



# Transformer-based modeling of abnormal driving events for freeway crash risk evaluation

Lei Han<sup>a</sup>, Rongjie Yu<sup>b,c,\*</sup>, Chenzhu Wang<sup>a</sup>, Mohamed Abdel-Aty<sup>a</sup>

<sup>a</sup> Department of Civil, Environmental & Construction Engineering, University of Central Florida, Orlando, FL 32816, United States

<sup>b</sup> The Key Laboratory of Road and Traffic Engineering, Ministry of Education, 4800 Cao'an Road, 201804, Shanghai, China

<sup>c</sup> College of Transportation Engineering, Tongji University, Shanghai 201804, China



## ARTICLE INFO

### Keywords:

Crash Risk Evaluation  
Abnormal Driving Event  
Transformer Model  
Time-decay Function  
Convolutional Neural Network

## ABSTRACT

A crash risk evaluation model aims to estimate crash occurrence possibility by establishing the relationships between traffic flow status and crash occurrence. Based upon which, Proactive Traffic Safety Management (PTSM) systems have been developed and implemented. The current crash risk evaluation models relied on high dense traffic detectors, which limited the applications of PTSM to infrastructures with enough sensing devices. To address such application limitation issue, this study employed the widespread abnormal driving event information that is generated by emerging driving monitoring and vehicle connection techniques to develop the crash risk evaluation model. Specifically, to characterize abnormal driving events, a six-tuple embedding method was proposed to store their space, time and kinetics features. Given their irregular and discrete distributions on roadways, a Transformer model with self-attention mechanism was proposed to extract the spatial distribution characteristics. In addition, a time-decay function was integrated to fit the temporal impacts of abnormal driving events on crash risk. Empirical data from a freeway in China were utilized for the analyses. The results showed that abnormal driving events with lower speed, larger acceleration and duration are more likely to cause crashes. The accumulation of multiple events in the time period of less than 3 min would lead to a sharp increase of crash risk. Besides, compared to the average metrics of the widely adopted Convolutional Neural Network (CNN), XGBoost, and logistic regression models, the proposed model achieved higher accuracy (0.841) and AUC (0.777), with average improvement of 2.5 % and 9.1 % respectively.

## 1. Introduction

Traffic crashes cause considerable loss in people's health and property world widely. According to WHO, approximately 1.3 million people die each year due to road crashes, which cost most countries 3 % of their gross domestic product (WHO, 2021). To improve road safety, Proactive Traffic Safety Managements (PTSM) systems have been suggested to actively evaluate the roadway crash risk and adopt multiple traffic management strategies to prevent crashes (Abdel-Aty et al., 2010). As the key component of a PTSM system, crash risk evaluation models establish the relationship between traffic flow status and crashes by comparing the pre-crash and normal

\* Corresponding author.

E-mail addresses: [le966091@ucf.edu](mailto:le966091@ucf.edu) (L. Han), [yurongjie@tongji.edu.cn](mailto:yurongjie@tongji.edu.cn) (R. Yu), [chenzhu.wang@ucf.edu](mailto:chenzhu.wang@ucf.edu) (C. Wang), [m.aty@ucf.edu](mailto:m.aty@ucf.edu) (M. Abdel-Aty).

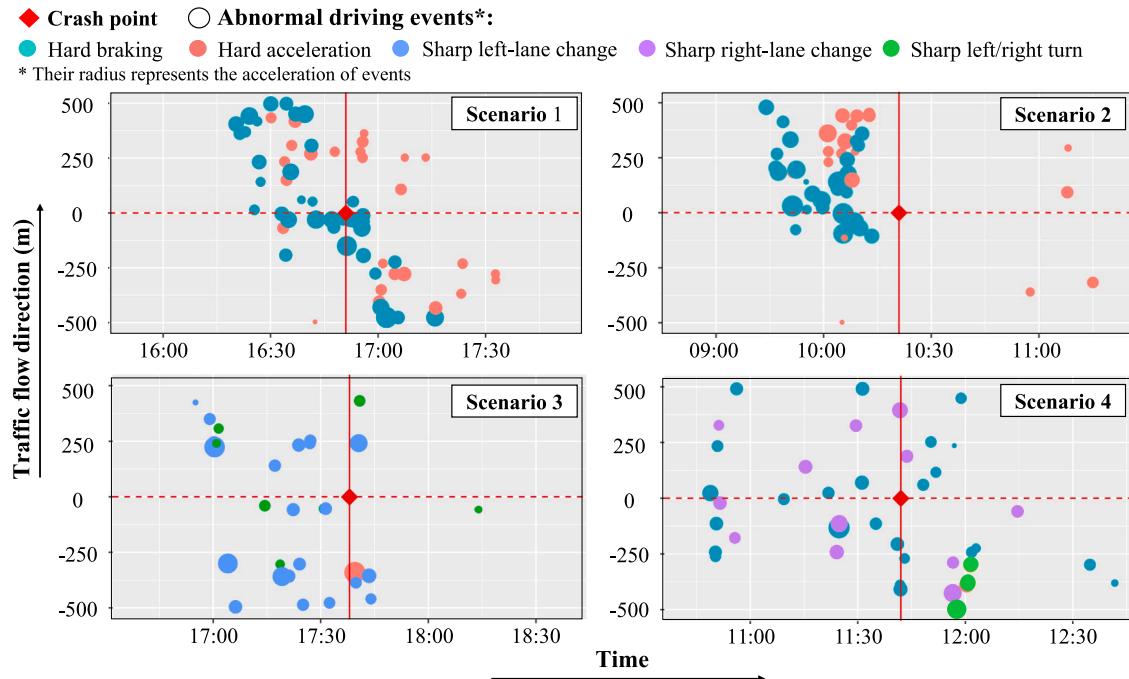
traffic flow conditions (Hossain et al., 2019). Based upon real-time traffic data, it can estimate crash occurrence possibility, which can be further used to trigger variable speed limit, ramp control and other PTSM strategies (Abdel-Aty et al., 2010; Abdel-Aty and Wang, 2017; Yu et al., 2021).

The existing crash risk evaluation studies mainly utilized the traffic data from roadside fixed detectors (e.g., loop detectors in Abdel-Aty et al., (2008); microwave radar in Xu et al., (2014), Abdel-Aty and Wang, (2017), and video detectors in Wang et al., (2019a)). According to Hossain et al (2019), the average detector spacing in most crash risk studies is 0.8 km (0.5mile), while such high-density layout conditions can only be satisfied on limited freeways. For instance, the majority of urban arterials, rural roads, and highways are not equipped with such traffic flow detectors. (Wang et al., 2020; Li and Abdel-Aty, 2022; Zhang et al., 2022). Meanwhile, the total cost of installation and regular maintenance of such detectors are substantial, which was estimated to be as high as \$2 to \$2.5 million per lane-mile (Bodvarsson and Muench, 2010). Therefore, the limited coverage and high installation/maintenance costs have strongly limited the applications of PTSM systems (Hossain et al., 2019; Zhang et al., 2022).

Recently, along with the applications of driving status monitoring and vehicle connection techniques, novel traffic status sensing data of abnormal driving events can be flexibly acquired from diverse types of equipment (e.g., vehicle on-board units and smart phones) in addition to traditional fixed detectors (Guo et al., 2022; Zhang and Abdel-Aty, 2022). For instance, AutoNavi, China's largest navigation company, can identify users' abnormal driving events through smart phone sensors. It has covered different kinds of roadways (e.g., highway, urban arterial, and rural road) in about 200 Chinese cities (AutoNavi, 2023); In the U.S., Wejo claimed that its real-time vehicle monitoring system has been applied to detect vehicle abnormal driving events in more than 7.4 million vehicles and covered 95 % of roads in U.S. (Wejo, 2023, Islam and Abdel-Aty, 2023). This new type of data can provide high-resolution time, location and kinematics information (e.g., speed, acceleration) of drivers' sudden braking, acceleration and other abnormal driving events, which are highly relevant to dangerous traffic conflicts and crashes (Petraki et al., 2020; Mathew et al., 2021). Therefore, it would show great potential to expand the applications of PTSM systems if the widespread abnormal driving event data are applied in crash risk evaluation.

Given the abovementioned advantages, this study aims to utilize the roadway abnormal driving event data to conduct the crash risk evaluation and explore their impact on crashes. However, unlike the aggregate traffic flow data, abnormal driving event data are spatial-temporally non-aggregated on the roadway. Fig. 1 shows the spatial-temporal distributions of abnormal driving events in four different real crash scenarios (the details of events and crashes matching can be found in **3. DATA PREPARATION**). In each crash scenario, the red diamond represents one crash, while the circles are different types of abnormal driving events with varied acceleration (i.e., the size of the circle radius). Specifically, in scenario 1 and 2, many sudden braking and acceleration events occurred and gathered in a short period of time before the crash. On the contrary, the events were mainly sharp left-lane changes and no obvious gathering pattern was showed in scenario 3. While scenario 4 was dominated by sudden braking and right-lane change events around the crash location.

Given such complex spatial-temporal distributions of abnormal driving events, two challenges need to be addressed to establish their correlations with crashes: First, as the occurrence of abnormal driving events depend on drivers' staccato operations and



**Fig. 1.** Comparison of abnormal driving events in different crash scenarios.

surrounding traffic flow status (Guo et al., 2021; Singh and Kathuria, 2021), their spatial locations are discretely distributed on the road and time intervals is random and irregular. This unconventional discrete data structures of abnormal driving events cannot be handled in existing aggregate modeling methods (e.g., logistic regression model and CNN). It is essential to explore new modeling methods to handle such non-aggregated features of abnormal driving events. Second, apart from the event frequency information used in existing studies (Guo et al., 2021; Zhang and Abdel-Aty, 2022), event type, acceleration, duration and other features of abnormal driving events would also affect the crash risk. However, the distributions of such features show significant differences in different crash situations. Therefore, how to extract the multidimensional features of abnormal driving events and establish their heterogeneous relationships with crashes needs to be investigated.

With the above-mentioned research gaps, this study aims to establish the relationship between road abnormal driving events and crashes, therefore expanding the applications of crash risk evaluation and PTSM to more roadways. Main contributions of this paper include:

- 1) A Transformer model was proposed to extract the irregular and discrete spatial features of abnormal driving events and establish their correlation with crashes.
- 2) For the first time, the impacts of the acceleration, speed, duration, and type of abnormal driving events on the crash risk were quantified through the six-tuple embedding and self-attention mechanism in Transformer.
- 3) A time-decay function was integrated to automatically fit the temporal impacts of abnormal driving events on crash risk, which reveals the crash risk spatial-temporal decay and collective superposition effect of multiple abnormal driving events.

The paper is organized into six sections. Following this section, [section 2](#) presents the literature review, followed by the data preparation described in [Section 3](#). [Section 4](#) shows the details of the proposed methodology and [section 5](#) illustrates the experiment results. Finally, the conclusion and discussion of this study is presented in [Section 6](#).

