LSTM + Transformer Real-Time Crash Risk Evaluation Using Traffic Flow and Risky Driving Behavior Data

Lei Han, Mohamed Abdel-Aty, *Senior Member, IEEE*, Rongjie Yu, and Chenzhu Wang, *Member, IEEE*

***Abstract*— Crash risk evaluation studies mainly established the relationship between the macro traffic status and crashes. However, the impact of risky driving behavior, a significant factor in crashes, has not been thoroughly investigated due to the data collection limitations of fixed detectors. In this study, the risky driving behavior data generated by Connected Vehicle (CV) techniques was introduced along with traffic flow data to develop the crash risk evaluation model. An LSTM + Transformer approach was developed, in which the Transformer could extract the non-aggregated spatial-temporal features of risky driving behaviors and LSTM learn the temporal patterns of traffic flow. An ensemble layer was proposed to integrate the macro traffic status features and micro driving behavior, and automatically fit their weights to optimize crash risk evaluation performance. Data from a Chinese freeway was used for empirical analysis. The results show that the proposed LSTM + Transformer model achieved high model accuracy (77.7%), recall (68.6%), and AUC (0.785), with average improvement of between 5.34%, 15.69%, and 5.97%, respectively, compared to existing LSTM, XGBoost, SVM and Logistic Regression (LR) models. Moreover, utilizing risky driving behavior data by incorporating the macro traffic status has proved to capture the pre-crash traffic flow turbulence more precisely. The model results explained by SHapley Additive exPlanations (SHAP) reveal that higher frequency, longer duration and greater acceleration of risky braking behavior increase the number of road vehicles affected, thereby heightening the crash risks. These findings could help the deployment of proactive traffic management and target CV control strategies to reduce crashes.**

***Index Terms*— Crash risk evaluation, risky driving behavior, Transformer, proactive traffic management.**

I. Introduction

TRAFFIC crashes have caused considerable losses in both health and property worldwide. The WHO [1] reported that approximately 1.19 million people died in 2021 due to these incidents, thus accounting for the loss of 10-12% of the global gross domestic product (about $1.8 trillion). For people aged 5–29, crashes were the leading cause of death [1]. To improve the road safety, Proactive Traffic Managements (PTM) systems were proposed to address the likelihood of crash occurrence, so as to implement corresponding measures to improve traffic operations and prevent crashes [2], [3], [4]. Serving as a basic component of PTM system, crash risk evaluation model aims to establish the relationship between road traffic flow status and crashes, therefore evaluating the likelihood of crashes based on real-time traffic data [5], [6].

The occurrence of crashes is often related to specific turbulences of traffic flow status, intensification of traffic conflicts, and risky driving behaviors [7], [8], [9], [10]. Due to data collection limitations, existing crash risk evaluation studies mainly adopted the traffic flow parameters from fixed detectors (e.g., loop detectors [11], [12], [13], roadside radars [14], [15], [16], and video detectors [17]) to quantify the crash likelihood. However, these parameters only represent the macro and aggregated traffic flow status at the road segment level, which cannot reflect the impact of detailed interaction of micro risky driving behaviors on the crash risk [5], [9], [18]. In addition, the traffic flow data from detectors upstream and downstream of crashes were utilized to represent the pre-crash traffic flow status. But the large spacing distances between the detectors and crashes (e.g., 800m-2km in Hossain et al. [5], Yu et al. [6]) would cause the deviation of traffic flow representation, that is, the data cannot precisely capture the changes/ turbulence of traffic flow around the crash point [19].

With the popularity and application of emerging connected vehicle and driving status monitoring technology, collecting and updating real time driving behavior data become possible. Such data not only contain a large number of single-vehicle kinematic parameters (e.g., location, speed, and acceleration), but also can provide rich risky driving behaviors such as sharp changes in lane and speed, which are highly relevant to dangerous traffic conflicts and crashes [7], [8], [9]. Compared to the data from the fixed-detectors, the distinct advantage of the risky driving behavior data is that they are obtained from individual vehicles that are not limited by the constraints of installation locations [20]. Thus, they can be high-resolution and discretely distributed throughout the whole roadway space, allowing for a more sensitive representation of driving behavior characteristics and traffic status [21]. Moreover, such vehicle-based risky driving behavior data provide comprehensive coverage, low acquisition cost, and

Manuscript received 1 January 2024; revised 7 June 2024; accepted 30 July 2024. Date of publication 12 August 2024; date of current version 1 November 2024. The Associate Editor for this article was L. Li. *(Corresponding author: Rongjie Yu.)*

Lei Han, Mohamed Abdel-Aty, and Chenzhu Wang are with the Smart and Safe Transportation Lab, Department of Civil, Environmental and Construction Engineering, University of Central Florida, Orlando, FL 32816 USA (e-mail: le966091@ucf.edu; m.aty@ucf.edu; chenzhu.wang@ucf.edu).

Rongjie Yu is with the Key Laboratory of Road and Traffic Engineering of Ministry of Education, Shanghai 201804, China, and also with the College of Transportation Engineering, Tongji University, Shanghai 201804, China (e-mail: yurongjie@tongji.edu.cn).

Digital Object Identifier 10.1109/TITS.2024.3438616

rich data variety [9], [10], [22]. Therefore, it would show great potential to comprehensively investigate the crash influence factors and improve crash risk evaluation performance using such multi-source data of traffic flow and risky driving behavior.

Given the abovementioned advantages, this research aims to utilize both macro traffic flow data and micro risky driving behavior to conduct crash risk evaluation and explore its impact on the occurrence of crashes. Despite recent attempts to combine the two datasets for crash risk evaluation, these studies have aggregated the risky driving behavior into large spatial-temporal units, potentially overlooking finer details. For example, Zhang and Abdel-Aty [9] and Ma et al. [7] utilized the frequency of risky driving behaviors measured in intervals of 1h and 5min as variables in their model. However, risky driving behavior may change significantly within a short time period (e.g., 1-5 min), especially before the occurrence of crashes. Existing aggregated methods would cause the loss of important non-aggregated behavior features such as their discrete spatial distribution and short-term temporal variations. In addition, previous studies directly put the traffic flow and risky driving behavior variables into one single model, but there are significant differences in their data scale and spatiotemporal resolution. How to fully extract the macro traffic flow and micro driving behavior characteristics in the model structure and combine them to improve the evaluation performance still needs to be further explored.

To bridge the research gaps mentioned above, this paper introduced a novel ensemble framework combing LSTM and Transformer models to evaluate crash risk incorporating both traffic flow and risky driving behavior data. The results of this study have the potential for improving the crash prediction accuracy and guiding the development of crash prevention measures. The main contributions of this paper are threefold:

1) A Transformer model was proposed to extract the non-aggregated spatial-temporal features of risky driving behaviors and establish their correlation with crashes.
2) An ensemble framework combining Transformer and LSTM was created to integrate the macro traffic flow and micro risky driving behavior characteristics, which can automatically adjust the feature weights for optimal crash risk evaluation performance.
3) The correlations between the traffic flow, risky driving behavior variables, and the crash risk were quantified using the SHAP algorithm.

The remaining sections of this paper are organized as follows. Section II provides a review of relevant literatures. Section III describes the dataset processing. Section IV shows the details of the proposed methodology, and section V illustrates the experiment results. Finally, Section VI offers conclusions and suggestions for future research and practices.

II. Literature Review

A. Crash Risk Evaluation Datasets

Different kinds of traffic data have been utilized in crash risk evaluation studies to represent the traffic status, which can be mainly divided into two types: infrastructure-based data and vehicle-based data.

