Traffic & Transportation

# Analysis of Traffic Conflicts at Roundabout Entrances and Exits – A Machine Learning Approach for Enhanced Safety

Yuzhou DUAN[1], Zhipeng LIN[2], Yulong WANG[3], Qiaowen BAI[4]

Review
Submitted: 6 June 2024
Accepted: 13 Dec 2024

[1] duanyz@haut.edu.cn, College of Civil Engineering, Henan University of Technology, Zhengzhou, China
[2] linzhipeng0318@163.com, College of Civil Engineering, Henan University of Technology, Zhengzhou, China
[3] k16657096@163.com, College of Civil Engineering, Henan University of Technology, Zhengzhou, China
[4] Corresponding author, victorbty@foxmail.com, College of Design and Engineering, National University of Singapore, Singapore



Publisher:
Faculty of Transport and Traffic Sciences, University of Zagreb

**ABSTRACT**

As a component of the urban road network, roundabouts play a crucial role in ensuring operational efficiency. The safety performance of roundabouts significantly impacts overall traffic safety, making it necessary to conduct safety analysis and evaluation. This study utilises UAV to capture video of vehicle trajectory at roundabouts, employing the time to collision (TTC) index and vehicle evasive actions to identify and analyse traffic conflicts. A real-time traffic safety evaluation method has been developed using machine learning algorithms, including random forest (RF), support vector machine (SVM), extreme gradient boosting (XGBoost) and decision tree (DT) model. This method aims to analyse the relationship between traffic states and conflicts, providing insights into potential safety risks in various traffic conditions. The four machine learning algorithms trained a total of 12 models, with RF demonstrating superior training effectiveness. It achieved high accuracy in predicting traffic conflict areas at the entrances and exits of a roundabout, with a prediction accuracy of 0.86 and an AUC (area under the receiver operating characteristic curve) of 0.88. In addition, this paper further explores the relationship between traffic conflict and state. The results show that traffic flow, speed, density, speed standard deviation and vehicle type ratio have a significant relationship to traffic conflict. This research provides valuable insights for transportation authorities to understand the nature of traffic conflicts at roundabouts, enabling them to implement appropriate early warning systems and management strategies.

**KEYWORDS**

roundabouts; machine learning algorithms; traffic states; traffic conflicts.

## 1. INTRODUCTION

All over the world, traffic safety has always been a major concern. In China, according to the National Bureau of Statistics of China in 2022, there were 256,409 road traffic accidents, resulting in 60,676 fatalities and 263,621 injuries, with direct property losses amounting to one billion two hundred million CNY. These statistics highlight the substantial threats to residents' lives and property posed by traffic accidents. Urban intersections, accounting for approximately 29.7% of road traffic accidents, pose significant safety challenges. The severity of safety issues at these intersections is particularly notable [1]. Based on the above background, research aimed at reducing and preventing traffic accidents at urban intersections has become a prominent topic within the field of transportation [2-6].

A roundabout is an at-grade intersection with a circular or elliptical central island in the centre. The accident rate of roundabouts is often lower than that of other types of at-grade intersections [7]. Accident records serve as the basic data for evaluating roundabout safety, enabling direct analysis through accident probability.

Existing methods for urban roundabout safety analysis and evaluation frequently rely on traffic accident data [8]. However, due to the randomness and scarcity of accidents, roundabouts often require a longer time to accumulate sufficient accident data to support evaluations compared to other intersections [9]. Therefore, safety evaluations of roundabouts cannot rely solely on traffic accident data. As an effective alternative, traffic conflicts are used for safety analysis at roundabouts. Traffic conflict is defined as the risk of collision between two or more road users while maintaining constant kinematics [10]. Real-time conflict prediction technology has been applied to intersection safety analysis and evaluation in recent years. Real-time conflict prediction is a highly effective measure to prevent accidents, as it can identify unsafe traffic states before accidents occur. When an intersection is determined to be in a dangerous condition, traffic management authorities can implement proactive safety measures, such as dynamic signal control, alert signs and other interventions, to reduce safety risks [11]. A large number of studies have demonstrated the effectiveness of real-time traffic conflict prediction technology in traffic safety analysis [12].

To address research gaps in acquiring roundabout accident records, the limited occurrence of accidents and the challenges in conducting safety evaluations, this paper utilises traffic conflict technology to assess the safety performance of roundabouts. A real-time traffic conflict prediction method is constructed by integrating traffic state and traffic conflict data. This method identifies the occurrence of traffic conflicts through real-time traffic states at the roundabout [13]. The established real-time traffic conflict prediction method offers practical advantages, as traffic state data are easier to obtain than traditional traffic trajectory data [14]. A conflict prediction model based on traffic status can be developed by acquiring traffic trajectory data for certain vehicles within the roundabout and extracting traffic states and traffic conflict data. This model allows the use of real-time traffic state data from detectors to predict traffic conflicts in specific spatial and temporal contexts, thereby enabling real-time safety evaluation of the roundabout [15].

Based on the above, the research objectives of this study can be summarised as follows:

— Explore the collision characteristics of vehicles at the roundabout
— Investigate the significance of different traffic state variables on vehicle conflicts at the roundabout
— Establish a real-time traffic conflict prediction model based on traffic state and traffic conflict for real-time safety evaluation of roundabout
— Investigate the relationship between traffic states and traffic conflicts at the roundabout.

## 2. LITERATURE REVIEW

To accurately establish the relationship between traffic states and conflicts at roundabouts, this study reviewed previous research in three key areas: studies on conflicts at roundabouts, studies on traffic conflict identification and studies on real-time traffic conflict prediction methods.

### 2.1 Research on traffic conflicts at roundabouts

Current research on traffic conflicts at roundabouts primarily focuses on exploring the distribution and types of traffic conflicts. Mohamed et al. used unmanned aerial vehicles (UAVs) to capture the traffic operations video at roundabouts, enabling the identification of traffic conflict types and their distribution. The highest probability and severity of traffic conflicts were observed in the entrance and exit areas of the roundabout, where the severity of conflicts was also the greatest [16]. Li et al. conducted a specific investigation into conflicts between vehicles and non-motorised vehicles at roundabouts. The findings revealed that traffic conflicts between vehicles and non-motorised vehicles primarily occur in the outer lane, specifically where vehicles enter the roundabout. This conflict is characterised by interactions between non-motorised vehicles in the entrance lane and vehicles circulating within the roundabout [17]. Zhang et al. analysed traffic conflicts at the roundabout using vehicle trajectory data for entry and exit. The study revealed a significant relationship between the occurrence of traffic conflicts and vehicle lane change behaviour [18]. When investigating vehicle operation at a roundabout, Zura et al. proposed that the traffic status of vehicles at the entering and exiting areas of the roundabout was often chaotic [19].

Vehicles frequently changed lanes when entering or exiting roundabouts due to varying traffic demands. This lane change behaviour is prone to causing traffic conflicts. Conversely, during the stage of navigating around the roundabout, the driving lane remained relatively fixed, and traffic conflicts were less likely to occur [20]. Therefore, this study primarily investigated traffic conflict situations within the entry and exit area of roundabout. Specific regional divisions are elaborated upon in the 4.1 section.

## 2.2 Research on traffic conflict identification

How to identify traffic conflict is one of the crucial steps in predicting traffic conflict. Presently, researchers commonly rely on spatial or temporal proximity as a standard for conflict identification. Discriminant indicators include time to collision (TTC) and its derivative indicators. This method provided an objective means to identify vehicle conflicts [21]. If a vehicle's TTC value or its derivative indicators is less than a certain threshold, it is concluded that the vehicle is involved in a conflict. Therefore, establishing an appropriate threshold for TTC is of particular importance. A large number of researchers have explored the TTC value of vehicle conflict at urban roundabouts, and finally, the threshold is determined to be 4 s [7]. That is, if the TTC value of the vehicle at the roundabout is less than 4 s, the traffic conflict will be identified. However, the identification of traffic conflicts by the TTC method is only applicable to rear-end collisions, and the TTC method is not applicable in the face of crossing conflicts [22].

Some studies have shown that vehicle evasive actions play a very important role in identifying traffic conflicts [23]. Vehicle evasive actions refer to the deceleration or steering behaviour of both or one of the vehicles to avoid collision [24]. Evasive actions are common in traffic operations, and drivers often avoid traffic accidents by evasive actions. Chai et al. extracted trajectory data between conflicting vehicles and built a vehicle conflict recognition model based on deceleration or steering behaviours between conflicting vehicles, which proved to be able to identify vehicle conflict through practice [25] accurately. Tag et al. used traffic data from different cities to prove the effectiveness of using evasive actions to identify traffic conflicts [26]. Mattas et al. reproduced the driver's evasive actions in the face of a possible collision. The study showed that when the driver adopts evasive actions in the face of an impending collision, the probability of the accident will be greatly reduced [27]. However, the evasive actions of some vehicles are small and difficult to observe, and it is easy to identify traffic conflict errors at this time.

