Accident Analysis and Prevention 124 (2019) 180–192

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Expressway crash risk prediction using back propagation neural network: A brief investigation on safety resilience

Junhua Wang[a], Yumeng Kong[a], Ting Fu[b,*]

[a] Key Laboratory of Road and Traffic Engineering of the Ministry of Education, 4800 Cao'an Highway, Shanghai 201804, China
[b] Department of Civil Engineering and Applied Mechanics, McGill University, Room 391, Macdonald Engineering Building, 817 Sherbrooke Street West, H3A 0C3 Montréal, Québec, Canada

ARTICLE INFO

Keywords:
Crash risk prediction
Expressway safety
Safety critical events
Safety resilience of traffic
Back-propagation neural network
Vehicle moving violation

ABSTRACT

This study presents the work in predicting crash risk on expressways with consideration of both the impact of safety critical events and traffic conditions. The traffic resilience theory is introduced to learn safety problems from the standpoint of *1)* considering safety critical events, such as traffic violations, as the safety disturbances, and *2)* considering safety resilience as the ability of the traffic, greatly associated with traffic conditions, to resist critical events turning into crashes. The concept of safety resilience was illustrated qualitatively through simulation experiments. Aimsun microsimulation software was used to simulate traffic conditions with safety critical events (vehicle violations, in this paper) involved based on the geometric design of the G15 Expressway in Shanghai. Based on data from the simulation experiment, a two-staged model was developed which classifies crash risk status into three types including no-risk, low-risk and high-risk status. Modeling approach that relies on the back propagation neural network method was applied. The performance of the model in prediction was validated through the Receiver Operating Characteristic (ROC) curve test. Results indicated that the model performed well in predicting crash risks in the simulated environment. After training the model, an extra simulation experiment involving six additional tests was conducted. Results show that the traffic resilience theory may work in explaining the relationship between traffic conditions, safety critical events and crash risk, which are the key elements in road safety field. The introduction of safety resilience may inspire further exploration on this topic in both research and practice. Meanwhile, the model can be used to predict and monitor risks on expressways in a potentially more precise way.

1. Introduction

1.1. Traffic safety on expressway

Crashes on the highways result in huge loss of life and property due to high lethality. For example, in 2016, 5947 of traffic fatalities and 11,956 of injuries happened on highways in China (PSM China, 2017). Bureau of Transportation Statistics in the US Department of Transportation has reported that 35,092 crashes happened on highways, which accounts for 96% percent of the total number (36,569) of road traffic fatalities nationwide (US DOT, 2017). In Canada, 57% of fatal collisions and 24% of injuries crashes in 2015 happened on rural locations which include highways, and local roads with the speed limit exceeding 60 km/h (Transport Canada, 2015). Expressways are probably the most dangerous among different types of highway in terms of severity due to their high speed limit. Despite great attempts having been made, safety issues on expressways are still a big concern in road safety area. For example, expressway crashes contribute to 9.43% of the total death roll and 31.49% of direct loss of property in China (PSM China, 2017). Learning to improve expressway safety has remained a challenge for researchers and practitioners.

Crash risk prediction helps investigate and solve safety issues, and assist in improving operations and safety management. Different studies have investigated different methodologies and proposed different models to predict crash risk on expressways (Lee et al., 2002) (Golob et al., 2004) (Golob et al., 2008) (Abdel-Aty et al., 2004) (Abdel-Aty et al., 2005) (Shi et al., 2015) (Cai et al., 2018). For instance, Lee et al. (2002) proposed an aggregated logarithmic linear crash risk prediction model based on crash information and traffic flow data before and after the crash. Golob et al. (2004) developed used a binary logistic model to model expressway crash risk. The rapid growth in the Intelligent Transportation Systems (ITS) industry, e.g. the development in new

* Corresponding author.
E-mail addresses: benwjh@163.com (J. Wang), 54.kym@163.com (Y. Kong), ting.fu.tj@gmail.com (T. Fu).

https://doi.org/10.1016/j.aap.2019.01.007
Received 4 May 2018; Received in revised form 9 December 2018; Accepted 4 January 2019
Available online 17 January 2019
0001-4575/

traffic sensing technologies, has made it easier for collection microscopic traffic parameters in better learning traffic safety. Ahmed et al. (2012) relied on information from the Automatic Vehicle Information (AVI) to predict crashes using a Bayesian Logistic regression model. To predict crash risk, a study conducted by Yu and Abdel-Aty (2013) used the Road Traffic Management System (RTMS) to collect traffic data and road alignment information such as gap time (Assad, 2013). Ahmed and Abdel-Aty (2013) used both RTMS and AVI data to predict crash risk.

Among the studies, machine learning techniques have been proved to be efficient in crash prediction (Golob et al., 2008) (Abdel-Aty et al., 2004) (Abdel-Aty et al., 2005) (Pande and Abdel-Aty, 2006). Golob et al. (2008) used the probabilistic neural network model based on non-parametric Bayesian to predict the probability of traffic crashes on expressways. Abdel-Aty et al. (2005) proposed a model that was trained using probabilistic neural network to predict crash risk based on the speed dispersion coefficient and the average occupancy rate. The model had an accuracy of over 70% in crash prediction. Pande and Abdel-Aty (2006) used the neural network classifier and was able to build a model which had an accuracy of 75% with 34% in false positive rate. Ahmed and Abdel-Aty (2013) introduced the stochastic gradient inference machine learning method to predict crash risk and the model performed well with an accuracy of 89% and a false alarm rate of 6.5%.

1.2. Traffic condition and human error

Previous research in crash prediction for expressway circumstances has been focused on traffic flow status whose characteristics such as the variation of the operational speed, the upstream occupancy, the variation of flow, among others, have been regarded as the main contributors to road safety (Lee et al., 2002) (Golob et al., 2004) (Golob et al., 2008) (Abdel-Aty et al., 2004) (Abdel-Aty et al., 2005) (Pande and Abdel-Aty, 2006) (Lee, 2003). However, safety critical events, which are trigger for traffic crashes in most cases, have been undervalued. Human errors lead to most safety critical events resulting 99% of vehicle crashes (Hendricks et al., 2000). Vehicle moving violation is one of the major types of human error that generates severe safety critical events. Statistical evidences have proved that vehicle moving violation is a major cause of expressway crashes, leading to 75% of road crashes in China (Zhang et al., 2013), 50% of all fatal crashes in Europe (Rothengatter, 1999), and 56% of fatal crashes in the United States (AAA Foundation for Traffic Safety, 2009). Therefore, vehicle moving violations should be considered when studying expressway safety.

However, not much has been done in studying expressway safety with consideration of vehicle moving violations. While capturing moving violations of vehicles and investigating their impact on traffic in real traffic is difficult and dangerous for observers, traffic simulation can be an effective approach to create virtual traffic scenarios which mimics reality with considering vehicle moving violations. Habtemichael and Santos (2014) used the VISSIM simulation software and the SSAM model (Chen, 2009) to simulate moving violations and investigate their impact on expressway safety. The study found that violations lead to crashes with increased severity. Wang et al. (2017) used Aimsun microsimulation software to simulate vehicle violations and investigate their impact on crash risk on expressways. Results indicated that the impact of violations increased with traffic volume. Despite these attempts, much remains to be studied in understanding the impact of vehicle moving violations on expressway safety.

The consequence of a vehicle moving violation under varied traffic conditions changes due to the difference in factors such as the vehicle speed that is closely associated to crash severity and time left for evasive maneuvers which is closely related to crash probability. The traffic condition and vehicle moving violation are two key elements that interactively contribute to expressway crashes, especially considering that the traffic condition on the expressway are always monotonous. Finding a method that considers both of these two elements with their relationship theoretically and statistically explained contributes greatly for expressway safety evaluation and crash prediction purposes.

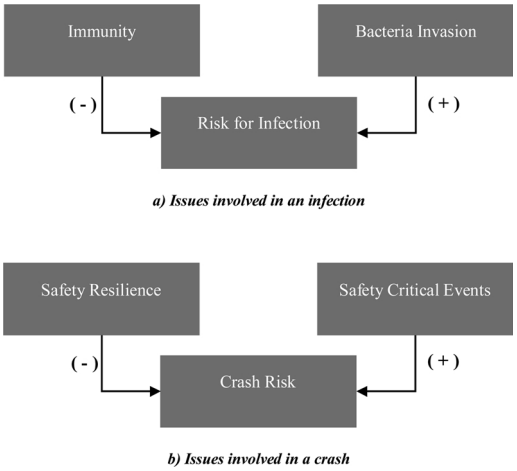


Fig. 1. Mechanism of infection and crash.