In the majority of crash risk evaluation studies, the infrastructure-based data from fixed detectors such as loop detectors [6], [11], [12], [23], automatic vehicle identifications [16], [24], Bluetooth [25], and cameras [17], [26] were commonly used. Such roadside sensors have been widely used to collect the aggregated traffic flow data (e.g., average speed and speed deviation) of road segments [5], [19]. The limitation of such data lies in their detection range, that is restricted to their locations. They only provide segment-level traffic parameters and are unable to precisely capture the traffic status at the exact point of a crash within the segment [7], [9]. Meanwhile, such high-density layout conditions of traffic sensors can only be satisfied on limited freeways. The majority of urban arterials, rural roads, and highways are still not equipped with such traffic detectors [18], [27]. Moreover, the total costs of installation and regular maintenance of such detectors are very high [4], [9], [28]. It is worth noting that some offline data such as historical crashes, road characteristics were also utilized to evaluate the crash risks at the road network level [29], [30], [31].

With the recent development of mobile sensing and vehicle connection technologies, it become available to obtain the vehicle-based traffic data. Some earlier studies attempted to evaluate crash risk using vehicle GPS data from floating cars such as taxis and buses [32]. However, those floating cars had certain picking-up/dropping-off patterns and only traveled on specific routes. Recently, the novel vehicle-based data from connected vehicles (CV) and in-vehicle devises become more accessible to overcome such shortcomings. Given such data are collected from CVs and drivers' phones, they represent mostly non-commercial trips with high penetration rates [9], [18], [33]. From such vehicle-based data, the details of drivers' risky driving behaviors can be extracted, which are highly relevant to dangerous traffic conflicts and crashes [7], [9]. Furthermore, such risky driving behavior data are continuously distributed along the roadways to present traffic parameters near crash points, which has great potential to improve crash risk evaluation performance [5], [19], [33].

Given the abovementioned benefits, some studies have begun to utilize the vehicle-based risky driving behavior data into traffic crash analysis research. Risky driving behaviors can be defined as specific driving actions that have negative consequences to surrounding traffic participants. They are also called as "risky/dangerous/hazardous driving behaviors" in existing studies [8], [21], [34], [35], [36]. As the key cause of traffic flow disorders, such risky driving behaviors have been found to be highly relevant to dangerous traffic conflicts and crashes. For example, Guo et al. [21] firstly introduced the risky driving behavior variables to detect crashes and achieved better accuracy than only use traffic flow dataset. Zhang and Abdel-Aty [9] combined the frequency of hard brakes and other risky driving behaviors with traffic flow variables to predict crash risk and analyzed the model transferability. Ma et al. [7] extracted the speed and numbers of risky acceleration and braking behaviors from smartphone data to predict the freeway crash risk and got good prediction performance. Although existing studies explored the advantages of applying risky driving behavior data, they only used the aggregate features of such behaviors, ignoring their type, severity, location, and other important information. In addition, the traffic flow and risky driving behavior data were simply combined at the variable level. How to fully exploit their spatiotemporal distribution and non-aggregated features to improve crash risk evaluation still need further investigation.

B. Crash Risk Evaluation Methods

The statistical and machine learning models are two main types of methods proposed to develop crash risk evaluation. Statistical methods, such as logistic regression [8], [11], [24] and Bayesian logistic regression [37] were widely used in earlier studies. While statistical methods offer ease of interpretation for examining the relationships between crashes and traffic flow variables, they typically require strong assumptions and are strongly dependent on data preparation techniques [5]. Compared with statistical methods, existing studies indicated that machine learning methods could achieve better predictive accuracy with fewer data assumption limitations [38], [39]. However, traditional machine learning methods such as Random Forest (RF), Support Vector Machine, Bayesian networks are still not able to handle massive high-dimensional traffic data [6], [9], [40], [41], [42].

In recent years, deep learning methods such as CNN [12], [43] and Recurrent Neural Network (RNN) [44], [45] have been applied in crash risk evaluation studies and achieved much better evaluation accuracy compared with statistical and machine learning methods. The advantage of deep learning models is that they can process massive amounts of data and deeply mine the nonlinear characteristics of variables [39]. Among them, the LSTM model has been widely used as its unique design of memory cells for time-series learning [9], [40], [46], [47]. Given that the temporal variation of pre-crash traffic flow features significantly differs from the normal state, the LSTM model has been proven effective in capturing such temporal fluctuation patterns, thereby enabling the precise identification of crash risks [39], [40], [47].

However, these methods require fixed-length input vectors, which limits their ability to handle non-aggregated risky driving behavior data. Unlike the traditional traffic flow data with a fixed collection period, the risky driving behaviors occur with uncertain time intervals, resulting in a sparse and irregular-length data structure. To solve such problem, a novel Transformer network was proposed to handle the irregular time-series data [48]. By encoding time-series as a set of observation tuples, 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. The Transformer model has been applied in mortality prediction and disease detection using clinical datasets and had better predictive accuracy than the RNN-based models [49], [50]. Given the good capability in handling irregular time-series data, the Transformer model was utilized to learn the risky driving behavior features and explore their impacts on crash risk.

III. Data Collection and Preparation

In this paper, three datasets (i.e., crash, traffic flow, and risky driving behavior data) were collected from a freeway segment of Zhejiang Huhangyong freeway in China. The freeway section is 10 km long between the Keqiao and Shaoxing interchange, which is separated into North bound and South bound directions. It has 4 lanes in each direction with a speed limit of 120 km/h. Since this freeway serves as a major link between two main cities, the studied segment experiences high traffic volumes and a significant number of crashes. Based on the data availability, the data from September 18, 2021, to September 19, 2022, were collected.

A. Raw Data

In this study, three types of data were collected:

1) Crash Data: The crash data were obtained from the Huhangyong Freeway Management Center. For each crash, the crash time, crash location stake, type (i.e., rear-end, side-crash), and severity were recorded. To make sure the crash data quality, the managers would use corresponding videos from roadside cameras to calibrate its time, location, severity and other information. In this study, single crashes and crashes caused by vehicle fires, damage, collisions with thrown objects were excluded as they were almost caused by drivers' wrong operations (e.g., distraction and fatigue). Finally, 409 two-vehicle and multi-vehicle crashes were included.

2) Traffic Flow Data: The Traffic flow data were collected from the 42 sets of roadside millimeter-wave radars with an average spacing of 250 m. The detected traffic flow features include the average traffic volume (vehicles/lane), speed (km/h) of each lane, and the number of big vehicles (e.g., heavy trucks and buses). The average collection accuracy can reach 90%, and the collection time interval is 1 minute.

3) Risky Driving Behavior Data: The risky driving behavior data ware provided by the AutoNavi Software Co., Ltd, which has more than 632 million monthly active users in China. The vehicle driving information such as speed, acceleration and driving angle were collected from the drivers' smartphone sensors and the risky driving behaviors were identified based on specific thresholds and rules [8], [36]:

Sharp left/right turn (type 1) & left/right-lane change (type 2): In the case of a phone staying in a certain position, the centripetal force of the original historical turn is judged. If the detection angle is greater than a certain threshold, it is identified as a sharp turn or a sharp lane-change.

Sharp acceleration (type 3) & brake (type 4): If the linear acceleration is greater than a certain threshold while the detection angle is less than the abovementioned threshold, a sharp acceleration or brake will be recognized.

