@foxmail.com (C. Han).

measurements proposed by Payyanadan et al. (2017). This type of methods has limitations in accounting for the dynamic changes of risk factors, and thus cannot provide a real-time route safety estimation. The real-time risk effects of dynamic traffic factors were considered in Jiang et al. (2020), but the cumulative risk effect of exposure was not accounted for. The result can only provide travelers a real-time traffic safety risk status, but incapable of making a reliable comparison of safety between alternative routes.

This paper intends to develop an approach for TRSE. The classical safety estimation function that defines the road safety as the product of traffic exposure and associated risk factors is held as the basis. Considering the randomness and rarity of crash occurrence, traffic conflicts is used as an indicator to characterize the real-time route safety level. The approach applies the route-based safety impedance estimation method, where two parallel safety estimations of both the component segment links and intersection turning links are conducted. The amount of exposure and the conflict risk are measured by the factors that represent the stationary features of road geometry and traffic control, and characteristics of dynamic traffic state that the traveler may encounter on the route. The fuzzy logic algorithm is used to aggregate the safety status of the components at the route level. For the purpose of illustration, the proposed TRSE approach is performed on the Vissim simulation platform. A four-horizontal and six-vertical network extracted from a real road network in China was established, and a series of calibration and validation works are conducted to ensure high reliability of the simulation.

2. Literature review

2.1. Route impedance

In the urban network, the travel impedance estimation methods can be divided into link-based and route-based methods. The link-based method either assumes that the route impedance is a simple addition of the travel impedances of its component links (Chen and Chien, 2001), or calculates the route impedance without considering the influence of intersection (Abdelhamid et al., 2017). For example, in the study of Jiang et al. (2014), the travel time of route is the sum of the travel times calculated from each group of detectors, and there is no obvious difference in the treatment of segments and intersections. Abdelhamid et al. (2017) focuses on scoring road segments on travel time, road health, and driver preferences to seek the most suitable route, and lacks consideration of intersections. In contrast, the route-based method estimates the impedance of segment and intersection from the view of the entire route, which provides a more scientific choice from the attributes of the route. For example, in the study of Jenelius and Koutsopoulos (2013), the travel time of a route is equal to the travel times of segments and the delay time of intersections, and in the Chandra's (2014) study, different forms of safety indicator were developed for the segment and intersection respectively, which were integrated to represent the route safety, Payyanadan et al. (2017) measured the safety risk of the entire route by quantifying driving risks such as road segments, left turns, and U-turns on the route. The difference between link-based method and route-based method lies in the special treatment of route elements.

The link-based method is prevailing in the earlier efficiency-based routing studies. The results mainly served for the traffic states identification for traffic management (Jiang et al., 2014; Chen and Chien, 2001). With the development of innovative data detection techniques, researchers have started to try generalized cost estimation directly from the route level (Chandra, 2014; Payyanadan et al., 2017). In addition to the applications in the field of efficiency and safety, both the link-based (Zou et al., 2020) and route-based (Alam et al., 2018) methods have been applied in the research of environment-concerning route planning.

2.2. Safety estimation

Crash frequency has long been taken as the major measure of road safety (Washington et al., 2020). Estimating the expected crash frequency given traffic exposure and associated risk factors is one common way to evaluate road safety performance (AASHTO, 2010). This technique aims to reflect the safety of underlying road characteristics objectively, which evaluate the long-term nature of road safety (Lord and Mannerling, 2010).

Alternatively, real-time crash prediction aims to predict the probability of a crash occurring on a specific road section within a very short time interval (Hossain et al., 2019). The traffic flow factors, such as the speed, volume, occupancy, and composition of the flow and their dynamic changes are the keys to the performance of the model (Basso et al., 2020; Orsini et al., 2021). Machine learning model has been used in many studies to build the real-time crash prediction model, due to its excellent performance in prediction accuracy and ability to deal with high-dimensional and nonlinear relationships in big data, such as decision tree model (Katrakazas et al., 2018), neural network (Jiang et al., 2020; Li et al., 2019), support vector machine (Lv et al., 2009), dynamic Bayesian networks (Katrakazas et al., 2019). Recently, deep learning models started to appear in this domain, such as the convolutional neural network (Huang et al., 2020), the long short-term memory neural network (Li et al., 2019), the deep convolutional generative adversarial network (Cai et al., 2020). The deep learning models have been found to have a better performance in prediction accuracy. For estimating real-time crash risk, traffic conflict have been commonly used as a surrogate measure of traffic safety (Lord et al., 2021). The traffic conflict indicators include the temporal-based proximity measures, such as the time-to-collision family (Hayward, 1971) and the encroachment time family (Cooper, 1984), and the distance-based proximity such as the proportion of stopping distance (Allen et al., 1978). Recently, Katrakazas et al. (2018) developed a real-time conflict prediction model using the data collected from a micro-simulation platform, while Formosa et al. (2020) applied the real-world traffic data collected with both loop detectors and in-vehicle sensors. Several other studies have also proposed real-time conflict prediction models for signalized intersections (Zheng and Sayed, 2020) and freeway section (Caleffi et al., 2017).

2.3. Influential factors

Crash occurrence is a complex process that is influenced by multiple factors (Lord et al., 2021). In traditional techniques, the aim is to provide a long-term safety performance evaluation for roadways facilities or safety countermeasures. Thus, the focus is laid on the roadway-related factors, such as road geometric features which including lane width (AASHTO, 2010), number of traffic lanes, widening approach (Park et al., 2015), road curvature (Haynes et al., 2007), intersection type (Cai et al., 2018), etc., and traffic control characteristics which including traffic signs and markings (Yao et al., 2019), traffic signals (Essa and Sayed, 2018), etc., and the factors reflecting the long-term traffic condition, such as annual average daily traffic (AADT) (Al-Omarie et al., 2021) and intersection total entering vehicles (TEV) (AASHTO, 2010).

In regard to the real-time safety estimation, an underlying assumption is that certain combinations of traffic conditions are relatively more 'crash-prone' than others (Roshandel et al., 2015). Therefore, by controlling for static geometric features and dynamic environmental status (Wu et al., 2018; Yasmin et al., 2018), the attentions are majorly paid to the factors reflecting the real-time traffic conditions such as the mean or deviation of speed, traffic density, traffic volume, and their combinations during a certain time interval (Shi and Abdel-Aty, 2015; Xu et al., 2013; Cai et al., 2020; Li et al., 2019). Furthermore, real-time traffic signals (Yuan et al., 2019), signal timing (Yuan et al., 2020), platoon ratio (Essa and Sayed, 2018), etc., were found to be the significant factors affecting the safety of signalized intersections. The application of

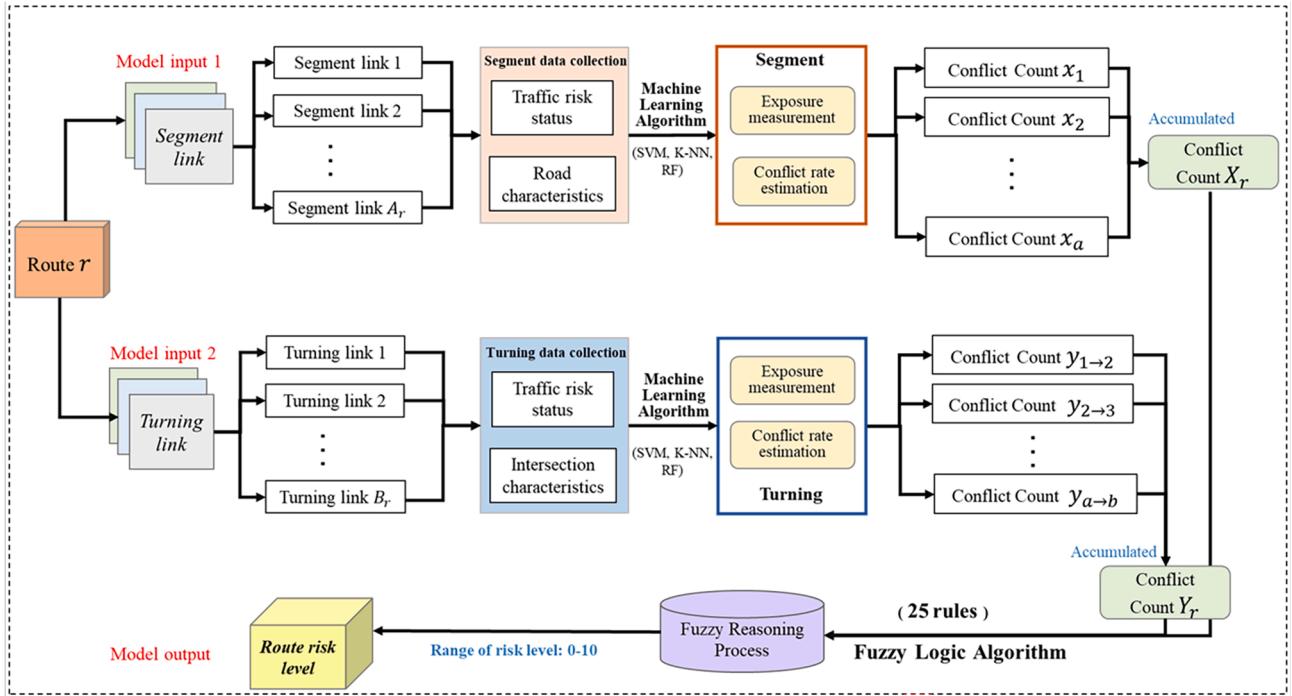


Fig. 1. Overall framework of TRSE.

real-time traffic features makes the field of real-time safety estimation develop rapidly.

3. Methodology

3.1. Overall modeling framework

As shown in Eq. (1), the safety of a system is defined as the product of the probability of having a crash (also called crash risk) under unit exposure and the observed exposure of the system during a specified period:

$$\text{Safety of the system} = \text{Crash risk of the system} \times \text{Number of exposure units of the system} \quad (1)$$

In this study, traffic conflict is used as the surrogate measure of route safety. Accordingly, conflict rate is predicted to represent the crash risk, and the amount of exposure is measured by the factors that contribute to the conflict occurrences.

A route-based method inspired by Jenelius and Koutsopoulos (2013) is proposed to calculate route safety impedance, which divides the route component into the segment links and intersection turning links, and calculates their safety by categories. This decomposition is meant to recognize the areas covered by the expected travel route and calculate the overall safety level of any combination of the areas which consist of the target route conveniently.

Figure 1 presents the overall framework of the proposed TRSE method, where A_r and B_r represent the total number of segment links and turning links contained in a route r , respectively; x_a and y_{a-b} denote the estimated conflict counts of segment link a and the turning link $a \rightarrow b$ connecting segment link a and segment link b , respectively; X_r and Y_r are the total conflict counts of all the segments and turnings of route r , respectively. Specifically, the estimation starts separately from two types of major route components (segment links and intersection turning links). Both the conflict rate and exposure are estimated for the segment link and the turning link, respectively. With the aim of pursuing a high estimation accuracy, a series of machine learning models are tested for selecting the best to conduct the conflict rate prediction. Then, the fuzzy

logic algorithm is used to make a reliable integration of segment links safety and turning links safety at the route level. The details of these techniques are described in the following sections.

3.2. Conflict exposure measurement

Driving exposure is defined as the “frequency of traffic events which create a risk of accident” (Hauer, 1982). In this study, two quantities including the traffic volume within a certain time interval and the distance traveled to finish a trip are selected as the exposure measures.

Specifically, note e_a and e_{a-b} be the conflict exposure of segment link a , and turning link $a \rightarrow b$. In this study, the segment link exposure e_a is measured by the miles that all the vehicles traveled through the segment a during a predefined time interval. That is the product of the traffic volume within the i th time interval and the length of segment:

$$e_a^i = v_a^i \times l_a^i, \forall a \in A; \forall i \in I \quad (2)$$

where v_a^i is the number of vehicles passing through segment a during the i th time interval (I is the total number of time intervals), and l_a^i is the length of the segment link a .