## 2. Literature review

### 2.1. Crash risk evaluation data

There were generally two kinds of data used in crash risk evaluation studies: infrastructure-based data and vehicle-based data. Most of crash risk evaluation studies rely on infrastructure-based data from fixed detectors such as loop detectors (Abdel-Aty et al., 2004, 2008; Xu et al., 2014; Yu et al., 2020), Automatic Vehicle Identifications (AVIs) (Ahmed and Abdel-Aty, 2011; Ahmed et al., 2012), Bluetooth detectors (Yuan et al., 2018), and cameras (Wang et al., 2017; Wang et al., 2019a). Such roadside sensors collect traffic data from a segment of the road. Thus, traffic flow parameters aggregated in road segment level such as average speed and speed deviation were used as modeling variables (Roshandel et al., 2015). However, the main drawback of such data is that they can only provide aggregated segment-level traffic parameters, which cannot accurately capture the traffic parameters on the crash point within the segment (Zhang and Abdel-Aty, 2022). Meanwhile, such high-density layout conditions of traffic sensors can only be satisfied on limited freeways. The majority urban arterials, rural roads, and highways are still not equipped with such traffic flow detectors (Li and Abdel-Aty, 2022; Zhang et al., 2022). Moreover, the total costs of installation and regular maintenance of such detectors are substantial (Bodvarsson and Muench, 2010; Islam and Abdel-Aty, 2023).

With the recent development of mobile sensing technologies and connected vehicles, it is now much easier to obtain vehicle-based data. Some research efforts were conducted to evaluate real-time crash risk using vehicle trajectory data from floating cars (Wang et al., 2019b; Xie et al., 2019). However, such data are collected mostly from taxis or buses. Taxis had certain picking-up/dropping-off patterns, while buses only traveled on specific routes (Zhang and Abdel-Aty, 2022). Recently, with the widespread application of driving status monitoring and vehicle connection techniques, vehicle trajectory data from connected vehicle (CV) and smartphones become more accessible to overcome such shortcomings. Such data are collected from the GPS of CV or smartphones of drivers, so it represents mostly non-commercial trips with high penetration rates, which could provide an acceptable representation of traffic flow (Day et al., 2017). The benefits of such vehicle-based data are that they can be efficiently obtained at a relatively low cost and present traffic parameters near crash points (Guo et al., 2021; Hu et al., 2022; Islam and Abdel-Aty, 2023), which has great potential to improve crash risk evaluation performance.

From the vehicle-based data, individual drivers' driving behaviors could be extracted instead of segment-level traffic flow parameters. Thus, it is possible to identify the abnormal driving events such as hard acceleration, hard braking on roads, which are highly relevant to dangerous traffic conflicts and crashes (Guo et al., 2010; Yao et al., 2019). Existing studies have analyzed the relationship between abnormal driving events and crash frequency and found that there was a significant positive correlation between them (Stipancic et al., 2018; Zhao et al., 2020; Desai et al., 2021). However, the relationship between abnormal driving events and crash risk is still unclear. Zhang and Abdel-Aty (2022) firstly attempted to use abnormal driving event and traffic state data to evaluate crash risk. They found that adding the abnormal driving events into traffic variables could improve the accuracy of model. However, they only used the aggregate number of abnormal driving events, but ignored their type, severity, location, and other information. Therefore, the use of vehicle-based abnormal driving event data to evaluate crash risk and their impacts on crash risk still need further investigation and exploration.

## 2.2. Crash risk evaluation methods

Recently, two types of methods have been developed to evaluate crash risk, statistical and machine learning methods. Statistical methods, such as logistics regression (Abdel-Aty et al., 2004; Pande and Abdel-Aty, 2005; Lee et al., 2006) and Bayesian logistics regression (Ahmed et al., 2012, 2014) were widely used in recent studies. Although statistical methods are easy to interpret to show the relationships between crashes and the traffic flow variables, these methods usually need strong assumptions or dependence on data preparation techniques (Hossain et al., 2019; Li and Abdel-Aty, 2022). Compared with statistical methods, existing studies indicated that machine learning methods could achieve better predictive accuracy with few limitations on data assumption (Wang et al., 2019b; Yu et al., 2021). However, traditional machine learning methods such as Random Forest (RF), Support Vector Machine, Bayesian networks are still not able to handle massive high-dimensional data (Li and Abdel-Aty, 2022; Yu et al., 2023). In recent years, deep learning methods such as CNN (Lu et al., 2017; Yu et al., 2020) and Recurrent Neural Network (RNN) (Yuan et al., 2019; Li and Abdel-Aty, 2022) have been applied in crash risk evaluation studies and achieved much better evaluation accuracy compared with the statistical and machine learning methods. Based on multi-layer convolution processing, CNN model is able to capture the high-dimensional spatial features of traffic data (Yu et al., 2020). The RNN model is especially useful for learning time-series data because of its unique design of memory cell (Yuan et al., 2019; Li and Abdel-Aty, 2022).

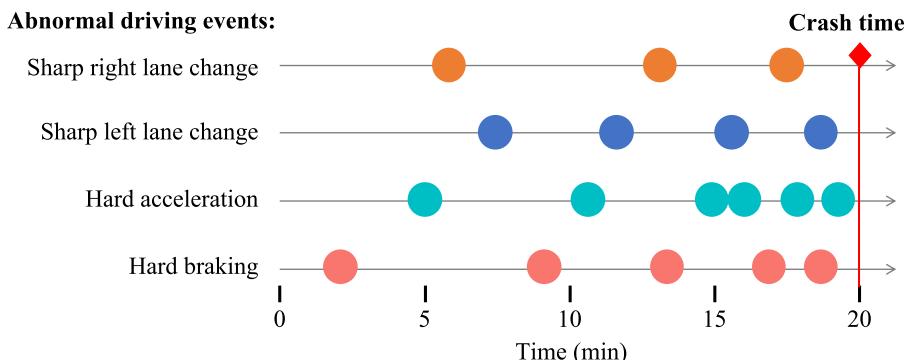
However, those methods need fixed-length input vectors, which makes them cannot handle the irregular abnormal driving events. Unlike the traditional traffic flow data with fixed collection period, the abnormal driving events occur with uncertain time intervals as shown in Fig. 2, resulting in relatively sparse and irregular-length inputs. To solve such problem, a novel Transformer network was proposed to handle the irregular time-series data (Tipirneni and Reddy, 2022). By treating time-series as a set of observation triplets, it can directly learn the contextual information from irregular inputs without data aggregation or imputation. And the attention mechanism in Transformer can help the network better capture the temporal-spatial correlations of such non-aggregate data (Vaswani et al., 2017). The Transformer model have been applied in mortality prediction and disease detection using clinical datasets and had better predictive accuracy than the RNN and Long-Short Term Memory (LSTM) models (Tipirneni, and Reddy, 2022; Chen et al., 2023; Lee et al., 2023). Therefore, given the good capability in handling irregular time-series data, the Transformer model was utilized in this study to learn the abnormal driving event features and explore their impacts on crash risk.

## 3. Data preparation

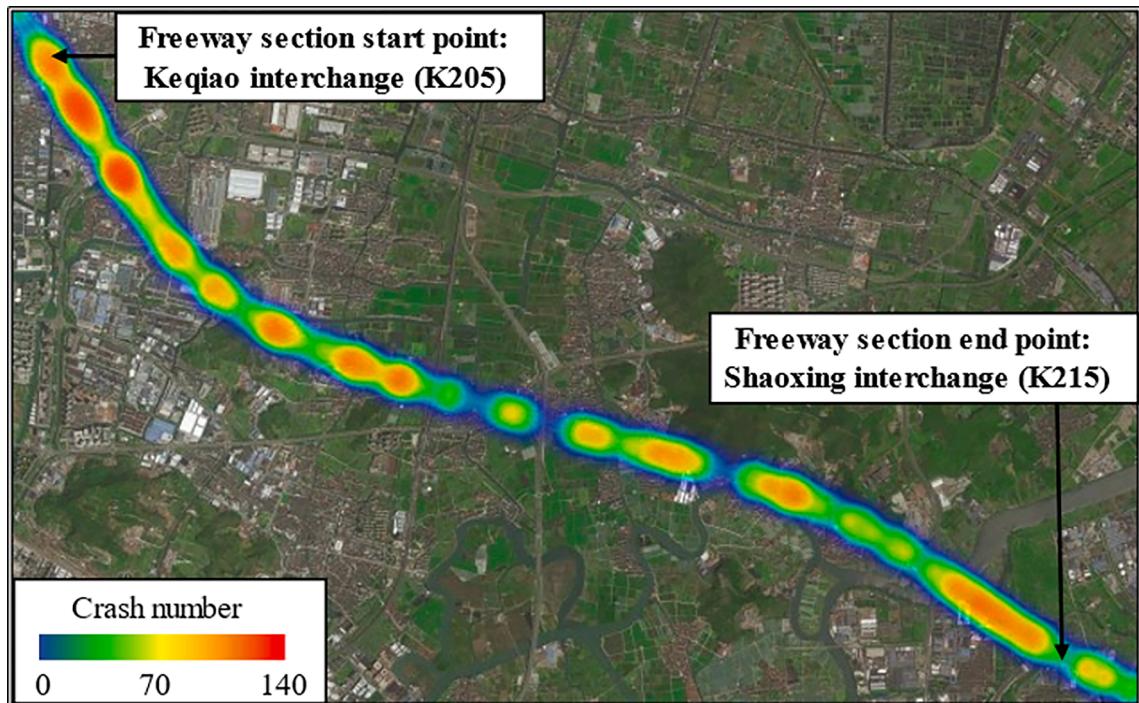
### 3.1. Crash and abnormal driving event dataset

In this study, the chosen freeway segment was part of the Zhejiang Huhangyong freeway in China. The freeway section is 10 km long and separated into North bound and South bound directions. It connects two main cities with large traffic volume and a large number of crashes. Based on the data availability, crash data and abnormal driving event data of 1 year (from September 18, 2021 to September 19, 2022) were employed:

- (1) **Crash data:** Crash data were obtained from the Huhangyong freeway management center and the crash distribution is shown in Fig. 3. For each crash, the crash time, crash location stake, type (i.e., rear-end, side-crash), and severity were recorded. Meanwhile, the managers would use corresponding videos by roadside cameras to calibrate such information to ensure the accuracy of the crash recordings. Finally, 409 two-vehicle and multivehicle crashes were included in this study.
- (2) **Abnormal driving event data:** The abnormal driving event data was provided by the AutoNavi Software Co., Ltd, which has more than 325 million monthly active users in China. The vehicle driving information such as speed, acceleration and driving angle was collected from the sensors on smart phones and the abnormal driving events were identified based on specific thresholds and rules (Yao et al., 2019; Guo et al., 2021). According to the AutoNavi reports (AutoNavi, 2023; AutoNavi open platform, 2024), The overall identification accuracy of abnormal driving behaviors can reach to 95 %-99.8 %. Meanwhile, the spatial positioning error is less than 5 m. The main indicators in the abnormal driving event data are shown in Table 1. Based on



**Fig. 2.** Illustration of the abnormal driving event data with irregular time points.



**Fig. 3.** Illustration of the crash distribution in Huhangyong freeway section.

**Table 1**  
Main indicators in the abnormal driving event data.

Indicators	Description	Unit
lat	The latitude of the abnormal driving event location	—
lon	The longitude of the abnormal driving event location	—
stake	The stake of the abnormal driving event location	m
direction	The traffic flow direction of the road section where the abnormal driving event occurred: 1- South bound, 2- North bound	—
event_type	The type of the abnormal driving event: 1-Sharp left turn, 2-Sharp right turn, 3-Sharp left-lane change, 4- Sharp right-lane change, 5-Sharp acceleration, 6-Sharp brake	—
start_time	The start time of the abnormal driving event	—
end_time	The end time of the abnormal driving event	—
ma	The maximum acceleration during the abnormal driving event	g(9.8 m/s <sup>2</sup> )
ms	The maximum speed during the abnormal driving event	m/s

such data, the type, location, time, speed, and acceleration of each abnormal driving event that occurred on the freeway can be comprehensively obtained. Finally, a total of 238,831 abnormal driving events were recorded (shown in Fig. 4).