1.3. Immunity and safety resilience

Calvert and Snelder (2015) has recently published a journal on the concept of traffic resilience. In their work, traffic resilience is considered as "the ability of a road section to resist and to recover from disturbances in traffic flow". Murray-Tuite (2006) considered safety as one of the 10 main dimensions considered in traffic resilience. To illustrate the role of safety resilience in transportation system, the study refers to immunity in medical science. In medical science, as given in Fig. 1a, *the invasion of bacteria or virus*[1] and the *immunity* are the two key elements for having an infection. *Immunity,* as defined in the Oxford Dictionaries, is "the ability of an organism to resist a particular infection or toxin …" (English Oxford Living Dictionaries, 2018). *Risk for infection*, defined in the Stedman's Medical Dictionary and accepted in the North American Nursing Diagnosis Association, is "a state in which an individual is at increased risk for being invaded by pathogenic organisms" (Stedman's Medical Dictionary, 2006). The improved immunity of an organism reduces the risk for infection. According to the definition given to resilience of traffic by Calvert and Snelder (2015), safety resilience of traffic can be defined as:

the ability of a road section to resist from safety disturbances, or in other words safety critical events, such as vehicle violations, driver maneuver and judgement errors, or any other disturbances in traffic which could result in a crash.

Traffic safety, presented in Fig. 1b, has a similar mechanism: safety resilience is the "immunity" of the traffic system, safety critical events are the "bacteria", and the outcome of a crash is the "infection". *Crash risk* can be defined as the state in which the traffic is at increased risk for having a crash. Safety critical events induce higher crash risk while they do not necessarily result in crashes because the traffic has its own ability, its safety resilience, to prevent crashes. Using traffic resilience, the relationship between traffic, safety critical events and expressway safety could be explained.

However, in building the theory of safety resilience, more conceptual, methodological and mathematical explorations need to be conducted. The mechanism between the ability of traffic in safety resilience, and the outcome of crashes and variables contributing to crashes needs further and comprehensive investigation. Quantifying the ability of safety resilience needs deeply exploration.

[1] It can be the invasion of bacteria or virus, the existence of bacteria or virus in the organism, or mutations in terms of cancer. The study just deals about the invasion of bacteria or virus to make it less confusing to illustrate.

This study aims to introduce the safety resilience theory to learn traffic safety on expressways. For the illustration purpose, traffic environments are simulated using Aimsun microsimulation software based on the work in (Wang et al., 2017) which allows adding vehicle moving violations (as the safety critical event investigated) in the simulation. This allows collecting data in sufficiently large quantity. To model crash risk in a precise way, the back propagation neural network machine learning method is applied in modeling crash risk status. The model is two staged with the first stage identifying risk and no-risk status, and the second stage identifying high-risk and low-risk status. The back propagation neural network method was trained using the traffic and risk information from a simulation experiment for each of the modeling stage. To validate the performance of the model in crash risk prediction, global accuracy, and the ROC curve metrics are tested. The concept of traffic resilience is applied in explaining why increased probability of crash (represented by averagely higher risk status) is detected under higher traffic volume with a same violation input.

2. Methodology

The methodology of the study consists of four steps: 1) definition of safety resilience of traffic; 2) the Aimsun Microsimulation software, which is introduced to simulate traffic and vehicle moving violations on the expressway; 3) modeling crash risk statues based on the traffic conflict technique (TCT) and back propagation neural network (BPNN).

2.1. Aimsun microsimulation software

As being hard to find an efficient and accurate way to obtain traffic information from entire road sections, traffic simulation was conducted. The Aimsun software was used to simulate traffic on expressway in this study (TSS, 1997-2014TSS, 1997-2014). Aimsun is widely used software for traffic simulation (Wang et al., 2017). The software makes use of the Gipps' vehicle-following and lane changing models which helps precisely mimic the behavior of vehicles in traffic (Chong et al., 2011) (Pakha et al., 2009) (Barceló, 2014). The Application Programming Interface (API) and the Aimsun Software Development Kit (SDK) modules of Aimsun provide users a powerful and flexible simulation platform. The API module offers users the option of improving the default simulation settings by loading user-defined and third party applications (TSS, 2011) (Vilarinho et al., 2014). Meanwhile, the SDK module enables users to apply user-defined plugins to override the default road user behavior models (TSS, 2014).

Traffic with vehicle violations was simulated based on the methodology proposed previously by the same authors of this work (Wang et al., 2017). Vehicle violation plugins, which is developed to assign different violation models to selected vehicles, were developed using the C++ programing platform, and loaded by the SDK module to the simulation. Fig. 2 presents how the SDK module loads external behavior model plugins (vehicle violation plugins). In each time step in the simulation, the program first selects vehicles for the behavior models. If external applications of behavior models are to be loaded, the SDK module will load the external applications and apply it to the behavior of the selected vehicles (otherwise, default internal models are used). The program checks if all the selected vehicles have been updated with the behavior model before moving on to the next time step. With these plugins vehicle violation behaviors can be assigned to random vehicles with an adjustable sample ratio or to targeted vehicles. More details including the explanations, the procedures and the algorithms for the vehicle violation simulation are provided in (Wang et al., 2017). According to (Wang et al., 2017), the safety impact of slow driving violations, referred to as violations made by drivers that they are driving below the minimum speed limit in a non-congested traffic (Kerner, 2009) and impeding traffic, is the most significant among the three main types of vehicle violations. The slow driving violation with vehicles assigned a speed equal to 50% of the maximum speed limit (100 km/h) was chosen as the vehicle violation type investigated in the study for illustration purposes. Risk of rear-end collisions, which are the most common type of expressway crashes (Khattak et al., 1998) and also the type of crash that slow driving violations are mostly likely related to (Wang et al., 2017), were explored.

2.2. Simulation calibration

The vehicle-following model in Aimsun is important to describe the rear-end collision course. The vehicle-following model considers two traits of the following vehicle: 1) intention to drive at an expected speed, and 2) actual speed of the following vehicle. Gipps' model in Aimsun considers human factors of driver expectation and reaction and accurately describes vehicle-following behavior in non-congested situations (Rakha and Wang, 2009). Meanwhile, it can be calibrated for non-steady-state conditions related to aggressive driving or violations (Rakha and Wang, 2009). To better describe the rear-end collision course, the vehicle-following model for this study needs to be calibrated based on data collected from the selected road section. The calibration involved several main parameters in the Gipps' vehicle-following model which include: 1) the speed, and the maximum desired deceleration rate (the maximum desired deceleration means most severe braking the driver wishes to take) of the leader; 2) the speed, desired speed, reaction time, and maximum desired deceleration of the follower; and 3) the distance between the follower and the leader, and follower's estimate of the leader's maximum deceleration (Vasconcelos et al., 2014). The model was calibrated for two different conditions: 1) regular vehicle-following conditions; 2) hard-braking conditions to avoid a crash. Details of the calibration of the vehicle-following model are included in the next section. Besides, as lane changing behavior is not investigated in this study, defaulted Gipps' lane-changing model in Aimsun is applied.

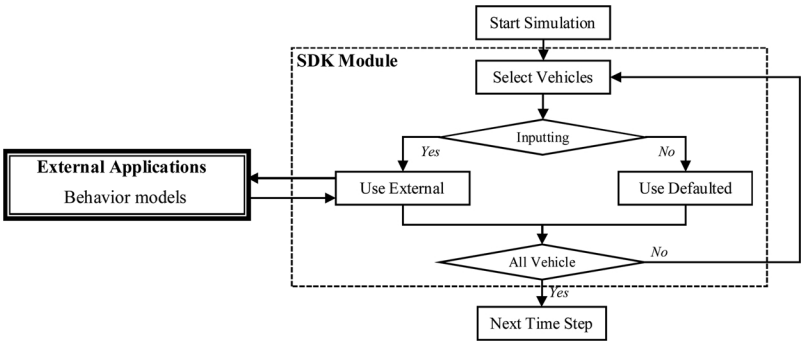


Fig. 2. Flowchart for the violation plugin (SDK module) (Wang et al., 2017).

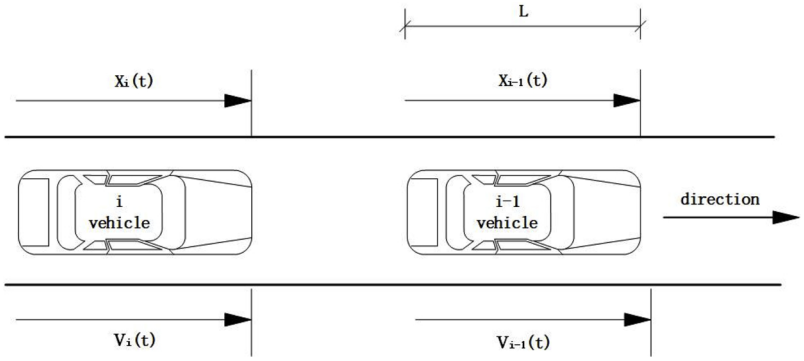


Fig. 3. Principle of TTC.

2.3. Modeling crash risk based on traffic conflict technique and back propagation neural network

The traffic conflict technology has been used in different studies in crash risk prediction. This section proposes a new crash risk prediction model based on TCT and BPNN. The crash risk status which relies on the observation of traffic conflicts is introduced as the measure of crash risk on road sections. The BPNN machine learning method was applied to model and determine crash risk status on different road sections. The performance of the model was validated.