Since the travel time that the vehicles spend on different turnings at the intersection did not show significant difference, the exposure measured for turning link $a \rightarrow b$ is only accounted for by the vehicle count (i.e., total entering vehicles) within the i th time interval, which can be expressed as:

$$e_{a-b}^i = v_{a-b}^i, \forall a \rightarrow b \in B; \forall i \in I. \quad (3)$$

3.3. Conflict rate estimation

Accordingly, the conflict rate of segment link a can be expressed as:

$$R_a^i = \frac{N_a^i}{v_a^i \times l_a^i}, \quad \forall a \in A; \forall i \in I \quad (4)$$

where R_a^i and N_a^i are the conflict rate and conflict frequency of segment link a in the i th time interval. Similarly, the conflict rate of turning link $a \rightarrow b$ is calculated as follows:

Table 1

Detailed description of the factors.

| | | Factors | Description and Use | Units |
|---------|---------------|--------------------------|--|------------------|
| Segment | Static | Number of lanes | The number of lanes of the segment | – |
| | | Width | The width of the lane | m |
| | | Length | The length of the segment | m |
| | | Access Density | Ratio of the number of access roads to the length of the main road | – |
| | | Bus Station | If there is a bus stop on the segment, it is set to 1, otherwise it is 0 | – |
| | Dynamic | Widened Approach | If there is a widening approach on the segment, it is set as 1, otherwise as 0 | – |
| | | Speed | The average speed of vehicles on a segment during the fixed time(t) | m/s |
| | | Acceleration | The average acceleration of vehicles on a segment during the fixed time | m/s ² |
| | | Traffic Volume | The number of vehicles passing the segment during the fixed time | vehicles/t |
| | | Variance of Speed | Variance of the speed of all vehicles on a segment during the fixed time | m/s |
| Turning | Static | Variance of Acceleration | Variance of the acceleration of all vehicles on a segment during the fixed time | m/s ² |
| | | Intersection Type | Four-leg intersection is set as 1 and three-leg intersection is set as 0 | – |
| | | Turning Type | Left turnings set as 0, straight turnings set as 1, right turnings set as 2 | – |
| | Dynamic | Inherent Conflict Points | The number of points crossed with other turnings | – |
| | | Speed | The average speed of all vehicles on the turning during the fixed time | m/s |
| | | Traffic Volume | The number of all vehicles pass the turning during the fixed time | vehicles/t |
| | Platoon Ratio | Difference of Speed | The difference in average speed between upstream segment and downstream segment during the fixed time | m/s |
| | | Difference of Volume | The difference in traffic volume between upstream segment and downstream segment during the fixed time | vehicles/t |
| | | Platoon Ratio | The count of vehicles arriving during the green light time divided by the ratio of the effective green time to the fixed time interval | – |

$$R_{a \rightarrow b}^i = \frac{N_{a \rightarrow b}^i}{v_{a \rightarrow b}^i}, \forall a \rightarrow b \in B; \forall i \in I \quad (5)$$

where $R_{a \rightarrow b}^i$ and $N_{a \rightarrow b}^i$ are the conflict rate and conflict frequency of turning link $a \rightarrow b$ in the i th time interval.

This risk effect of unit exposure on safety relates to multiple factors representing the driver-vehicle-roadway/environment characteristics. For the modeling purpose, three machine learning models could be selected, including random forest (RF), k-nearest neighbor (K-NN), and support vector machine (SVM).

RF belongs to the group of ensemble learners and more specifically to the group of bagging algorithms (Katrakazas et al., 2018). RF classifier uses Gini coefficient or information gain as the loss function, and RF regressor uses mean square error (MSE) or absolute value difference (MAE) as the loss function. In this paper, the input samples for RF are represented as $\{\mathbf{x} = [x_1, x_2, \dots, x_n]\}$, $i = 1, 2, 3, \dots, m$, and m represents the size of data samples, n is the number of features adopted in the study. The conflict rate, as the output of the model, is a continuously changing label. Therefore, this paper chooses the RF to conduct the regression task, and the average of the output of all decision trees is the final output of the model.

SVM is a supervised learning method, which constructs one or a set of hyperplanes in a high-dimensional or infinite-dimensional space. These hyperplanes can be used for classification, regression, or other tasks. In the classification problems, SVM is to maximize the “distance” to the nearest sample point of the hyperplane (Kusakapan et al., 2021). The sparse solution and good generalization of support vector machines are conducive to adapting to regression problems (Basak et al., 2007), support vector regression (SVR) as the important branch of the application of SVM, which is to minimize the “distance” to the farthest sample point of the hyperplane. For SVR, the relationship between the input $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ and the corresponding output can be expressed by the following function:

$$R_{svr}(\mathbf{x}) = \omega^T \varphi(\mathbf{x}) + b \quad (6)$$

where R_{svr} refers to the output of the model, i.e., the conflict rate, ω is the weight vector, and $\varphi(\mathbf{x})$ indicates the kernel function, defining a nonlinear relationship between features, b refers to the bias.

K-nearest neighbor is a non-parametric model that can solve classification and regression problems. For classification problems, if most of the k neighbors of a sample in the feature space belong to a class, the sample also belongs to this particular class. For regression problems, the output result is to take the average of its k nearest neighbor samples.

3.4. Conflict risk factors of segment links

Both static and dynamic factors are selected as the estimators of the model. The static factors include the number of lanes and lane width which have commonly used to represent basic characteristics of road segment (AASHTO, 2010). Widening approach that may reduce conflicts at approach (Park et al., 2015) and bus stops which have a significant impact on segment crash risk are also take into consideration. Besides, access density which have been proved to have impact on safety (Eisele & Frawley, 2005) is also considered as the key factor, and is defined as:

$$d_a = \frac{z_a}{l_a} \quad (7)$$

where z_a is the number of road access on segment link a . As for the dynamic factors, traffic volume and the mean and variance of speed which have long been taken as the key dynamic factors in the real-time safety estimation, and vehicle acceleration and deceleration rate that often associate with the segment conflict frequency (Formosa et al., 2020), are involved in the segment conflict estimation model.

3.5. Conflict risk factors of intersection turning links

Since intersection's structure directly determines the streams of traffic, and the risk of having a conflict varies with different turning movements (Cai et al., 2018; Huang et al., 2016), the characteristics of the intersection where the route passing through, including its type and inherent conflict points, and the type of the turning where the route occupies are considered as the stationary contributing factors of conflicts. In regard to the dynamic factors, except the factors related flow and speed, the difference of average speed and the average acceleration between upstream and downstream traffic of the joint segment links of the turning are also considered. This study also consider platoon ratio as

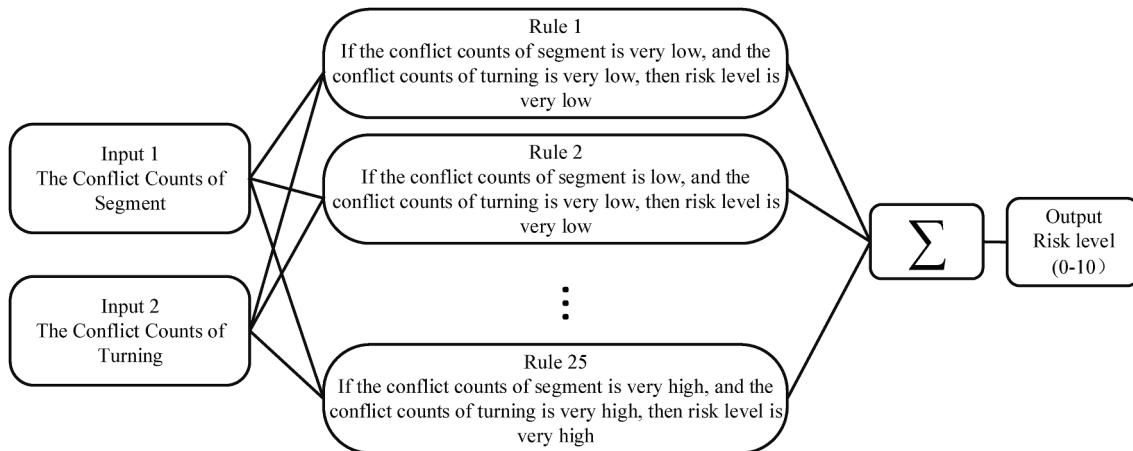


Fig. 2. Fuzzy reasoning process.

Table 2
Fuzzy rule matrix.

| $S \backslash T$ | VL | L | M | H | VH |
|------------------|----|----|---|----|----|
| VL | VL | VL | L | M | M |
| L | VL | L | L | M | M |
| M | L | L | M | M | H |
| H | M | M | M | H | VH |
| VH | M | M | H | VH | VH |

a factor, which has been proved to have effect on rear-end conflicts frequency at signalized intersections (Essa & Sayed, 2018). According to Highway Capacity Manual (2010), the platoon ratio of a turning can be expressed as follows:

$$r_{a \rightarrow b}^i = \frac{p^i}{g^i/C} \quad (8)$$

where p^i is the count of vehicles arriving at turning link $a \rightarrow b$ during the total green light time within the i th time interval; g^i denotes effective green time (s) of the i th time interval; and C is the length of the time interval (s), which is a fixed value.

Table 1 summarizes all the factors used in this study, which lays a foundation for the implementation of TRSE.

3.6. Route safety evaluation

Given conflict rate of each route component, the conflict counts of the segments and turnings can be obtained by multiplying with the respective exposure. The next step is to evaluate the safety at the route level on the basis of the acquired conflict counts of component links.

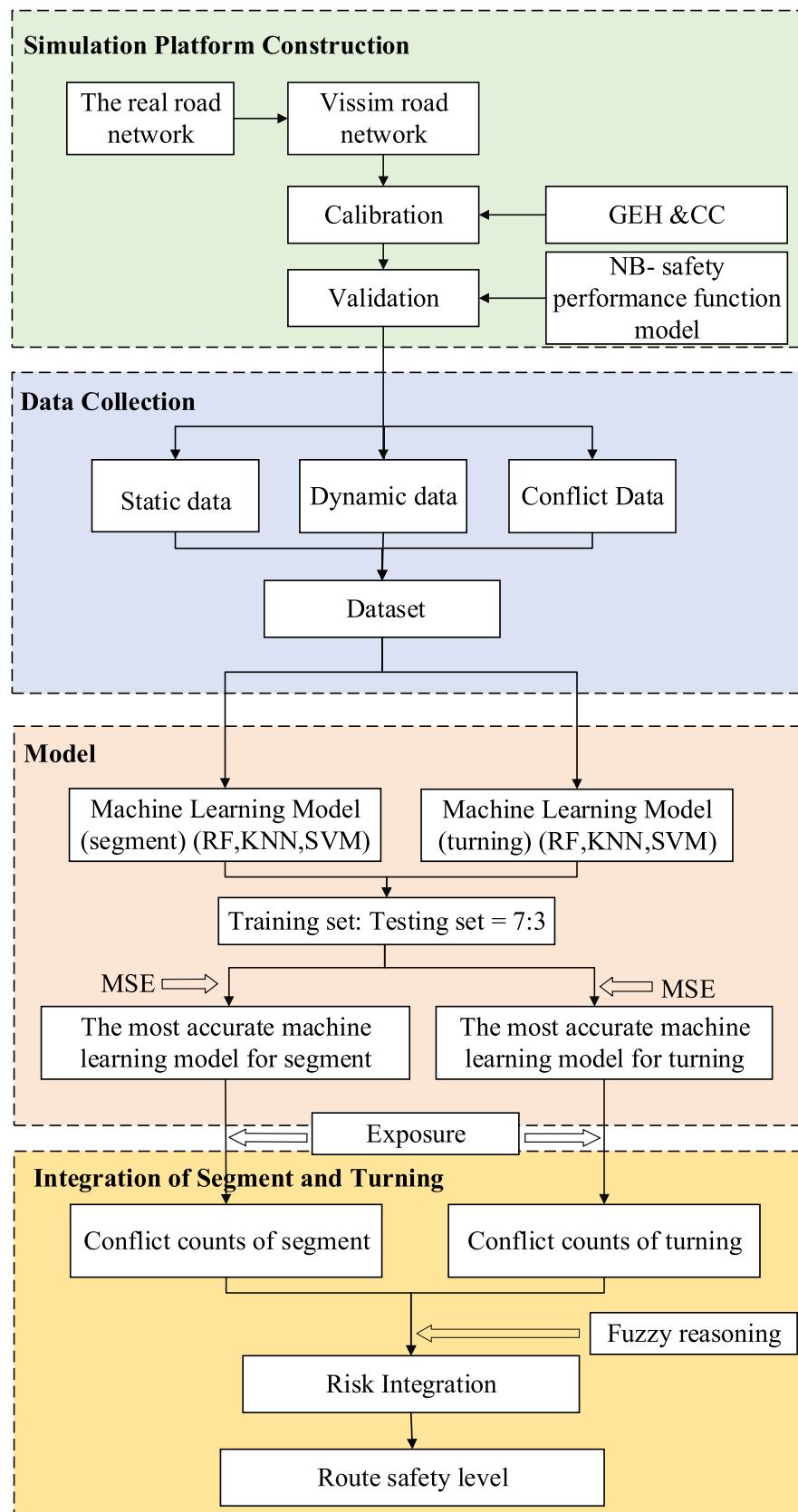
Since the factors used to measure the exposure and predict the conflict rate of segment link and turning link are different, the same count of conflicts being estimated for them may represent the different degrees of safety. Segment link and turning link should have their respective criteria in judging safety. It would be biased if the route safety is indicated by simple addition of the conflicts of segment link and turning link. Fuzzy logic has a good performance in the tolerance of the multiple-dimension data, and is flexible in dealing with nonlinear relationships among the complex data by fuzzy rules (Jiang et al. 2020; Zadeh, 2008). By explaining the relationship between inputs (total conflict count of segments, total conflict count of turnings) and output (risk level), fuzzy rules infer the most probable outcome under alternative prevailing conditions in the process of associating multiple