### 3.2. Temporal-Spatial matching of crashes and abnormal driving events

Referring to the existing studies (Guo et al., 2021; Zhang and Abdel-Aty, 2022), the temporal-spatial matching rule of crash and abnormal driving events is shown in Fig. 5. For each crash sample, in terms of the temporal range, the recording crash time is set as zero time. A total of 30-min interval from -5 to -35 min is set as the matching temporal period. The data between -5 to 0 min is discarded to avoid the crash time record deviation. As for the spatial range, the 1.5 km upstream and downstream of the crash point is set as the influence range, which can be further subdivided into crash section, upper section and down section with a length of 1 km. Based on the proposed temporal-spatial matching range, the abnormal driving event data from any vehicle within the range would be extracted as shown in Table 2. Apart from the event features (e.g., event type, speed), the time and space of abnormal events were also included. To be specific, the "Space" means the spatial distance between one abnormal driving event with the crash location. The "Time" means the timing differences between the abnormal driving events and crash. For example, if an abnormal driving event happened 11 min prior to the crash at 0.553 km upstream as shown in the Fig. 5. The corresponding time would be -11 and space would be 0.553.

As for the non-crash condition extractions, the matched case-control method was adopted to balance the proportion of crash and non-crash samples. For each crash, 4 non-crash cases were collected in consistent with the majority studies (Abdel-Aty et al., 2004; Hossain et al., 2019; Yu et al., 2019, 2021), considering the same time of day, day of week, and location but different weeks (two weeks

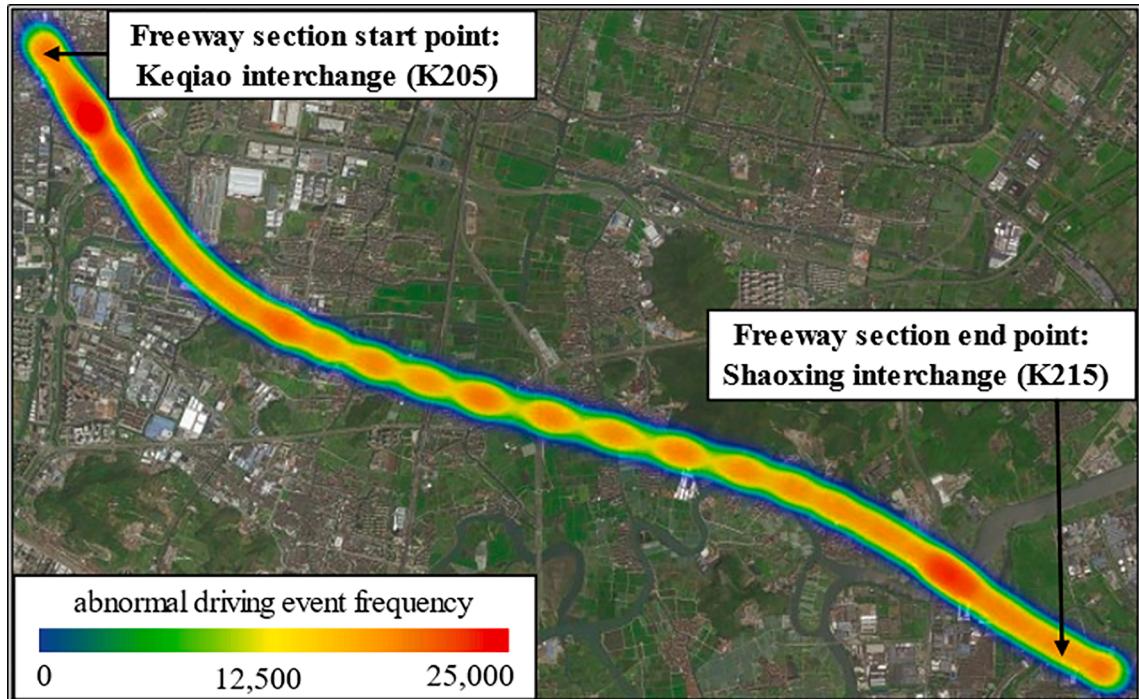


Fig. 4. Illustration of the abnormal driving event distribution in Huhangyong freeway section.

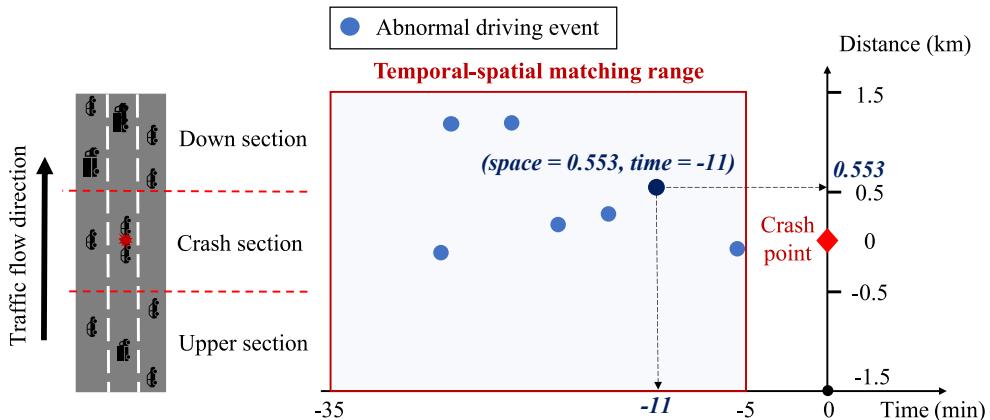


Fig. 5. The temporal-spatial matching rule of crash and abnormal driving events.

**Table 2**  
Illustration of extracted abnormal driving event data.

ID	Time (min)	Space (km)	Event type	Speed (m/s)	Acceleration ( $g = 9.8 \text{ m/s}^2$ )	Duration (s)
110	-26	-0.063	Acceleration	22.873	0.213	7
110	-27	1.326	Acceleration	32.967	0.198	18
110	-18	0.128	Brake	34.173	0.213	6
110	-14	0.110	Brake	22.233	0.231	5
110	-11	0.553	Brake	29.568	0.202	4
110	-5	0.182	Brake	27.893	0.663	10

before and two weeks after the crash). For example, if a crash occurred on April 21, 2022, at 9:35 a.m. on K210 + 500, abnormal driving event data from 9:00 to 9:30 a.m. (30-min interval) on K209 + 000 to K212 + 000 were extracted as the crash sample (Crash = 1). The corresponding non-crash conditions (Crash = 0) were collected from the same section in the same time period of April 7, 14, 28, and May 5.

After the data processing, the final dataset contains 342 crash and 1351 non-crash cases (the non-exact 1:4 ratio of crash and non-crash is due to lack of abnormal driving event data in some crash and non-crash cases). 70 % of the dataset was used for model training and the remaining 30 % of dataset was utilized to test the model performance.

## 4. Methodology

### 4.1. Crash risk evaluation transformer model

In this study, a Transformer model (Tipirneni and Reddy, 2022) was proposed to evaluate crash risk based on non-aggregate abnormal driving events. Formally, an abnormal driving event can be defined as a six-tuple vector  $(t, s, f, vs, va, vd)$ , which respectively represent its time, space, type, value of speed, value of acceleration and value of duration. Each crash or non-crash sample may contain several abnormal driving events, so it can be recorded as a multivariate time-series vector  $T = \{(t_i, s_i, f_i, vs_i, va_i, vd_i)\}_{i=1}^n$ , where  $n$  is the number of abnormal driving events. It worth noting that  $n$  would vary among different samples due to the randomness and sparsity of the abnormal driving events. The corresponding crash label can be defined as  $y \in \{0, 1\}$ . Finally, the model dataset can be defined as  $D = \{(T^k, y^k)\}_{k=1}^N$  with  $N$  observed samples, where the  $k$ th sample contains an abnormal driving event vector  $T^k$ , and a crash label  $y^k$ . Thus, the target of the Transformer model is to evaluate  $y^k$  given  $T^k$ .

The architecture of the proposed Transformer model is illustrated in Fig. 6. First, each abnormal driving event vector is embedded by the Six-tuple Embedding module, which can encode model inputs of different lengths into the same-dimension variable space. Then the initial embeddings are passed through a Feature Learning Transformer module to extract the spatial-temporal and severity features of abnormal driving events. Finally, in the Fusion Self-attention and Evaluation module, these features are combined via self-attention mechanism and passed through a dense layer to evaluate the crash risk result.

- (1) **Six-tuple Embedding:** Given an input time-series  $T = \{(t_i, s_i, f_i, vs_i, va_i, vd_i)\}_{i=1}^n$ , the initial embedding for the  $i$ -th vector  $e_i \in \mathbb{R}^d$  is computed by summing the following component embeddings:  $e_i = e_i^t + e_i^s + e_i^f + e_i^{vs} + e_i^{va} + e_i^{vd}$ . Since the event type variable  $f_i$  is a categorical object (e.g., Acceleration, Brake, and Left-lane changes), embeddings  $e_i^f(\bullet)$  are obtained from a lookup table similar to word embeddings. While the other variables such as time, space are continuous, they are embedded by the Continuous Value Embedding (CVE) technique (Tipirneni and Reddy, 2022) using a one-to-many feed-forward network (FFN) with learnable parameters i.e.,  $e_i^t = FFN^T(t_i)$ , and  $e_i^s = FFN^S(s_i)$ . Each FFN has one input neuron,  $d$  output neurons, a single hidden layer with  $\lfloor \sqrt{d} \rfloor$  neurons, and  $tanh(\cdot)$  activation. They follow the form  $FFN(x) = Utanh(Wx+b)$  where the dimensions of weights  $\{W, b, U\}$  can be inferred from the size of hidden and output layers of the FFN.
- (2) **Feature Learning Transformer:** The initial embeddings  $\{e_1, \dots, e_n\}$  are then passed through a Transformer network (Vaswani et al., 2017) with  $M$  blocks, each containing a Multi-Head Attention (MHA) layer with  $h$  attention heads and a FFN with one hidden layer. Each block takes  $n$  input embeddings  $E \in \mathbb{R}^{n \times d}$  and outputs the corresponding  $n$  output embeddings  $C \in \mathbb{R}^{n \times d}$  that capture the contextual information. MHA layers use multiple attention heads to attend to information contained in different embedding projections in parallel. The computations of the MHA layer are given by:

$$H_j = \text{softmax}\left(\frac{\left(\mathbf{EW}_j^q\right)\left(\mathbf{EW}_j^k\right)^T}{\sqrt{d/h}}\right)(\mathbf{EW}_j^v) j = 1, \dots, h \quad (1)$$

$$MHA(E) = (H_1 \circ \dots \circ H_h) W_C \quad (2)$$

Each head projects the input embeddings into query, key, and value subspaces using matrices  $\{\mathbf{W}_j^q, \mathbf{W}_j^k, \mathbf{W}_j^v\} \subset \mathbb{R}^{d \times dh}$ . The queries and keys are then used to compute the attention weights which are used to compute weighted averages of value vectors. Finally, the outputs of all heads are concatenated and projected to original dimension with  $W_C \in \mathbb{R}^{hd_h \times d}$ . The FFN layer takes the form:

$$F(X) = \text{ReLU}\left(XW_1^f + b_1^f\right)W_2^f + b_2^f \quad (3)$$

with weights  $\mathbf{W}_1^f \in \mathbb{R}^{d \times 2d}$ ,  $\mathbf{b}_1^f \in \mathbb{R}^{2d}$ ,  $\mathbf{W}_2^f \in \mathbb{R}^{2d \times d}$ ,  $\mathbf{b}_2^f \in \mathbb{R}^d$ . During network training, dropout, residual connections, and layer normalization are added for every MHA and FFN layer. Also, attention dropout randomly masks out some positions in the attention matrix before the SoftMax computation. The output of each block is fed as input to the succeeding one, and the output of the last block gives the contextual embeddings  $\{c_1, \dots, c_n\}$ , which characterize the spatial-temporal and severity features of abnormal driving events.