2.3.1. Traffic conflict and crash risk status

In TCT, time-to-collision (TTC) has been used as an important indicator in crash risk analysis. For rear-end collisions (Fig. 3), once the speed of the following vehicle exceeds the leading vehicle, there generates the possibility of a crash if they maintain their current speeds. Then, TTC can be calculated by Eq. (2):

$$TTC_i = \frac{x_{i-1}(t) - x_i(t) - l_{i-1}}{v_i(t) - v_{i-1}(t)} \tag{2}$$

where, TTC_i is the TTC value for the *i*-th vehicle that could have a rear-end collision with the leading vehicle, *i-1*, $x_{i-1}(t)$ and $x_i(t)$ are the locations of the vehicles *i* and *i-1* at time *t*, $v_i(t)$ and $v_{i-1}(t)$ are the speeds of the vehicles, $v_i(t) > v_{i-1}(t)$, and l_{i-1} is the length of the leading vehicle.

TTC thresholds have been used to determine traffic conflicts. In the literature, suggestions of TTC the threshold range from 1.5 s to 5 s (Vogel, 2003). For instance, Hydén (1987) and Green (2000) set the critical value of 1.5 s to identify severe traffic conflicts. Van der Horst (1991) used a TTC threshold of 4 s to distinguish safety and unsafety situations on the road. Li et al. (2016) suggested a TTC threshold of 2.3 s to determine conflicts related to rear-end collisions, and a threshold of 2.8 s to determine severe conflicts related to side-impact collisions. Since the perception-reaction procedure of the drivers is crucial for the outcome of a road crash, the perception-reaction time threshold of 2.5 s from the AASHTO (2015) was used as the TTC threshold in the study. In other words, traffic conflicts are determined as situations where the following vehicle has a chance of colliding with the leading vehicle with a TTC of less than 2.5 s.

2.3.2. Crash risk status

For a given section of traffic or road, the risk can be presented based on number of conflicts on this section during a certain period of time. This study classifies crash risk status (Eq. (3)) of the road section studied into *no-risk, low-risk and high-risk* according to numbers of conflicts on this section:

$$Crash\ Risk\ Status = \{no\ risk,\ low\ risk,\ high\ risk\} \tag{3}$$

According to the crash risk status, safety situations on expressways can be analyzed.

The crash risk status of zero risk is easily determined as situations with no conflicts observed. Classifying low and high-risk is conducted through the k-means clustering method using Matlab. The k-means clustering method is an unsupervised machine learning algorithm that classifies the targeted data (here, observed numbers of observed conflicts) into a given number (k, k = 2 in this study) of groups according to similarities in their features (Li et al., 2004). Based on the threshold determined from the clustering results, the crash risk status for road section can be defined as:

$$Crash\ Risk\ Status = \begin{cases} \text{if no. of observed conflicts} = 0: & \textbf{No Risk} \\ \text{if no. of observed conflicts} \neq 0: & \begin{cases} \text{if no. of observed conflicts} < n & o_{threashold}: \text{Low Risk} \\ \text{if no. of observed conflicts} > n & o_{threashold}: \text{High Risk} \end{cases} \end{cases} \tag{4}$$

where, *no. of observed conflicts* is the number of conflicts observed on the road section, $no_{threashold}$ is the threshold generated from the k-means clustering algorithm for determining the high-risk status and the low-risk status.

2.3.3. Crash risk modeling and prediction

After crash risk status information being generated from the conflict information of the road section studied, the crash risk can be modeled. The study used the BPNN to model the crash risk.

2.3.3.1. Description of the back propagation neural network algorithm. The BPNN algorithm is applied in building the crash prediction model. BPNN is a multi-layer feed-forward network which belongs to machine learning modeling approach. It is trained based on error back propagation algorithm (Oh, 2011). The structure of the model consists of three components including the input layer, the output layer and a number of hidden layers with each of whom composed of countless nodes that are interleaved and connected.

For solving engineering problems, a three-layer (one-hidden layer) BPNN model can be sufficient in achieving satisfactory results. This study relies on the three-layer multiple-layer feedforward perception structure of BPNN algorithm for crash risk modeling. The structure of the three-layer BPNN model is presented in Fig. 4.

2.3.3.2. Imbalanced data and SMOTE technique. In the field of machine learning, imbalanced data leads to significant biases, which result in low classification accuracy for the minor class. Therefore, having balanced data is important in the data preparation for training purpose. To deal with the imbalanced data problem, this paper uses the Synthetic Minority Over-sampling Technique (SMOTE), which is an over-sampling method to deal with the imbalanced data (Chawla et al., 2002).

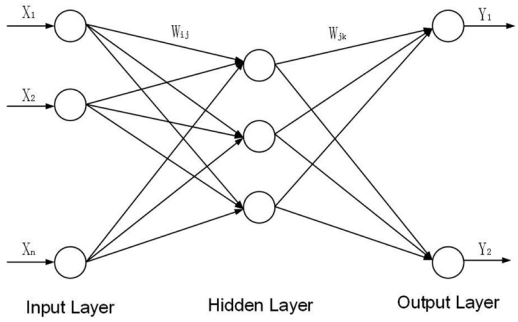


Fig. 4. Structure of the three-layer BPNN.

To solve the problem of imbalanced data, the SMOTE method works by "creating synthetic samples from the minor class" (Chawla et al., 2002). The basic assumption of the SMOTE method is that samples of the same type are close to each other in the pattern space. Based on the assumption, synthetic samples are created in the area where the minor class, $X = \{X_i\}$, $(i = 1,2, ...,k$, where k is the number of samples in the minor class), falls in the pattern space. For a sample in the minor class, X_i, there are similar instances around this sample, $Y = \{Y_{ij}\}$, $(j = 1,2, ...,m$, where m is the number of selected instances). For interpolating the synthetic samples, the method using distance measure (the "distance" in the pattern space) is applied, presented in the number line (one-dimensional space) as:

$$P_{ij} = x_i + rand(0,1) \times (y_{ij} - x_i) \quad (5)$$

where, P_{ij} is the position of the synthetic sample, x_i is the position of the sample, X_i, y_{ij} is the position of the instance close to the sample, Y_{ij}, $rand(0,1)$ is a number randomly picked from 0 to 1.

In this study, the procedure of data balancing is conducted using the Matlab SMOTE algorithm.

2.3.3.3. Modeling crash risk. Considering the three different status level determined in two different steps, the modeling approach is two staged: *1)* stage one – classifying the no-risk status and the risk status; *2)* stage two – classifying the risk status into the low-risk status and high-risk status. The BPNN model is applied in each stage. Traffic information including traffic volume, vehicle average speed, and average occupancy rate are the inputs, while classification results are the outputs.

In this study, to model the crash risk status, the BPNN algorithm was trained and tested based on the traffic and the crash risk status data from the traffic simulation for both of the two stages. The Matlab Neural Network Toolbox was used in modeling. For each stage, records were randomly divided into the training set (70%), the validation set (15%) and the test set (15%). The BPNN algorithm was trained based on the training and validation sets using the holdout validation method, as illustrated in Fig. 5, for each of the modeling stage. The model which had the best prediction performance on the validation set was selected as the trained model. Then the performance of the model was tested on the test set.

2.3.4. Model validation

For model validation, the ROC curve test is used to investigate the model prediction performance and derive metrics. In the study, the performance of the trained model for the two stages was validated on the test set. Some existing studies in modeling crash risk using machine learning techniques have validated performance of their models based on the overall dataset (Abdel-Aty et al., 2004) (Ahmed et al., 2012) (Pande and Abdel-Aty, 2006). For comparison purpose, the performance of the model on the overall dataset in each modeling stage was also validated.

The confusion matrix is introduced. In both the two modeling stages, only two classes are included (stage 1, no-risk and risk; stage 2, high-risk and low-risk). In other words, they are considered as binary classification problems. Table 1 presents the confusion matrix in binary case.

The commonly applied metric is the accuracy defined as the proportion of correct predictions which is presented in Eq. (4).

$$\text{Accuracy} = \frac{TP + TN}{P + N} \times 100\% \quad (6)$$

However, the accuracy is not sufficient to report the performance for two-class classification (binary case). The performance of the model is also evaluated using the Receiver Operation Characteristic (ROC) curve (Xu et al., 2015). ROC curve relies on the true positive rate (TPR) and the false positive rate (FPR) in the confusion matrix. Area Under Curve (AUC) of the ROC curve, which has been often used in the machine learning field for model validation (You et al., 2017), is used to present the performance of the classifier (the trained model). In performance validation, an increased AUC value indicates better prediction accuracy.

In the study: *1)* In stage 1 - risk status is the positive, while no-risk is the negative; situations with risk status being detected as no-risk status are false negatives, while situations with no-risk status being detected as risk status are false positives. *2)* In stage 2 - high-risk status is the positive, while low-risk is the negative; situations with high-risk status being detected as low-risk status are false negatives, while situations with low-risk status being detected as high-risk status are false positives. Results for both the two stages are included.