variables, rather than simply adding quantitative measures to get output. Furthermore, quantitative measures used in this paper are vague in determining the level of risk, and people's subjectivity leads to different understandings of vague things. In contrast, a precise risk value can be obtained through the reasoning of fuzzy rules, that is, the conflict counts of segment link and turning link contained in route r can be integrated into a single performance measure – risk level. Fig. 2 depicts the detail of the fuzzy reasoning process.

The input and output of the fuzzy reasoning process are the specified values within a certain range. For the same set of inputs, it is firstly fuzzified by the membership functions, and then are evaluated by all the fuzzy rules in parallel. The evaluation results of all rules will be combined for defuzzification to get a specific risk level. In this paper, we limit the risk level between 0 and 10. Gaussian membership function is used for fuzzification of input variables and defuzzification of output variables, which is based on its increased ability to model datasets of nonlinear "nature" (Dimitriou and Vlahogianni, 2015). Moreover, we set five membership functions for each fuzzy set (conflict counts of segment and turning, risk level of route). These functions are used to describe the membership of input fuzzy set to linguistic variables, which include "very low" (VL), "low" (L), "medium" (M), "high" (H), and "very high" (VH). The $\text{Min}(\bullet)$ operator is used for AND fuzzy operation and implication operation, while the $\text{Max}(\bullet)$ operator is used for OR fuzzy operation and aggregation operation. The centroid algorithm will be used to determine the final output value in the aggregated shape, namely defuzzification.

Table 2 shows the matrix representation of all fuzzy rules for risk level. S and T are fuzzy sets representing the conflict counts of segment links and turning links respectively. A total of 25 fuzzy rules are defined. For example, the interpretation of cell (1, 5) is as follows: if the conflict counts of segment link is very low and the conflict counts of turning link is very high, then the risk level of the route is medium.

It is worth noting that for the segment conflict counts and the turning

**Fig. 3.** Overall framework of the illustration.

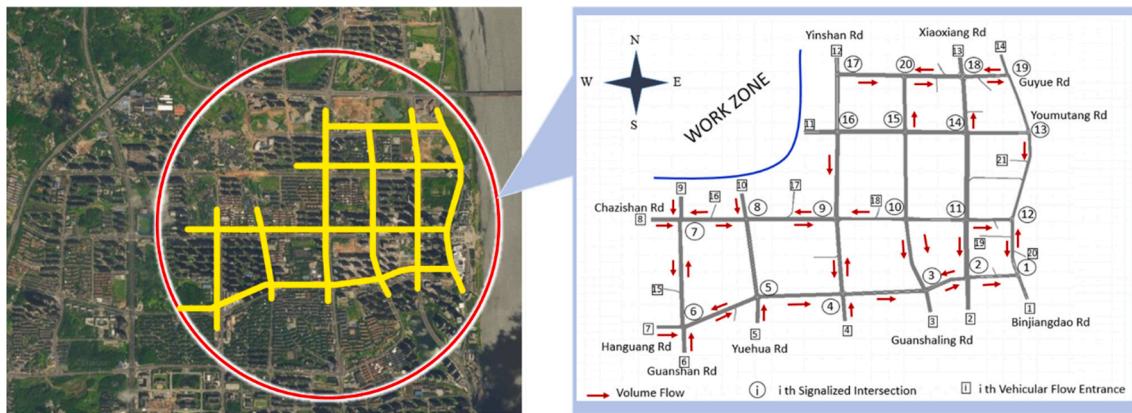


Fig. 4. Road network for illustration.

Table 3
Model fitting result of segment crash.

| | Estimate | Std.error | T-Statistic | p-Value |
|-------------------------------|----------|-----------|-------------|---------|
| (Intercept) α | -1.004 | 0.405 | -2.478 | 0.018 |
| (Conflicts) β | 0.631 | 0.074 | 8.500 | <0.001 |
| Overdispersion parameters k | | 1.908 | | |
| R-squared | | 0.697 | | |

Table 4
Model fitting result of turning crash.

| | Estimate | Std.error | T-Statistic | p-Value |
|-------------------------------|----------|-----------|-------------|---------|
| (Intercept) α | 0.413 | 0.433 | 0.954 | 0.353 |
| (Conflicts) β | 0.755 | 0.148 | 5.088 | <0.001 |
| Overdispersion parameters k | | 2.879 | | |
| R-squared | | 0.695 | | |

conflict counts in any pair of inputs, the membership of VL , L , M , H , and VH will be obtained respectively, and then the 25 rules will be used for the parallel evaluation. Finally, a specific risk level between 0 and 10 will be obtained.

4. Illustration example

In this section, we use simulation to show how the TRSE method works. Relying on the Vissim simulation platform and SSAM, a simulated road network based on the real road network is built. We display the detailed construction and application process of TRSE approach. To validate the effectiveness and sensitivity of TRSE, we simulated the

traffic conditions of a peak hour and an off-peak hour, respectively. The risk level and the exposure of the specific routes, the distribution of the conflict counts of the components were compared and analyzed. Fig. 3 shows the overall framework of applying the TRSE approach by means of simulation.

4.1. Simulation modeling

A four-horizontal and six-vertical urban road network in Changsha, China was built to simulate the traffic (Fig. 4). The network includes 20 signalized intersections (with a total of 106 signal phases), 21 vehicle entrances, 22 vehicle exits, and 105 road links. The vehicle inputs data of the simulation comes from the traffic monitoring data, that is the simulation traffic flow which runs the entire road network, furthermore, the signal phase data comes from Changsha traffic police brigade. The traffic monitoring data of 7:00–10:00 a.m., March 11, 2019, was used to

Table 5
Machine learning models of segment.

| Measure | RF | K-NN | SVM |
|---------|------|------|------|
| R^2 | 0.85 | 0.75 | 0.64 |
| MSE | 0.23 | 0.53 | 0.46 |

Table 6
Machine learning models of turning.

| Measure | RF | K-NN | SVM |
|---------|------|------|------|
| R^2 | 0.81 | 0.51 | 0.59 |
| MSE | 0.02 | 0.05 | 0.04 |

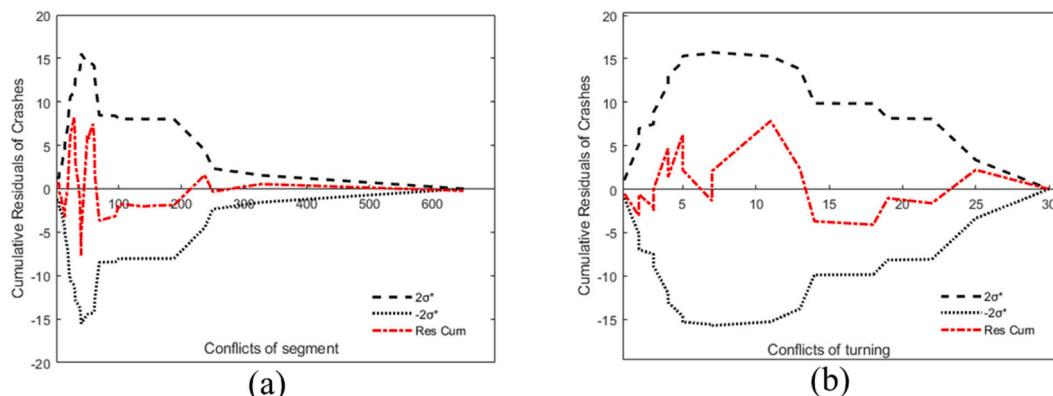


Fig. 5. CURE plots of segment (a) and turning (b).

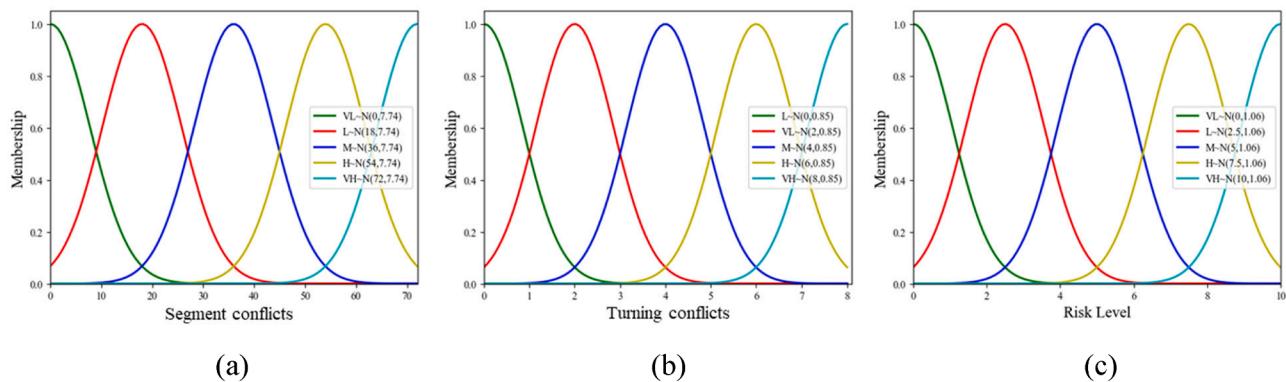


Fig. 6. Image of the membership functions of segment conflicts (a), turning conflicts (b), and risk level (c).