For commercial reasons, the threshold applied by AutoNavi Software Co, could not be obtained and published. The information of speed, acceleration, time, and location when the behavior occurs will be also recorded. According to the AutoNavi, the overall identification accuracy of risky driving behaviors reaches 95%. Meanwhile, the spatial positioning error is less than 5m. However, due to the privacy issues, the vehicle and driver information corresponding to this behavior

TABLE I
MAIN INDICATORS IN RISKY DRIVING BEHAVIOR (RDB) DATA

Indicators	Description	Unit
Lat	The latitude of the RDB location	-
Lon	The longitude of the RDB location	-
Direction	The traffic flow direction of roadway where RDBs occurred: 1- South bound 2- North bound	-
Type	The type of the RDBs: 1- Sharp left/right turn 2- Sharp left/right-lane change 3- Sharp acceleration 4- Sharp brake	-
Duration	The duration of the RDBs	s
Ma	The maximum acceleration of the RDBs	g=9.8m/s^2
Ms	The maximum speed of the RDBs	m/s

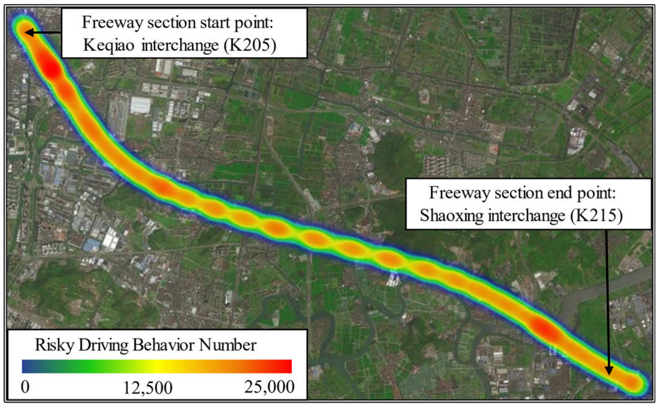

Freeway section start point:
Keqiao interchange (K205)
Freeway section end point:
Shaoxing interchange (K215)
Risky Driving Behavior Number
0
12,500
25,000

Fig. 1. Risky driving behavior distribution on studied freeway.

cannot be recorded. A total of 4 types of risky driving behaviors were included, and the main indicators in the risky driving behavior data are shown in Table I. Based on such data, the risky driving behaviors that occurred on the studied freeway segment can be comprehensively obtained. Finally, a total of 238,831 risky driving behaviors that occurred on this road during this year were recorded as shown in Fig. 1.

B. Data Processing

To extract the pre-crash traffic flow and risky driving behaviors, the temporal-spatial matching rule was set referring to existing studies [5], [10], [12]. Meanwhile, the temporal-spatial distribution of pre-crash risky driving behaviors was also checked to refine the matching range. As shown in Fig. 2, for the spatial range, the distance of one basic road section was set to 1km to avoid the highly similar traffic data of adjacent radars within short distances (e.g., 250m and 500m). Therefore, each crash was matched to its nearest radar and Crash Section was set from the location of the nearest upstream radar to 1km downstream. Data for its 1km upstream and downstream were also collected. As for the temporal range, the recording crash time was denoted as zero time. A total of 25-min interval from -5 to -30 min was set as the matching temporal period. The data between -5 to 0 min need to be discarded to avoid the crash time record deviation.

The macro traffic flow data from radars were aggregated into five 5-minutes interval. Six traffic parameters were calculated including the Average and Standard deviation of Speed (AS, SS), Speed Difference among each lane (DS), Average and Standard deviation of Flow (AF, SF), and the Ratio of heavy vehicles (HR). Therefore, 90 (i.e., 6 parameters * 3 sections *5 time slices) aggregated traffic flow variables were derived. For example, ASC1 means the average speed of crash section in -5 to -10 min.

As for the micro risky driving behaviors, they are believed to cause the traffic flow fluctuations and crashes [7], [9], [10]. Thus, pre-crash risky driving behaviors that occurred on the road were extracted and compared with that in non-crash scenarios to explore their relationships with crash risks. For the non-aggregation of behavioral data, a novel method was conducted to directly extract their space, time and kinetics features. Specifically, if several risky driving behaviors occurred on the road section before a crash as shown in Fig. 3. Existing methods only calculate aggregate features of such behaviors (e.g., frequency, average speed, etc.) [7], [9], [21]. While in our method, each behavior would be recorded as a piece of data, including its time, space, type, speed, acceleration (Acc), and duration as shown in Fig. 3. Given such risky driving behaviors occurred one after another before the same crash, they would be matched with the same Crash ID. Meanwhile, the risky driving behaviors may be recorded repeatedly if several people use AutoNavi Software in the same vehicle. Thus, data double-checking had been done and no duplicate risky driving behavior records were found in the modeling dataset.

As for the non-crash sample extraction, 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 of studies [10], [12], considering the same time of day, day of week, and location but different weeks (2 weeks before and after crash). For example, if a crash occurred on April 22, 2022, at 10:35 a.m. on K210+500, abnormal driving event data from 10:00 to 10:30 a.m. (30-min interval) on K209+000 to K212+000 were extracted as the crash sample (Crash=1). The corresponding non-crash samples (Crash=0) could be collected from the same section in the same period of April 8, 15, 29, and May 6.

As the nearby radar or risky driving behavior data had some malfunctions or distortions, 145 crashes cannot be matched with traffic data and therefore eliminated. Finally, after the data filtering and temporal-spatial matching, 264 crash and 1056 non-crash cases were extracted as the modeling dataset. 70% of them was used for model training and the remaining 30% of dataset was utilized to test the model performance.

IV. METHODOLOGY

A. Overall Framework of Crash Risk Evaluation Model

An LSTM + Transformer method was proposed to combine the macro traffic flow and micro risky driving behavior features for crash risk evaluation. As shown in Fig. 4, the overall model framework includes two parts: the traffic flow variables