3. Simulation experiment and data preparation

3.1. Simulation experiment

3.1.1. Simulation scenario

To illustrate the methodology in crash risk modeling and the concept of traffic resilience, a simulation experiment was conducted by Aimsun. The simulation scenario, as shown in Fig. 6, was based on the geometric design of a section of the G15 Shenhai Expressway in Shanghai, China. The simulated section is 5 miles in length with two lanes in each direction. The simulation only investigates traffic travelling from south to north. The simulation experiment excluded any on- or off-ramps to avoid the impact of merging vehicles on traffic. This expressway section has a design speed of 120 km/h with a posted speed limit of 100 km/h. 10 virtual loops were placed on the section at intervals of 0.5 miles for collecting traffic information in the simulation experiment. In the experiment, three different traffic conditions were included: the low volume condition (2000 veh/h), the moderate volume condition (2500 veh/h), and the high volume condition (3000 veh/h). The rates for different types of vehicles in the simulation were set according to the real-road traffic data collected from the G15 Shenhai Expressway. To illustrate safety resilience of traffic, vehicles conducting slow driving violations were simulated and distributed in the traffic using the violation plugins proposed in (Wang et al., 2017). For details of the simulation progresses and the plugins for simulating violating vehicles, one can referred to (Wang et al., 2017).

3.1.2. Calibration of vehicle-following model

3.1.2.1. Regular vehicle-following conditions. The Gipps' vehicle-following model was calibrated in the simulation for the regular-vehicle following conditions. Table 2 presents the variables, provided in Section 2.2 and their units and definitions included in the Gipps' vehicle-following model. The model was calibrated using data collected from the selected road section. To collect the data for calibration purpose, five participants were employed to drive two equipped vehicles along the selected road section and asked to follow each other. The vehicles were equipped with contactless fifth wheel sensors (FMCSA, 2003), as shown in Fig. 7, to record their speeds during the data collection. Participants took turns for the conducting the trips,

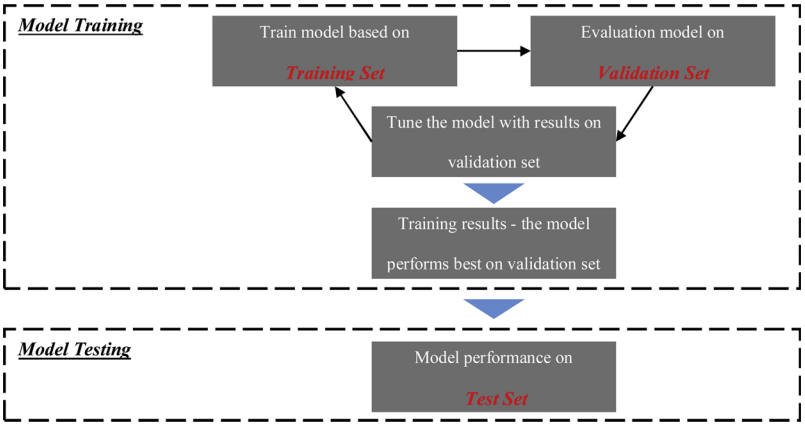


Fig. 5. Model training with holdout validation.

Table 1
Confusion Matrix and Basic Terms.

		Predicted Class		
		Positive	Negative	
True Class	Positive	True Positives (TP)	False Negatives (FN)	
	Negative	False Positives (FP)	True Negatives (TN)	**False Positive Rate (FPR)** $= \frac{FP}{FP+TN}$
		Positive Predictive Value (PPV) $= \frac{TP}{TP+FP}$		

Table 2
Main Variables in the Gipps' Vehicle-following Model.

Variables	Unit	Definition
vl	m/s	initial speed of the leader
Vf	m/s	initial speed of the follower
vf_d	m/s	desired speed of the follower
vl_e *	m/s	follower's estimation of the leader's speed
L	m	distance between the two vehicles
r_tr *	s	reaction time of the follower
r_al *	m/s^2	maximum desired deceleration of the leader
r_af *	m/s^2	maximum desired deceleration of the follower

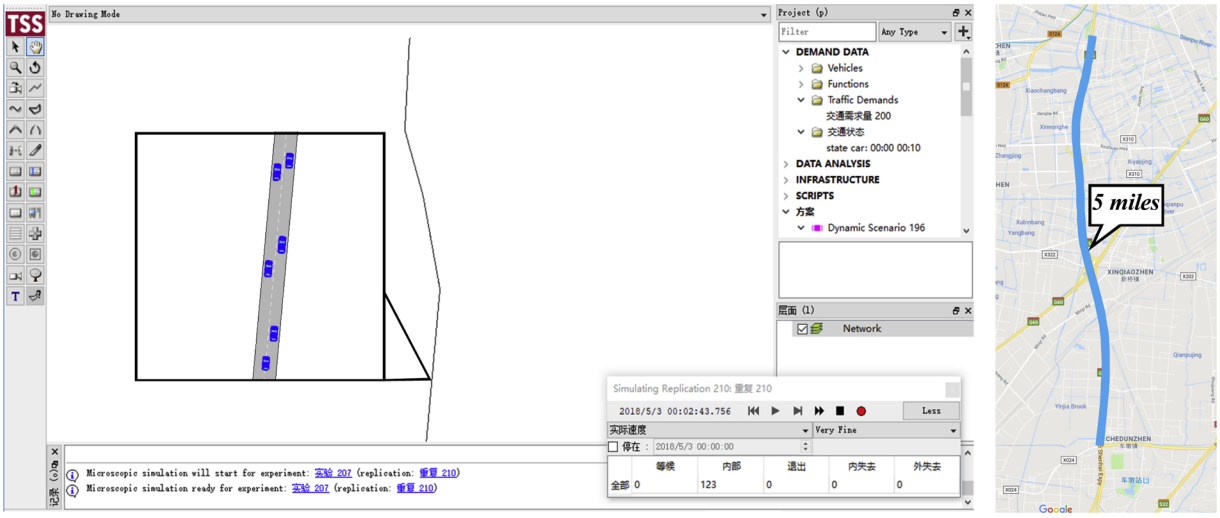
Note: * represents that the market variables need to be calibrated; other variables are automatically generated during the simulation.

totalling to 10 trips (two participants each trip with one in the leading vehicle and the other in the following vehicle, non-repeatedly). Each of the participants drove twice – the first time in the leading vehicle and the second time in the following vehicle.

In each trip, the driver in the leading vehicle was asked to drive at a regular speed of 70–110 km/h. The actual speeds of the leading vehicle (*vl*) and the following vehicle (*vf*) were recorded using the fifth wheel sensor installed on them. The distance between the two vehicles (*L*) was measured using a LiDAR gun and recorded manually. Table 3 shows a sample of the collected speed and distance data.

To get the leader's speed that the follower estimated (*vl_e*), the driver in the following vehicle was asked to speak out his estimation of the leading vehicle by referring to the speedometer every 15 s, which was designed to allow an observer sitting in the leading vehicle to record the actual (*vl*) and estimated speeds (*vl_e*) of the leading vehicle at the same time.

In the uncalibrated Gipps' vehicle-following model, the speed of the leading vehicle estimated by the follower (*vl_e*) is defaulted to be the same as the actual speed of the leading vehicle (*vl*). Crash risks will be



a) Snapshot from the simulation scenario *b) Map of the site*

Fig. 6. Aimsun interface of the simulated expressway section.

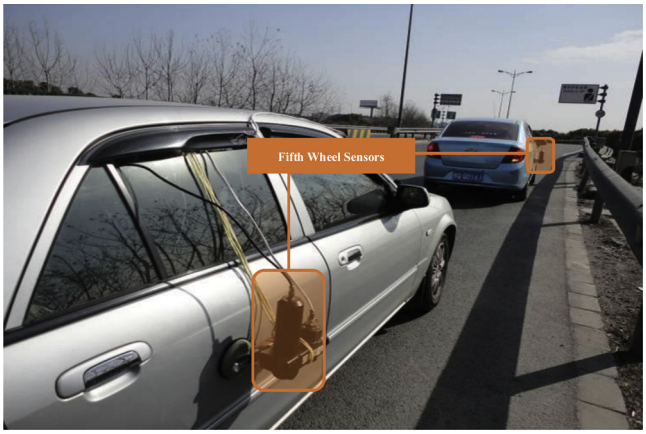


Fig. 7. Test vehicles equipped with fifth wheel sensors.

lower compared to real traffic conditions because the follower can always react properly to the leading vehicle based on the uncalibrated model. However, errors exist for the following driver to estimate the speed of the leading vehicle (*vl_e*). Such errors in estimation bring increased crash risks. The study took consideration of the existence of the follower's estimation errors in speeds of the leaders by investigating the distribution of estimation errors using data collected from real traffic conditions. Table 4 shows a small sample of the speed information and errors between the actual and estimated speeds of the leading vehicles. Based on the data, the overall relative error is 8.7%. The distribution of the relative estimated speed (the estimated speed divided by the actual speed, $\frac{vl\_e}{vl}$) approximates a normal distribution, $\frac{vl\_e}{vl} \tilde{} N(0{,}8.7\%)$. This normal distribution was applied in the calibrated model to predict the estimated speed of the leading vehicle by the driver in the following vehicle (*vl_e*), in predicting the speed of the following vehicle, given the actual speed of the leading vehicle. For more calibration details, one can refer to (Wang et al., 2010)

Other parameters were also calibrated. As information of driver reaction time is hard to collect, driver reaction time of the following vehicle (*r_tr*) in regular vehicle-following conditions was assigned a normal distribution, $r\_tr \tilde{} N$ (1.1,0.16), according to a Chinese study (Xu and Tian, 2007). For regular vehicle-following conditions, maximum desirable decelerations of both the leading vehicle (*r_al*) and the following vehicle (*r_af*) in the simulation were set to be 2.5 m/s^2, the comfort deceleration above which the deceleration maneuver is considered as non-comfort (Liu et al., 2018).