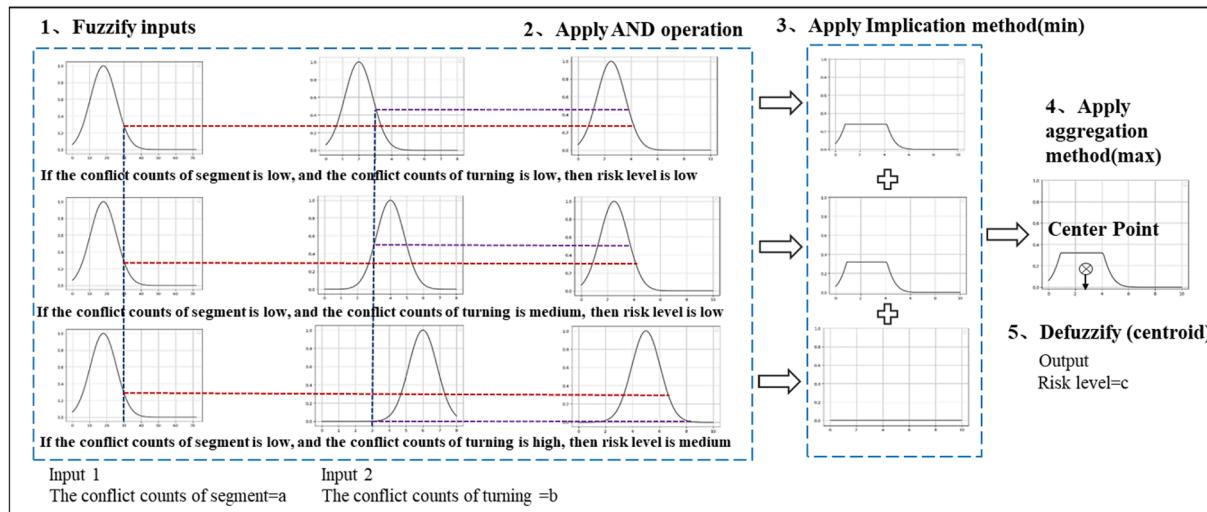


Fig. 7. Fuzzy reasoning process.

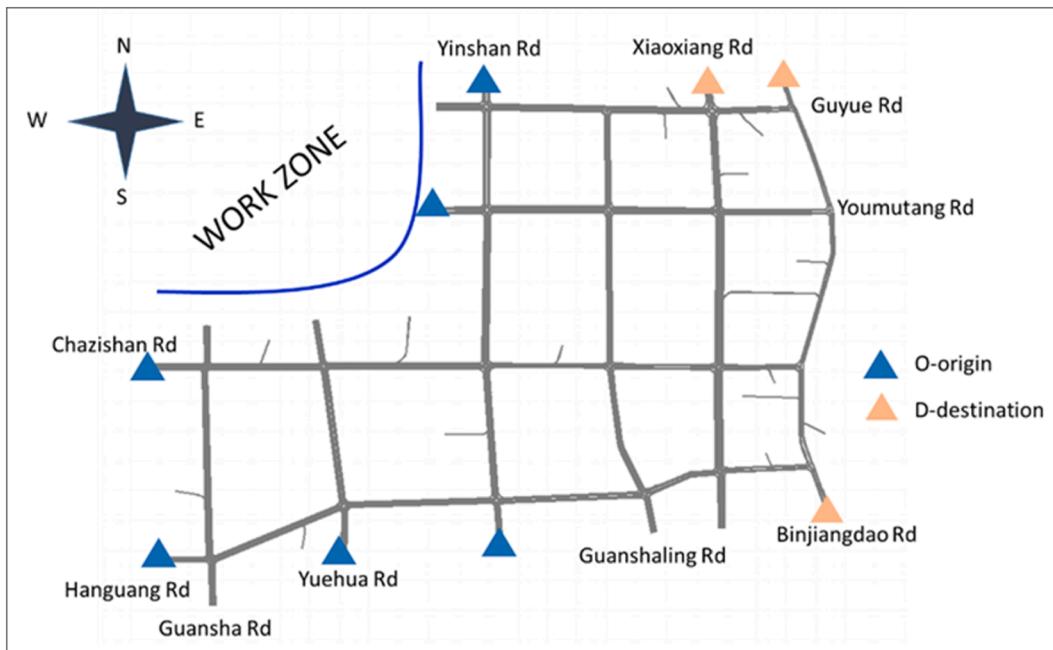


Fig. 8. Distribution of O-D positions.

Table 7
Route conflict counts - risk level during 7:30–7:31 a.m.

| O-D | Route | Segment conflict counts | Turning conflict counts | Risk level |
|-----|-------|-------------------------|-------------------------|------------|
| 1 | 1 | 3 | 1 | 1.81 |
| | 2 | 28 | 6 | 4.53 |
| | 3 | 1 | 0 | 1.20 |
| 2 | 1 | 10 | 3 | 2.31 |
| | 2 | 4 | 2 | 1.64 |
| | 3 | 7 | 3 | 2.16 |
| 3 | 1 | 5 | 3 | 2.02 |
| | 2 | 2 | 4 | 2.77 |
| | 3 | 4 | 6 | 4.42 |
| 4 | 1 | 3 | 5 | 3.56 |
| | 2 | 1 | 2 | 1.30 |
| | 3 | 9 | 5 | 3.78 |
| 5 | 1 | 2 | 4 | 2.77 |
| | 2 | 3 | 2 | 1.52 |
| | 3 | 4 | 2 | 1.64 |
| 6 | 1 | 3 | 3 | 1.96 |
| | 2 | 5 | 3 | 2.02 |
| | 3 | 5 | 4 | 2.77 |

Table 8
Route conflict counts- risk level during 9:30–9:31 a.m.

| O-D | Route | Segment conflict counts | Turning conflict counts | Risk level |
|-----|-------|-------------------------|-------------------------|------------|
| 1 | 1 | 0 | 2 | 1.21 |
| | 2 | 14 | 4 | 2.77 |
| | 3 | 0 | 3 | 1.96 |
| 2 | 1 | 2 | 1 | 1.70 |
| | 2 | 0 | 0 | 1.20 |
| | 3 | 0 | 1 | 1.48 |
| 3 | 1 | 2 | 1 | 1.70 |
| | 2 | 0 | 1 | 1.48 |
| | 3 | 7 | 5 | 3.65 |
| 4 | 1 | 3 | 2 | 1.52 |
| | 2 | 2 | 4 | 2.77 |
| | 3 | 7 | 6 | 4.36 |
| 5 | 1 | 0 | 1 | 1.48 |
| | 2 | 1 | 2 | 1.30 |
| | 3 | 1 | 2 | 1.30 |
| 6 | 1 | 6 | 2 | 1.89 |
| | 2 | 4 | 2 | 1.64 |
| | 3 | 4 | 4 | 2.77 |

run the simulated network with the simulation steps is 5, and the first 30 min is used to warm up the network. Before carrying out the illustration work of this paper, the first step is to calibrate and validate the simulated road network. Traffic volume is used as the object of calibration (Rahman et al., 2019), which is aggregated into 5 min intervals.

4.1.1. Simulation calibration

In the car-following model of Wiedemann-74 in Vissim simulation, three important parameters need to be well-calibrated so as to properly mimic the real flow scenarios, including average standstill distance, additive part of safety distance, and multiplicative part of safety

distance. In addition, Vissim allocates traffic flow evenly for each turning by default, which is inconsistent with most actual situation, therefore, we adjusted the traffic volume proportions of each turning of the intersections through the route decision function to minimize the difference of traffic volume between the real scenario and simulation model. In this study, Geoffrey E. Heavers (GEH) statistic and correlation coefficient (CC) were used to test the adjusted traffic volume (Rahman et al., 2019).

GEH statistic is a modified chi-square statistic containing relative and absolute differences (Rahman et al., 2018), which are used to compare traffic volume of the real scenario and simulation model. The definition of GEH statistic is as follows:

$$\text{GEH} = \sqrt{\frac{2 \times (V_{obs} - V_{sim})^2}{(V_{obs} + V_{sim})^2}} \quad (9)$$

where V_{obs} is the half-hour traffic volume of the field detectors, V_{sim} is the half-hour traffic volume of the simulation. If the cases where the GEH value is less than 5 accounting for 85%, it is considered that the simulated traffic well replicates the actual traffic (Abdel-Aty and Wang, 2017).

Correlation coefficient is used as an index to measure the goodness of fit, which indicates the degree of linear correlation between the field volume and the simulated volume (Rahman et al., 2019). The definition of CC is as follows:

$$CC = \frac{1}{n-1} \sum_{i=1}^n \frac{(v_{i,sim} - v_{sim}^\wedge)(v_{i,obs} - v_{obs}^\wedge)}{S_{sim} S_{obs}} \quad (10)$$

where n is the total number of traffic volume observations, v_{sim}^\wedge and v_{obs}^\wedge are respectively the average traffic volume of simulation model and field detectors within the five-minute interval. S_{sim} and S_{obs} are the standard deviations of the traffic volume of simulation model and field detectors in the five-minute interval, respectively. When the CC value is 0.85, the calibration model is considered acceptable (Rahman et al., 2019). In order to eliminate the influence of random effects, different random seeds are set to run the simulation 10 times, and used the average traffic volume of five-minute intervals to calculate the calibration result. The results show that 94.4% of the cases have a GEH value less than 5 and the CC value is 0.9. Correspondingly, the values of the three important parameters in the car-following model, average standstill distance, additive part of safety distance, and multiplicative part of safety distance are respectively to be 4.34 (m), 2.34, and 2.33.

4.1.2. Simulation validation

Through the interface between Vissim and SSAM, the conflict data of the road network is obtained by analyzing the trajectory data of simulated vehicles. In order to verify the reliability of the simulation in terms of safety performance, a conflict-based negative binomial safety performance function model (Cafiso et al., 2018; El-Basyouny and Sayed, 2013) was applied to prove whether the SSAM conflicts are representative of the actual crashes. All types of conflicts (rear-end, lane-change, crossing) will be considered, but rear-end conflicts account for the largest proportion. The validation of conflict is carried out separately for segment and turning. The specific form of the conflict-based negative binomial safety performance function model is in Eq. (11).

$$E(y_{crash}) = e^\alpha \times T^\beta \quad (11)$$

where y_{crash} is the expected crash frequency of segment/turning in the study area from 2018 to 2019, T is the conflict counts of segment/turning obtained by the SSAM model, α and β are regression parameters.

The results of the validation are shown in Table 3 and Table 4. The R-squared of the segment model and turning model are 0.697 and 0.695 respectively. The p -value of parameters of the segment model are less than 0.05, showing a significant effect on the y_{crash} . Although the p -value