After the literature review, it was found that only using the TTC method or evasive actions to identify traffic conflict recognition has limitations in the scene. However, the combination of evasive actions and the TTC method can greatly improve the accuracy of traffic conflict identification. Because the advantages and disadvantages of these two methods complement each other, the vehicle often has relatively large evasive actions when the crossing conflict occurs, and the method of evasive actions can be used to identify it. The TTC method can be used to identify the rear-end conflicts in the face of some smaller evasive actions.

## 2.3 Research on real-time traffic conflict prediction techniques

Using traditional traffic crash analysis methods to evaluate traffic safety has notable limitations, making it necessary to adopt new traffic conflict techniques. Current traditional traffic crash analysis methods fall into two categories: (1) studying the link between crash data (frequency or severity) and macroscopic traffic conditions (volume, speed, occupancy), and (2) analysing conflicts using individual vehicle trajectory data. Both methods face practical challenges – crash data may be unavailable for certain roads, and extracting trajectory data from traffic videos is time-consuming and requires offline processing. These issues make real-time evaluation difficult, highlighting the need for more efficient conflict-based techniques.

The essence of real-time traffic conflict prediction technology lies in establishing a connection between traffic state indicators and traffic conflict indicators, using the former to predict the latter [18]. However, there is often no linear relationship between these indicators. Traffic state indicators may be continuous, while traffic conflict events are relatively rare, leading to an imbalanced correspondence between the two [28]. Machine learning algorithms offer accurate predictions and general applicability to unknown data, demonstrating robustness in handling high-dimensional data and complex features. Moreover, these algorithms can manage nonlinear data relations and exhibit good tolerance for missing values and outliers. Therefore, to evaluate the safety status of intersections in real time, researchers employ various machine learning algorithms to construct real-time traffic conflict prediction models. These models predict the probability of conflict occurrence based on short-term traffic state data. Based on different kinds of machine learning algorithms, Gore, N. et al. developed a comprehensive traffic conflict assessment framework using macroscopic traffic state variables. The model results indicate a significant relationship between the traffic conflict rate and traffic state. Among different machine learning models, the RF model was observed as the best-fitted model to predict traffic conflict based on macroscopic traffic variables [29]. Orsini, F. et al. extracted vehicle trajectory data from the diversion area of the expressway, subsequently developing a real-time traffic conflict prediction model utilising the SVM algorithm within a machine learning framework, which yielded significant results [30]. Yang et al. use a combination of tracking algorithm, inverse perspective mapping (IPM), and trajectory prediction mechanism of machine learning to improve the accuracy of traffic conflict identification [31].

Compared with the traditional crash-based models, the method of establishing the relationship between traffic state and traffic conflict indicators using machine learning algorithms to evaluate traffic safety can identify potential traffic safety hazards more effectively. However, the correlation between these indicators varies across different environments. Therefore, to ensure the accuracy of the real-time traffic conflict prediction model, this paper explores the correlation between various traffic state indicators and traffic conflict indicators at roundabouts. It employs machine learning algorithms to process the relationship between traffic state and traffic conflict, ultimately constructing a real-time vehicle conflict prediction model for roundabouts.

## 3. METHODS

The methodology of this paper is divided into three parts to establish the real-time conflict prediction model of roundabout accurately. The first part is to identify the traffic conflict accurately and establish the traffic conflict identification method. The second part utilises vehicle trajectory data captured by UAVs at the roundabout. It employs the conflict identification method from the first part to identify traffic conflicts and extract traffic status and conflict data (the data collection method will be introduced in detail in Section 4). A binary logistic model was used to screen traffic state indicators significantly associated with traffic conflicts. The third part is to use machine learning algorithms to train traffic state data and traffic conflict data to build a real-time vehicle conflict prediction model. The method system of the paper is shown in *Figure 1*.

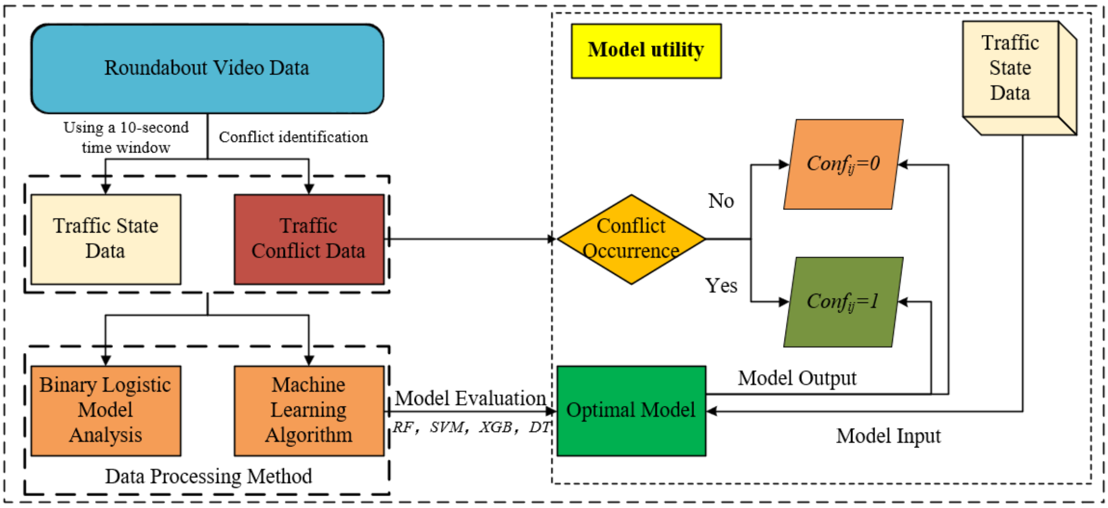



*Figure 1 – Methodological framework*

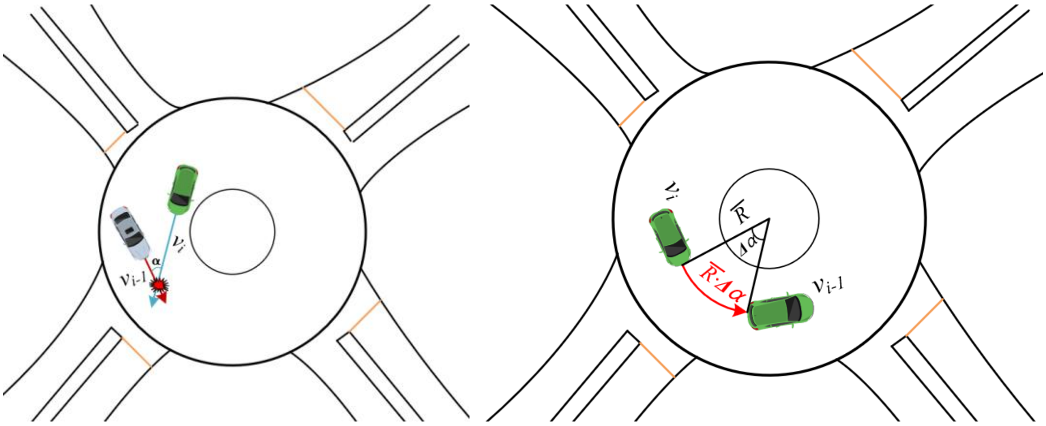
### 3.1 Traffic conflict identification

This paper employs a combination of evasive actions and the TTC index to determine traffic conflicts. TTC refers to the remaining time before a collision between a following vehicle and a leading vehicle if both maintain their current states of motion. This index can objectively describe collision behaviour between vehicles on the same track, such as rear-end collisions. When the TTC value of a vehicle falls below a certain threshold, a collision is deemed imminent, indicating that the vehicle is in a dangerous state. According to the literature review by Li et al., the TTC threshold for vehicle traffic conflicts at roundabouts is 4 s [7].

Vehicles in a roundabout can be classified into two scenarios: straight segment operation and curved segment operation, as shown in *Figure 2*. The TTC calculation method for vehicles in the straight segment is illustrated in *Figure 2a*. The calculation is represented by *Equation 1*;

$$TTC = \frac{x_{i-1}(t) - x_i(t) - l_{i-1}}{v_i(t) - v_{i-1}(t)} \tag{1}$$

where, $x_{i-1}(t)$ and $x_i(t)$ represent the coordinates of the leading vehicle $i$-1 and the following vehicle $i$ on the straight segment at time $t$, measured in m. $v_{i-1}(t)$ and $v_i(t)$ represent instantaneous speeds of the leading vehicle $i$-1 and the following vehicle $i$ on the straight segment at time $t$, measured in m/s. $l_{i-1}$ is the length of the leading vehicle $i$-1, measured in m.