3.1.2.2. Hard-braking conditions. Traffic conflicts happen when some safety critical events (e.g. vehicle moving violations) happen in a car following situation. Avoiding a collision in the traffic conflict context always requires the driver of the following vehicle to decelerate at the maximum possible rate. Such situations can be defined as hard-braking conditions. In this simulation experiment, hard-braking conditions happen when other vehicles come close to, and get involved in a vehicle-following situation with vehicles conducting a slow-driving violation.

Table 4
Sample of Errors in Follower's Estimation of the Leader's Speed.

vl (m/s)	vl_e (m/s)	Error (m/s)	Relative Error (%)
87.1	80	7.1	8.9
84.8	85	−0.2	0.2
97.1	90	7.1	7.9
77.6	85	−7.4	8.7
79.0	80	−1.0	1,3
84.8	85	−0.2	0.2
70.6	80	−9.4	11.8
91..4	80	11.4	14.3
99.1	85	14.1	16.6
102.3	100	2.3	2.3

In hard-braking conditions, neither the Aimsun defaulted vehicle-following model, nor the calibrated model for regular vehicle-following conditions, works properly. Different calibrations were conducted. A normal distribution was used to describe the deceleration of the following vehicle (*a_l*). The deceleration of following vehicle can be described as $a\_l \tilde{} N$ (5.2,1), with a mean of 5.2 m/s^2 and a variance of 1 m/s^2 according to (Carbaugh et al., 1998). Besides, driver reaction time of the following vehicle (*r_tr*) was assigned a log-normal distribution of *LN* (0.17, 0.44), suggested by (Wang, 2002).

3.1.3. Simulation test

The three traffic conditions with slow driving violations were simulated. Each situation was simulated 30 times, of which it takes 10 min apiece. Each 10-minute test was assigned with one violating vehicle which enters from the south of the section in the beginning of the test (the 10 min length was set to allow the violating vehicle to cross the whole section). Considering the road section was empty when starting the simulation, there remained some warm-up time in the beginning of each test but was excluded from the 10 min to guarantee the quality of the study. Within each test, traffic information and number of conflicts on the entire road section during each 30 s were gathered from the recorded virtual loop and vehicle trajectory data. This results in 20 observations for each of the 30 tests, for the 3 traffic conditions, total of 1800 observations.

3.1.4. Data collection

Raw data were collected from the simulation including the traffic information and vehicle trajectory data. This section gives detailed descriptions about the data extracted from the simulation experiment, including both traffic data and vehicle trajectory data.

3.1.4.1. Traffic information. Traffic data were collected from the 10 virtual loop detectors. From the loop data, traffic information of the road section at each 30-second interval was collected. Traffic information includes the location (the number of the detector), time, and average traffic volume, average speed and average occupancy rate during each 30-second interval. Table 5 shows a sample of the traffic information.

3.1.4.2. Vehicle trajectory information. Traffic conflict analysis is

Table 3
Sample of Collected Data in Vehicle-following Conditions.

vl (m/s)	vf (m/s)	L (m)	vl (m/s)	vf (m/s)	L (m)	vl (m/s)	vf (m/s)	L (m)
75.6	72.3	42.8	82.8	80.8	48.3	85.6	89.5	45.5
76.9	74.5	43.7	83.7	81.0	49.0	85.9	88.7	44.5
77.8	75.0	44.3	83.7	83.4	49.3	85.6	89.2	43.7
77.1	77.0	44.8	85.3	84.5	49.3	85.1	86.6	43.0
79.2	76.6	45.2	85.7	86.7	48.9	85.3	86.5	42.6
80.5	77.4	45.7	86.8	87.5	48.3	85.5	84.8	42.6
…	…	…	…	…	…	…	…	…

Table 5
Sample of Traffic Information.

Data	Average Speed(km/h)			Volume(/veh)			Average Occupancy Rate (%)		
Time	*1*	…	*10*	*1*	…	*10*	*1*	…	*10*
1st 30 sec	107.552	…	98.629	20	…	12	3.449	…	3.705
2nd 30 sec	104.384	…	107.990	13	…	16	5.927	…	4.752
…	…	…	…		…	…	…	…	…
20th 30 sec	97.677	…	106.356	15	…	17	3.780	…	6.558

Table 6
Sample of Vehicle Information.

Time (/0.1 sec)	Vehicle ID	Position (m)	Speed (km/h)	Lane	Time of Entrance (/sec)
…	…	…	…	…	…
3952	392	861.042	105.137	1	443.817
3952	393	752.267	101.485	1	446.615
…	…	…	…	…	…

Table 7
Descriptive Statistics of Conflict Number Observations.

No. of Obs.	Median	Std. Dev.	Max.	Min.	No. of conf. = 0	No. of conf. > 0
1800	4	3.4	18	0	1543	257

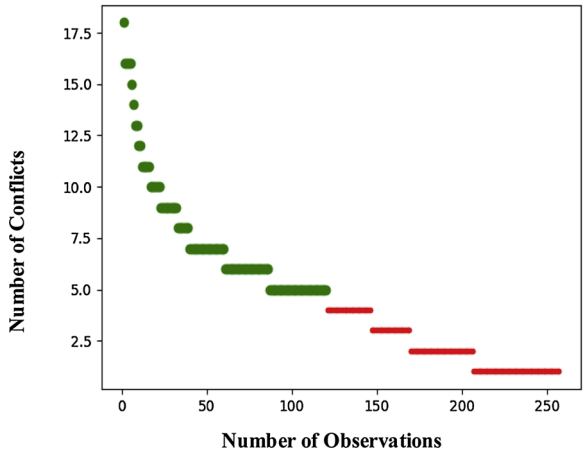


Fig. 8. Result from k-means clustering.

Table 8
Descriptive Statistics of Crash Risk Status.

No. of Obs.	No. of No-risk Status	No. of Low-risk Status	No. of High-risk Status
1800	1543	137	120

conducted based on the position and speed information of the vehicles; therefore, vehicle trajectory information is required. For ensuring the quality of the trajectory information, the simulation step in the experiment was set to be 0.1 s, which is the minimum in the Aimsun program. Information of the vehicles on the entire road section, including the time (the number of the step), vehicle ID, position, speed, number of the lane and the time it enters the road section, was recorded for each step. A sample of vehicle information was provided in Table 6.

Table 9
Mann-Whitney U test results for SMOTE samples.

Variable	Significance - p value	Mann-Whitney U test* Result – 0, pass; 1, fail
V1	0.330	0
V2	0.589	0
V3	0.874	0
V4	0.771	0
V5	0.701	0
V6	0.941	0
V7	0.901	0
V8	0.855	0
V9	0.745	0
V10	0.586	0
V11	0.706	0
V12	0.746	0
V13	0.680	0
V14	0.782	0
V15	0.926	0
V16	0.739	0
V17	0.793	0
V18	0.865	0
V19	0.798	0
V20	0.912	0
V21	0.966	0
V22	0.767	0
V23	0.723	0
V24	0.700	0
V25	0.870	0
V26	0.710	0
V27	0.786	0
V28	0.989	0
V29	0.807	0
V30	0.970	0
V31	0.977	0

3.2. Data preparation

3.2.1. Conflict information and risk status clustering

Based on the trajectory data, TTC values between vehicles in traffic were calculated and numbers of conflicts occurred on the road section during every 30-sec period were extracted. Descriptive statistics of the observations were provided in Table 7. In total, 1800 observations were obtained. Among them, 1543 had no conflict, while the rest 257 were with conflicts detected.

Based on the number of conflicts, the k-Means clustering method was applied to define the threshold for determining the high and low risks statuses. The result from clustering is presented in Fig. 8. From the result, situations with 5 or more conflicts observed were determined as those with high-risk status, while situations with conflicts observed but the number of conflicts less than 5 were those with low-risk status. Table 8 gives the descriptive statistics of the risk status observations.

3.2.2. Data balancing

After risk status data were obtained for analysis purpose, balanced data are required in the machine learning process. In the first modeling stage in this study, when using the BPNN algorithm to determine risk and no-risk status, the two sample classes were not balanced with an approximate ratio of 1:6 (risk: no-risk), as from Table 8. Therefore, data should be balanced for the first stage. Checking it for the modeling in stage 2, the ratio between the sample sizes of high-risk status and low-risk status is about 1:1, indicating quite well-balanced data.