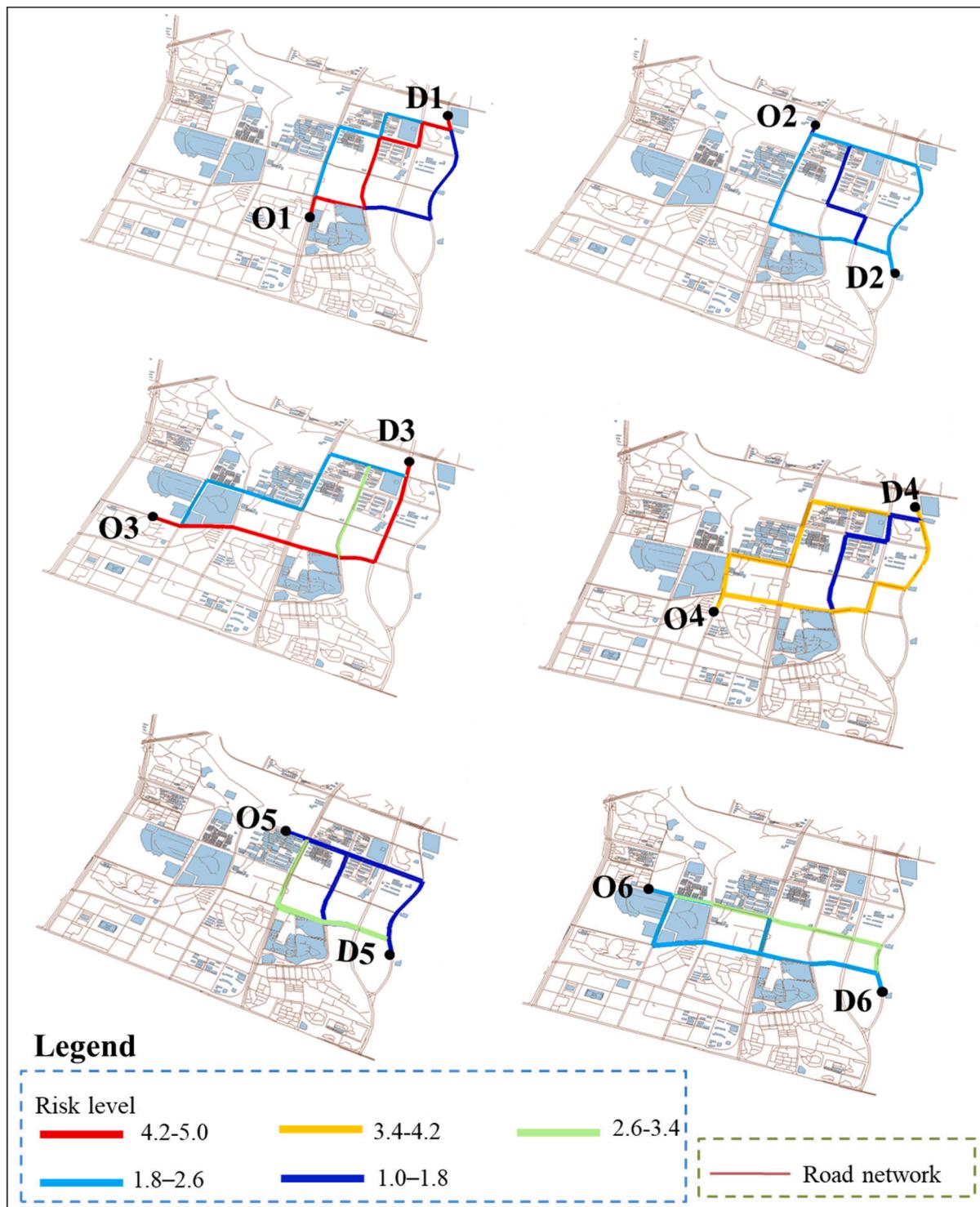


Fig. 9. Route risk level of six O-D pairs at peak period.

of α in the turning model is 0.353, the p -value of the parameter β that has an important influence on the model is less than 0.001, which indicates the acceptability of the model result. In addition, the CURE plots (Hauer and Bamfo, 1997) for judging the adequacy of the model are present in Fig. 5. Cumulative residuals of the models fluctuate up and down around 0, and do not exceed the boundary of $\pm 2 \sigma^*$, showing that the models are well adapted to the data. That is, there is a certain nonlinear relationship between simulation conflict and real crash, which indirectly verifies the safety performance of simulation.

4.2. Data collection

The static data, dynamic data and conflict data mentioned in Fig. 3 were collected on the basis of the completion of the calibration and validation work. First, static data are collected based on Google Map. Second, the vehicle record file output after the operation of the Vissim simulation provides a way to track the traffic state of segments and turnings. Specifically, through simulation, vehicle record file of 8:00–10:00 a.m., March 11, 2019 were collected, and dynamic traffic

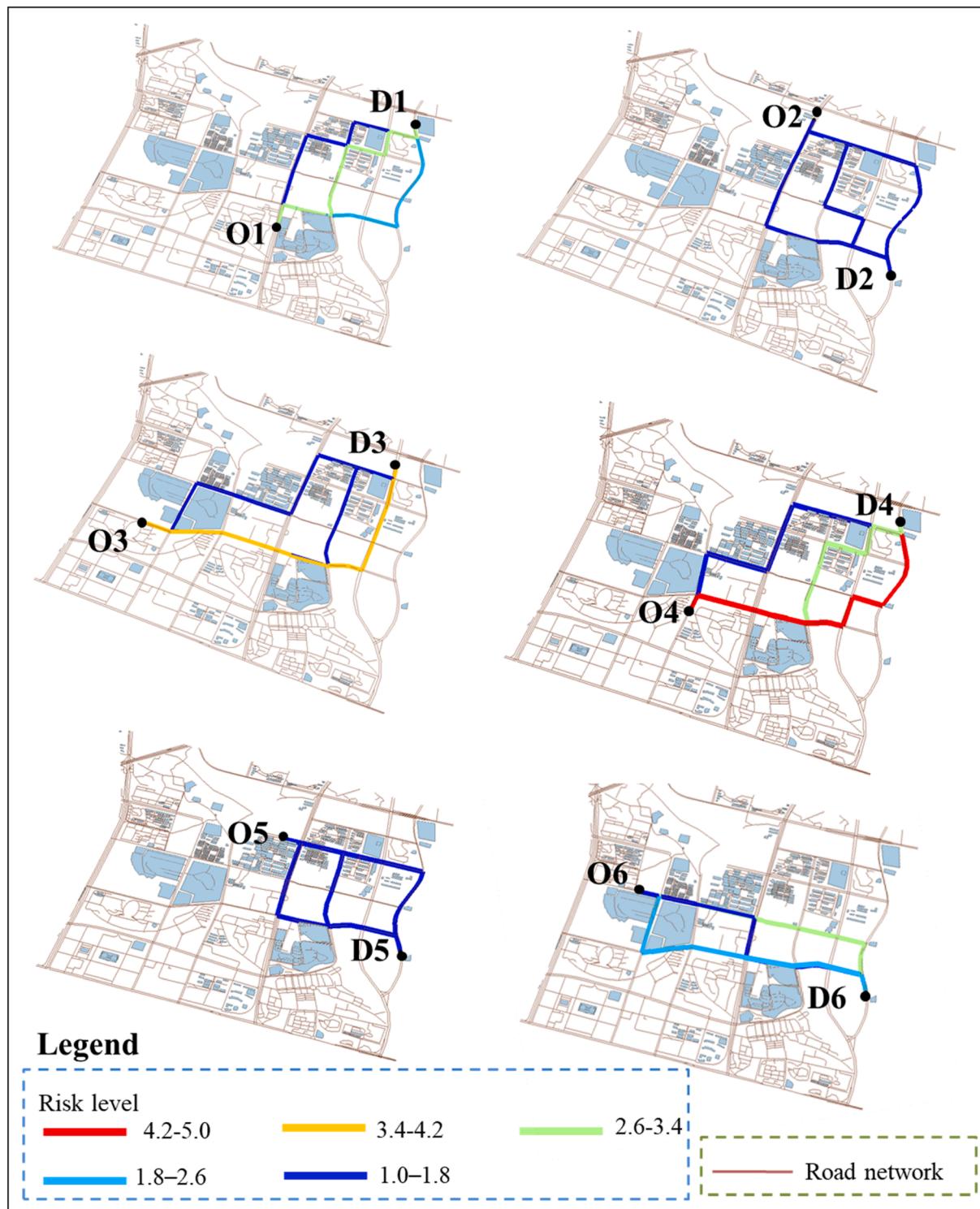


Fig. 10. Route risk level of six O-D pairs at off-peak period.

data such as speed, acceleration, flow, etc. within one minute are obtained. Third, the vehicle trajectory file output by Vissim is used for conflict analysis. In SSAM, the conflict identification indicators are time to collision (TTC) and post encroachment time (PET), and we set the thresholds to 4 s (Xing et al., 2019) and 5 s (Sohail et al., 2016), respectively. Namely, if the TTC of the two vehicles is less than 4 s and the PET of the two vehicles is less than 5 s, SSAM classifies the interaction of the two vehicles as a conflict. After classification and summary, conflict counts of segment and turning are collected. According to Eq.

(4) and Eq. (5), the conflict rate of segment and turning within one-minute time interval were obtained.

4.3. Conflict estimation

As mentioned in the previous section, the conflict rate is used as the label in machine learning model. Correspondingly, R^2 and MSE indicators are used to evaluate model performance. Training set and testing set are delimited by the ratio of 7:3, and the model performances

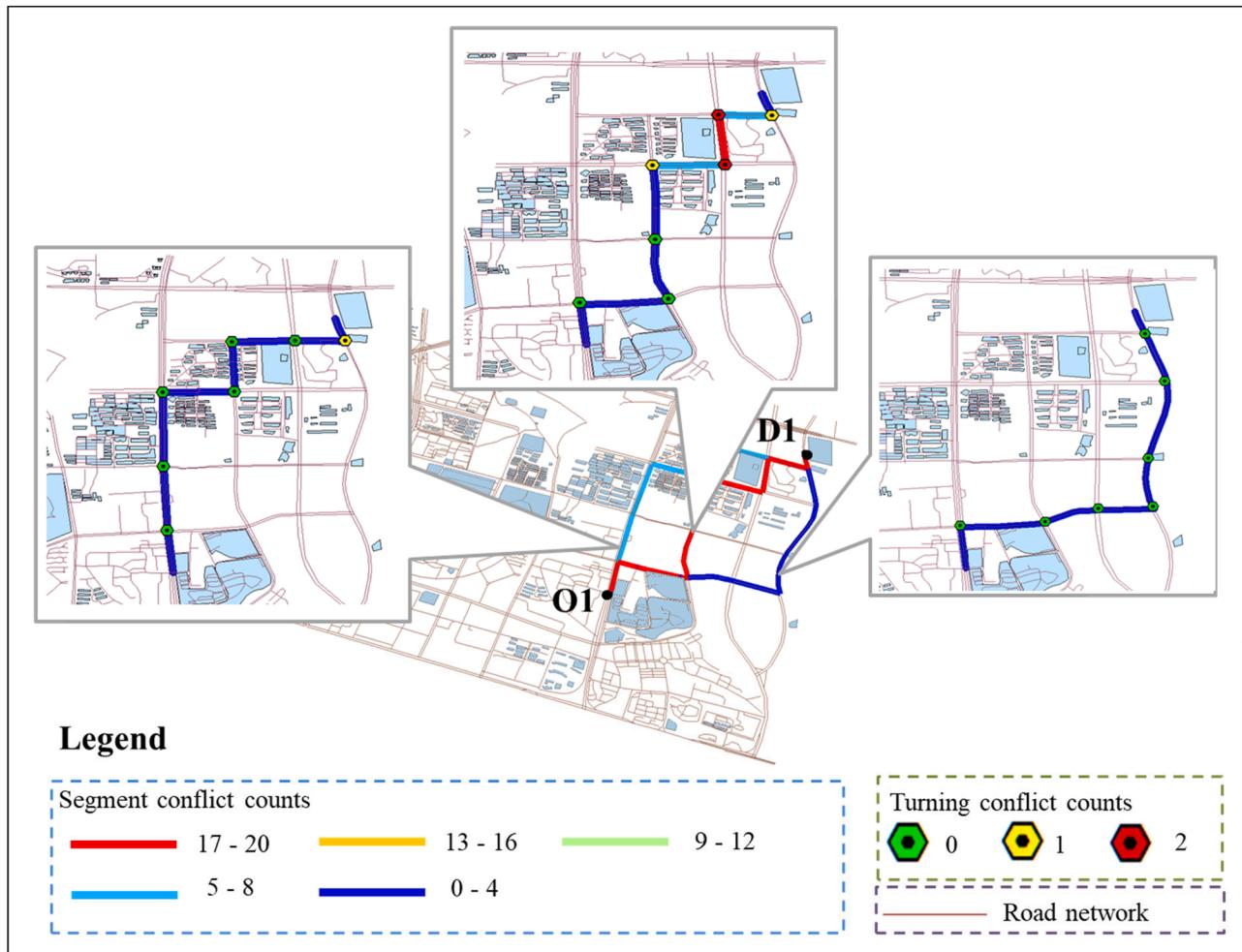


Fig. 11. Distribution of the segment and turning conflict counts for the three routes between O-D 1 at peak period.

of segment and turning are shown in Table 5 and Table 6. Given its best performance in R^2 and MSE, the trained random forest algorithm was selected to construct the conflict prediction model.