*(a)* *(b)*

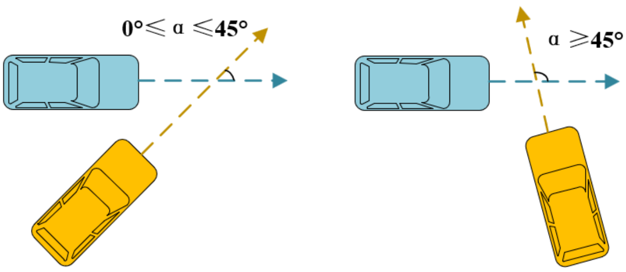
*Figure 2 – Method for calculating TTC: a) Calculation method of TTC for straight segments at the roundabout; b) Calculation method of TTC for curved segments at the roundabout*

The TTC for straight segments assumes that the leading and following vehicles maintain their current direction and speed until a collision occurs. However, in the curved segment of a roundabout, vehicles travel along a curved path, resulting in continuous changes in their direction of motion. This makes the straight-segment TTC algorithm unsuitable for roundabout scenarios. Therefore, the calculation method for TTC in the curved segment of a roundabout must account for the characteristics of vehicle motion along curves. This paper presents an improved method for calculating vehicle TTC at roundabouts, as illustrated in *Figure 2b*. This method assumes that the leading and following vehicles follow an average trajectory at the point of potential conflict, thereby enhancing the accuracy of the TTC calculation. The equation for the improved TTC is as follows:

$$TTC = \frac{[R_{i-1}(t) + R_i(t)] * [\alpha_{i-1}(t) - \alpha_i(t)] - 2l_{i-1}}{2[v_i(t) - v_{i-1}(t)]} = \frac{\bar{R} * \Delta\alpha - l_{i-1}}{v_i(t) - v_{i-1}(t)} \quad (2)$$

where, $R_{i-1}(t)$ and $R_i(t)$ represent the radius of the circular motion around the centre of the roundabout for the leading vehicle *i*-1 and the following vehicle *i* at time *t*, measured in m. $\alpha_{i-1}(t)$ and $\alpha_{i-1}(t)$ represent the radian that the leading vehicle *i*-1 and the following vehicle *i* have turned around the centre of the roundabout at time *t*, measured in rad. $v_{i-1}(t)$ and $v_i(t)$represent the instantaneous speed of the circular motion around the centre of the roundabout for the leading vehicle *i*-1 and the following vehicle *i* at time *t* in the direction of their respective tangent along the motion trajectory, measured in m/s. $\bar{R}$ is the average radius corresponding to the circular motion of the two vehicles around the centre of the roundabout, measured in m. $\Delta\alpha$ is the radian difference between the two vehicles, measured in rad.

However, the TTC method is not suitable for identifying traffic crossing conflicts, making it necessary to use evasive actions as an alternative approach for conflict identification. In crossing conflicts, one or both drivers typically exhibit significant evasive actions, which can be observed. *Figure 3* illustrates the schematic diagrams of rear-end and crossing collisions, while *Figure 4* depicts the identification process of traffic conflicts utilised in this study.




a. Rear-end conflict b. Crossing conflict

*Figure 3 – Traffic conflict schematic diagrams*

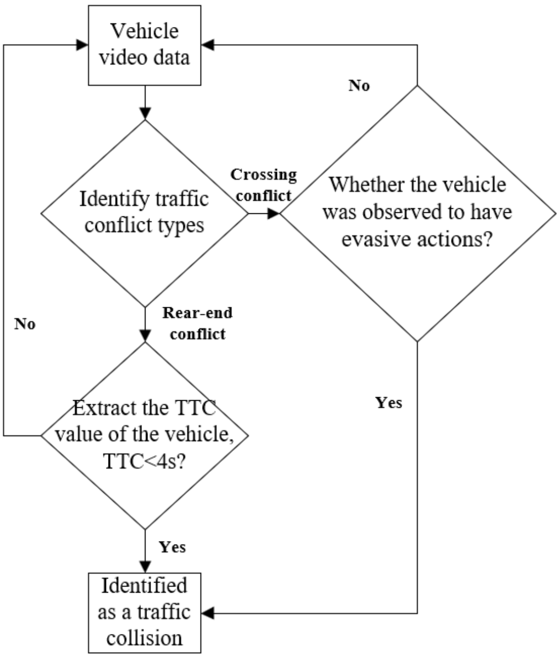



*Figure 4 – Identification process of traffic conflicts*

## 3.2 Binary logistic model

To accurately develop a real-time collision prediction model for vehicles at roundabouts using machine learning algorithms, it is essential to test the input traffic state variables. Variables unrelated to traffic conflicts are eliminated to ensure the accuracy of the input variables in the machine learning algorithm. The binary logistic model effectively identifies significant relationships between input traffic state variables and traffic conflicts. It achieves this through probability modelling, the logistic function, maximum likelihood estimation and significance testing. The binary logistic model can also enhance interpretability by examining statistical significance and the signs of the corresponding coefficients, making the model's output easier to explain and understand. Compared to machine learning classifiers, this further improves its interpretability. Therefore, this paper employs the binary logistic model to test the correlation between the selected traffic state variables and traffic conflicts. The equation of the binary logistic model is shown in *Equations 3* to *4*;

$$f(x) = \ln\left(\frac{p}{1-p}\right) = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n \tag{3}$$

$$Odds\ Ratio(OR) = \frac{\mathrm{p}}{1-p} = e^{f(x)} \tag{4}$$

where $\alpha_n$ is the linear coefficient. $x_n$ is the input variable. $p$ represents the probability of a conflict event. $f(x)$ represents the linear relationship between traffic state variables and traffic conflicts. $OR$ ($Odds\ Ratio$) represents the constant influence of each traffic state variable on traffic conflict.

## 3.3 Machine learning algorithm

This paper employs machine learning algorithms, including random forest (RF), support vector machine (SVM), extreme gradient boosting (XGBoost) and decision tree (DT) to train the relationship between traffic state variables and traffic conflicts. SVM and DT algorithms are non-ensemble methods. The core idea of the SVM algorithm is to find an optimal hyperplane in the feature space to separate different classes of samples, distinguishing between conflicting and non-conflicting instances. The DT algorithm's core idea is to train on traffic state and conflict data, generate a decision tree, and use it to predict new traffic states. RF and XGBoost are ensemble algorithms that combine multiple weak learners into strong predictive models. The aforementioned machine learning algorithms exhibit excellent robustness in handling high-dimensional data and complex features, managing nonlinear data relationships, and tolerating missing values and outliers. Traffic conflicts at roundabouts entering and exiting areas may relate to various traffic state variables, such as vehicle speed, speed standard deviation and density, making the data both high-dimensional and voluminous. Therefore, this paper utilises RF, SVM, XGBoost and DT algorithms to train the relationship between traffic state data and traffic conflicts. 70 per cent of the data is randomly selected as the training set, and the remaining 30 per cent serves as the test set. The data are input into the RF, SVM, XGBoost and DT algorithms and trained using Python programming. To effectively prevent model overfitting, this paper uses the k-fold cross-validation method. Specifically, the dataset is divided into k subsets. One subset is used as the validation set, while the remaining subsets are used as the training set. By training the model on multiple training and validation sets, the model's performance can be effectively evaluated for overfitting. If the model shows significant performance differences across subsets, especially performing well on the training set but poorly on the validation set, it indicates a risk of overfitting.

In addition, to address dataset imbalance and improve classifier performance, data resampling techniques were employed. Random under-sampling and SMOTE (synthetic minority over-sampling technique) were compared for data rebalancing.