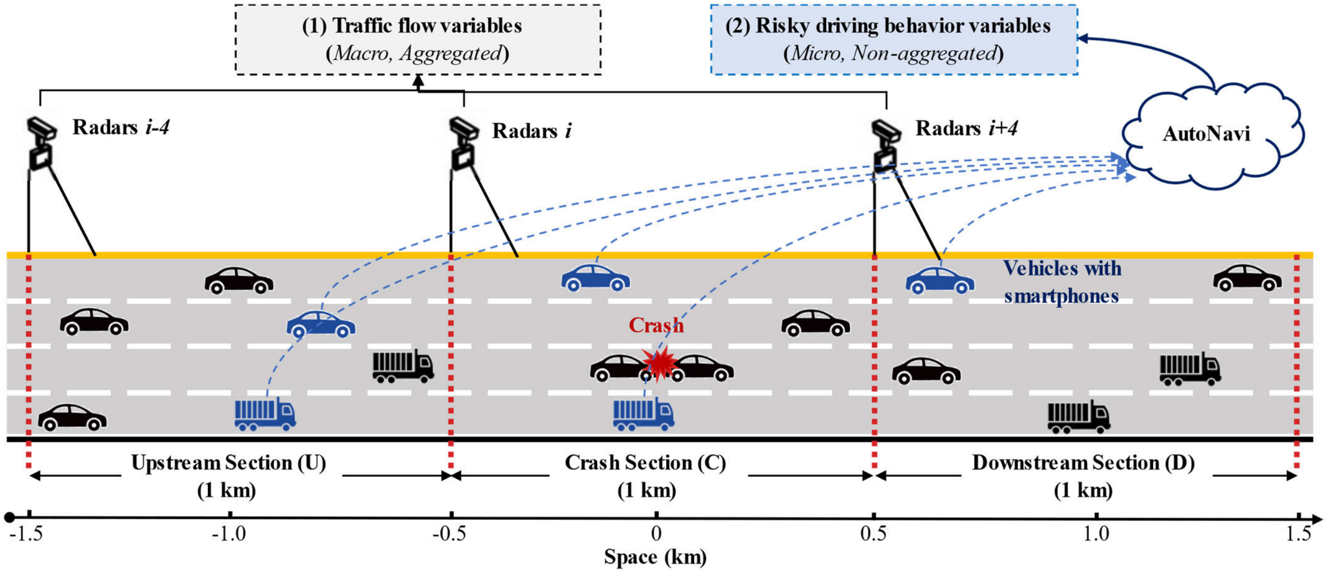
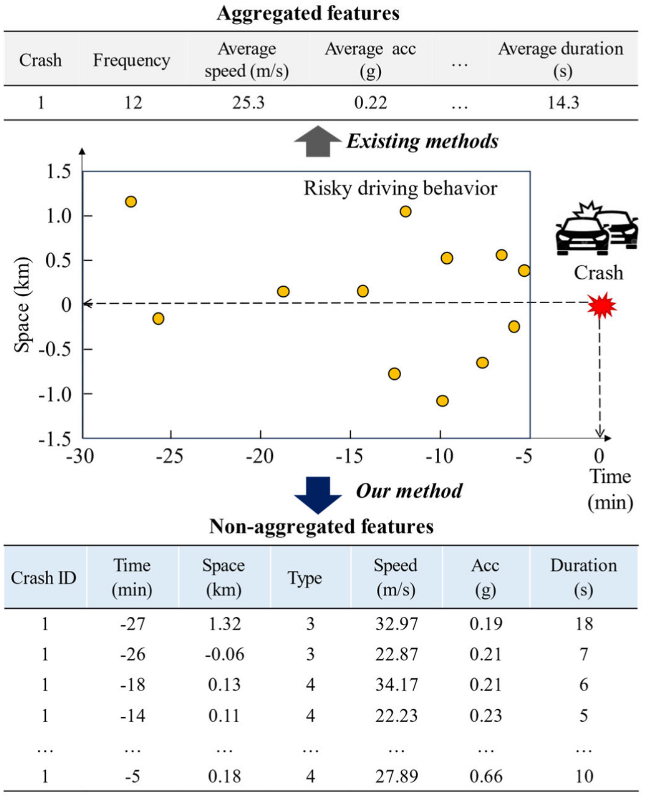


Fig. 2. The temporal-spatial matching rule of crashes, traffic flow data, and risky driving behaviors.


Aggregated features

Crash	Frequency	Average speed (m/s)	Average acc (g)	…	Average duration (s)
1	12	25.3	0.22	…	14.3

Non-aggregated features

Crash ID	Time (min)	Space (km)	Type	Speed (m/s)	Acc (g)	Duration (s)
1	-27	1.32	3	32.97	0.19	18
1	-26	-0.06	3	22.87	0.21	7
1	-18	0.13	4	34.17	0.21	6
1	-14	0.11	4	22.23	0.23	5
…	…	…	…	…	…	…
1	-5	0.18	4	27.89	0.66	10

Fig. 3. Comparation of risky driving behaviors features extraction between existing and our methods.

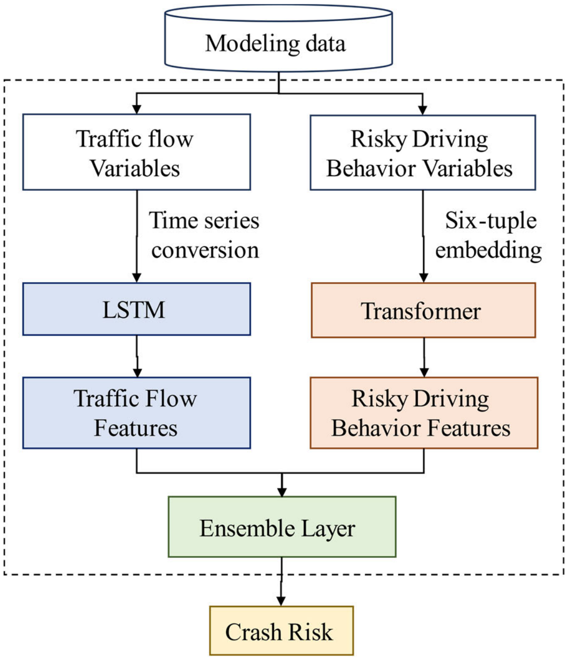


Fig. 4. Overall framework of LSTM + Transformer model.

would be converted into the time-series tensors and input into an LSTM network [40], [51]. Then the LSTM will learn the temporal characteristics of different traffic flow variables and output the extracted traffic flow features. While the risky driving behavior variables would be embedded into six-tuples which record their space, time, and kinetic information. These tuples are then input into a Transformer network [52] to learn their non-aggregated behavioral characteristics and generate the abstract risky driving behavior features. Finally, an ensemble layer [53] is introduced to fuse the two feature vectors based on specific weights and predict the final crash risk. It is worth noting that during the model training process, the LSTM and Transformer networks are trained in parallel, and the weight of the ensemble layer can be dynamically adjusted according to the importance of features to achieve optimal model performance.

B. Long Short-Term Memory Network

The LSTM network was utilized to extract the temporal characteristics of traffic flow variables in this paper. Compared with traditional neuron nodes, LSTM can use the memory

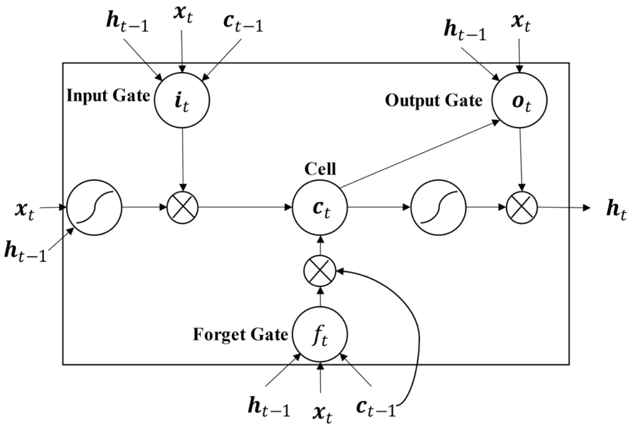


Fig. 5. LSTM unit structure.

blocks in hidden layers to capture the sequential information. In view of the advantages of LSTM in temporal feature extraction, it has been applied in recent crash risk evaluation studies and achieve good model accuracy [39].