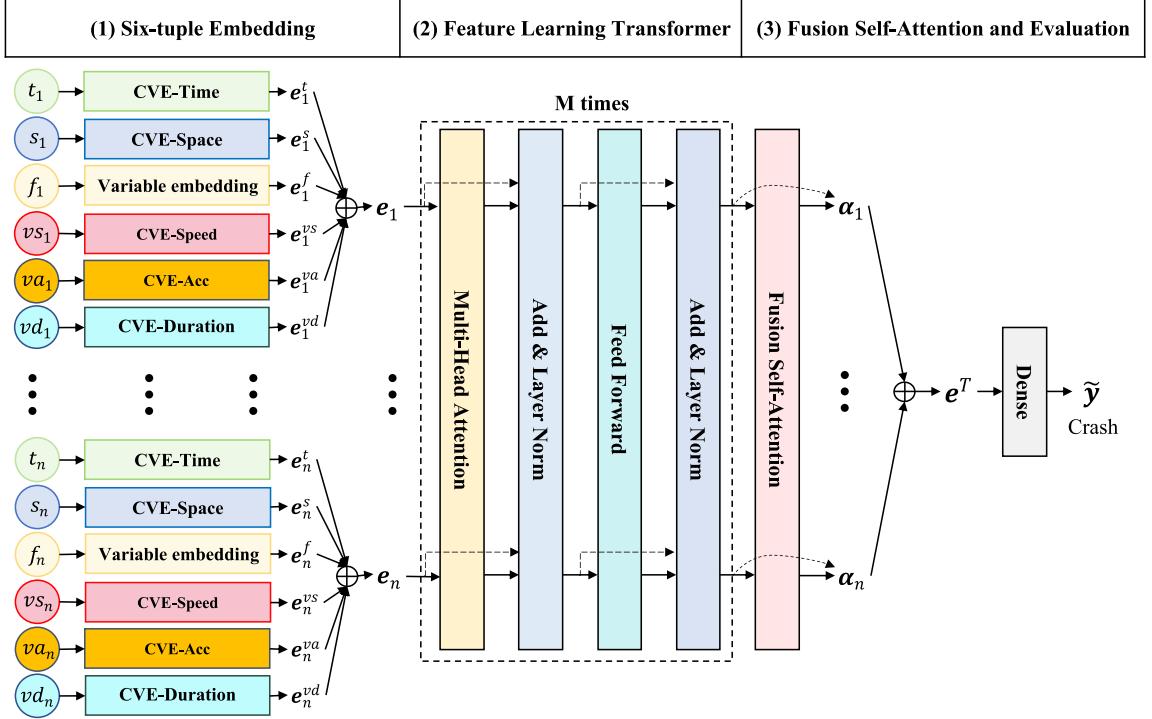


Fig. 6. The architecture of the crash risk evaluation Transformer model.

(3) **Fusion Self-attention and Evaluation:** After computing the feature embeddings, a self-attention layer is used to fuse them to compute time-series embedding  $e^T \in \mathbb{R}^d$ . This layer first computes attention weights  $\{\alpha_1, \dots, \alpha_n\}$  by passing each contextual embedding through a FFN and computing a SoftMax over all the FFN outputs.

$$a_i = \mathbf{u}_a^T \tanh(\mathbf{W}_a \mathbf{c}_i + \mathbf{b}_a) \quad (4)$$

$$\alpha_i = \frac{\exp(a_i)}{\sum_{j=1}^n \exp(a_j)} \forall i = 1, \dots, n \quad (5)$$

$\mathbf{W}_a \in \mathbb{R}^{d_a \times d}$ ,  $\mathbf{b}_a \in \mathbb{R}^{d_a}$ ,  $\mathbf{u}_a \in \mathbb{R}^{d_a}$  are the weights of this attention network which has  $d_a$  neurons in the hidden layer. The time-series embedding is then computed as

$$e^T = \sum_{i=1}^n \alpha_i \mathbf{c}_i \quad (6)$$

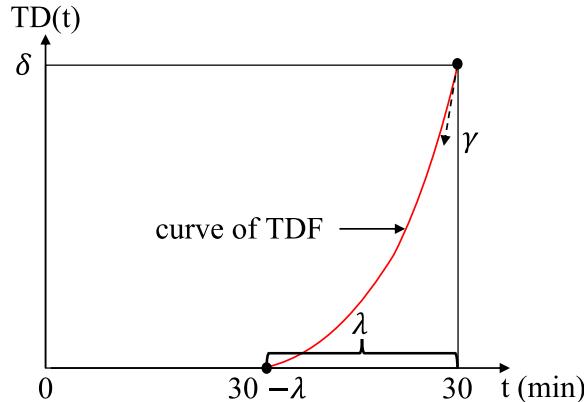


Fig. 7. The TDF curve of the impact of abnormal driving events.

The final evaluation for crash risk is obtained by passing the time-series embeddings through a dense layer with weights  $\mathbf{w}_o^T \in \mathbb{R}^d$ ,  $\mathbf{b}_o \in \mathbb{R}$  and sigmoid activation.

$$\tilde{\mathbf{y}} = \text{sigmoid}(\mathbf{w}_o^T \mathbf{e}^T + \mathbf{b}_o) \quad (7)$$

The model is trained on the target task using cross-entropy loss.

#### 4.2. Time-decay function of abnormal driving events

Existing studies have shown that abnormal driving events can cause traffic flow disorder, aggravate congestion and other negative influences, thereby significantly increasing the crash risk on the road (Guo et al., 2021; Baikejuli and Shi, 2023). However, the duration of their effects should not be relatively too long or short, which are still unclear. From empirical observations, the number of abnormal driving events in the 0–10 min before the crash is significantly more than that in other periods (e.g., 10–30 min before the crash). Such phenomenon may imply that the time-decay effect of abnormal driving events, that is, the closer to the crash, the greater the impact of abnormal driving events on it, and vice versa. In other words, the impact of roadway abnormal driving events on the crash risk gradually decays over time until it ends.

In order to capture such time-decay effect, referring to existing research (Su et al., 2018), a time-decay function (TDF) of the impact of abnormal driving events is proposed as shown in Formula (8) and its function curve image is shown in Fig. 7.

$$TD(t) = \delta^* \left\{ \frac{\max[(t - (30 - \lambda)), 0]}{\lambda} \right\}^\gamma \quad (8)$$

where  $TD(t)$  is the time-decay value of time  $t \in [0, 30]$ ;  $\lambda \in (0, 30]$  is the max-decay time, which indicates the maximum impact duration of an event;  $\gamma > 0$  is the decay exponent to control the speed of time decay; And  $\delta > 0$  is the time decay coefficient, which represents the weight of time value when integrated into the model. For the convenience of calculation, the time of an abnormal driving event  $t$  is recoded to 0–30 as the time of 35 min before a crash is set as the zero time. Therefore, the TDF can exhibit such time-decay trend: the impact of an abnormal driving event on the crash risk is the largest when it occurs close to the crash time, and gradually decays as it occurs far from the crash time.

The time-delay effect was integrated into the Transformer model as shown in Fig. 8. The only difference is that the CVE encoding for time value was replaced by TDF and the time decay value  $TD(t_i)$  was then multiplied by the sum embeddings of the remaining five variables. It is worth noting that during the process of model training, the three parameters of TDF can be automatically adjusted to fit the optimal time decay patterns.

## 5. Results

### 5.1. Model performance of crash risk evaluation

During the model training, Grid Search method was chosen to find the optimal parameter settings for the models. The final hyperparameters of the model are tuned as shown in Table 3. For instance, the input six-tuple embedding dimension is set as 50; The Transformer has 2 MHA layers and each layer has 4 attention heads; In the initial setting of TDF, the time decay exponent is selected as 1, the max-decay time is selected as 30 and the time decay coefficient is selected as 1. For model training, the optimizer is selected as Adam, compared with Stochastic Gradient Boosting (SGD) and Root Mean Squared Propagation (RMSprop). The learning rate is set as 0.001, the batch size is selected as 64. The model was trained with 100 epochs.

For comparison, some aggregated modeling methods in existing studies (Guo et al., 2022; Zhang and Abdel-Aty, 2022) were developed. Specifically, the aggregate characteristics of abnormal driving events were calculated based on the rules as shown in the Fig. 9. The 25-minute interval from −5 to −30 min was selected and split into five 5-minute time slices. The space locations were divided into Upstream (U), Crash (C), and Downstream (D) sections following the ranges in Fig. 5. According to the event type, the number, average speed, average acceleration, and average duration of abnormal driving events in each aggregated space-time windows were calculated. Finally,  $4*2*3*5 = 120$  aggregated variables were extracted as the model inputs. For example, “NBC1” means the number of sharp braking events in the Crash section in the first 5-minute time slices. As for the models, besides the proposed models, all used models in this study are shown below:

- (1) (Non-aggregated) **Transformer model with TDF**;
- (2) (Non-aggregated) **Transformer model without TDF**;
- (3) (Aggregated) **CNN model**: two convolutional layers (with Batch Normalization and ReLU function), average pooling layer, and a fully connected dense layer (Yu et al., 2020; Zhang and Abdel-Aty, 2022);
- (4) (Aggregated) **eXtreme Gradient Boosting (XGBoost) model** (Guo et al., 2022).
- (5) (Aggregated) **Logistic Regression (LR) model** (Ahmed et al., 2012; Guo et al., 2022).