### 3.4 Model evaluation

The accuracy rate (A), precision rate (P), recall rate (R), comprehensive evaluation index (F1) and the area under the ROC (receiver operating characteristic) curve (AUC) serve as evaluation indicators to assess the reliability of the aforementioned three algorithms for conflict identification prediction. Among them, P and R are contradictory indicators. It is necessary to use F1 to calculate the harmonic mean of the accuracy rate and the recall rate. AUC serves as a comprehensive indicator for model performance, with values ranging from 0 to 1. A higher AUC value indicates a more ideal predictive performance of the model. A, P, R and F1 can be directly calculated using the equations presented in *Equations 5* to *8*;

$$A = \frac{\mathrm{T}_{cor} - T_{noc-c}}{\mathrm{T}_{cor} + T_{noc-c} + F_{cor} + F_{noc-c}} \tag{5}$$

$$P = \frac{\mathrm{T}_{cor}}{\mathrm{T}_{cor} + F_{cor}} \tag{6}$$

$$R = \frac{\mathrm{T}_{cor}}{\mathrm{T}_{cor} + F_{noc-c}} \tag{7}$$

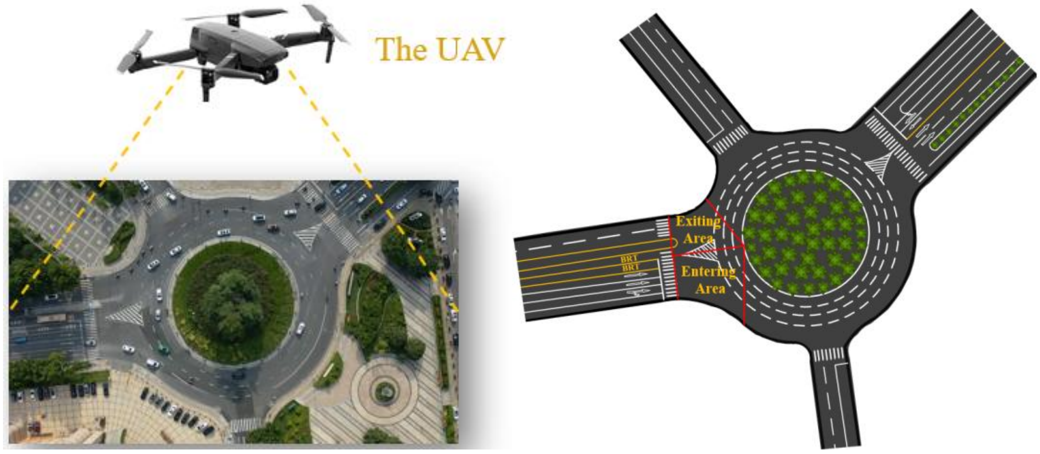
$$F1 = \frac{2 * P * R}{\mathrm{P} + R} \tag{8}$$

where $\mathrm{T}_{cor}$ is the correct prediction of a traffic conflict. $T_{noc-c}$ is the correct prediction of a non-traffic conflict. $F_{cor}$ is the incorrect prediction of a traffic conflict. $F_{noc-c}$ is the incorrect prediction of a non-traffic conflict.

## 4. DATA COLLECTION AND DESCRIPTIVE STATISTICS

### 4.1 Data collection

The vehicle trajectory video dataset for the roundabout was obtained from 4K HD recordings of vehicle operations at the Ruida Road – Hehuan Street roundabout in Zhengzhou City during morning and evening peak hours, captured by a UAV. This roundabout is a signal-controlled intersection, reflecting the current state of roundabouts in China [32]. Additionally, the proximity of large shopping malls to the roundabout results in substantial daily traffic flow, meeting the requirements for data extraction. The schematic diagram of the video capture setup is shown in *Figure 5a*. Literature review indicates that vehicle conflicts at roundabouts predominantly occur in the entrance and exit areas. Therefore, to accurately extract vehicle states and conflict data, this study delineates the entrance and exit areas of the roundabout as shown in *Figure 5b*. The entrance and exit areas of the remaining lanes in the roundabout's direction are divided according to *Figure 5b*.




(a) (b)

*Figure 5 – Roundabout: a) Schematic diagram of the video capture; b) Entering and exiting area of the roundabout*

This paper aims to extract two types of data from the video: traffic state data and traffic conflict data. Initially, traffic conflict data will be extracted using the method outlined by Yuan et al. [33]. The traffic conflict data extraction method is shown in *Figure 6*. To anticipate conflicts proactively, we will create a real-time risk prediction model. This model will use features from the 10-second window before the zero time marker. Including non-conflict observations is crucial, as they offer distinctive features that help differentiate between conflict-prone situations. In this study, non-conflict observations refer to scenarios where no conflicts occurred during the corresponding time frame or immediately after. As shown in *Figure 6*, we exclude the intervals from 10 to 20 seconds before the zero time marker and the 10-second period after a conflict to avoid any potential influence from the emergence of conflicts. After partitioning the time slices, we segment the entire trajectory dataset into multiple segments. Each observation is uniquely identified by a start and end timestamp and labelled with a binary indicator: 0 for no conflict and 1 for conflict. This is illustrated in *Equation 9*. The extracted data are categorised according to the aforementioned roundabout entry and exit areas.

$$\begin{Bmatrix} Conf_{if} = 0, & if\ on\ conflict \\ Conf_{if} = 1, & if\ conflict \end{Bmatrix} \tag{9}$$

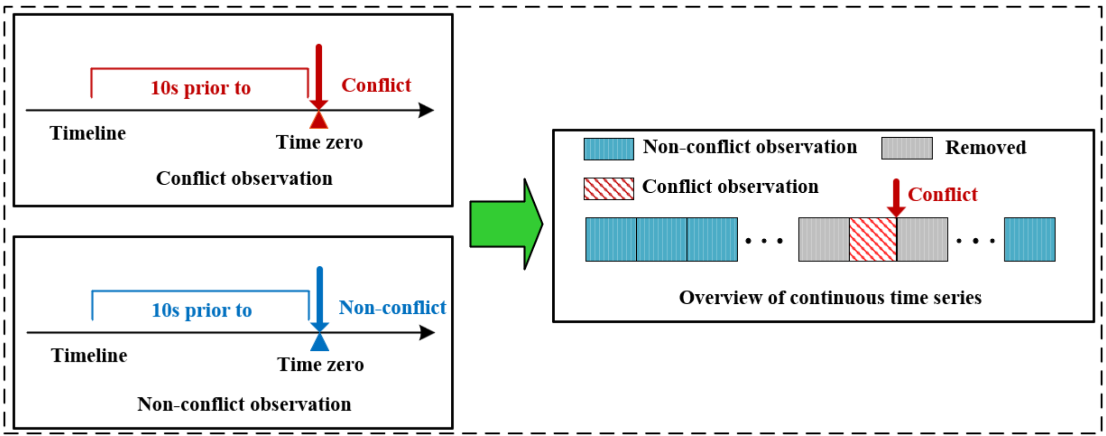



*Figure 6 – Traffic conflict data extraction method*

As the research object of this research is the urban roundabout, they typically accommodate various traffic flows, including both vehicles and non-motorised vehicles. The presence of non-motorised vehicles and large vehicles can influence the capacity and service level of the roundabout, thereby impacting the safety of vehicular operations. Therefore, during the extraction of traffic state data, indicators such as traffic volume, vehicle speed, speed standard deviation, traffic density, the percentage of small vehicles, the percentage of large vehicles and the percentage of non-motorised vehicles are included, as shown in *Table 1*.

*Table 1 – Traffic variable interpretation*

| Category | Variables | Meanings |
|---|---|---|
| *Volume (pcu/10s)* | *Q* | Traffic volume of entering and exiting area of the roundabout |
| *Speed (m/s)* | *Avg_speed* | The average speed of entering and exiting area of the roundabout |
| | *Stds_speed* | Standard deviation of *Avg_speed* |
| *Density (pcu/m)* | *D* | Traffic density of entering and exiting area of the roundabout |
| *Traffic composition* | $P_{car}$ | Proportion of cars entering and exiting area of the roundabout |
| | $P_{truck}$ | Proportion of large vehicles entering and exiting area of the roundabout |
| | $P_{n\text{-}mv}$ | Proportion of non-motorised vehicles entering and exiting area of the roundabout |

*Where Q is equivalent to traffic volume, with a conversion factor of 0.2 for non-motorised vehicles, 1 for small vehicles and 2 for large vehicles such as BRT [34].*

## 4.2 Descriptive statistics and conflict characteristics analysis

Using Tracker software for video data extraction, the dataset included information on 1,544 vehicles and 2,338 non-motorised vehicles.

The extracted traffic conflict data are presented in *Table 2* and *Figure 7* (a, b, c) to understand the vehicle's operational state during traffic conflicts. The LogNormal function is used to describe the statistical distribution of a random variable in a specific domain. It was applied to fit the distribution of each indicator, with the fitting results shown in *Table 2*. The $R^2$ exceeding 0.80 indicates a satisfactory fitting effect. The speed, acceleration and collision angle value of the vehicle in conflict followed a lognormal distribution. This suggests that the values of these indicators are distributed near the mean value when the vehicle is in conflict, with a low probability of extreme values.