4.4. Route safety estimation

In this section, the fuzzy logic reasoning process aggregates segment conflicts and turning conflicts to get the risk level of the entire route. In order to solve the problem of the difference in the number of components of the route, conflict counts of all segments and all turnings in the route are used as the inputs of fuzzy logic reasoning process. According to the SSAM conflict analysis, it can be concluded that the maximum of total conflict counts in the segment link of a route is 72, and the maximum of total conflict counts in the turning link of a route is 8. Accordingly, the range of segment conflict input and turning conflict input as [0, 72] and [0, 8] respectively. In addition, the following Fig. 6 (a)–(c) display the five gauss membership function images of the segment conflicts, turning conflicts, and the risk level, respectively.

25 rules will carry out fuzzy reasoning on the same set of inputs at the same time. For the purpose of illustration, Fig. 7 shows the detailed process of fuzzy logic reasoning with three fuzzy rules (the conflict counts of segment is a, the conflict counts of turning is b, and the risk level is c).

4.5. Illustrations

We selected six O-D pairs from the six origins and three destinations

in the study area. Fig. 8 presents the spatial distribution of each O-D position. Since the current section is just to perform the TRSE approach, three routes in each O-D pair were randomly selected for analysis.

For each O-D pair, the selected three routes are numbered as 1, 2, and 3 from left to right as shown in Fig. 9 and Fig. 10. For sensitivity analysis, Vissim simulation runs in a peak hour (7:00–8:00 a.m.) and an off-peak hour (9:00–10:00 a.m.), and outputs vehicle record files and vehicle trajectory files. Data of route components in 7:30–7:31 a.m. (for the convenience of explanation, the following uses peak period instead) and in 9:30–9:31 a.m. (the following uses off-peak period instead) are selected to perform the TRSE approach. Table 7 and Table 8 list the calculated results of the route conflict counts-risk level during the peak period and off-peak period respectively.

Due to the phenomenon of route coverage, some routes cannot be fully displayed, but according to the location of the O-D pairs, the direction of the route can be easily understood. Fig. 9 and Fig. 10 show in detail the direction of the selected route and its risk levels during the peak period and the off-peak period, with different colors indicating different ranges of route risk levels. At the peak period, between the 18 selected routes, the third route of O-D 3 has the highest risk level of 4.53, but for the off-peak period, the highest risk level of the route is 4.36, which is positioned on the third route of O-D 4. For the lowest route risk level, both the peak period and the off-peak period are 1.2, which are the third route of O-D 1 at the peak period and O-D 2 at the off-peak period. Overall, during the off-peak period, 81.25% of the routes have a lower risk level than that during the peak period.

In addition, we select three routes in O-D 1 to show the specific

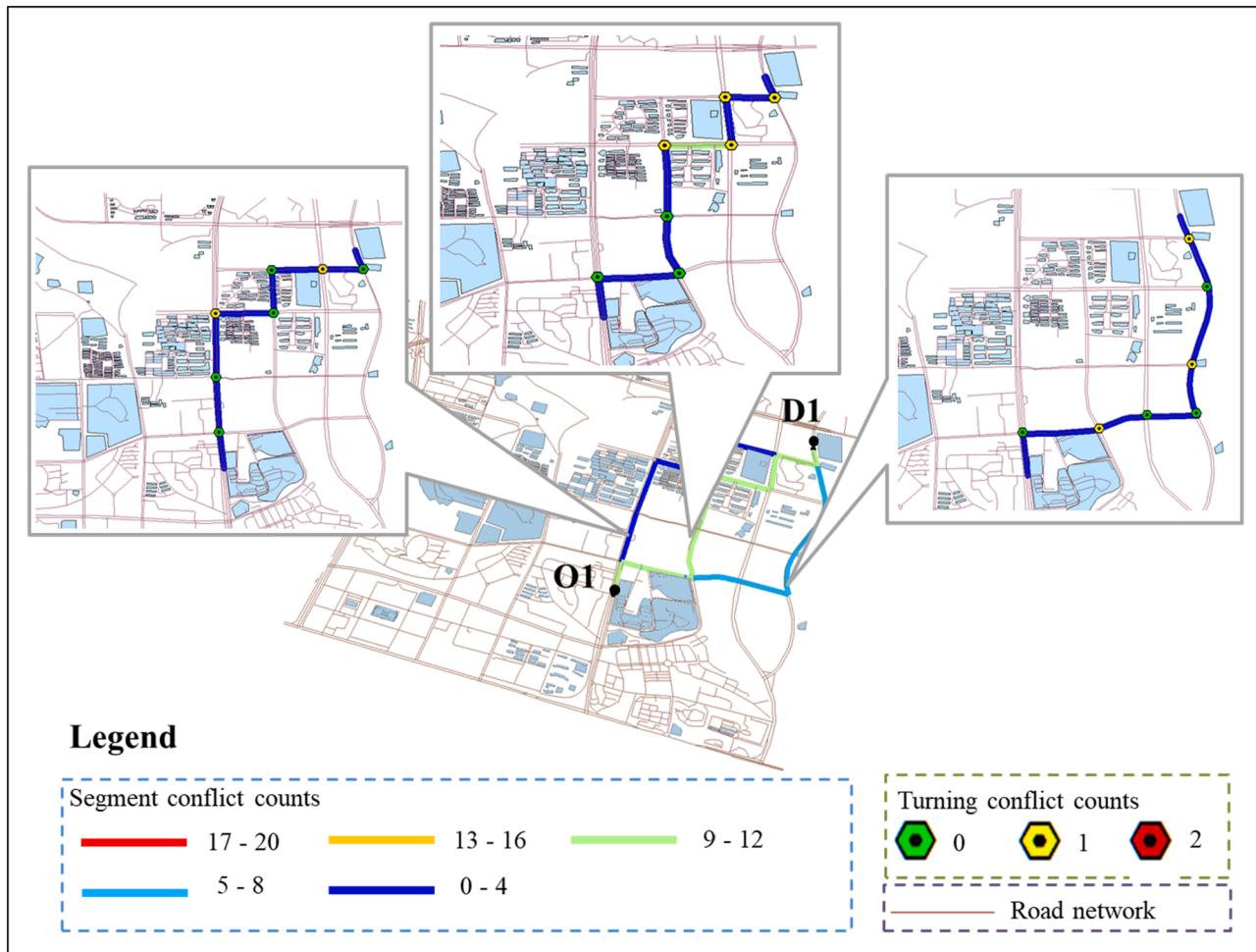


Fig. 12. Distribution of the segment and turning conflict counts for the three routes between O-D 1 at off-peak period.

conflict counts distribution of route components (Fig. 11 and Fig. 12). Herein each route consists of eight segments and seven turnings, broken lines and hexagons represent segment links and turning links respectively, and different colors distinguish the specific conflict frequency range.

In the case of the peak period, the conflict counts in the sixth segment (counting from O to D, the sixth segment of the route) of route 2 reached the maximum value of 18. Similarly, the maximum conflict counts of turning which is 2 was also located in route 2. Within the off-peak period, route 2 still generated the maximum conflicts of segment and turning, but the risk level of route 2 decreased significantly compared to the peak period. However, relative to the peak period, the conflict counts of turning in route 3 increased, so the risk level increased slightly, which was considered to be caused by the phenomenon of high speed at the intersection due to the relatively low vehicle volume at the off-peak period.

Cumulative exposure is used to illustrate the changes in exposure along with the progress of route travel. Due to the dimensional difference of the exposure of the segment and the turning, the exposure of the entire route cannot be measured with a unified index. Standardization addressed this problem by limiting exposure of segment links and turning links to 0–1 and has achieved the goal of de-dimensionalization. Based on this, the step-by-step cumulative addition of the exposure is used to represent the change of the route exposure. Fig. 13 and Fig. 14 show in detail the exposure accumulation of the peak period and off-peak period of the three routes of O-D1. The differences in height and color of the exposure rectangles in the figures demonstrate different amount of cumulative risk exposure of each route. At the peak period,

the cumulative exposure of route 1 and route 3 slowly increased. However, the cumulative exposure of route 2 in the sixth segment link and the sixth turning link increases sharply. As reflected in Fig. 11, the conflict counts of the sixth segment link and the sixth turning link also reached the highest. In addition, the total cumulative exposure of route 2 is the highest among three routes, which is consistent with the result where the risk level of route 2 is the highest among the three selected routes. At off-peak period, the cumulative exposure of route 2 in the first 5 segment links and turning links is lower than that at the peak period, but the cumulative exposure of the first 6 segment link and turning links is higher than that at the peak period. It is worth noting that the total cumulative amount of the entire route 2 is higher than that at the peak period, which may be due to the low traffic density at the off-peak period and the number of vehicles passing through the segment link and turning link is increased.

4.6. Interpretation of contributing factors

Although the present study is focused on the model construction, a good understanding of the contributing factors of traffic conflict occurrence is also important to justify the model validity. RF model has the ability to generate variables' importance ranking, which can help us to identify the factors that are imperative in traffic conflict estimation.

Figure 15 depicts the variables' relative importance ranking on conflict estimation for segment links. As we can see, the dynamic factors contribute more to conflict occurrence than the stationary factors do. The result is in accordance with other studies (Guo et al., 2020; Katrakazas et al., 2018; Orsini et al., 2021). It indicates that the changes of

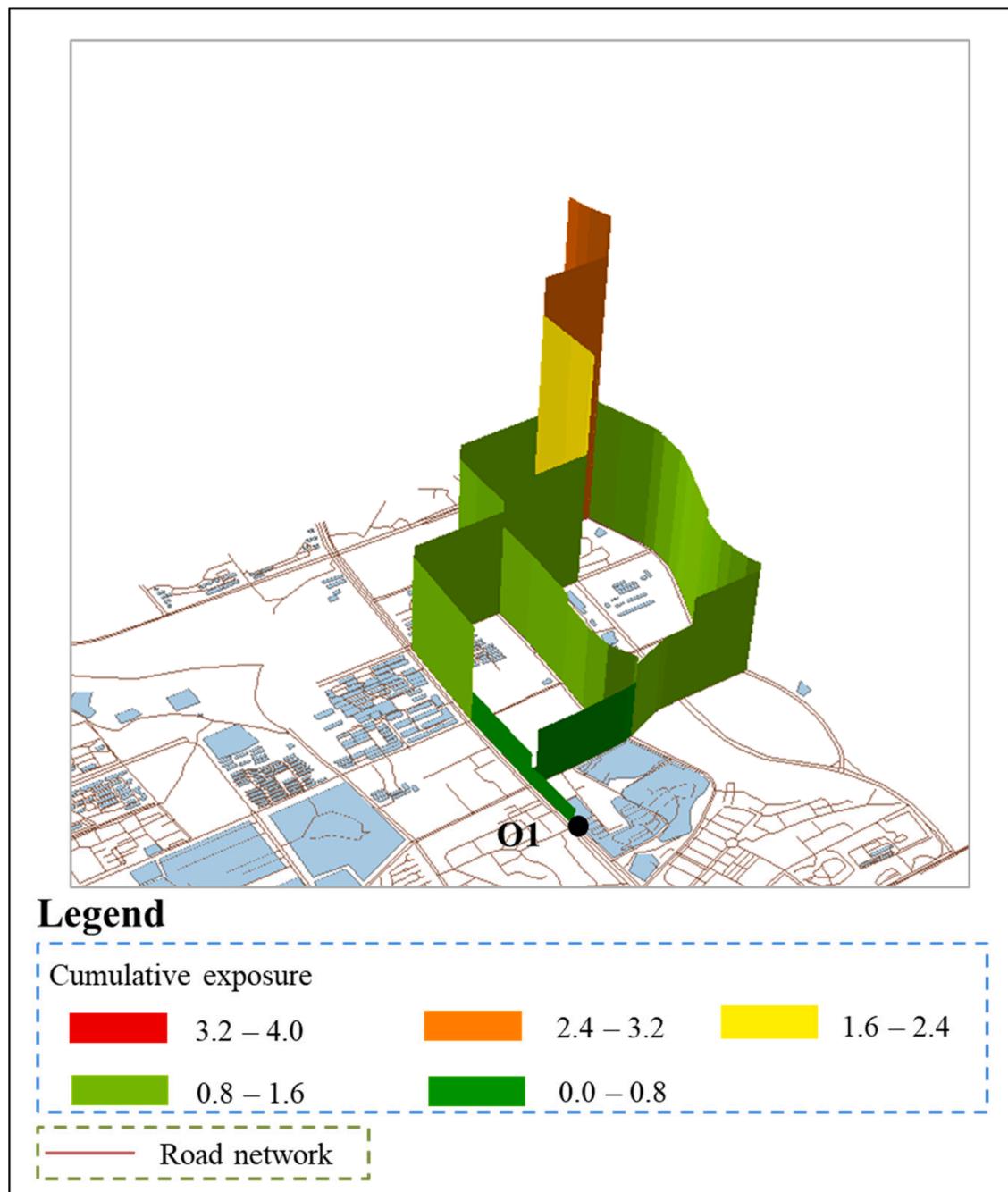


Fig. 13. Cumulative exposure of the three routes between O-D1 at peak period.