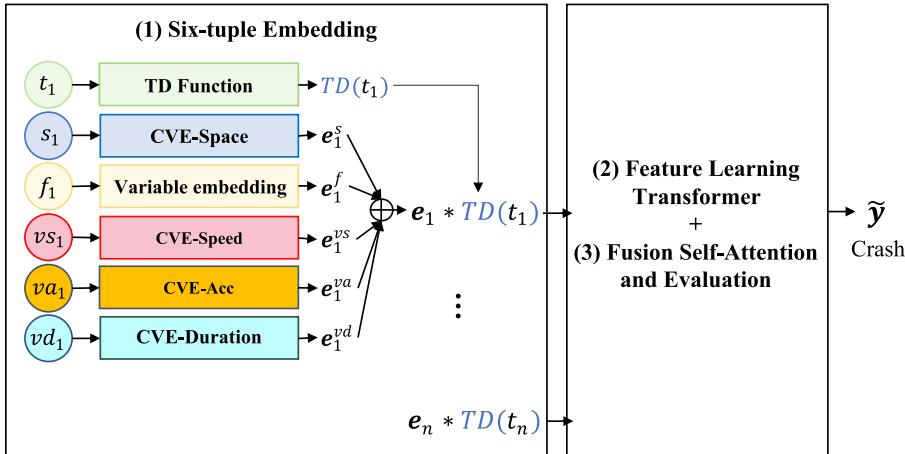


Fig. 8. The architecture of the Transformer model with time-delay effect.

**Table 3**  
The hyperparameters tuning of proposed Transformer model.

Hyperparameter	Tuning range	Selected value
Transformer architecture		
Six-tuple embedding dimension $d$	25, 50, 100	50
Number of MHA layers $M$	1, 2, 3, 4	2
Number of attention heads in MHA $h$	1, 2, 4, 8	4
TDF initial setting*		
Time decay exponent $\gamma$	0.5, 1, 2	1
Max-decay time $\lambda$	10, 15, 20, 25, 30	30
Time decay coefficient $\delta$	0, 0.5, 1, 2	1
Model training		
Optimizer	SGD, Adam, RMSprop	Adam
Learning rate	$1 \times 10^{-5}, 2 \times 10^{-5}, 4 \times 10^{-5}$	$4 \times 10^{-5}$
Batch size	16, 32, 64, 128	32
Dropout rate	0, 0.1, 0.2, 0.6	0.2

\*: These three parameters of TDF will be automatically learned and changed during model training.

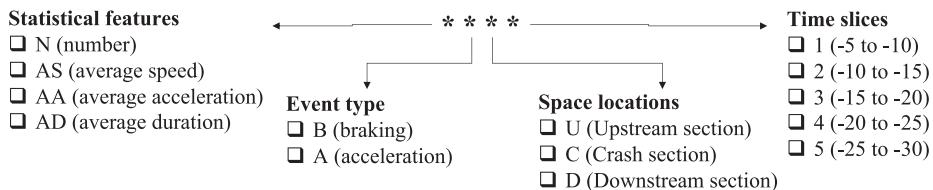


Fig. 9. Aggregate rules for abnormal driving event features.

In terms of the evaluation metrics, accuracy (ACC), false alarm rate (FAR), and the Area under ROC curve (AUC) were used, which can be calculated by Equation (9) and (10) given the classification threshold of 0.5.<sup>1</sup> ACC represents the ratio of correctly classified samples to all samples; A low FAR means the model can correctly classify most negative samples; AUC measures the area under the Receiver Operating Characteristic curve, the higher the AUC, the better the model is at distinguishing between the two classes.

$$FAR = FP / (TN + FP) \quad (9)$$

$$ACC = (TP + TN) / (TP + TN + FP + FN) \quad (10)$$

The experiment results are listed in Table 4. The AUC values on training data set range from 0.683 to 0.785, which denote all five models were well trained. On the test dataset, the Transformer model with TDF has the best performance with highest ACC (0.841), AUC (0.777) and very low FAR (0.095). While the Transformer model has similar model metrics. The above result indicates that the time-decay mechanism can help to capture the time-varying pattern of abnormal driving events, thereby improving the model performance. On the other hand, the proposed Transformer models achieve the state-of-the-art performance. Compared with the best

<sup>1</sup> TP: true positive, FP: false positive, TN: true negative, FN: false negative.

**Table 4**

Experiment results of five models.

Model	Train	Test	FAR	AUC
	AUC	ACC		
<b>Aggregated modeling methods (Existing studies)</b>				
LR	0.683	0.806	<b>0.050</b>	0.665
XGboost	0.722	0.812	0.053	0.710
CNN	0.755	0.822	0.101	0.739
<b>Non-aggregated modeling methods (Ours)</b>				
Transformer	0.775	0.827	0.112	0.761
<b>Transformer + TDF</b>	<b>0.785</b>	<b>0.841</b>	0.095	<b>0.777</b>

model (CNN) in aggregated models, the AUC improvement of Transformer model with TDF is  $(0.777 - 0.739)/0.739 = 5.1\%$  and the ACC improvement is  $(0.841 - 0.822)/0.822 = 2.3\%$ . Besides, the FAR decreases from 0.101 to 0.095. The above results show that the Transformer can extract the non-aggregated features of abnormal driving events (i.e., their time intervals, spatial locations near crash point), which also have important impacts on road crashes.

### 5.2. Impact of the acceleration, speed, duration and event type of abnormal driving events on crash risk

An abnormal driving event has six inherent features, including its type, speed, acceleration, duration, occurrence time, and spatial location. In order to explore the impact of such features on the crash risk, the control variates method was used in this study. The main idea is to analyze the change of crash risk with different values of one abnormal driving event feature while keeping other feature values unchanged. For example, to comprehend the impact of the event acceleration on the crash risk, the acceleration is continually changing and other features (i.e., type, speed, duration, time, and spatial location) are fixed. Such abnormal driving event samples are then input into the well-trained Transformer with TDF model to observe the change of the output crash risk.

In this section, the impact of the acceleration, speed, duration, and type of abnormal driving events on crash risk are discussed. Fig. 10 shows the real-world observation distributions of these event features between crash and non-crash samples (left side: (a), (c), (e)) and their corresponding crash risk in the model (right side: (b), (d), (f)). Based on the results, main findings can be summarized as follows:

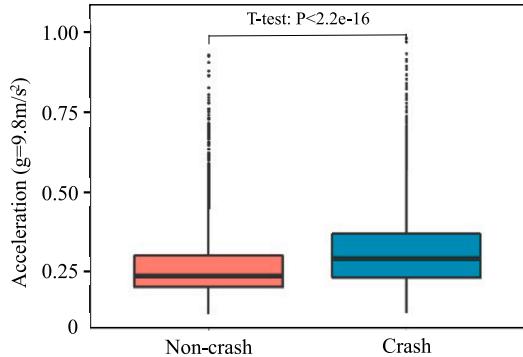
- (1) **Acceleration:** Fig. 10 (b) illustrates that the acceleration of abnormal driving events is positively correlated with the crash risk, which is confirmed by its real-world distribution on the crash and non-crash samples in Fig. 10 (a). It indicates that an abnormal driving event with higher acceleration would have greater impacts on the surrounding traffic flow, which is more likely to lead to crashes.
- (2) **Speed:** There is a negative relationship between the speed of abnormal driving events and crash risk in Fig. 10 (d), which consist with the observation distribution (Fig. 10 (c)). It shows that in congested traffic, abnormal driving events are more likely to cause crashes. While the crash risk is relatively low in a high-speed traffic flow.
- (3) **Duration:** A positive correlation can be seen between the duration of abnormal driving events and the crash risk in Fig. 10 (f), and the same trend can be seen in the corresponding real-world distribution in Fig. 10 (e). It indicates that the longer an abnormal driving event persists, the more surrounding vehicles and traffic flow would be affected, consequently heightening the crash risk.
- (4) **Event type**<sup>2</sup>: Given identical acceleration, speed, and duration, the crash risk associated with sharp braking consistently surpasses that of sharp acceleration. It implies that the sharp braking events pose greater dangers because they may cause the rear vehicles to fail to avoid in time, resulting in rear-end, side-collision and other types of crashes.

### 5.3. Crash risk temporal-spatial decay of single abnormal driving event

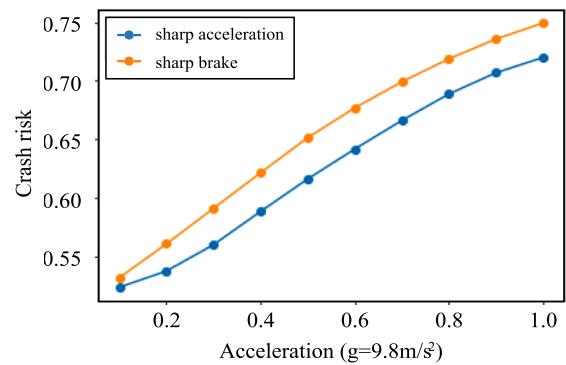
Though the results given by the Transformer with TDF model (Fig. 11), the spatial-temporal impact of a single abnormal driving event on the crash risk can be analyzed:

- (1) **Space:** From Fig. 11(a), it can be seen that the crash risk is negatively correlated with the distance from the abnormal driving event to crash point. The highest crash risk is observed on the crash point (space = 0 km), implying that when an abnormal driving event occurs on the road, the traffic flow around it would be mainly affected. As this adverse disturbance gradually diminish upstream and downstream, it is worth noting that the impact at downstream would be slightly higher than upstream segment.
- (2) **Time:** The crash risk caused by an abnormal driving event tends to decrease over time (seen in Fig. 11(b)). Specifically, the crash risk gradually decreases within 10 min after an abnormal driving event occurs until it does not change. It shows the time-

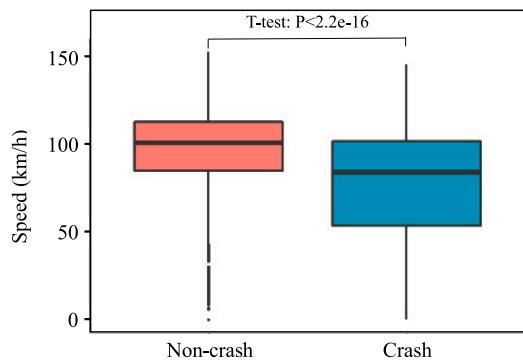
<sup>2</sup> Data of sharp left/right turns and sharp left/right-lane turn are too few in extracted samples, thus not be included.