*Table 2 – Traffic conflict characteristics*

| Variable | Mean | Std. | Min. | Max. | P15% | P85% | $R^2$ |
|---|---|---|---|---|---|---|---|
| Speed (m/s) | 4.56 | 0.14 | 0.56 | 9.56 | 2.40 | 5.71 | 0.96 |
| Acceleration ($m/s^2$) | -0.95 | 0.09 | -6.23 | 5.34 | -3.21 | 0.29 | 0.98 |
| Angle (°) | 23.03 | 4.45 | 2.30 | 85.00 | 4.17 | 34.03 | 0.94 |

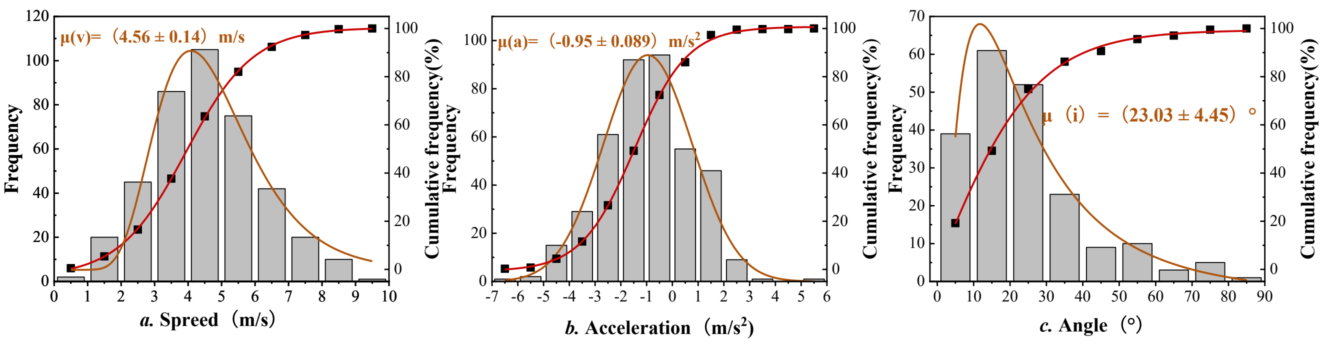



*Figure 7 – Traffic conflict characteristics: a) Speed; b) Acceleration; c) Angle*

As shown in *Table 2* and *Figure 6*, the maximum speed of vehicles during conflicts at the entering and exiting areas of the roundabout is 9.56 m/s, though this occurrence is infrequent. The standard deviation of the average speed is 0.14, indicating that the speed distribution during conflicts is relatively concentrated. The average speed of both vehicles involved in the collision is 4.56 m/s, and the average acceleration is -0.95 $m/s^2$, suggesting a high probability of deceleration behaviour from one or both vehicles. The mean collision angle is 23.03°, with most collision angles below 40°, indicating that rear-end collisions are the predominant type of conflict in these areas. However, collision angles above 45° suggest a small probability of crossing collisions between vehicles in these zones.

## 5. RESULTS

### 5.1 Results of logistic regression models

The correlation relationship between traffic state indicators and traffic conflict indicators was verified through binary logistic model analysis. The results are presented in *Table 3*, including a list of estimated coefficients (Coef.), standard error of the estimation (Std.Err.), corresponding significance level (P-value) and odds ratio (OR).

*Table 3 – Results of binary Logistic model*

| Variable | Coef. | Std.Err. | P-value | OR |
|---|---|---|---|---|
| Volume | 1.260 | 0.231 | <0.001 | 3.525 |
| Speed | -1.320 | 0.140 | 0.014 | 0.267 |
| Std_speed | 1.436 | 0.198 | <0.001 | 4.204 |
| Density | 0.454 | 0.121 | <0.001 | 1.575 |
| $P_{car}$ | 0.226 | 0.021 | 0.029 | 1.254 |
| $P_{truck}$ | -0.434 | 0.023 | 0.021 | 0.648 |
| $P_{n\text{-}mv}$ | -0.420 | 0.014 | 0.034 | 0.657 |

***P<0.001, indicates a extremely significant. **P<0.05 indicates a significant.*

The model estimation results show that significant variables along with the sign of estimated coefficients in the two models are consistent. The coef. value of speed, $P_{truck}$, and $P_{n\text{-}mv}$ are negative, indicating that the likelihood of conflicts decreases as speed increases or as the $P_{truck}$, and $P_{n\text{-}mv}$ increase. Conversely, the coefficients for other selected traffic state variables are positive, suggesting that the probability of conflict rises with increases in volume, std_speed, density and $P_{car}$ values. These findings are further discussed in Section 6. As shown in *Table 4*, the P-values for all selected traffic state variables are less than 0.05, demonstrating that each of these variables has a significant impact on traffic conflicts.

### 5.2 Results of machine learning models

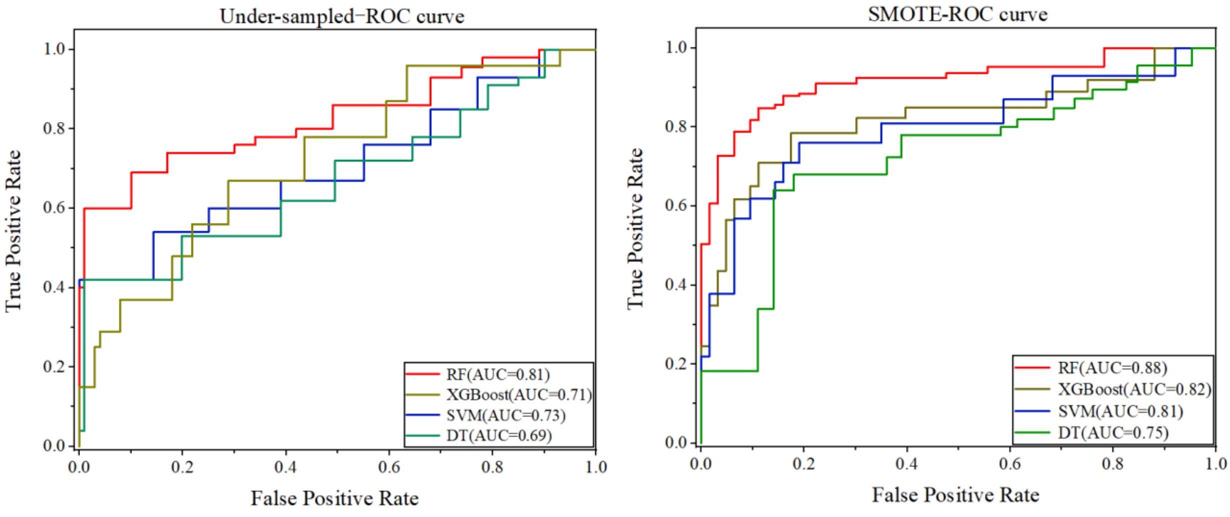
Using the above traffic state indicators as input variables for machine learning, and considering the occurrence of a conflict as a binary output variable, the model was trained. The use of k-fold cross-validation indicates that the trained model does not exhibit overfitting. *Table 4* presents the results of three machine learning models in predicting conflicts, including A, P, R, F1 and AUC.

*Table 4 – Comparison results of classifiers using resampling techniques*

| | Origin data | | | | | Under-sampled | | | | | SMOTE | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | A | P | R | F1 | AUC | A | P | R | F1 | AUC | A | P | R | F1 | AUC |
| RF | 0.86 | 0.64 | 0.66 | 0.65 | 0.79 | 0.83 | 0.92 | 0.90 | 0.82 | 0.81 | 0.86 | 0.86 | 0.88 | 0.87 | 0.88 |
| SVM | 0.82 | 0.68 | 0.70 | 0.69 | 0.70 | 0.72 | 0.73 | 0.92 | 0.81 | 0.73 | 0.79 | 0.83 | 0.88 | 0.85 | 0.81 |
| XGBoost | 0.84 | 0.66 | 0.71 | 0.68 | 0.74 | 0.74 | 0.85 | 0.94 | 0.89 | 0.71 | 0.82 | 0.85 | 0.88 | 0.86 | 0.82 |
| DT | 0.71 | 0.61 | 0.68 | 0.64 | 0.65 | 0.69 | 0.73 | 0.91 | 0.81 | 0.69 | 0.74 | 0.78 | 0.96 | 0.83 | 0.75 |

*Table 4* shows the prediction accuracy, precision, recall, F1 score and AUC values of four machine learning models on the original dataset, the under-sampled dataset and the SMOTE-processed dataset. For the original

dataset, the F1 scores of all four models are below 0.7, indicating poor performance. This suggests that the models often predict unsafe situations as safe, resulting in suboptimal prediction performance. Therefore, this study applied under-sampling and SMOTE processing to the dataset. When comparing the three metrics – accuracy, F1 score and AUC – the dataset processed with SMOTE performed better, with all four models achieving prediction accuracies above 0.7. Among the four traffic conflict prediction models at roundabouts, the RF model outperformed the others, with a prediction accuracy of 0.86, an F1 score of 0.87 and an AUC of 0.87. In contrast, the DT model had the lowest prediction accuracy at 0.74. *Figure 8* presents the ROC curves of the four models after under-sampling and SMOTE processing. Compared to SVM, XGBoost and DT, the AUC of the RF ROC curve is closer to 1, with a value of 0.88. This indicates that the RF model has the best training performance among the four algorithms.