Data balancing was conducted for the first modeling stage using the SMOTE sampling method. After data balancing, the sample size of risk status was increased to 1542 which brought and the ratio between the sample sizes of risk status and no-risk status to be 1:1. The total number of the sample increased to 3085. The Mann-Whitney U nonparametric test was conducted to validate the results from the oversampling (You et al., 2017). Results from the Mann-Whitney U nonparametric test are given in Table 9. Data has 31 variables, which include the risk status

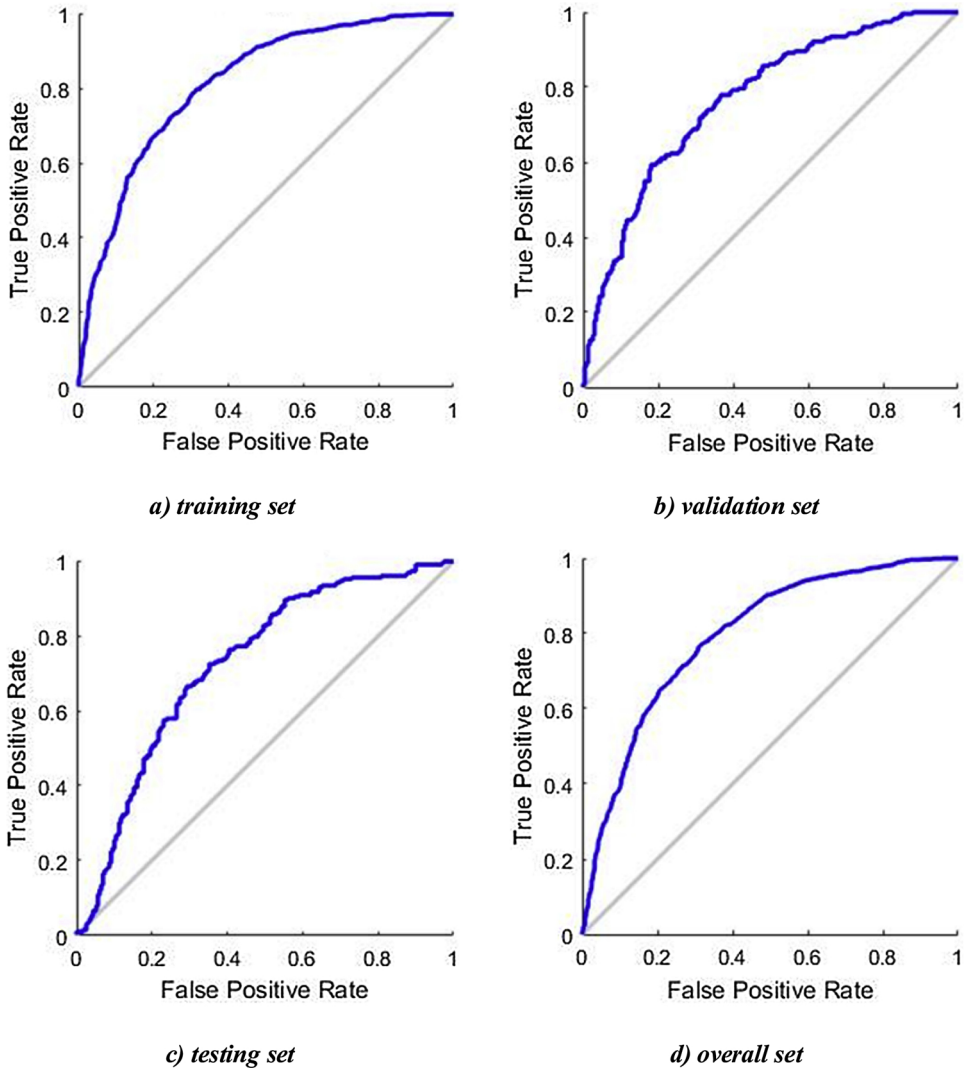


Fig. 9. ROC curve of risk and no-risk model.

variable (V1 in the table) and the 30 traffic information variables (average speed, volume, and average occupancy rate collected from the 10 virtual loop detectors, V2 – V31 in the table). Using the Mann-Whitney U test, the significance of the difference for each variable is compared. Results show the p values for all variables are greater than 0.05, indicating a good oversampling result.

4. Modeling results

4.1. Modeling performance

Based on the 3085 records including the risk status information, as the dependant variable (output layer), and the traffic information (as the input layer), the BPNN model was trained for the two modeling stages. In stage one, 3085 records were randomly divided into a training set with the size of 2159, a validation set with the size of 463, and a test set with the size of 463. In stage two, the 257 records with risk status were divided into a training set with the size of 180, a validation set with the size of 39, and a test set with the size of 38. Models for the two stages were trained based on the training and validation sets. The performance of the trained models was validated on both the test sets, and the overall datasets.

Figs. 9 and 10 present the ROC curve for the two modeling stages. Checking from the figures, the performance of the models in both the two stages stays constantly reliable on the training, validation and test sets. The metrics from ROC curve analysis for the model performance on the test sets are given in Table 10. The trained models performed well in predicting crash risk status for both two stages. For the model in identifying risk and no-risk status, TPR, AUC and accuracy reaches 70.79%, 0.75 and 67.79% respectively. FPR value is 35.22% which is reasonable. For the model in identifying high-risk and low-risk status, TPR, AUC and accuracy values are 82.33%, 0.93, and 86.62%, and FPR is 11.18%, showing a good modeling performance.

Table 11 presents the metrics from the ROC curve analysis the overall datasets for both the two stages. The TPR values for both of the two models (75.51% for modeling risk and no-risk status, and 83.02% for modeling high-risk and low-risk status) are either similar or higher than what was given in the previous studies on crash prediction modeling (Abdel-Aty et al., 2004) (Ahmed et al., 2012) (Pande and Abdel-Aty, 2006). Meanwhile, the FPR values, which are 30.71% for the model in identifying risk and no-risk status and 7.61% for the model in identifying high-risk and low-risk status, are remarkably lower than results from previous studies (Abdel-Aty et al., 2004) (Ahmed et al., 2012) (Pande and Abdel-Aty, 2006). AUC values from the ROC curve test reach 0.78 for modeling risk and no-risk status, and 0.95 for modeling high-risk and low-risk status, also indicating outstanding performances of the models compared with other existing literature (Yu and Abdel-Aty, 2013). From validation results, the two staged modeling method and the models generated perform reliably in predicting crash risk status.

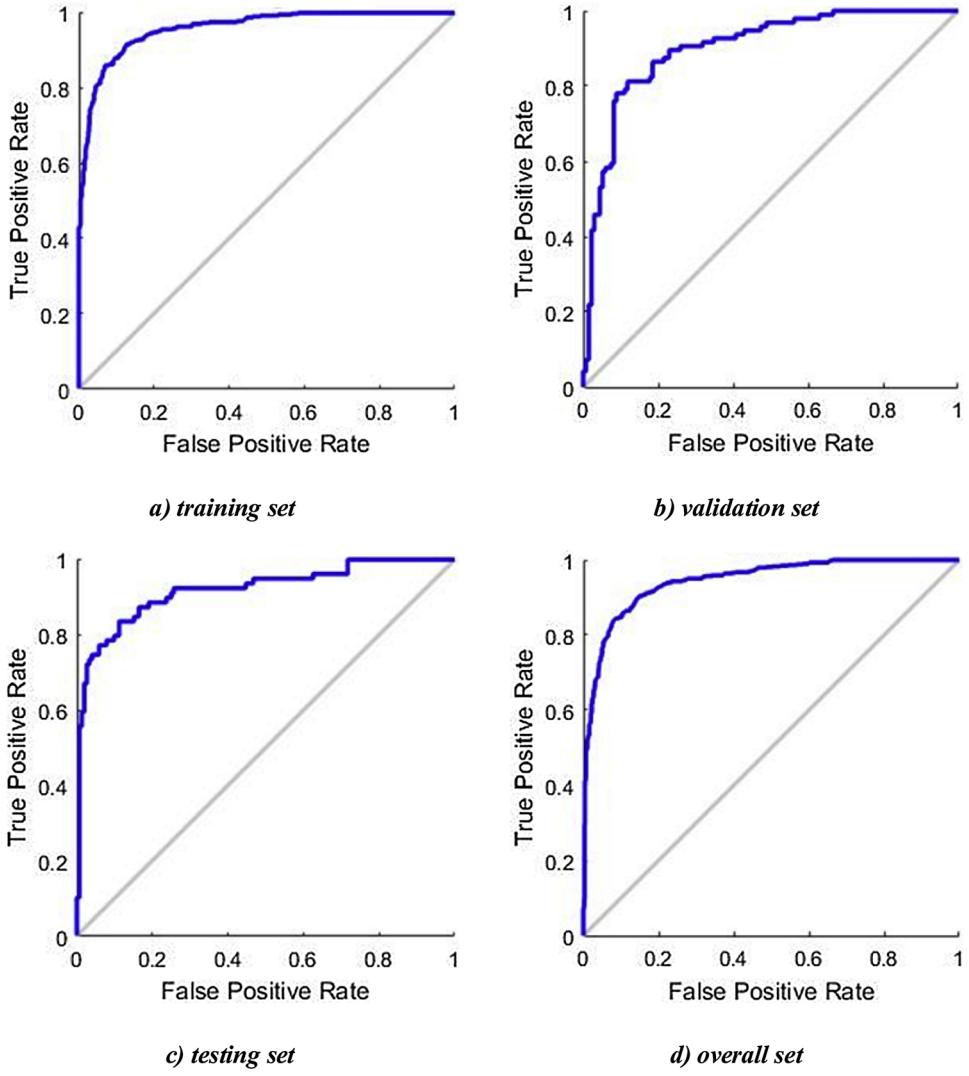


Fig. 10. ROC curve of high-risk and low-risk model.

Table 10
Validation Results on Test Sets - Metrics from ROC Curve Method.