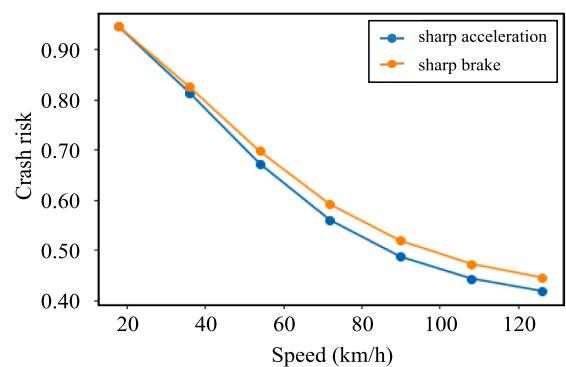
(a) Acceleration distribution



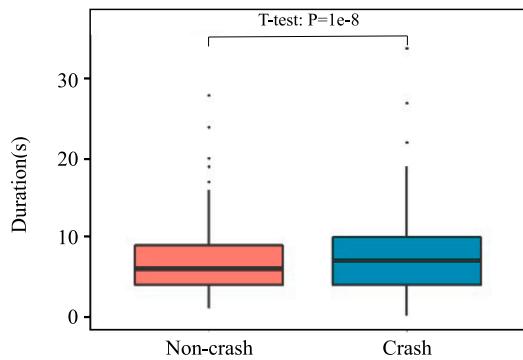
(b) The crash risk in different acceleration



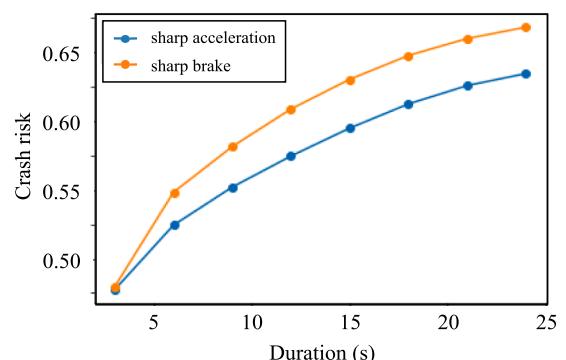
(c) Speed distribution



(d) The crash risk in different speed



(e) Duration distribution

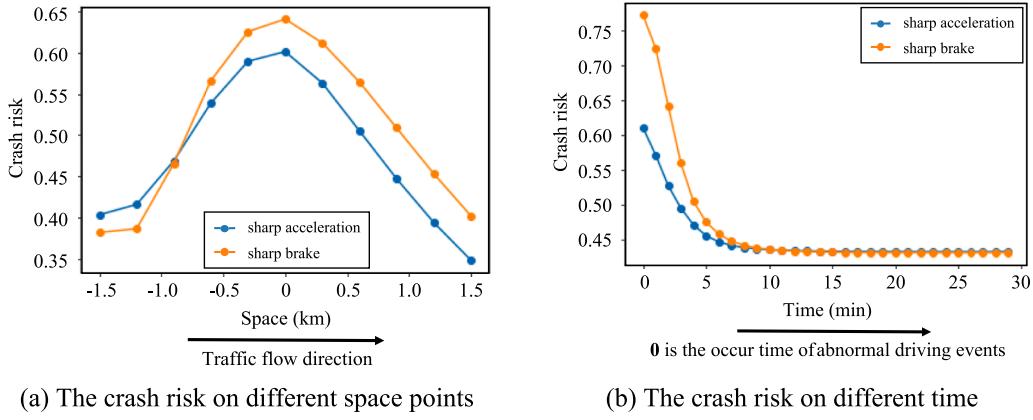


(f) The crash risk in different duration

**Fig. 10.** Distributions of abnormal driving event features and their corresponding crash risk\*. \*Setting of other control variables when one variable is changeable: acceleration = 0.3 g; speed = 80 km/h; duration = 10 s; space = -0.5 km; time = 25 min (10 min before the crash time). a) Acceleration distribution. b) The crash risk in different acceleration. c) Speed distribution. d) The crash risk in different speed. e) Duration distribution. f) The crash risk in different duration.

decay effect of an abnormal driving event on crash risk: once a single abnormal driving event occur, it instantly generates adverse effects on the surrounding vehicles and traffic flow. Then, this negative impact begins to fade within the subsequent 10 min if there are no other abnormal driving events happen. Therefore, the impact duration of abnormal driving events is estimated to be up to 5–10 min.

Based on the above analysis, the spatial-temporal distribution of the crash risk caused by an abnormal driving event can be obtained. Taking sharp acceleration and braking at different acceleration levels as examples, Table 5 illustrates their respective spatial-temporal distributions. It can be seen that:



**Fig. 11.** The crash risk of a single abnormal driving event on different spatial-temporal points. (a) The crash risk on different space points. (b) The crash risk on different time.

- At an acceleration of 0.1 g (a modest level), neither sharp acceleration nor sharp braking poses a notably heightened crash risk.
- When acceleration escalates to 0.3 g, each abnormal driving event would cause an obvious crash risk peak at its occurrence point. Such crash risk would then gradually decay over time and decrease toward the upstream and downstream of the road. Overall, the peak of crash risk caused by the sharp brake is larger, and the scope of its spatial-temporal impact range is much wider.
- If the acceleration reaches 0.5 g, both sharp acceleration and sharp brake would cause a significant crash risk peak ( $>0.8$ ). Following the onset of these events, similar spatial-temporal decay patterns of crash risk can be observed. Moreover, the crash risk in the downstream is significantly higher than that in the upstream. Such results indicate that when a serious abnormal driving event occurs, its negative impact would significantly propagate downstream along the traffic flow.

#### 5.4. Crash risk collective superposition effect of multiple abnormal driving events

In most pre-crash scenarios, there are more than one abnormal driving event that occur continuously on the road. The combined impact of multiple abnormal driving events on crash risk needs to be further explored. In this section, the crash risk of multiple abnormal driving events at different time intervals are compared with that of a single event. For the sake of discussion, three abnormal driving events (their information are in Table 6) were selected and the crash risk under different scenarios are shown in Fig. 12.

From the Fig. 12, three crash risk patterns can be summarized:

- (1) If three abnormal driving events occur at the same time (Fig. 12(a), time interval = 0) on the road, the crash risk caused by them is higher than that caused by any one alone. It indicates that the simultaneous occurrence of multiple abnormal driving events has a higher impact on traffic flow than a single one, thus causing a greater crash risk.
- (2) If the abnormal driving events occur within a short time interval (Fig. 12(b), (c), time interval = 1, 2 min), the crash risk will superimpose to increase as new events occur. It indicates that in a short time period, the traffic flow disorder caused by the last abnormal driving event continues and superimposes with the crash risk at this time, resulting in the higher crash risk of multiple events than that of a single one.
- (3) However, if a more extended gap, like 4 min (Fig. 12(d)), separates these events, the impacts caused by the previous events would decay to be much little on the traffic. As a result, the overall crash risk would be very similar to the crash risk caused by the abnormal driving event at the current moment.

From the above analysis, the collective superposition effect of abnormal driving events can be concluded:

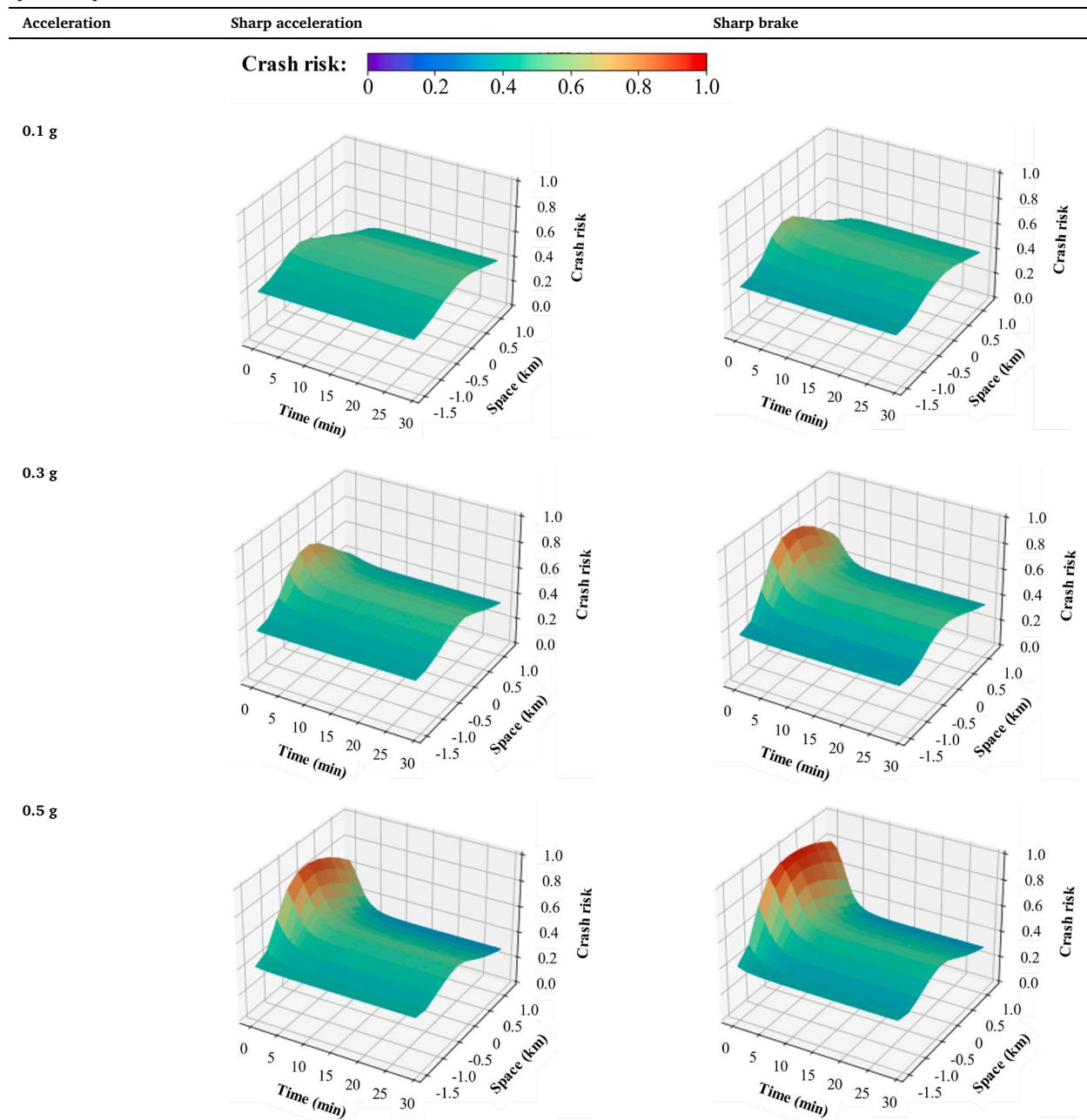
*If multiple abnormal driving events gather in a short time period (0–3 min), the negative impact of those events on traffic flow would be superimposed, resulting in a significantly higher crash risk than a single one.*

Such effect can also be clearly observed and verified in the real crash samples. For example, Fig. 13 illustrates the process of the crash risk collective superposition on a road section. The upper part of it shows the crash risk in the pre-crash 30 min, and the lower part is the corresponding spatial-temporal distribution of abnormal driving events on the road. It can be seen that:

- In **stage 1** (7–9 min) and **stage 2** (11–12 min), two abnormal driving events caused a high crash risk. However, the crash risk gradually subsided as no new abnormal driving events occurred in the subsequent short time period.
- While in **stage 3** (17–30 min), more than 15 abnormal driving events occurred around the crash point. The adverse impacts on traffic flow by those events were gradually superposed. In such situation, the crash risk increased rapidly and remained at an extremely high level ( $>0.8$ ), eventually leading to the occurrence of the crash.