(a) *(b)*

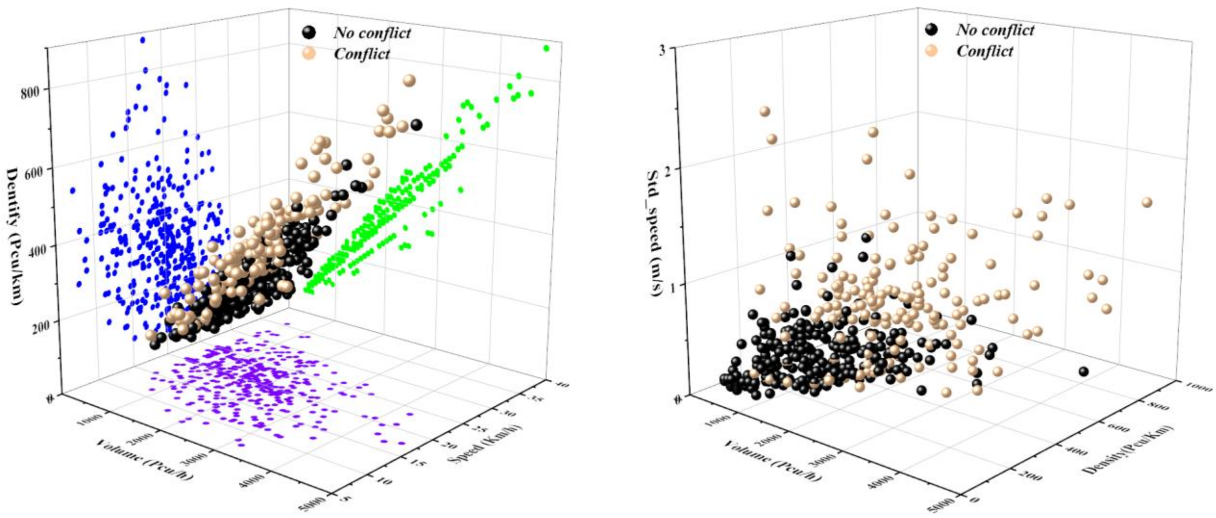
*Figure 8 – Comparison of ROC curves and AUC: a) Under-sampled ROC curve; b) SMOTE ROC curve*

## 6. DISCUSSION

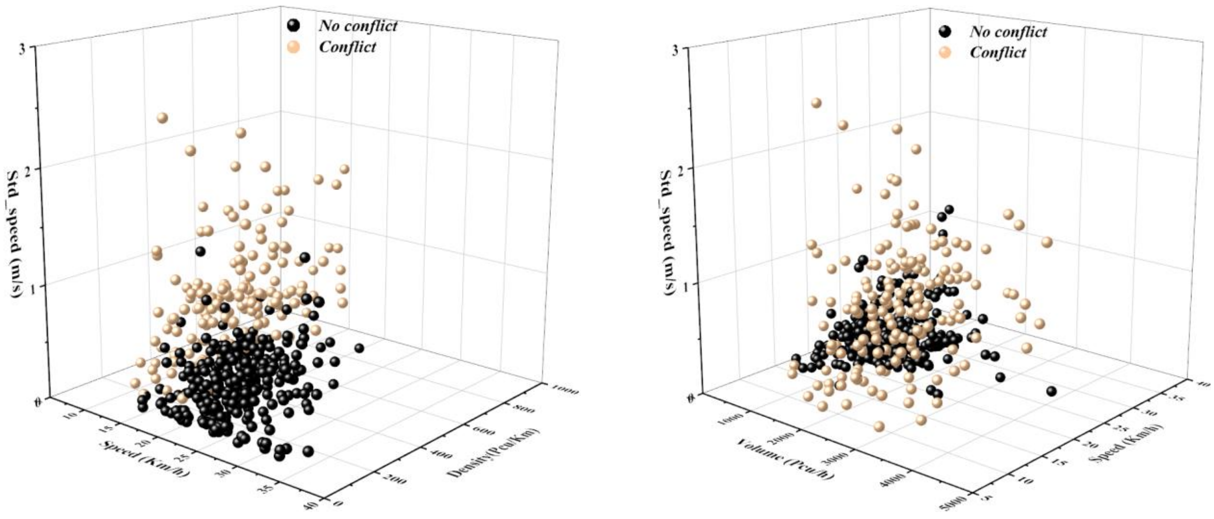
Binary logistic model and machine learning results show that traffic conditions and the proportions of large and non-motorised vehicles significantly impact conflict risk at roundabouts. Increased traffic volume and higher speed variability are linked to a higher conflict risk, while higher average speeds and a greater share of large vehicles reduce this risk. This study will further examine the relationship between traffic conditions and conflict risk.

### 6.1 Effects of traffic state variables

*Figure 9 (a, b, c, d)* illustrates traffic conflict situations under different traffic conditions.




*(a)* *(b)*




*(c)* *(d)*

*Figure 9 – The relationship between traffic state and traffic conflict: a) Conflict interacts for "Volume", "Speed'' and "Density"; b) Conflict interacts for "Volume", Speed" and "Std_speed"; c) Conflict interacts for "Density", "Speed'' and "Std_speed"; d) Conflict interact for "Volume", "Density'' and "Std_speed"*

*Figure 9a* illustrates the relationship between the conflict situation and the flow, speed and density of vehicles at the entering and exiting areas of the roundabout. The risk of traffic conflicts rises with increased traffic flow, especially in conditions of high traffic flow and low speed. *Figure 9b* illustrates the relationship between traffic conflicts and the flow, speed and speed standard deviation. As speed standard deviation increases, the likelihood of collisions also rises, although some conflicts occur even at low-speed standard deviation. According to *Figure 9c*, this phenomenon happens under conditions of low vehicle speed. Higher average vehicle speeds suggest a better driving environment, reducing the likelihood of conflicts; conversely, low-speed conditions indicate poor traffic states where conflicts are more frequent. *Figure 9d* shows that traffic conflicts primarily occur under conditions of high traffic density, flow and speed standard deviation. These results corroborate the conclusions of the binary logistic model in Section 5.1. Thus, traffic conflicts at roundabout entrances and exit areas are closely related to fluctuating flow, density, speed and speed standard deviation, with conflicts being more likely under conditions of high flow, high density, low speed and high-speed variability.

## 6.2 Effects of the proportion of vehicle types

Previous studies on traffic conflicts typically include large vehicles in analyses of crash severity. However, research on the influence of large vehicle presence on road safety within specific spatial and temporal intervals remains limited [35]. Non-motor vehicles are a critical part of urban traffic, and crashes with motor vehicles often lead to severe injuries for non-motor vehicle users. Therefore, it is essential to consider non-motorised vehicles when analysing traffic conflicts at roundabouts. However, most urban intersection safety studies focus exclusively on motor-vehicle interactions, often overlooking non-motor vehicles [36].

According to the binary logistic model, non-motor vehicles and large vehicles significantly impact traffic conflicts. To explore this relationship, this paper analyses data on the proportion of non-motor vehicles and large vehicles relative to the traffic state. *Figure 10* illustrates the relationship between non-motor vehicles, large vehicles and the average speed of vehicles. *Figure 10a* shows that conflicts primarily occur when the proportion of non-motor vehicles is below 0.5, with fewer conflicts when this proportion exceeds 0.5. Due to their small size and manoeuvrability, non-motor vehicles drive across roundabouts more easily than motor vehicles. A low proportion of non-motor vehicle traffic might not alarm motor vehicle drivers, leading to conflicts due to the unpredictable paths of non-motor vehicles. However, as the proportion of non-motor vehicles increases, motor vehicle drivers become more vigilant, reducing the probability of conflicts. When all vehicles at the roundabout are non-motor vehicles, the likelihood of conflicts significantly diminishes, and their speed is relatively high, indicating that traffic conditions are better when only non-motor vehicles are present compared to mixed traffic.

For large vehicles, the study examines the relationship between the proportion of large vehicles and their speed. *Figure 10b* shows that as the proportion of large vehicles increases, the average speed of all vehicles gradually decreases. This is due to the significant impact large vehicles have on the road traffic environment. Large vehicles, with their considerable size, obstruct other drivers' visibility, making smaller vehicles and non-motorised vehicles more cautious in assessing road conditions. Consequently, they reduce speed to ensure safety. Additionally, large vehicles accelerate slowly and require longer braking distances. At roundabouts, they tend to slow down the traffic behind them, forcing other vehicles to adjust to their lower speeds. Smaller vehicles travelling near large vehicles typically maintain a greater following distance, further increasing gaps between vehicles, reducing overall flow efficiency and leading to a decrease in average speed. The presence of large vehicles prompts smaller vehicles and non-motor vehicles to be more cautious, reducing their speed to avoid conflicts.