Stage 1 – Modeling Risk and No-risk			
TPR	*FPR*	*AUC*	*Accuracy*
70.79%	35.22%	0.75	67.79%
Stage 2 – Modeling High-risk and Low-risk			
TPR	*FPR*	*AUC*	*Accuracy*
82.33%	11.18%	0.93	86.62%

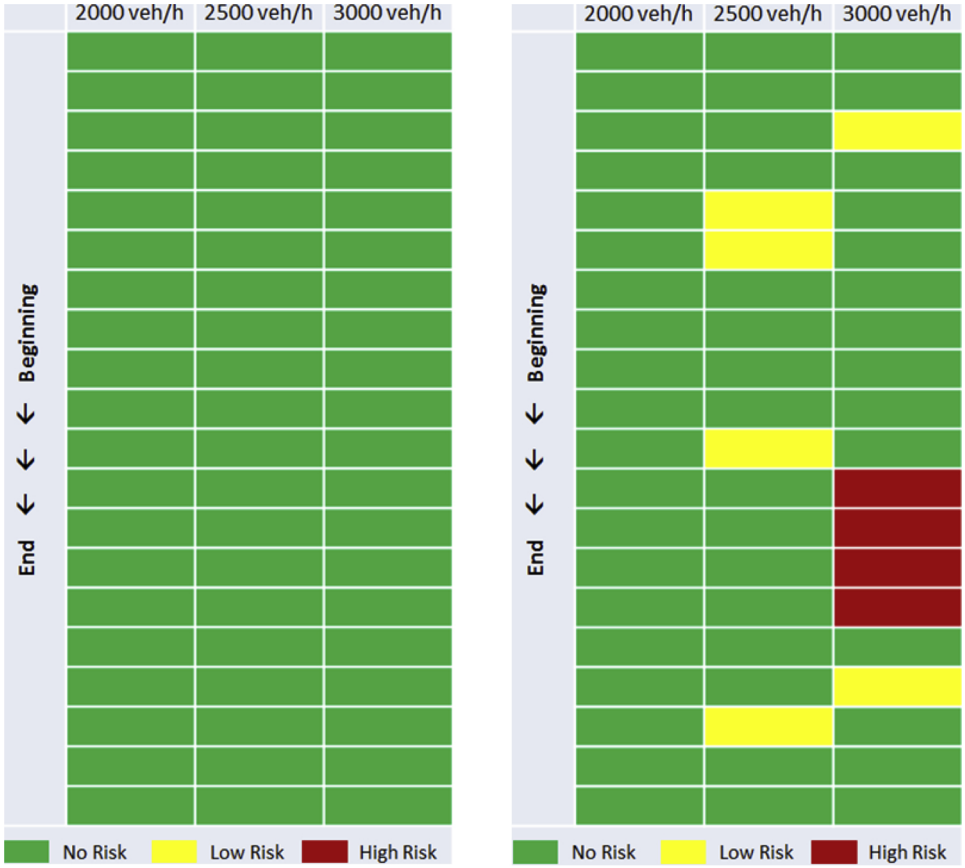
Table 11
Validation Results on Overall Datasets - Metrics from ROC Curve Method.

Stage 1 – Modeling Risk and No-risk			
TPR	*FPR*	*AUC*	*Accuracy*
75.51%	30.71%	0.78	72.43%
Stage 2 – Modeling High-risk and Low-risk			
TPR	*FPR*	*AUC*	*Accuracy*
83.02%	7.61%	0.95	88.85%

4.2. Model output and discussion on safety resilience

For the purpose of qualitatively investigation the relationship between traffic environment, crash risk and traffic resilience, an extra simulation experiment was conducted. The experiment involved three 10-minute tests for the three different traffic conditions (2000 veh/h, 2500 veh/h and 3000 veh/h) with a violating vehicle (a vehicle conducting slow driving violation) driving through the road section, and three additional 10-minute tests without the violating vehicle. Based on the proposed methodology and the models, crash risk status on the road section through the 10-minute tests can be predicted. For comparison, another set of tests was included involving three tests without violations. Results of the crash risk status for the road segment under the three different traffic conditions with and without violation can be visualized using a colored chart, as presented in Fig. 11. Here in the chart, time is set on the vertical axis from top to bottom. Each block in the vertical direction represents a 30-sec interval, which is the time interval for crash risk analysis. Scenarios with different traffic volume are assigned along the horizontal axis. In this chart, observations with no-risk status are marked in green, those with low-risk status are colored in yellow, and the ones with high-risk status are painted in red.

Checking from the chart, with the defaulted settings of the Aimsun simulation which considers no safety critical events, traffic safety remained quite stable with no-risk status observed throughout the tests.



a) without driving violation *b) with driving violation*

Fig. 11. Crash risk status under three traffic situations over a 10-min simulation test.

However, after adding violating vehicles, crash risk changes significantly overtime and varies greatly for the three traffic flow conditions. The test with high volume condition had the most critical situation with several occurrences of high-risk status, while the one with low volume condition remained constantly safe with no-risk status observed. No crash happened during the simulation. Violations bring increased risk in traffic. The safety resilience of traffic helps reduce crash risks and prevent crashes from happening at the occurrence of violations. However, because of the change in the safety resilience for different traffic and environment conditions, the impact of violations on crash risk varies among different conditions. In this case, the low traffic condition has the highest resilience resulting in lowest crash risk, while the high traffic condition has the lowest resilience resulting in highest crash risk.

5. Conclusion

This paper proposed the concept of safety resilience to learn traffic safety on expressways. For illustration, a simulation experiment was conducted to generate expressway traffic scenarios. Slow driving violations are selected as the representative safety critical events. To build a model in describing crash risk status on expressways, a methodology based on the BPNN method is presented. A two-staged modelling approach was applied: traffic conditions with risk and no-risk status were modeled in the first stage; among the conditions with risk, those with high-risk and low-risk status were identified using the model in the second stage. The models were trained based on data from a traffic simulation experiment conducted on the Aimsun simulation software. The modeling results were tested through the ROC curve method. The concept of safety resilience of traffic was proposed to explain the relationship between traffic conditions, safety critical events (of which vehicle moving violation is a major type), and crash risks. An extra simulation experiment involving six additional simulation tests was conducted to illustrate qualitatively the safety resilience. Several key conclusions can be drawn:

- Validation results show that the two-staged model performs well in predicting crash risk on expressways. The model outperformed the ones included in most of the previous studies relying on traffic simulation experiments, as it considers vehicle violations in predicting crash risk with adding traffic violations as a simulation input.
- Idea of traffic resilience seems to work well in explaining the relationship between traffic conditions, safety critical events and crash risk, which are the key elements in road safety field. Safety critical events brought about the risk in traffic, while the traffic environment has its own immune system, the resilience, to prevent these events developing into crashes. Therefore, safety critical events and traffic resilience function together, and are two main factors that decide crash risk on the road.
- Results from the experiment show that scenarios with different traffic conditions but same vehicle moving violation added have observably different crash risk outcomes. This indicates the difference in the ability of different traffic environments in resisting to traffic violations.

As key contributions, the paper proposed a model that considers both violations and traffic conditions in crash prediction on

expressways, which are two key elements in road safety studies but yet not much been explored with both their impacts considered together. The model trained on the machine learning method of BPNN performed well in predicting crash risk. This model can be used for researchers and practitioners to predict and monitor risks on expressways in a potentially more precise way, if it is trained on real-road traffic data and calibrated using historical crash data. By introducing traffic resilience, the study offers an interesting and promising insight in learning road safety problems. Traffic resilience functions as the immunity of the road network that helps prevent crashes.

However, limitations exist in this study and should be further included in future studies. Without traffic data collected from the road section used in the simulation, we were not able to fully calibrate the simulation. Though the use of the virtual and simplified simulation scenario is able to serve the purpose of the study in illustrating the model and the concept of traffic resilience, the simulation should be better calibrated and the model should be further trained and validated using traffic data from real road environments with the help of new technologies such as computer vision and LiDAR techniques (Atev et al., 2005; St-Aubin et al., 2015; Tarko et al., 2017). Slow moving vehicles with different speeds may have different impacts on freeway safety, but such impacts have not been extensively investigated in this study. A sensitivity analysis for different speeds of slow moving vehicles can be performed using the simulation software. In addition, environments with other types of moving violations will be further investigated with developments in simulation tools. Despite rear-end collisions are the most common type of crashes on expressways, other collision types need to be considered in future work. As being promising in crash prediction, the approach using machine learning techniques in crash risk analysis will be further explored with different advanced machine learning methods applied and tested. The safety resilience of traffic will be deeper explored both theoretically and practically. Finding proper methods to quantify and identify indicators to explain the safety resilience can be an interesting topic to study.

Acknowledgments

Funding for this project was provided in part by the Chinese National Natural Science Foundation (71871161) and Fundamental Research Funds for the Central Universities.