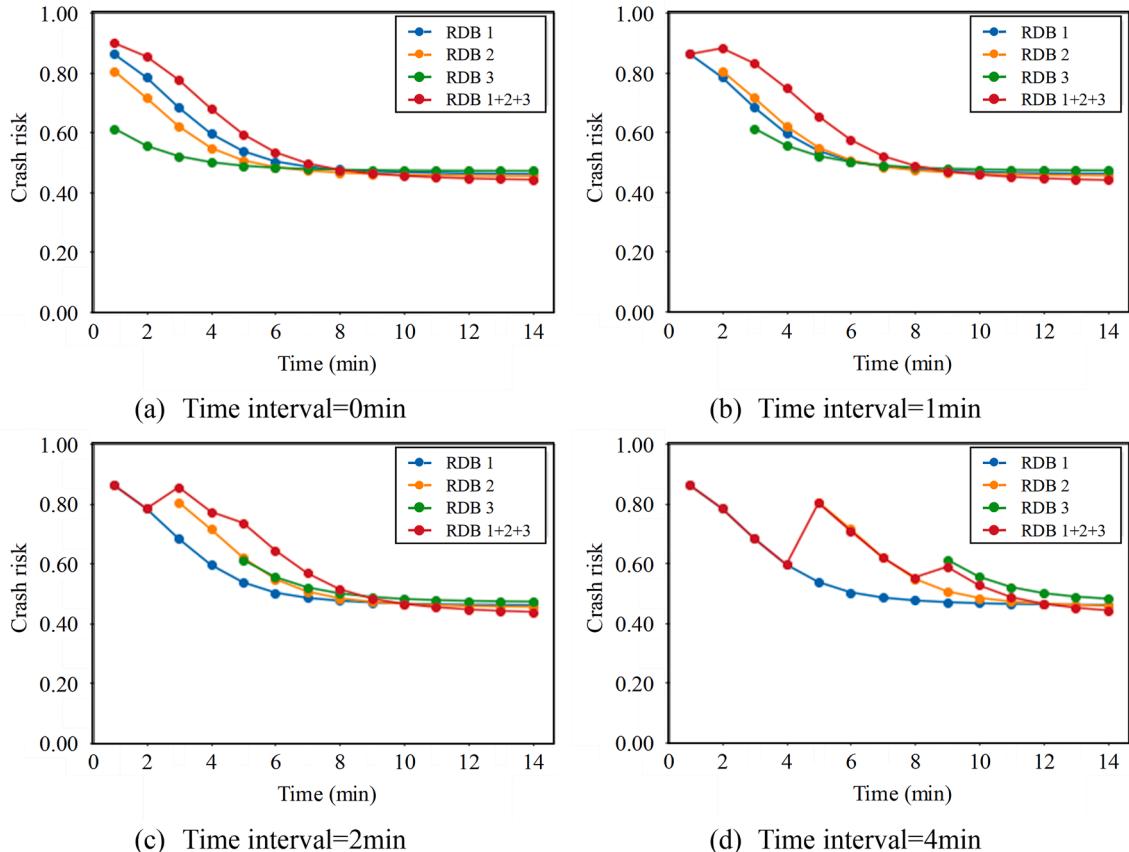
**Table 5**

Spatial-temporal distribution of the crash risk under different accelerations.

**Table 6**

Information of three abnormal driving events.

Abnormal driving event (RDB)	Space (km)	Event type	Acceleration (g)	Speed (km/h)	Duration (s)
RDB 1	-0.5	Brake	0.4	80	10
RDB 2	-0.6	Brake	0.3	85	15
RDB 3	-0.5	Acceleration	0.2	90	8



**Fig. 12.** The crash risk comparison of single and multiple abnormal driving. \* “RDB 1 + 2 + 3” means that the three abnormal driving events continue to occur at specific time intervals. a)Time interval = 0 min. b)Time interval = 1 min. c)Time interval = 2 min. d)Time interval = 4 min.

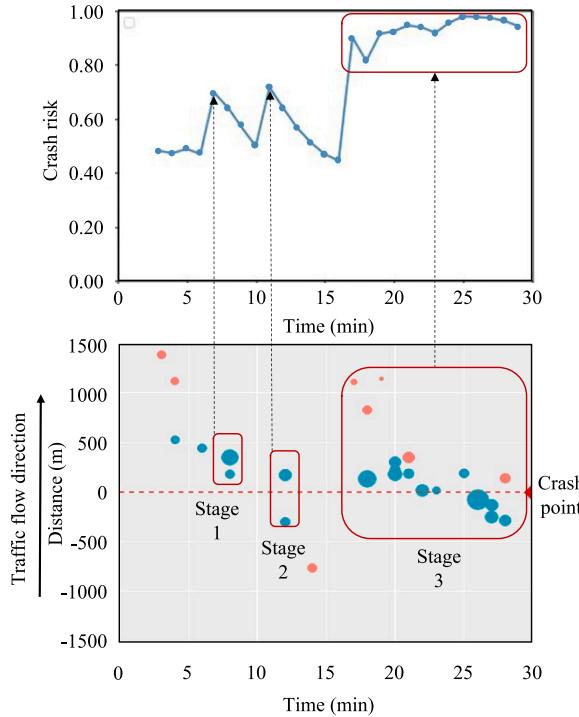
## 6. Conclusions and discussion

The current crash risk evaluation models relied on high dense traffic detectors, which limited the applications of PTS to the infrastructure with enough sensing devices. With the development of emerging driving status monitoring and vehicle connection technologies, the vehicle-based abnormal driving event data can be obtained in recent years (Guo et al., 2021). Compared to traditional fixed-detector traffic data, the widespread abnormal driving events data can be flexibly acquired from diverse types of equipment and provide high-resolution driving behavior information (Zhang and Abdel-Aty, 2022; Li and Abdel-Aty, 2022; Islam and Abdel-Aty, 2023). Given such abnormal driving events (e.g., sudden braking, acceleration) are highly relevant to dangerous driving interactions and crashes (Yao et al., 2019; Zhang et al., 2019; Mathew et al., 2021), the application of widespread abnormal driving event data in crash risk evaluation would show great potential in expanding the applications of crash risk evaluation to more roadways.

In this study, the abnormal driving events data from smart phones were used to evaluate crash risk on freeways and their impact on real-time crash risk were explored. Specifically, abnormal driving event data including the spatial-temporal position, type, acceleration, speed, and duration were first extracted to construct the modeling variables; Then, a Transformer model with self-attention mechanism was proposed to extract the spatially irregular and discrete distribution of events. And a time-decay function was integrated to fit the temporal impacts of abnormal driving events on crash risk. The results showed that the proposed Transformer model could achieve higher ACC of 0.841 with low FAR of 0.095 than existing aggregated modeling methods.

Based on the experimental results, the main findings of the study can be summarized as:

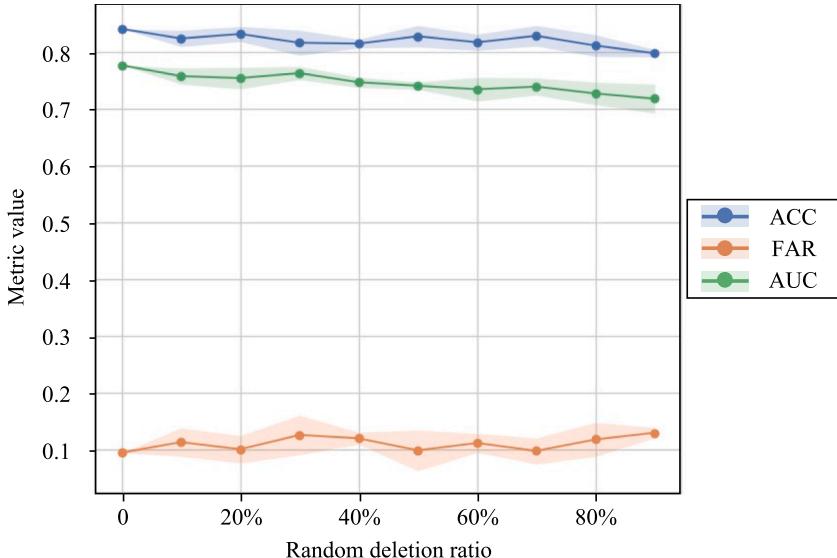
- (1) The acceleration and duration of an abnormal driving event is positively correlated with the crash risk. While the speed is significantly negatively correlated with the crash risk. For the same acceleration, speed, and duration, the crash risk of a sharp braking event is always higher than that of sharp acceleration.
- (2) The crash risk of a single abnormal driving event shows a temporal-spatial decay trend: Temporally, the crash risk of an abnormal driving event will gradually decrease in approximately 10 min. Spatially, the crash risk around an abnormal driving event is highest and decays to the upstream and downstream.



**Fig. 13.** An example of the process of the crash risk collective superposition.

- (3) In scenarios involving multiple abnormal driving events, a clear collective superposition effect can be observed: If multiple abnormal driving events gather in a short time period (0–3 min), their negative impact on surrounding vehicles and traffic flow would be superimposed, leading to a significantly higher crash risk than a single event.

Furthermore, previous studies (Guo et al., 2021; Zhang and Abdel-Aty, 2022) estimated crash risks mainly based on the frequency of abnormal driving events, which may be affected by the proportion of vehicles using navigation. In the contrary, ablation experiments have been performed to show the high robustness of the proposed Transformer model to the proportion of vehicles using navigation. Specifically, we randomly deleted specific ratio (denoted by random deletion rate) of abnormal driving events for each sample to simulate different proportions of vehicles using navigation. Then, the processed samples were fed back into the proposed



**Fig. 14.** The model performance under different random deletion ratio.

Transformer model to observe the changes of model performance. In the ablation experiments, the random deletion rate gradually increases from 10 %, 20 % to 90 %. Due to the randomness in the deletion operations, five sets of experiments were repeated with different random seeds at each random deletion rate. Fig. 14 shows the model metrics and their fluctuation ranges across different random deletion rates. Notably, despite substantial increase in the random deletion rate, the changing of ACC, AUC, and FAR remain slight. When the random deletion rate reaches even 90 %, ACC drops by approximately 5.11 % (0.841 to 0.798) and AUC by 7.59 % (0.777 to 0.718). FAR rises slightly from 0.098 to 0.130. The above results show that the model evaluation performance is barely affected by the navigation usage proportion. In the other word, the proposed Transformer model is highly robust to proportion of vehicles using navigation as the evaluated crash risk does not rely on the navigation usage proportion.

The results proved the advantage of using abnormal driving events to expand the application scopes of crash risk evaluation. With the proposed crash risk evaluation model, the study can be used to implement the safety potential prediction components in PTSIM. Nonetheless, there are still a few limitations in the current study. Firstly, the used abnormal driving event data only have the penetration rate of 5–10 %. The usage rate and penetration rate of navigation systems may impact data quality and introduce bias in crash analysis. With the popularization of connected and autonomous driving technology, multi-sources and higher penetration of vehicle-based data will become possible in the future, therefore reducing the data bias and enhancing the model performance. Secondly, this study mainly established the relationship between abnormal driving events and crashes, other factors such as weather, lighting, traffic composition that might have an impact on the crash risk were not included. Thirdly, more data can be collected to test the model transferability to other expressways.

#### CRediT authorship contribution statement

**Lei Han:** Conceptualization, Data curation, Formal analysis, Methodology, Validation, Writing – original draft, Writing – review & editing. **Rongjie Yu:** Methodology, Validation, Writing – original draft, Writing – review & editing, Conceptualization, Data curation, Formal analysis. **Chenzhu Wang:** Formal analysis, Validation, Writing – original draft, Writing – review & editing. **Mohamed Abdel-Aty:** Formal analysis, Supervision, Validation, Writing – original draft, Writing – review & editing.

#### Declaration of competing interest

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

#### Acknowledgements

This study was supported by the Chinese National Natural Science Foundation (NSFC 52172349), the Belt and Road Cooperation Program under the 2023 Shanghai Action Plan for Science, Technology and Innovation (No. 23210750500), and the Science and Technology Plan Project of Zhejiang Provincial Department of Transport. The authors also acknowledge the support of UCF's Smart and Safe Transportation lab.