these dynamic variables are highly associated with the variation of segment conflict. Among them, the most important variables are those related to speed, such as the mean and the variance of speed and acceleration. One exception is that the length of the segment link ranked third. It verifies our assumption that longer links provide travelers more chances of being involved into conflicts. In regard to turning links, as shown in Fig. 16, the dynamic variables related to traffic speed, volume, their difference between upstream and downstream, and traffic control feature (platoon ratio) also show higher importance in conflicts estimation than the static variables such as turning types, inherent conflict point, and intersection types. This is consistent with previous studies that at signalized intersections conflict rate is significantly associated with the traffic condition within the signal cycle (Guo et al., 2020; Zheng and Sayed, 2020).

The findings above demonstrate the underlying assumption that

certain combinations of traffic conditions (which are normally represented by traffic flow parameters, traffic control features, and road site characteristics) are relatively more ‘conflict prone’ than others, and also shows that detect and quantify conflict-prone traffic conditions is the core of real-time conflict prediction.

5. Conclusion and future work

In this paper, we developed a travel route safety estimation (TRSE) approach, which can provide a technical foundation for safer route planning. The classic safety evaluation model was taken as the basis to build the framework of TRSE. We used the traffic conflict counts that the travelers may encounter on the route to estimate the level of route safety. The amount of exposure and conflict rate were estimated using a combination of a set of dynamic and static factors that have been found

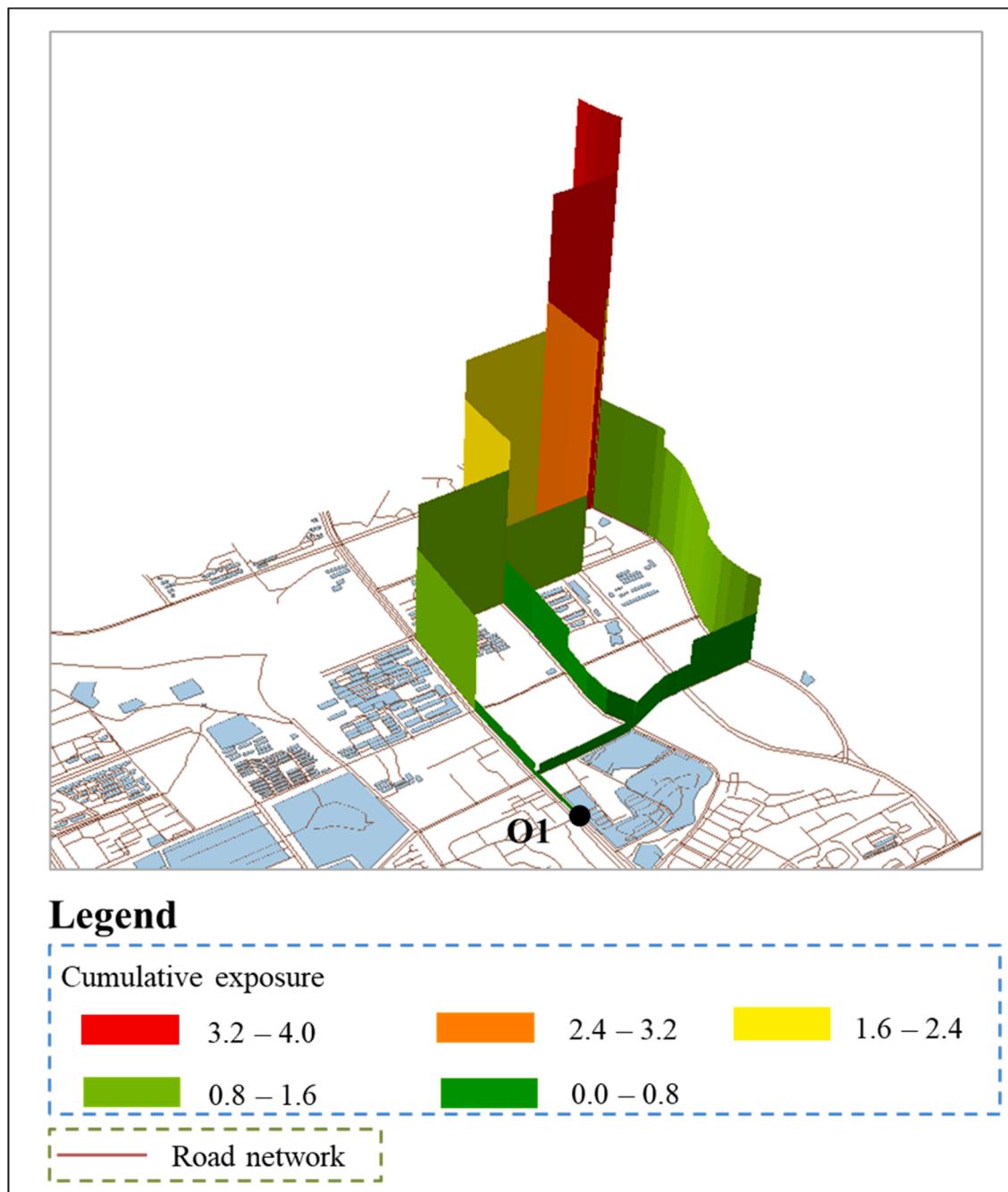


Fig. 14. Cumulative exposure of the three routes between O-D1 at off-peak period.

to have significant effects on traffic conflicts. A route-based method was employed where two parallel estimations of conflict were conducted for both the component segments links and intersection turning links. We built three machine learning models, including RF, SVM, and K-NN, and selected the best fitting one to perform the estimation. A fuzzy reasoning process based on the fuzzy logic algorithm was employed to conduct the route level safety estimation.

A simulated road network was built in the Vissim environment, with the aim to validate the proposed TRSE method. Six O-D pairs of the network were selected to illustrate the estimation process of risk level of the routes. The estimations were performed for the peak hour and off-peak hour. In addition, we selected three routes between O-D 1 to show and analyze the overall conflict exposures and the conflict counts estimated for each route component. In the end, with the help of RF's ability of generating variables' importance ranking, an analysis of the

effects of selected traffic and roadway factors was conducted. The dynamic factors showed high significance both in the conflict estimations of segment links and turning links.

It is worth mentioning that the current technology is very difficult to collect safety indicators for large-scale road networks, as a result, simulation is used as an application tool in this paper and is not part of the TRSE approach. For the practical application of TRSE approach, in the Internet era, it can be realized by real-time data collection and transmission through roadside (infrastructure) sensors or on-board equipment (radar, camera, sensor, etc.), as well as the real-time operation and processing of model to achieve real-time acquisition of conflict data. Real-time data collection and processing technology will enable the implementation of TRSE approach based on more accurate risk factors and safety indicators, which will benefit the application of TRSE approach in the intelligent and connected vehicle (ICV) environment.

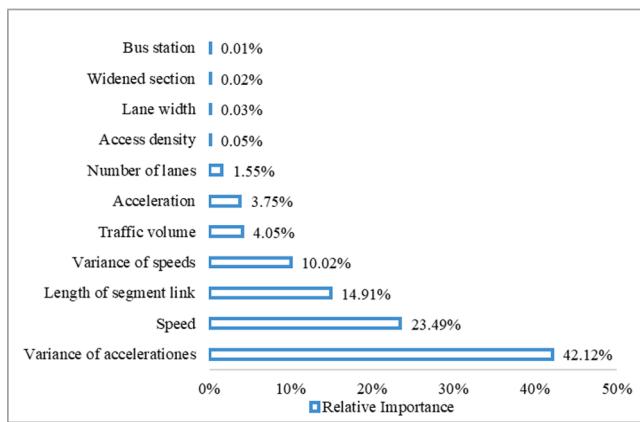


Fig. 15. Variables' relative importance ranking of segment link.

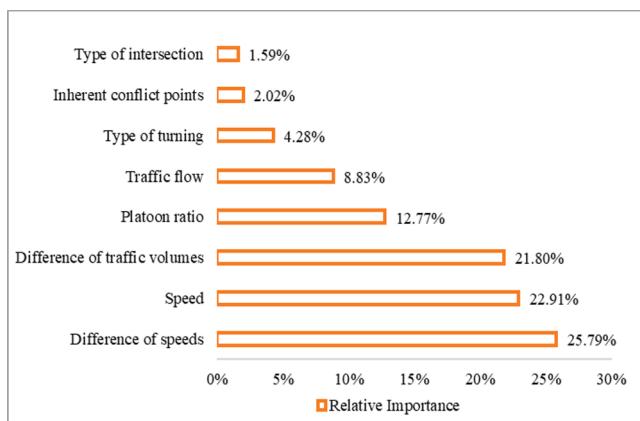


Fig. 16. Variables' relative importance ranking of turning link.

In this study, we develop a modeling framework of route safety estimation, which is helpful to promote the development of safer route planning. There are several further directions noted as follows: First, the TRSE approach proposed in this study is an off-line estimation to some extent. There is still much work to do to develop the on-line safer route navigation system for instant route safety information. The proposed TRSE approach can be combined with the current route navigation system to achieve that goal. Second, considering that the risks of different locations on the segment are dynamic, future research can split the segment to provide more detailed route safety information. Third, this paper mainly focused on safety-related issues, and future research can take travel time into account for travel route evaluation. However, how to balance the relationship between safety and travel time to make appropriate route selection? Given that the recipient of safety and travel time information is the driver, perhaps doing the above based on the driver's preference would be a feasible solution.

CRediT authorship contribution statement

Helai Huang: Conceptualization, Supervision, Writing – review & editing. **Yulu Wei:** Investigation, Methodology, Data curation, Simulation, Software, Writing – original draft. **Chunyang Han:** Conceptualization, Investigation, Methodology, Data curation, Writing – original draft, Writing – review & editing. **Jaeyoung Lee:** Writing – review & editing. **Suyi Mao:** Simulation. **Fan Gao:** Software.

Acknowledgements

This work was supported by: 1) the National Natural Science

Foundation of China (No. 71971222); 2) the China Postdoctoral Science Foundation (No. 2021M701898).