References

AAA Foundation for Traffic Safety, 2009. Aggressive driving: Research. Foundation for Traffic Safety, Washington DC.
AASHTO, 2015. Human Factors in Traffic Safety. Lawyers and Judges Publishing Company Inc.
Abdel-Aty, M., Uddin, N., Pande, A., Abdalla, F., Liang, H., 2004. Predicting freeway crashes from loop detector data by matched case-control logistic regression. Transp. Res. Rec.: J. Transp. Res. Board 1897 (1), 88–95.
Abdel-Aty, M., Pande, A., Liang, Y.H., Abdalla, F., 2005. The potential for real-time traffic crash prediction. ITE J. Web 75.
Ahmed, M., Abdel-Aty, M., 2013. A data fusion framework for real-time risk assessment on freeways. Transp. Res. Part C Emerg.Technolog. 26 (1), 203–213.
Ahmed, M.M., Abdel-Aty, M., Yu, R., 2012. Bayesian updating approach for real-time safety evaluation with automatic vehicle identification data. Transp. Res. Rec.: J. Transp. Res. Board 12-0368 (2280), 60–67.
Assad, M., 2013. Investigating the different characteristics of weekday and weekend crashes. J. Safety Res. 46 (6), 91–97.
Atev, S., et al., 2005. A vision-based approach to collision prediction at traffic intersections. IEEE Trans. Intell. Transp. Syst. 6 (4), 416–423.
Barceló, J., 2014. Aimsun Microscopic Traffic Simulator: a Tool for the Analysis and Assessment of Its Systems. Technical Notes. TSS, Barcelona, Spain.
Cai, Q., Mohamed Abdel-Aty, M.S., Yuan, Jinghui, Jaeyoung, Lee., 2018. Safety impact of weaving distance on freeway facilities with managed lanes using both microscopic traffic and driving simulations. Transp. Res. Rec.: J. Transp. Res. Board.
Calvert, S.C., Snelder, M., 2015. A methodology for road traffic resilience analysis and review of related concepts. Transp. A Transp. Sci.
Carbaugh, J., Godbole, D.N., Sengupta, R., 1998. Safety and capacity analysis of automated and manual highway systems. Transp. Res. Part C Emerg. Technol. 6 (1), 69–99.
Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16 (1), 321–357.
Chen, C., 2009. Surrogate Safety Assessment Model (SSAM).
Chong, L., Abbas, M.M., Medina, A., 2011. Simulation of driver behavior with agent-based back-propagation neural network. Transp. Res. Rec.: J. Transp. Res. Board 2249 (-1), 44–51.
English Oxford Living Dictionaries, 2018. https://en.oxforddictionaries.com/definition/immunity.
FMCSA, 2003. On-board Sensors For Determining Brake System Performance Report. Federal Motor Carrier Safety Administration., Washington, DC.
Golob, T.F., Recker, W.W., Alvarez, V.M., 2004. Freeway safety as a function of traffic flow. Accid. Anal. Prev. 36 (6), 933–946.
Golob, T.F., Recker, W., Pavlis, Y., 2008. Probabilistic models of freeway safety performance using traffic flow data as predictors. Saf. Sci. 46 (9), 1306–1333.
Green, M., 2000. How long does it take to stop - methodological analysis of driver perception-brake times. Transp. Hum. Factors 195–216.
Habtemichael, F.G., Santos, L.D.P., 2014. Crash risk evaluation of aggressive driving on motorways: microscopic traffic simulation approach. Transp. Res. Part F Psychol. Behav. 23 (23), 101–112.
Hendricks, D.L., Fell, J.C., Freedman, M., 2000. The relative frequency of unsafe driving acts in serious traffic crashes. Presented at the Proceedings of the 44th Annual Conference of the Association for the Advancement of Automotive Medicine.
Hydén, C., 1987. The Development Of A Method For Traffic Safety Evaluation: The Swedish Traffic Conflicts Technique. Lund University, Lund, Sweden.
Khattak, A., Kantor, P., Council, F., 1998. Role of adverse weather in key crash types on limited-access roadways implications for advanced weather systems. Transp. Res. Rec.: J. Transp. Res. Board 10–19.
Lee, C., 2003. Real-time crash prediction model for application to crash prevention in freeway traffic. Transp. Res. Rec.: J. Transp. Res. Board 1840 (1).
Lee, C., Saccomanno, F., Hellinga, B., 2002. Analysis of crash precursors on instrumented freeways. Transp. Res. Rec.: J. Transp. Res. Board 1784 (1).
Li, D., Deogun, J., Spaulding, W., Shuart, B., 2004. Towards Missing Data Imputation: A Study of Fuzzy K-means Clustering Method. Springer, Berlin Heidelberg.
Li, S., Xiang, Q., Ma, Y., Li, H., 2016. Crash risk prediction modeling based on the traffic conflict technique and a microscopic simulation for freeway interchange merging areas. Int. J. Environ. Res. Public Health 13 (November (11)).
Liu, B., Xia, J., Zhang, Z., Ding, N., Peng, W., 2018. Influence Mechanism of Visual Perception on Driver's Speed Control and Steering Behaviors. Scientific Research Publishing.
Murray-Tuite, P.M., 2006. A comparison of transportation network resilience under simulated system optimum and user equilibrium conditions. Paper Presented at *the Simulation Conference*, 2006. WSC 06. Proceedings of the Winter.
Oh, S.H., 2011. Error back-propagation algorithm for classification of imbalanced data. Neurocomputing 74 (6), 1058–1061.
Pande, A., Abdel-Aty, M., 2006. Comprehensive analysis of the relationship between real-time traffic surveillance data and rear-end crashes on freeways. Transp. Res. Rec.: J. Transp. Res. Board 1953 (1), 31–40.
PSM China, 2017. Compilation of Road Traffic Accident Statistical Data of the People's Republic of China, 2016. Traffic Management Bureau of the Public Security Ministry of the People's Republic of China.
Rakha, H., Wang, W., 2009. Procedure for calibrating gipps car-following model. Transp. Res. Rec.: J. Transp. Res. Board 2124 (2124), 113–124.
Rothengatter, J.A., 1999. Police Enforcement Strategies to Reduce Traffic Casualties in Europe.
Shi, Q., Abdel-Aty, M., Yu, R., 2015. Multi-level bayesian safety analysis with unprocessed automatic vehicle identification data for an urban expressway. Accid. Anal. Prev. 88, 68.
St-Aubin, P., Saunier, N., Miranda-Moreno, L., 2015. Large-scale automated proactive road safety analysis using video data. Transp. Res. Part C Emerg. Technol. 58 (Part B), 363–379.
Stedman's Medical Dictionary, 2006. 28th Edition.
Tarko, A., et al., 2017. Feasibility of tracking vehicles at intersections with a Low-end LiDAR. Paper Presented at the 96th Annual Meeting of the Transportation Research.
Transport Canada, 2015. T. Canadian Motor Vehicle Traffic Collision Statistics. https://www.tc.gc.ca/eng/motorvehiclesafety/tp-tp3322-2015-1487.html.
TSS, 2011. Aimsun Microsimulator API Manual Barcelona: Transport Simulation Systems.
TSS, 2014. Aimsun 8 Users' Manual Barcelona: Transport Simulation Systems. ..
TSS, 2014. Aimsun Dynamic Simulators Users' Manual Barcelona: Transport Simulation Systems.Aimsun Dynamic Simulators Users' Manual Barcelona: Transport Simulation Systems.
US DOT. B. o. T, 2017. Transportation Fatalities by Mode. Bureau of Transportation Statistics, US Department of Transportation.
Vasconcelos, L., Neto, L., Santos, S., Silva, A.B., Seco, Á., 2014. Calibration of the gipps car-following model using trajectory data. Transp. Res. Procedia 3 (2014), 952–961.
Vilarinho, C., Soares, G., Macedo, J., Tavares, J.P., Rossetti, R.J.F., 2014. Capability-enhanced AIMSUN with real-time signal timing control. Procedia - Soc. Behav. Sci. 111, 262–271.
Vogel, K., 2003. A comparison of headway and time to collision as safety indicators. Accid. Anal. Prev. 35 (3), 427–433.

Wang, D., 2002. Traffic Flow Theory. China Communications Press, Beijing, pp. 13–18.
Wang, H., Wang, W., Chen, J., 2010. Using trajectory data to analyze intradriver heterogeneity in car-following. Transp. Res. Rec.: J. Transp. Res. Board 2188, 85–95.
Wang, J., Kong, Y., Fu, T., Stipancic, J., 2017. The impact of vehicle moving violations and freeway traffic flow on crash risk: an application of plugin development for microsimulation. PLoS One 12 (9).
Xu, X., Tian, H., 2007. Kuaisulu Jiashiyuan Fanying Shijian de Queding 4. Xinxi Jishu yu Xinxihua (Information Technology & Informatization), pp. 105–108.
Xu, C., Wang, W., Liu, P., Li, Z., 2015. Calibration of crash risk models on freeways with limited real-time traffic data using bayesian meta-analysis and bayesian inference approach. Accid. Anal. Prev. 85, 207.
You, J., Wang, J., Fang, S., Guo, J., 2017. An optimized real-time crash prediction model on freeway with over-sampling techniques based on support vector machine. J. Intell. Fuzzy Syst. 33 (1), 555–562.
Yu, R., Abdel-Aty, M., 2013. Utilizing support vector machine in real-time crash risk evaluation. Accid. Anal. Prev. 51 (2), 252–259.
Zhang, G., Yau, K.K., Chen, G., 2013. Risk factors associated with traffic violations and accident severity in China. Accid. Anal. Prev. 59 (10), 18.