#### References

- Abdel-Aty, M., Uddin, N., Pande, A., Abdalla, M.F., Hsia, L., 2004. Predicting freeway crashes from loop detector data by matched case-control logistic regression. *Transp. Res.* 1897 (1), 88–95.
- Abdel-Aty, M., Pande, A., Das, A., Knibbe, W.J., 2008. Assessing safety on Dutch freeways with data from infrastructure-based intelligent transportation systems. *Transp. Res. Rec.* 2083 (1), 153–161.
- Abdel-Aty, M., Pande, A., Hsia, L., 2010. The concept of proactive traffic management for enhancing freeway safety and operation. *ITE Journal* 80 (4), 34.
- Abdel-Aty, M., Wang, L., 2017. Implementation of variable speed limits to improve safety of congested expressway weaving segments in microsimulation. *Transp. Res. Procedia* 27, 577–584.
- Ahmed, M.M., Abdel-Aty, M.A., 2011. The viability of using automatic vehicle identification data for real-time crash prediction. *IEEE Trans. Intell. Transp. Syst.* 13 (2), 459–468.
- Ahmed, M.M., Abdel-Aty, M., Yu, R., 2012. Assessment of interaction of crash occurrence, mountainous freeway geometry, real-time weather, and traffic data. *Transp. Res. Rec.* 2280 (1), 51–59.
- Ahmed, M.M., Abdel-Aty, M., Lee, J., Yu, R., 2014. Real-time assessment of fog-related crashes using airport weather data: a feasibility analysis. *Accid. Anal. Prev.* 72, 309–317.
- AutoNavi open platform. <https://lbs.amap.com/product/locate/#/>. Accessed March 2024.
- AutoNavi, 2023. China's Major Cities Traffic Analysis Report in 2022.
- Baikejuli, M., Shi, J., 2023. A cellular automata model for car–truck heterogeneous traffic flow considering drivers' risky driving behaviors. *Internat. J. Modern Phys. D* 23, 2350154.
- Bodvarsson, G.A., Muench, S.T. (2010). Effects of loop detector installation on the Portland cement concrete pavement lifespan: case study on I-5 (No. WA-RD 744.5). Washington (State). *Dept. of Transportation. Office of Research and Library Services*.
- Chen, L.C., Hung, K.H., Tseng, Y.J., Wang, H.Y., Lu, T.M., Huang, W.C., Tsao, Y., 2023. Self-supervised based general laboratory progress pretrained model for cardiovascular event detection. *arXiv preprint arXiv:2303.06980*.
- Day, C.M., Li, H., Richardson, L.M., Howard, J., Platte, T., Sturdevant, J.R., Bullock, D.M., 2017. Detector-free optimization of traffic signal offsets with connected vehicle data. *Transp. Res. Rec.* 2620 (1), 54–68.
- Desai, J., Li, H., Mathew, J.K., Cheng, Y.T., Habib, A., Bullock, D.M., 2021. Correlating hard-braking activity with crash occurrences on interstate construction projects in Indiana. *Journal of Big Data Analytics in Transportation* 3 (1), 27–41.
- Guo, F., Klauer, S.G., Hankey, J.M., Dingus, T.A., 2010. Near crashes as crash surrogate for naturalistic driving studies. *Transp. Res. Rec.* 2147 (1), 66–74.
- Guo, M., Zhao, X., Yao, Y., Yan, P., Su, Y., Bi, C., Wu, D., 2021. A study of freeway crash risk prediction and interpretation based on risky driving behavior and traffic flow data. *Accid. Anal. Prev.* 160, 106328.

- Guo, M., Zhao, X., Yao, Y., Bi, C., Su, Y., 2022. Application of risky driving behavior in crash detection and analysis. *Physica A* 591, 126808.
- Hossain, M., Abdel-Aty, M., Quddus, M.A., Miromachi, Y., Sadeek, S.N., 2019. Real-time crash prediction models: state-of-the-art, design pathways and ubiquitous requirements. *Accid. Anal. Prev.* 124, 66–84.
- Hu, Y., Li, Y., Huang, H., Lee, J., Yuan, C., Zou, G., 2022. A high-resolution trajectory data driven method for real-time evaluation of traffic safety. *Accid. Anal. Prev.* 165, 106503.
- Islam, Z., Abdel-Aty, M., 2023. Traffic conflict prediction using connected vehicle data. *Analytic Methods in Accident Research* 39, 100275.
- Lee, C., Abdel-Aty, M., Hsia, L., 2006. Potential real-time indicators of sideswipe crashes on freeways. *Transp. Res. Rec.* 1953 (1), 41–49.
- Lee, K., Lee, S., Hahn, S., Hyun, H., Choi, E., Ahn, B., Lee, J., 2023. Learning Missing Modal Electronic Health Records with Unified Multi-modal Data Embedding and Modality-Aware Attention. *arXiv preprint arXiv:2305.02504*.
- Li, P., Abdel-Aty, M., 2022. Real-time crash likelihood prediction using temporal attention-based deep learning and trajectory fusion. *Journal of Transportation Engineering. Part a: Systems* 148 (7), 04022043.
- Lu, W., Luo, D., Yan, M., 2017. A model of traffic accident prediction based on convolutional neural network. In: In 2017 2nd IEEE International Conference on Intelligent Transportation Engineering. IEEE, pp. 198–202.
- Mathew, J.K., Desai, J., Li, H., Bullock, D.M., 2021. Using anonymous connected vehicle data to evaluate impact of speed feedback displays, speed limit signs and roadway features on interstate work zone speeds. *Journal of Transportation Technologies* 11 (4), 545–560.
- Pande, A., Abdel-Aty, M., 2005. A freeway safety strategy for advanced proactive traffic management. *J. Intell. Transp. Syst.* 9 (3), 145–158.
- Petraiki, V., Ziakopoulos, A., Yannis, G., 2020. Combined impact of road and traffic characteristic on driver behavior using smartphone sensor data. *Accid. Anal. Prev.* 144, 105657.
- Roshandel, S., Zheng, Z., Washington, S., 2015. Impact of real-time traffic characteristics on freeway crash occurrence: Systematic review and meta-analysis. *Accid. Anal. Prev.* 79, 198–211.
- Singh, H., Kathuria, A., 2021. Analyzing driver behavior under naturalistic driving conditions: a review. *Accid. Anal. Prev.* 150, 105908.
- Stipancic, J., Miranda-Moreno, L., Saunier, N., Labbe, A., 2018. Surrogate safety and network screening: modelling crash frequency using GPS travel data and latent Gaussian Spatial Models. *Accid. Anal. Prev.* 120, 174–187.
- Su, S.Y., Poon, P.C., Chen, Y.N., 2018. How time matters: Learning time-decay attention for contextual spoken language understanding in dialogues. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1* (pp. 2133–2142).
- Tipirneni, S., Reddy, C.K., 2022. Self-supervised transformer for sparse and irregularly sampled multivariate clinical time-series. *ACM Trans. Knowl. Discov. Data* 16 (6), 1–17.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Polosukhin, I., 2017. Attention is all you need. *Adv. Neural Inf. Proces. Syst.* 30.
- Wang, L., Abdel-Aty, M., Lee, J., 2017. Implementation of active traffic management strategies for safety on congested expressway weaving segments. *Transp. Res. Rec.* 2635 (1), 28–35.
- Wang, P., Lai, J., Huang, Z., Tan, Q., Lin, T., 2020. Estimating traffic flow in large road networks based on multi-source traffic data. *IEEE Trans. Intell. Transp. Syst.* 22 (9), 5672–5683.
- Wang, J., Luo, T., Fu, T., 2019b. Crash prediction based on traffic platoon characteristics using floating car trajectory data and the machine learning approach. *Accid. Anal. Prev.* 133, 105320.
- Wang, C., Xu, C., Dai, Y., 2019a. A crash prediction method based on bivariate extreme value theory and video-based vehicle trajectory data. *Accid. Anal. Prev.* 123, 365–373.
- Wejo, 2023. Remote Diagnostics Solutions. Real-Time Vehicle Monitoring and Maintenance with Wejo. <https://www.wejo.com/uses/remote-diagnostics>.
- WHO, 2021. Road traffic injuries.
- Xie, K., Ozbay, K., Yang, H., Li, C., 2019. Mining automatically extracted vehicle trajectory data for proactive safety analytics. *Transp. Res. Part C: Emerg. Technol.* 106, 61–72.
- Xu, C., Liu, P., Wang, W., Li, Z., 2014. Identification of freeway crash-prone traffic conditions for traffic flow at different levels of service. *Transp. Res. Part A: Policy Practice* 19, 58–70.
- Yao, Y., Zhao, X., Zhang, Y., Ma, J., Rong, J., Bi, C., Wang, Y., 2019. Development of urban road order index based on driving behavior and speed variation. *Transp. Res. Rec.* 2673 (7), 466–478.
- Yu, R., Zheng, Y., Qin, Y., Peng, Y., 2019. Utilizing partial least-squares path modeling to analyze crash risk contributing factors for Shanghai urban expressway system. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering* 5 (4), 05019001.
- Yu, R., Wang, Y., Zou, Z., Wang, L., 2020. Convolutional neural networks with refined loss functions for the real-time crash risk analysis. *Transp. Res. Part C: Emerg. Technol.* 119, 102740.
- Yu, R., Han, L., Zhang, H., 2021. Trajectory data-based freeway high-risk events prediction and its influencing factors analyses. *Accid. Anal. Prev.* 154, 106085.
- Yu, R., Han, L., Wang, L., Zou, Z., 2023. Improve model robustness of traffic crash risk evaluation via adversarial mix-up under traffic flow fundamental diagram. *Accid. Anal. Prev.* 194, 107360.
- Yuan, J., Abdel-Aty, M., Wang, L., Lee, J., Yu, R., Wang, X., 2018. Utilizing bluetooth and adaptive signal control data for real-time safety analysis on urban arterials. *Transp. Res. Part C: Emerg. Technol.* 97, 114–127.
- Yuan, J., Abdel-Aty, M., Gong, Y., Cai, Q., 2019. Real-time crash risk prediction using long short-term memory recurrent neural network. *Transp. Res. Rec.* 2673 (4), 314–326.
- Zhang, S., Abdel-Aty, M., 2022. Real-time crash potential prediction on freeways using connected vehicle data. *Analytic Methods in Accident Research* 36, 100239.
- Zhang, T., Chan, A.H., Xue, H., Zhang, X., Tao, D., 2019. Driving anger, aberrant driving behaviors, and road crash risk: Testing of a mediated model. *Int. J. Environ. Res. Public Health* 16 (3), 297.
- Zhang, Z., Nie, Q., Liu, J., Hainen, A., Islam, N., Yang, C., 2022. Machine learning based real-time prediction of freeway crash risk using crowdsourced probe vehicle data. *J. Intell. Transp. Syst.* 1–19.
- Zhao, X., Ju, Y., Li, H., Zhang, C., Ma, J., 2020. Safety of raised pavement markers in freeway tunnels based on driving behavior. *Accid. Anal. Prev.* 145, 105708.