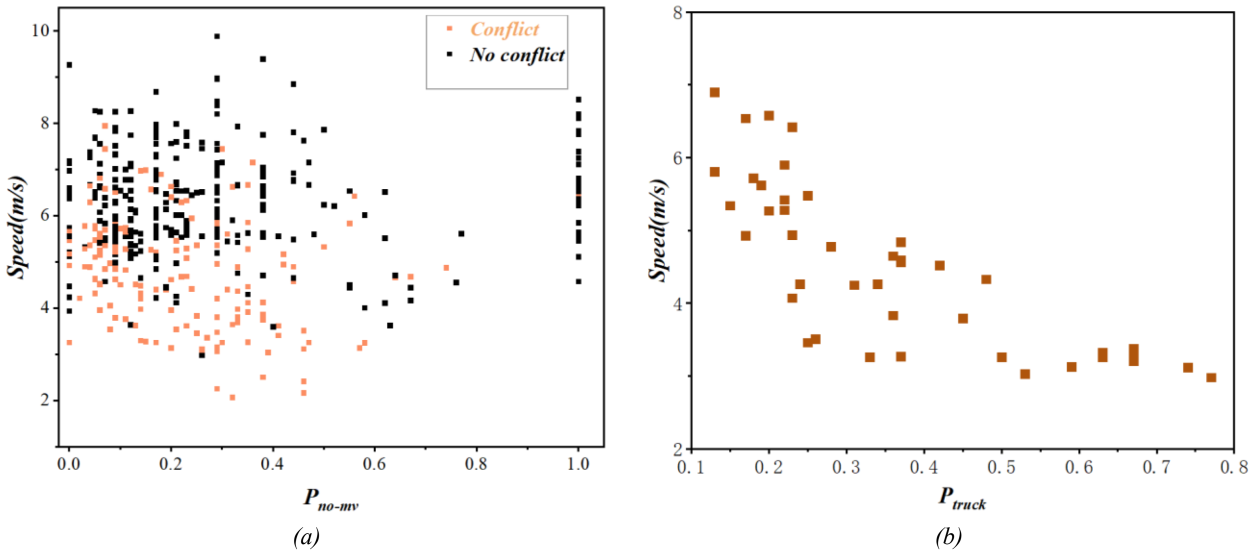



*Figure 10 – Conflict interaction for $P_{no\text{-}mv}$ and $P_{truck}$: a) Conflict interaction for $P_{no\text{-}mv}$; b) Conflict interaction for $P_{truck}$*

## 6.3 Interpretability of RF model

This study applied machine learning methods to predict traffic safety at roundabouts. The modelling results show that RF, as an ensemble algorithm, outperforms single algorithms like SVM and DT. Additionally, whether using the original or resampled dataset, RF consistently demonstrates better predictive performance than XGBoost. The advantages of the RF method are reflected in the following aspects: (1) RF is easier to implement in real-world applications, offering high prediction accuracy and fast training; (2) RF has strong noise resistance and adaptability across different datasets; (3) its predictive accuracy and performance stability surpass single algorithms, and its running speed is faster than more complex methods like deep learning.

The RF algorithm generates a feature importance ranking table for input indicators in different scenarios, as illustrated in *Figure 11*. Since the prediction performance of various models improved to varying degrees after data were processed with SMOTE, this paper focuses on discussing the importance of indicators estimated by RF after SMOTE processing. Features influencing conflict occurrence are ranked based on their importance. The RF model finds speed-related variables (standard deviation of speed and average speed) to be the most critical in predicting traffic conflicts, followed by volume, density and vehicle type proportions.

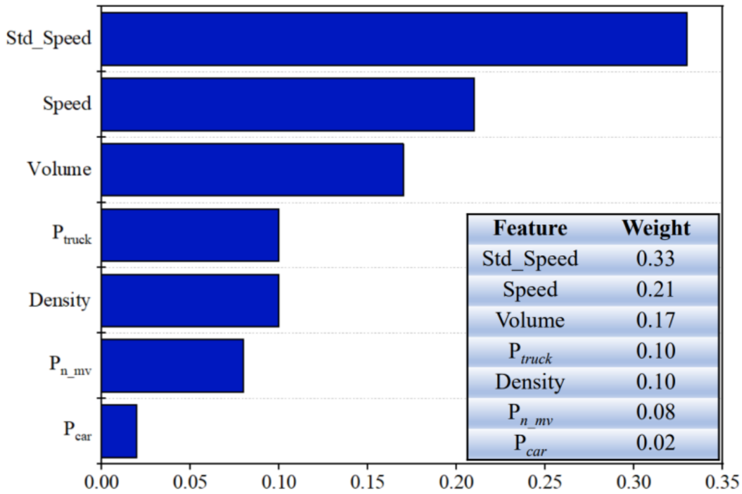



*Figure 11 – Features importance of SMOTE-RF*

The importance ranking of predictive indicators provided by the RF model is consistent with the binary logit model. In the binary logistic model, the OR value of std_speed is 4.204, and the coef. value is 1.436, ranking first among positively significant indicators. This indicates that std_speed has the greatest positive impact on traffic conflicts at roundabout entrances and exits. For a one-unit increase in std_speed, the probability of traffic conflict significantly increases by 43.6%. The OR value of speed is 0.267, and the coef. value is -1.320, ranking first among negatively significant indicators. This means that for a one-unit increase in speed, the probability of traffic conflict significantly decreases by 32%. In the RF model, speed-related variables (standard deviation of speed and average speed) are also identified as the most critical factors in predicting traffic conflicts, consistent with the binary logistic model.

## 7. CONCLUSION

Using vehicle trajectory data collected at the roundabout, this paper combines TTC and avoidance behaviour to identify traffic conflicts and analyse vehicle collision characteristics. A real-time traffic conflict prediction method is then developed, integrating traffic status and conflict data to explore their interrelationship. Specifically, a binary logistic regression model quantifies the relationship between traffic state and conflicts. Additionally, machine learning algorithms – including RF, SVM, XGBoost and DT – are employed for the prediction model. The conclusions of this study are as follows.

— The established binary logistic model demonstrates a significant relationship between traffic state variables (i.e. traffic volume, speed, density, standard deviation of speed, proportion of motor vehicles, proportion of non-motor vehicles, proportion of large vehicles) and traffic conflicts. Among these, speed, proportion of large vehicles and proportion of non-motor vehicles have a significant negative relationship with traffic conflict, whereas other traffic state variables exhibit a significant positive relationship.
— The performance of the RF, SVM, XGBoost and DT models on the dataset processed with SMOTE is better than that on the unprocessed dataset or the dataset processed with under-sampling techniques. The RF model achieves the best performance, particularly on the dataset processed with SMOTE. Traffic state factors used as model inputs can effectively predict real-time conflicts. The prediction accuracy of the RF algorithm in the entering and exiting area of the roundabout is 0.86, with AUC values of 0.88.
— Data statistics have shown that rear-end collisions are the predominant type of traffic conflict in these areas.
— Traffic conflict at the entering and exiting areas of the roundabout is closely related to time-varying flow, density, speed and speed standard deviation. Traffic conflicts are more likely to occur in traffic states of high traffic flow, high density, low speed and high-speed standard deviation. When the proportion of non-motor vehicles is low, the probability of traffic conflict is greater than the traffic state with a high proportion of non-motor vehicles. As the proportion of large vehicles increases, the average vehicle speed decreases, leading to a decrease in the probability of conflict.

## ACKNOWLEDGEMENT

This research was funded by the International Science and Technology Cooperation Program of Henan Province, China, grant number 242102520030. This research was funded by the Foundation of Henan University of Technology for Outstanding Young Teachers, grant number 21420157.

## REFERENCES

[1] Fu L, et al. Road traffic accidents (2022). *China Statistical Yearbook*. 2023:4-5.

[2] Tao W, et al. A conflict risk graph approach to modeling spatio-temporal dynamics of intersection safety. *Transportation Research Part C: Emerging Technologies*. 2024;169:104874. DOI: 10.1016/j.trc.2024.104874.

[3] Wei W, et al. Enhancing intersection safety in autonomous traffic: A grid-based approach with risk quantification. *Accident Analysis and Prevention.* 2024;200:107559. DOI: 10.1016/j.aap.2024.107559.

[4] Lu Y, et al. An improved object detection and trajectory prediction method for traffic conflicts analysis. *Promet - Traffic&Transportation*. 2023;35(4):462-484. DOI: 10.7307/ptt.v35i4.173.