To input the traffic flow data into the LSTM network, such aggregated traffic flow variables were first converted into a time-series vector $\{(ASC_i, SSC_i, \ldots, HRD_i)\}_{i=1}^{5}$, where ASC, SSC, ..., HRD are the 18 (i.e., 6 parameters $*$ 3 sections) traffic variables, and i is the corresponding time window index from 1 (-5 to -10 min) to 5 (-25 to -30 min). At each time step, the LSTM network has several self-connected LSTM units as shown in Fig.5. Each unit is composed of input gate $\boldsymbol{i}_t$, forget gates $\boldsymbol{f}_t$, memory cell $\boldsymbol{c}_t$, output gate $\boldsymbol{o}_t$, and hidden state $\boldsymbol{h}_t$. The hidden state in the current LSTM layer is used as input for the next layer. Taking the sequence vectors $\boldsymbol{X} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n\}$ as inputs, the LSTM network could generate an output temporal feature vector $\boldsymbol{e}^L$=$\{\boldsymbol{e}_1$,$\boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ though calculating the unit activations using the following equations, iterated from $t = 1, \ldots, n$:

$$\boldsymbol{i}_t = \sigma(\boldsymbol{W}_{xi}\boldsymbol{X}_t + \boldsymbol{W}_{hi}\boldsymbol{h}_{t-1} + \boldsymbol{W}_{ci}\boldsymbol{c}_{t-1} + \boldsymbol{b}_i) \tag{1}$$

$$\boldsymbol{f}_t = \sigma(\boldsymbol{W}_{xf}x_t + \boldsymbol{W}_{hf}\boldsymbol{h}_{t-1} + \boldsymbol{W}_{cf}\boldsymbol{c}_{t-1} + \boldsymbol{b}_f) \tag{2}$$

$$\boldsymbol{c}_t = \boldsymbol{f}_t\boldsymbol{c}_{t-1} + \boldsymbol{i}_t tanh(\boldsymbol{W}_{xc}\boldsymbol{X}_t + \boldsymbol{W}_{hc}\boldsymbol{h}_{t-1} + \boldsymbol{b}_c) \tag{3}$$

$$\boldsymbol{o}_t = \sigma(\boldsymbol{W}_{xo}\boldsymbol{X}_t + \boldsymbol{W}_{ho}\boldsymbol{h}_{t-1} + \boldsymbol{W}_{co}\boldsymbol{c}_t + \boldsymbol{b}_o) \tag{4}$$

$$\boldsymbol{h}_t = \boldsymbol{o}_t \tanh(\boldsymbol{c}_t) \tag{5}$$

$$\boldsymbol{e}_t = \boldsymbol{W}_{hy}\boldsymbol{h}_t + \boldsymbol{b}_y \tag{6}$$

where σ is the sigmoid function, $\boldsymbol{W}$ is the weight matrices, for example, $\boldsymbol{W}_{xi}$ denotes weight matrix from the input gate to the input.

C. Transformer Network

In this paper, a Transformer model [48], [52] was proposed to extract the non-aggregate risky driving behavior features. Formally, a risky driving behavior can be defined as a six-variable vector (t, s, f, vs, va, vd), which respectively represent its time, space, type, value of speed, acceleration, and duration. Therefore, the risky driving behaviors in each sample can be form a time-series vector $\boldsymbol{T} = \{(t_i, s_i, f_i, vs_i, va_i, vd_i)\}_{i=1}^{k}$, where k is the number of risky driving behaviors in each crash and non-crash scenario. It worth noting that k would vary among different samples due to the randomness of the risky driving behaviors on the roads. Finally, the modeling dataset is defined as $\boldsymbol{D} = \{\boldsymbol{T}_1, \boldsymbol{T}_2, \ldots, \boldsymbol{T}_N\}$ with N observed samples, where the jth sample contains a risky driving behavior vector $\boldsymbol{T}_j$.

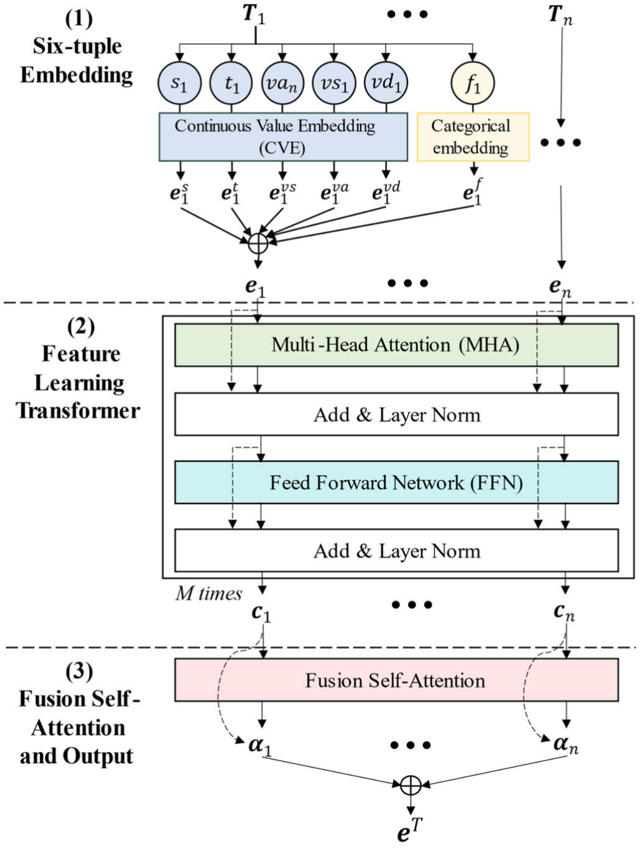


Fig. 6. The architecture of the transformer model.

The architecture of the Transformer network is illustrated in Fig.6. First, each risky driving behavior vector is embedded by the Six-tuple Embedding module. The initial embeddings are then fed into a Feature Learning Transformer module to extract their spatial-temporal and severity features. Finally, in the Fusion Self-attention and Output module, these features are combined via self-attention mechanism to provide the final risky driving behavior features.

1) Six-Tuple Embedding: Given the time-series vector T, the initial embedding for the ith vector $\boldsymbol{e_i}$ is calculated by adding together the following component embeddings: $\boldsymbol{e}_i = \boldsymbol{e}_i^t + \boldsymbol{e}_i^s + \boldsymbol{e}_i^f + \boldsymbol{e}_i^{vs} + \boldsymbol{e}_i^{va} + \boldsymbol{e}_i^{vd}$. Sine the behavior type variable f_i is a categorical object (e.g., Acceleration, and Brake), embeddings $\boldsymbol{e}_i^f(\cdot)$ are obtained from a lookup table like word embeddings. While the other continuous variables such as time and space, they are embedded by the Continuous Value Embedding (CVE) approach [48]. In CVE, each variable is mapped to a multidimensional vector by a feed-forward network (FFN) such as $\boldsymbol{e}_i^t = FFN^T(t_i)$. Each FFN has one input neuron, d output neurons, a single hidden layer with $\lfloor\sqrt{d}\rfloor$ neurons, and the $tanh(\cdot)$ activation. They follow the form:

$$FFN(\boldsymbol{x}) = Utanh(W\boldsymbol{x} + b) \tag{7}$$

where W represents the weights of inputs, b is the bias factor, and U is the overall weights for FFN layer.