References

- AASHTO, 2010. Highway Safety Manual. 1st ed.. Transportation Research Board, Washington D.C.
- Abdel-Aty, M., Wang, L., 2017. Implementation of variable speed limits to improve safety of congested expressway weaving segments in microsimulation. *Transp. Res. Procedia* 27, 577–584.
- Abdelhamid, S., Elsayed, S.A., Abuali, N., Hassanein, H.S., 2017. Driver-Centric Route Guidance. In: 2016 IEEE Global Communications Conference (GLOBECOM) (pp. 1–6). IEEE.
- Al-Omari, M.M.A., Abdel-Aty, M., Cai, Q., 2021. Crash analysis and development of safety performance functions for Florida roads in the framework of the context classification system. *J. Saf. Res.* 79, 1–13.
- Alam, M.S., Perugu, H., Mcnabola, A., 2018. A comparison of route-choice navigation across air pollution exposure, CO₂ emission and traditional travel cost factors. *Transp. Res. Part D: Transp. Environ.* 65 (DEC.), 82–100.
- Allen, B.L., Shin, B.T., Cooper, P.J., 1978. Analysis of traffic conflicts and collisions. *Transp. Res. Rec.* 667, 67–74.
- Basak, D., Srimanta, P., Patranbis, D.C., 2007. Support Vector Regression. In: neural information processing letters and reviews, p. 11.
- Basso, F., Basso, L.J., Pezoa, R., 2020. The importance of flow composition in real-time crash prediction. *Accid. Anal. Prev.* 137, 105436.
- Cafiso, S., D'Agostino, C., Kieć, M., Bak, R., 2018. Safety assessment of passing relief lanes using microsimulation-based conflicts analysis. *Accid. Anal. Prev.* 116 (JUL.), 94–102.
- Cai, Q., Abdel-Aty, M., Yuan, J., Lee, J., Wu, Y., 2020. Real-time crash prediction on expressways using deep generative models. *Transp. Res. Part C: Emerg. Technol.* 117, 102697.
- Cai, Q., Abdel-Aty, M., Lee, J., Wang, L., Wang, X., 2018. Developing a grouped random parameters multivariate spatial model to explore zonal effects for segment and intersection crash modeling. *Analytic Methods Accid. Res.* 19, 1–15.
- Caleffi, F., Anzanello, M.J., Bettella Cybis, H.B., 2017. A multivariate-based conflict prediction model for a Brazilian freeway. *Accid. Anal. Prev.* 98 (JAN.), 295–302.
- Chandra, S., 2014. Safety-based path finding in urban areas for older drivers and bicyclists. *Transp. Res. Part C: Emerg. Technol.* 48, 143–157.
- Chen, M., Chien, S.I.J., 2001. Dynamic freeway travel-time prediction with probe vehicle data: link based versus path based. *Transp. Res. Rec. J. Transp. Res. Board* 1768 (1), 157–161.
- Choi, Y., Kang, D., Bahk, S., 2014, Oct 22–24. Improvement of AODV routing protocol through dynamic route change using hello message. International Conference on Information and Communication Technology Convergence 5th International Conference on Information and Communication Technology Convergence (ICTC), Minist Sci ICT and Future Planning, Busan, SOUTH KOREA.
- Cooper, P.J., 1984. Experience with traffic conflicts in canada with emphasis on “post encroachment time” techniques. *Int. Calibration Study Traffic Conflicts, NATA ASI Series F5*, 75–96.
- Dimitriou, L., Vlahogianni, E.I., 2015. Fuzzy modeling of freeway accident duration with rainfall and traffic flow interactions. *Analytic Methods Accid. Res.* 5–6, 59–71.
- Eisele, W.L., Frawley, W.E., 2005. Estimating the safety and operational impact of raised medians and driveway density: experiences from Texas and Oklahoma case studies. *Transp. Res. Rec. J. Transp. Res. Board* 1931 (1931), 108–116.
- El-Basyouny, K., Sayed, T., 2013. Safety performance functions using traffic conflicts. *Saf. Sci.* 51 (1), 160–164.
- Essa, M., Sayed, T., 2018. Traffic conflict models to evaluate the safety of signalized intersections at the cycle level. *Transp. Res. Part C: Emerg. Technol.* 89, 289–302.
- Formosa, N., Quddus, M., Ison, S., Abdel-Aty, M., Yuan, J., 2020. Predicting real-time traffic conflicts using deep learning. *Accid. Anal. Prev.* 136, 105429.
- George, B., Kim, S., 2013. Shortest Path Algorithms for a Fixed Start Time. In Spatio-temporal Networks. Springer.
- Guo, Y., Sayed, T., Essa, M., 2020. Real-time conflict-based Bayesian Tobit models for safety evaluation of signalized intersections. *Accid. Anal. Prev.* 144, 105660.
- Han, C., Huang, H., Lee, J., Wang, J., 2018. Investigating varying effect of road-level factors on crash frequency across regions: a Bayesian hierarchical random parameter modeling approach. *Analytic Methods Accid. Res.* 20, 81–91.
- Hauer, E., 1982. Traffic conflicts and exposure. *Accid. Anal. Prev.* 14 (5), 359–364.
- Hauer, E., Bamfo, J., 1997. Two tools for finding what function links the dependent variable to the explanatory variables . In Proceedings of the ICTCT 1997 Conference, Lund, Sweden.
- Hayward, J.C., 1971. Near misses as a measure of safety at urban intersections. Pennsylvania Transportation and Traffic Safety Center.
- Highway Capacity Manual, 2010. Transportation Research Board. Washington, D.C.
- Hossain, M., Abdel-Aty, M., Quddus, M.A., Muromachi, Y., Sadeek, S.N., 2019. Real-time crash prediction models: State-of-the-art, design pathways and ubiquitous requirements. *Accid. Anal. Prev.* 124, 66–84.
- Huang, H., Abdel-Aty, M., 2010. Multilevel data and Bayesian analysis in traffic safety. *Accid. Anal. Prev.* 42 (6), 1556–1565.
- Huang, H., Song, B., Xu, P., Zeng, Q., Lee, J., Abdel-Aty, M., 2016. Macro and micro models for zonal crash prediction with application in hot zones identification. *J. Transp. Geogr.* 54 (JUN.), 248–256.
- Huang, T., Wang, S., Sharma, A., 2020. Highway crash detection and risk estimation using deep learning. *Accid. Anal. Prev.* 135, 105392.

- Haynes, R., Jones, A., Kennedy, V., Harvey, I., Jewell, T., 2007. District variations in road curvature in England and Wales and their association with road-traffic crashes. *Environ. Plann. A* 39 (5), 1222–1237.
- Jenelius, E., Koutsopoulos, H., 2013. Travel time estimation for urban road networks using low frequency probe vehicle data. *Transp. Res. Part B: Methodol.* 53, 64–81.
- Jiang, S., Jafari, M., Kharbeche, M., Jalayer, M., Al-Khalifa, K.N., 2020. Safe route mapping of roadways using multiple sourced data. *IEEE Trans. Intell. Transp. Syst.* 99, 1–11.
- Jiang, Z., Zhang, C., Xia, Y., 2014. Travel time prediction model for urban road network based on multi-source data. *Procedia-Social and Behavioral Sciences* 138, 811–818.
- Katrakazas, C., Quddus, M., Chen, W.-H., 2018. A simulation study of predicting real-time conflict-prone traffic conditions. *IEEE Trans. Intell. Transp. Syst.* 19 (10), 3196–3207.
- Katrakazas, C., Quddus, M., Chen, W.H., 2019. A new integrated collision risk assessment methodology for autonomous vehicles. *Accid. Anal. Prev.* 127, 61–79.
- Kuskapan, E., Çodur, M.Y., Atalay, A., 2021. Speed violation analysis of heavy vehicles on highways using spatial analysis and machine learning algorithms. *Accid. Anal. Prev.* 155, 106098.
- Li, P., Abdel-Aty, M., Yuan, J., 2019. Real-time crash risk prediction on arterials based on LSTM-CNN. *Accid. Anal. Prevent.* 135, 105371.
- Lord, D., Manning, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. *Transp. Res. Part A: Policy Pract.* 44 (5), 291–305.
- Lord, D., Qin, X., Geedipally, S.R., 2021. *Highway Safety Analytics and Modeling*. Elsevier.
- Lv, Y., Tang, S., Zhao, H., Li, S., 2009. Real-time highway accident prediction based on support vector machines. *2009 Chinese Control and Decision Conference*.
- Nääätänen, R., Summala, H., 1974. A model for the role of motivational factors in drivers' decision-making. *Accid. Anal. Prev.* 6 (3–4), 243–261.
- Orsini, F., Geccheli, G., Rossi, R., Gastaldi, M., 2021. A conflict-based approach for real-time road safety analysis: comparative evaluation with crash-based models. *Accid. Anal. Prev.* 161, 106382.
- Park, J., Abdel-Aty, M., Wang, J.H., Lee, C., 2015. Assessment of safety effects for widening urban roadways in developing crash modification functions using nonlinearizing link functions. *Accid. Anal. Prev.* 79 (JUN.), 80–87.
- Payyanadan, R.P., Sanchez, F.A., Lee, J.D., 2017. Assessing route choice to mitigate older driver risk. *IEEE Trans. Intell. Transp. Syst.* 18 (3), 527–536.
- Rahman, M.H., Abdel-Aty, M., Lee, J., Rahman, M.S., 2019. Enhancing traffic safety at school zones by operation and engineering countermeasures: A microscopic simulation approach. *Simul. Model. Pract. Theory* 94, 334–348.
- Rahman, M.S., Abdel-Aty, M., Wang, L., Lee, J., 2018. Understanding the highway safety benefits of different approaches of connected vehicles in reduced visibility conditions. *Transp. Res. Rec.: J. Transp. Res. Board* 2672 (19), 91–101.
- Roshandel, S., Zheng, Z., Washington, S., 2015. Impact of real-time traffic characteristics on freeway crash occurrence: Systematic review and meta-analysis. *Accid. Anal. Prev.* 79 (JUN.), 198–211.
- Shi, Q., Abdel-Aty, M., 2015. Big data applications in real-time traffic operation and safety monitoring and improvement on urban expressways. *Transp. Res. Part C: Emerg. Technol.* 58, 380–394.
- Sohail, Zangenehpour, Jillian, Strauss, Luis, F., Miranda-Moreno, Nicolas, Saunier, 2016. Are signalized intersections with cycle tracks safer? A case-control study based on automated surrogate safety analysis using video data. *Accid. Anal. Prevent.* 86, 161–172.
- Washington, S., Karlaftis, M., Manning, F., Anastasopoulos, P., 2020. *Statistical and Econometric Methods for Transportation Data Analysis*. Chapman and Hall/CRC.
- Wu, Y., Abdel-Aty, M., Lee, J., 2018. Crash risk analysis during fog conditions using real-time traffic data. *Accid. Anal. Prev.* 114, 4–11.
- Asashio Highway Safety Manual 2010 *Transportation Research Board Washington*, D.C.
- Xing, L., He, J., Abdel-Aty, M., Cai, Q., Li, Y., Zheng, O., 2019. Examining traffic conflicts of up stream toll plaza area using vehicles' trajectory data. *Accid. Anal. Prevent.* 125 (APR.), 174–187.
- Xu, C., Tarko, A.P., Wei, W., Pan, L., 2013. Predicting crash likelihood and severity on freeways with real-time loop detector data. *Accid. Anal. Prev.* 57 (AUG.), 30–39.
- Yao, X., Zhao, X., Liu, H., Huang, L., Yin, J., 2019. An approach for evaluating the effectiveness of traffic guide signs at intersections. *Accid. Anal. Prev.* 129, 7–20.
- Yasmin, S., Eluru, N., Wang, L., Abdel-Aty, M.A., 2018. A joint framework for static and real-time crash risk analysis. *Analyt. Methods Accid. Res.* 18, 45–56.
- 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.
- Yuan, J., Abdel-Aty, M.A., Yue, L., Cai, Q., 2020. Modeling real-time cycle-level crash risk at signalized intersections based on high-resolution event-based data. *IEEE Trans. Intell. Transp. Syst.* 22 (11), 6700–6715.
- Zadeh, L.A., 2008. Fuzzy logic. *Computer* 21 (4), 83–93.
- Zheng, L., Sayed, T., 2020. A novel approach for real time crash prediction at signalized intersections. *Transp. Res. Part C: Emerg. Technol.* 117.
- Zou, B., Li, S., Zheng, Z., Zhan, B.F., Yang, Z., Wan, N., 2020. Healthier routes planning: A new method and online implementation for minimizing air pollution exposure risk. *Comput. Environ. Urban Syst.* 80.