[5] Paul AB, et al. Investigating and modeling the influence of PET-types on crossing conflicts at urban unsignalized intersections in India. *International Journal of Injury Control and Safety Promotion*. 2022;30(2):239-254. DOI: 10.1080/17457300.2022.2147194.

[6] Shahana A, et al. Spatiotemporal variation of traffic conflict at signalized intersections under non-lane-based traffic condition. *Transportation Research Record*. 2023;2677(1):326-340. DOI: 10.1177/03611981221099288.

[7] Li L, et al. The role of traffic conflicts in roundabout safety evaluation: A review. *Accident Analysis and Prevention*. 2024;196:107430. DOI:10.1016/j.aap.2023.107430.

[8] Johnson MT. Effects of phi and view angle geometric principles on safety of multi-lane roundabouts. *Transportation Research Record*. 2023;2677:362-371. DOI:10.1177/0361198123115912.

[9] Hu Y, et al. A high-resolution trajectory data driven method for real-time evaluation of traffic safety. *Accident Analysis and Prevention*. 2022;165:106503. DOI:10.1016/j.aap.2021.106503.

[10] Cao JY, et al. Trajectory data-based severe conflict prediction for expressways under different traffic states. *Physica a-Statistical Mechanics and Its Applications*. 2023;621. DOI:10.1016/j.physa.2023.128595.

[11] Arun A, et al. A systematic mapping review of surrogate safety assessment using traffic conflict techniques. *Accident Analysis and Prevention*. 2021;153:106016. DOI:10.1016/j.aap.2021.106016.

[12] Lu J, et al. Learning the representation of surrogate safety measures to identify traffic conflict. *Accident Analysis and Prevention*. 2022;174:106755. DOI:10.1016/j.aap.2022.106755.

[13] Ray Sarkar D, et al. A review of surrogate safety measures on road safety at unsignalized intersections in developing countries. *Accident Analysis and Prevention*. 2024;195:107380. DOI: 10.1016/j.aap.2023.107380.

[14] Swanson JM, et al. Observing pedestrian-vehicle traffic conflicts in school zones to evaluate the effectiveness of road safety interventions and reduce injuries in Ghana, Vietnam, and Mexico, 2019-2021. *Journal of injury & violence research*. 2022;14:3. DOI: 10.5249/jivr.v14i3.1710.

[15] Adjenughwure K, et al. Monte carlo-based microsimulation approach for estimating the collision probability of real traffic conflicts. *Transportation research record*. 2023;2677(9):314-326. DOI: 10.1177/03611981231159124.

[16] Shawky M, et al. Traffic safety assessment for roundabout intersections using drone photography and conflict technique. *Ain Shams Engineering Journal*. 2023;14. DOI:10.1016/j.asej.2023.102115.

[17] Li R, et al. Analysis of conflict distribution characteristics between motor vehicles and e-bikes at roundabouts. *Applied Sciences-Basel*. 2023;13. DOI:10.3390/app13063475.

[18] Zhang S, et al. Prediction of pedestrian-vehicle conflicts at signalized intersections based on long short-term memory neural network. *Accident Analysis and Prevention*. 2020;148:105799. DOI:10.1016/j.aap.2020.105799.

[19] Žura M, et al. Roundabout capacity estimation model considering driver behaviour on the exiting and entry flows. *Promet-Traffic & Transportation*. 2022;34:397-405. DOI: 10.7307/ptt.v34i3.3973.

[20] Leonardi S, Distefano N. Turbo-Roundabouts as an instrument for improving the efficiency and safety in urban area: an italian case study. *Sustainability*. 2023;15(4). DOI:10.3390/su15043223.

[21] Ortiz FM, et al. Road traffic safety assessment in self-driving vehicles based on time-to-collision with motion orientation. *Accident Analysis and Prevention*. 2023;191:107172. DOI:10.1016/j.aap.2023.107172.

[22] Formosa N, et al. Predicting real-time traffic conflicts using deep learning. *Accident Analysis and Prevention*. 2020;136:105429. DOI:10.1016/j.aap.2019.105429.

[23] Hang J, et al. An improved automated braking system for rear-end collisions: A study based on a driving simulator experiment. *Journal of safety research*. 2022;80:416-427. DOI:10.1016/j.jsr.2021.12.023.

[24] Li D, et al. An integrated approach of machine learning and Bayesian spatial Poisson model for large-scale real-time traffic conflict prediction. *Accident Analysis and Prevention*. 2023;192:107286. DOI:10.1016/j.aap.2023.107286.

[25] Chai H, et al. Trajectory-based conflict investigations involving two-wheelers and cars at non-signalized intersections with computer vision. *Expert Systems with Applications*. 2023;230. DOI:10.1016/j.eswa.2023.120590.

[26] Tageldin A. and T. Sayed. Models to evaluate the severity of pedestrian-vehicle conflicts in five cities. *Transportmetrica a-Transport Science*. 2019;15:354-375. DOI:10.1080/23249935.2018.1477853.

[27] Mattas K, et al. Driver models for the definition of safety requirements of automated vehicles in international regulations. Application to motorway driving conditions. *Accident Analysis and Prevention*. 2022;174:106743. DOI:10.1016/j.aap.2022.106743.

[28] Patel H, et al. Novel traffic conflict-based framework for real-time traffic safety evaluation under heterogeneous and weak lane-discipline traffic. *Transportation Research Record*. 2024;2678:118-134. DOI:10.1177/03611981231172962.

[29] Gore N, et al. Traffic conflict assessment using macroscopic traffic flow variables: A novel framework for real-time applications. *Accident Analysis and Prevention*. 2023;185:107020. DOI: 10.1016/j.aap.2023.107020.

[30] Orsini F, et al. Conflict-based real-time road safety analysis: Sensitivity to data collection duration and its implications for model resilience. *Transportation Research Record*. 2024;2678(1):460-422. DOI: 10.1177/03611981231171151.

[31] Yang L, et al. An improved object detection and trajectory prediction method for traffic conflicts analysis. *Promet-Traffic & Transportation*. 2023;35:462-484. DOI:10.7307/ptt.v35i4.173.

[32] Hu L, et al. Research on identification model of vehicle conflict risk area at road roundabout. *Transportation Systems Engineering and Information*. 2024;24(01):168-178. DOI:10.16097/j.cnki.1009-6744.2024.01.017.

[33] Yuan C, et al. Using traffic flow characteristics to predict real-time conflict risk: A novel method for trajectory data analysis. *Analytic Methods in Accident Research.* 2022;35:100217.

[34] Abdur-Rouf K, Shaaban K. Comparing traffic noise levels between signalized intersections and roundabouts in an urban environment. *Transportation Research Record*. 2022;2676:608-619. DOI:10.1177/03611981221088219.

[35] Wen HY, et al. Analyzing the impact of curve and slope on multi-vehicle truck crash severity on mountainous freeways. *Accident Analysis and Prevention.* 2023;181:106951. DOI:10.1016/j.aap.2022.106951.

[36] Russo B, et al. Analyzing the impacts of intersection treatments and traffic characteristics on bicyclist safety: development of data-driven guidance on the application of bike boxes, mixing zones, and bicycle signals. *Transportation Research Record.* 2023;2677:187-200. DOI:10.1177/03611981231167414.

段宇洲，林志朋，王玉龙，白乔文（通讯作者）

环形交叉口出入口交通冲突分析：一种提高安全性的机器学习方法

摘要
环形交叉口作为城市道路网络的重要组成部分之一，在保障交通运行效率方面发挥着关键作用。其安全性能对交通网络安全具有重要影响，因此对环形交叉口进行安全性分析与评估具有重要意义。本研究利用无人机（UAV）采集环形交叉口内的车辆轨迹视频，结合碰撞时间（TTC）指标和车辆避险行为，对交通冲突进行识别与分析。提出了一种基于随机森林（RF）、支持向量机（SVM）、极端梯度提升（XGBoost）和决策树（DT）机器学习算法的实时交通安全评估方法，用于分析交通状态与交通冲突之间的关系，并识别不同交通条件下的潜在安全风险。研究中使用四种机器学习算法共训练了 12 个模型，其中 RF 算法表现最佳。在预测环形交叉口出入口区域交通冲突时，RF 算法的预测准确率达到 0.86，ROC 曲线下面积（AUC）为 0.88。此外，本文进一步探讨了交通状态与交通冲突之间的关系，发现交通流量、车速、密度、车速标准差及车型比例等因素与交通冲突显著相关。本研究为交通管理部门了解环形交叉口交通冲突特性提供了重要参考，可用于制定高效的安全预警系统和管理策略，从而提高交通安全水平。