2) Feature Learning Transformer: Six-tuple embeddings $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ are processed through a transformer network, which consists of M blocks. Each block contains a Multi-Head Attention (MHA) layer with h attention heads and a FFN with a hidden layer. In this network, every block receives n input embeddings $\boldsymbol{E}$ and generates n corresponding output embeddings $\boldsymbol{C}$**.** The MHA layer use multiple attention heads to focus on information contained in different embedding projections. The computations of the MHA are detailed as follows:

$$\boldsymbol{H}_j = softmax(\frac{\left(\boldsymbol{E}\boldsymbol{W}_j^q\right)\left(\boldsymbol{E}\boldsymbol{W}_j^k\right)^T}{\sqrt{d/h}})(\boldsymbol{E}\boldsymbol{W}_j^v) \quad j = 1, \ldots, h \tag{8}$$

where $\boldsymbol{W}_j^q$, $\boldsymbol{W}_j^k$, $\boldsymbol{W}_j^v$ are the query, key, and value weights in each attention heads. Then the outputs of all heads are concatenated and projected to their original dimensionality using the weights $\boldsymbol{W}_C$:

$$MHA(\boldsymbol{E}) = (\boldsymbol{H}_1 \circ \ldots \circ \boldsymbol{H}_H)\boldsymbol{W}_C \tag{9}$$

And the FFN layer takes the form:

$$F\left(X\right) = ReLU\left(\boldsymbol{X}\boldsymbol{W}_1^f + \boldsymbol{b}_1^f\right)\boldsymbol{W}_2^f + \boldsymbol{b}_2^f \tag{10}$$

where $\boldsymbol{W}_1^f$, $\boldsymbol{b}_1^f$, $\boldsymbol{W}_2^f$, $\boldsymbol{b}_2^f$ are different FFN weights.

During network training, dropout and layer normalization techniques are added for every MHA and FFN layer to avoid model overfitting. The output generated by each block is fed as the input to its subsequent block. Therefore, the output of the last block provides the final feature embeddings $\boldsymbol{C} = \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n\}$.

3) Fusion Self-Attention and Output: A self-attention Layer is used to fuse the feature embeddings $\boldsymbol{C} = \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n\}$ to compute the final feature embedding $\boldsymbol{e}^{\boldsymbol{T}}$. In this layer, attention weights $\{\alpha_1, \ldots, \alpha_n\}$ are first computed by passing $\boldsymbol{C}$ through an FFN. Then, a SoftMax function is applied across all the FFN outputs:

$$a_i = \boldsymbol{u}_a^T tanh(\boldsymbol{W}_a \boldsymbol{c}_i + \boldsymbol{b}_a) \tag{11}$$

$$\alpha_i = \frac{exp(a_i)}{\sum_{j=1}^n exp(a_j)} \quad \forall i = 1, \ldots, n \tag{12}$$

$$\boldsymbol{e}^T = \sum\nolimits_{\boldsymbol{i=1}}^{\boldsymbol{n}} \alpha_i \boldsymbol{c}_i \tag{13}$$

where $\boldsymbol{W}_a$, $\boldsymbol{b}_a$, $\boldsymbol{u}_a^T$ are weights of the self-attention layer.

D. Ensemble Layer and Evaluation

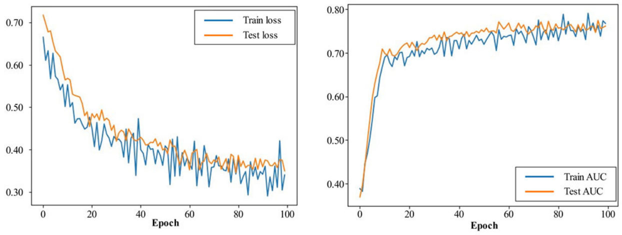
An ensemble layer is established to combine the extracted traffic flow features $\boldsymbol{e}^L$ from LSTM and risky driving behavior features $\boldsymbol{e}^T$ from Transformer into the final evaluation features $\boldsymbol{e}^{all}$ with different weights:

$$\boldsymbol{e}^{all} = W_L \boldsymbol{e}^L \oplus W_T \boldsymbol{e}^T \tag{14}$$

where W_L, W_T are weights for the traffic flow and risky driving behavior features respectively with initial values of 0.5, 0.5. They are set to be trainable to adjust their values according to the model loss. Therefore, the ensemble layer can gradually get the optimal balance of the importance of two kinds of features.

TABLE II
THE CONFUSION MATRIX

		Prediction	
		Crash	**Noncrash**
True	**Crash**	True positive (TP)	False Negative (FN)
	Noncrash	False Positive (FP)	True Negative (TN)



(a) Loss (b) AUC

Fig. 7. Model indicators during training process.

Finally, the evaluation features are passed through a dense layer with sigmoid function to obtain the final crash risk $\tilde{y}$:

$$\tilde{y} = Sigmiod(\boldsymbol{W}_{all}\boldsymbol{e}^{all} + \boldsymbol{b}_{all}) \tag{15}$$

As for the model training, the cross-entropy loss and Adam optimizer were utilized to get the best model.

E. Evaluation Criteria and Other Baseline Models

To evaluate the model performance, accuracy (ACC), recall, false alarm rate (FAR), and the area under ROC curve (AUC) were used, which can be calculated based on Table II and (16)-(18). ACC is the ratio of correctly classified samples to all samples; A high recall and low FAR means the model can correctly classify most crash samples with low prediction errors; 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. Apart from the proposed method, the widely used Logistic Regression (LR), Support Vector Machine (SVM), and XGBoost models were also employed as baseline models for comparison.

$$\text{ACC} = (TP + TN)/(TP + TN + FP + FN) \tag{16}$$

$$Recall = TP/(TP + FN) \tag{17}$$

$$FAR = FP/(TN + FP) \tag{18}$$

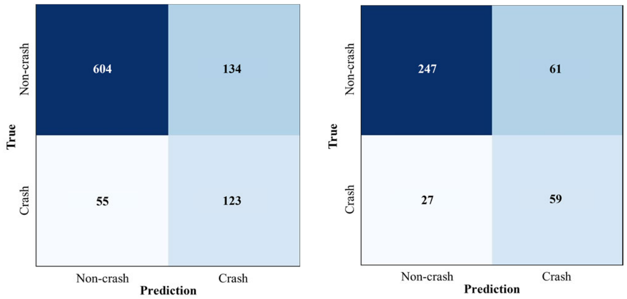
V. RESULTS

A. Crash Risk Evaluation Results

During the model training, grid searching was utilized to choose the best model hyperparameters and layer numbers for LSTM and Transformer. Finally, the optimal model hyperparameters are displayed in Table III.

TABLE III
THE OPTIMAL MODEL HYPERPARAMETERS SETTING

Hyperparameter		Value
LSTM	Number of LSTM layers	2
	Number of LSTM units in each layer	50
Transformer	Six-tuple embedding dimension	50
	Number of MHA layers	2
	Number of attention heads in MHA	4
Model training	Batch size	32
	Learning rate	5*e*-5
	Dropout rate	0.2



(a) Train dataset (b) Test dataset

Fig. 8. Confusion matrix of train and test dataset.

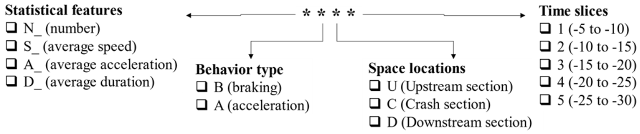


Fig. 9. Aggregate rules for risky driving behaviors.

Fig.7 illustrates the changes of main indicators (i.e., model loss and AUC) during the model training. In the early stage of training (<20 epochs), model loss dropped significantly, and AUC increased sharply from 0.40 to 0.70. After the epoch exceeds 80, the loss and AUC show no significant changes at 0.35 and 0.78, meaning that the model achieved convergence. During the whole training process, very similar losses and AUC were observed between the train and test datasets, showing that the model could gradually learn the relationship between the predictor variables and the targets without causing overfitting problem.

Fig.8 shows the final confusion matrix of crash evaluation results in both the train and test datasets. The best classification threshold was determined to keep the FAR within a reasonable range (near 0.2). In the train dataset, the model can correctly identify 604 non-crashes and 123 crashes. The ACC, recall, and FAR are 0.794, 0.691, and 0.182, respectively. In the test dataset, this model correctly classified 247 non-crashes and 59 crashes, achieving similar ACC (0.777), recall (0.686) and FAR (0.198). The above results show that the proposed LSTM + Transformer model can capture the macro traffic flow features and micro driving behavior characteristics, thereby sensitively identifying the high-risk traffic operating status and risky driving behaviors and achieving good crash risk evaluation performances.

TABLE IV
MODEL CRITERIA OF FIVE MODELS

Model	Model Evaluation Criteria			
	ACC	Recall	FAR	AUC
Aggregated modeling methods				
LR	0.711	0.523	0.237	0.677
SVM	0.739	0.616	0.227	0.732
XGBoost	0.741	0.605	0.221	0.776
LSTM	0.759	0.628	0.205	0.778
Average	*0.737*	*0.593*	*0.222*	*0.741*
Non-aggregated modeling methods				
LSTM + Transformer	**0.777**	**0.686**	**0.198**	**0.785**
Improvement	*+5.34%*	*+15.69%*	*-10.95%*	*+5.97%*

B. Models Performance Comparation

For comparison, some aggregated modeling methods in existing studies [7], [9], [21] were also developed. Given that such models need the fixed-length input variables for each sample, the non-aggregated data structure of risky driving behaviors would not be suitable. Thus, their aggregated variables were extracted as shown in Fig.9. Specifically, the risky driving behavior data were aggregated into 5-minute time slices and U/C/D sections following the same space-time windows as the traffic flow variables. According to their type, the number (N_), average speed (S_), average acceleration (A_), and average duration (D_) of risky driving behaviors were calculated. Therefore, 4*2*3*5 = 120 aggregated variables were obtained. For example, "N_BC1" means the number of sharp braking in the crash section in the first 5-minute time slices. Finally, a total of 210 variables (120 risky driving behavior variables + 90 traffic flow variables) were used to establish the aggregated models.

The final model evaluation criteria on the test datasets are listed in Table IV. To ensure the result credibility, those criteria are the average values in a 5-fold test experiment. The proposed LSTM + Transformer model achieves the best model performance with highest ACC (0.777), recall (0.686), AUC (0.785) and lowest FAR (0.198). In the aggregated models, the LSTM and XGBoost models show better ACC (0.759, 0.741) and AUC (0.778, 0.776). The worst model is the LR model with the highest FAR (0.237) and the lowest ACC (0.711) and AUC (0.677). Compared to the average criteria of aggregated models, the ACC, recall, and AUC are improved at 5.34%, 15.69%, and 5.97% respectively in the proposed model. Also, the FAR decreases from 0.222 to 0.198. The above results show that:

1) The proposed LSTM + Transformer model can combine both the features of macro traffic flow status and micro risky driving behaviors to achieve high-accuracy crash risk evaluation.
2) Compared with existing models, the Transformer in the proposed model can directly extract the non-aggregated features of behaviors. Therefore, the spatial distributions, temporal varying patterns and other more precise

TABLE V
RESULTS OF MODELS USING DIFFERENT DATASETS

Data (Model)	Model Evaluation Criteria			
	ACC	Recall	FAR	AUC
Traffic flow dataset (LSTM)	0.756	0.570	**0.192**	0.746
Risky driving dataset (Transformer)	0.754	0.605	0.205	0.762
Both datasets (LSTM + Transformer)	**0.777**	**0.686**	0.198	**0.785**

behavioral features can be extracted, thus significantly improving the crash risk evaluation performance.

Furthermore, two models using only traffic flow dataset or risky driving behavior dataset were constructed to compare the impact of different kind of data on crash risk evaluation. Among them, an LSTM model was trained to establish the relationship between aggregated traffic flow data and crashes. A Transformer model was developed to extract the non-aggregated risky driving behavior features and predict crash. Their network structures were set to be consistent with the corresponding parts of the proposed LSTM + Transformer.

Table V compares their model results on the test datasets. The model using both two datasets has the best model evaluation performance, which is consistent with existing studies [9], [21]. The reason may be that this model can combine the features of both traffic flow and risky driving behaviors to better capture the traffic flow turbulence and high-risk driving behavior anomalies on the roadways. Compared to traffic flow data, the model based on risky driving behavior shows better recall (0.605) and AUC (0.762), indicating that the use of risky driving behavior data is benefit of capturing the pre-crash traffic flow turbulence. The above advantages may be due that the risky driving behaviors are distributed through roadways and can be quite close to crash points, therefore precisely representing the traffic flow disorders from drivers' behaviors. Instead, the traffic flow data from roadside sensors can only provide macroscopic traffic flow characteristics.

C. Crash Risk Influence Variables Explanation

The explanation of the crash risk evaluation model is prominent for crash prevention and active traffic management. To analyze the influences of the traffic flow and risky driving behavior features to crash risk, a SHapley Additive exPlanations (SHAP) approach [54] was employed to explain the results of LSTM + Transformer model. As the SHAP framework cannot handle the non-aggregated risky driving behavior data structure (i.e., each sample contains multiple behavior data), their corresponding aggregated variables was used to replace them for model explanation.

As the results show that the impact of variables in the time slice of -5 to -10 minutes are significantly higher than that of other periods, the importance rank of the top 10 variables during that time slice is illustrated in Fig. 10. Among them, most variables are risky driving behavior variables, indicating that road risky driving behaviors have significant impacts on crashes. The most important variable is the average speed of crash section (ASC). The average traffic flow of crash section (AFC), the number of braking on crash section (N_BC), the speed standard deviation of crash section (SSC), and the duration of braking on upstream section (D_BU) are also very important to the crash risk.

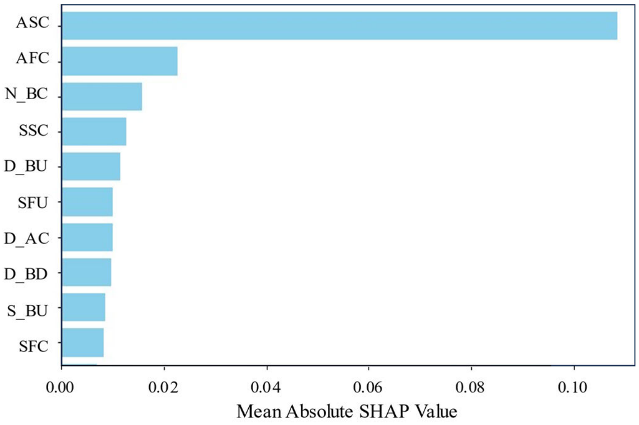


Fig. 10. The top-10 important crash risk influence variables.

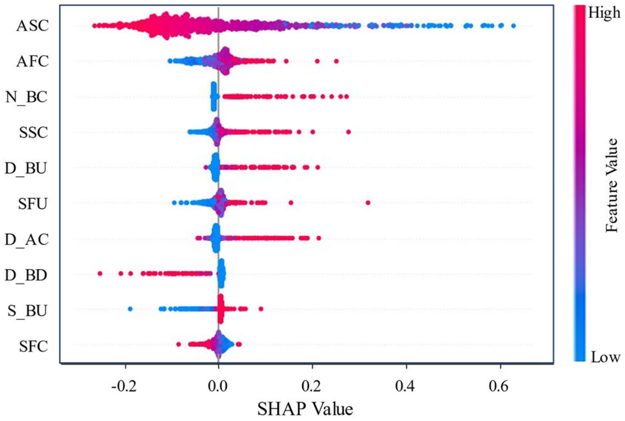


Fig. 11. The SHAP values of the top-10 important variables.

Fig. 11 shows the detailed SHAP value distributions of the top 10 variables to analyze their effect on real-time crash risk. Several interesting results can be found:

1) For the average speed of crash section (ASC), the red feature value points concentrate on the negative SHAP value portion, meaning that the contribution to the crash being true increases as the average speed value on crash section reduces, which is consistent with existing studies [14], [19]. It indicates that a low road speed, which may be related to traffic congestion or slow-flow traffic, is more likely to cause crashes in the freeway.
2) For the average traffic flow of crash section (AFC) and speed standard deviation of crash section (SSC), their red feature value points concentrate on the positive SHAP value portion, indicating that they are highly positive with crash risk. As road traffic flow increases, high-density traffic flow may cause speed slowdowns and vehicle space distance reduction, thereby leading to crashes. While the high standard deviation of speed means high traffic fluctuations, which are typical pre-crash precursors need to be prevented [5], [19], [38].

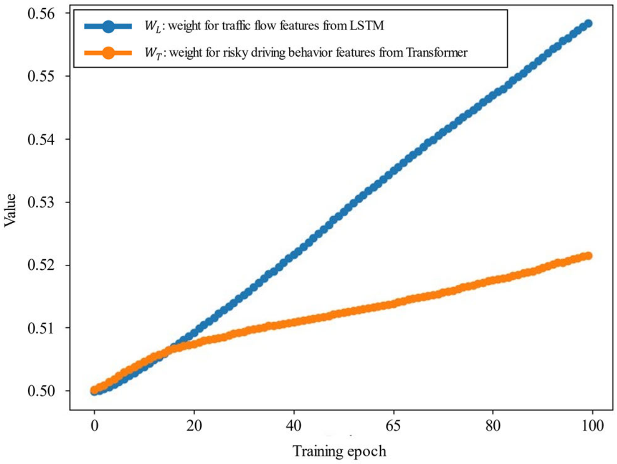


Fig. 12. Ensemble weights during model training.

3) As for the variables related to risky driving behavior, the red data points representing the number of risky braking on the crash section (N_BC) predominantly exhibit positive SHAP values. This indicates a significant positive relationship between this variable and crash risk. In other words, a higher frequency of risky braking in the crash section significantly impacts the traffic flow, thereby increasing the likelihood of crashes. The conclusions outlined above align with findings from existing researches [7], [8], [9].
4) Meanwhile, the duration of risky braking in the upstream section (D_BU) and the duration of risky accelerations in the crash section (D_AC) both exhibit a positive impact on crash risk. The high data points for these variables are predominantly located in the positive SHAP value range. This reveals that if the duration of sharp braking and acceleration behaviors on upstream and crash sections persists longer, the more surrounding vehicles and traffic flow would be affected, consequently heightening the crash risk.

Overall, the ensemble weights for traffic flow and risky driving behavior features were also updated with their variable importance during model training as shown in Fig. 12. The results shows that the weight for traffic flow features become 0.56 from 0.50, which is higher than that of risky driving behavior features. It means that the traffic flow features become more important to the crash risk evaluations, which is consistent with the variable importance ranking results. For instance, the traffic flow variables (e.g., ASC, AFC) are more important among all variables. Therefore, their corresponding model weight would be relatively higher.

To further verify the effectiveness of the proposed model in analyzing the relationship between risky driving behaviors and crashes, the distributions of risky driving behaviors in typical real-world crash and non-crash scenarios are shown in Fig. 13. It can be seen that:

1) In the crash accident scenarios, risky driving behaviors on the studied road increased significantly before each crash and lasted for a relatively long time. Hard braking and accelerations were concentrated near the crash point with high frequency, eventually leading to the occurrence of crashes.
2) While in the normal scenarios, although several risky driving behaviors happened on the roads, these behaviors were low in frequency with short duration. Temporal-spatial clustering of such behaviors could not be observed.

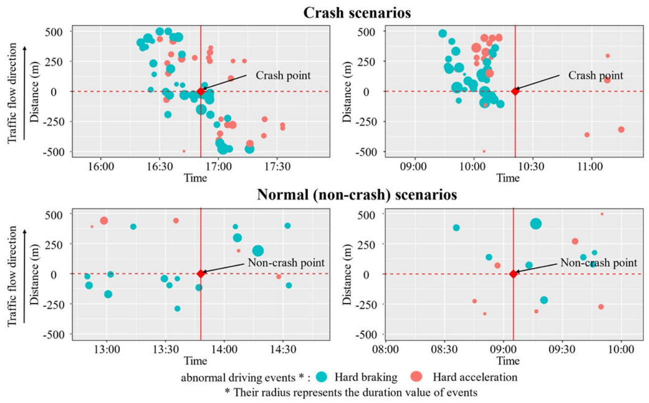


Fig. 13. Distributions of risky driving behaviors in real-world crash and non-crash scenarios.

The above results confirmed the consistency between actual observations and the analytical results of the proposed model. It shows that the proposed Transformer model can capture the frequency, duration and other characteristics of risky driving behaviors and establish their association with the crash risks on the road.

VI. Discussion and Conclusion

With the recent development of emerging CV and driving status monitoring technology, it is now much easier to obtain vehicle-based risky driving behavior data. Compared with the traditional fixed-detector data, the significant advantage of such data is that they can be flexibly obtained from vehicles without the installation location limitation, which can distribute throughout the entire roadway space to sensitively represent driving behavior characteristics and traffic status.

Leveraging the above advantage, this paper has utilized both the macro traffic flow and micro risky driving behavior data to evaluate crash risk and explore their impact on crashes. To address the research gaps in existing studies of ignoring the non-aggregated natures of risky driving behaviors and poor capability in feature combining, an LSTM + Transformer ensemble framework was developed to evaluate crash risk. The proposed Transformer could extract the non-aggregated spatial-temporal features of risky driving behaviors and LSTM could learn temporal varying patterns of pre-crash traffic flow. The ensemble framework of the two models can combine the extracted macro traffic flow and micro risky driving behavior characteristics and automatically adjust the feature weights to achieve the best crash risk evaluation performance.

Data from a Chinese freeway was used for empirical analysis. Based on the experiment results, main conclusions of the paper can be summarized as:

1) The proposed LSTM + Transformer model had achieved improvements of ACC (5.34%), recall (15.69%), AUC

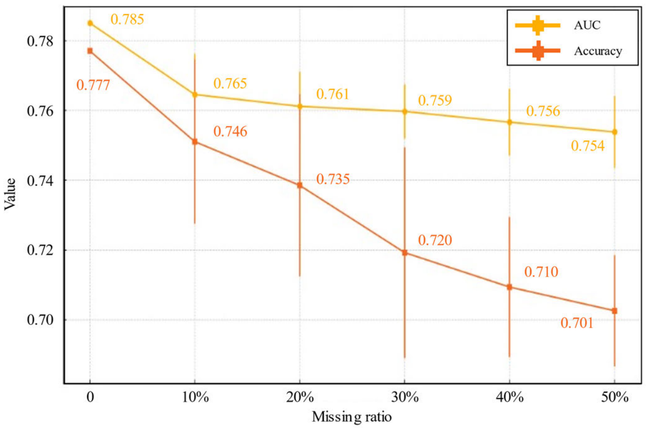


Fig. 14. Model performances under different missing ratios.

(5.97%), and FAR (10.95%) compared to the aggregated models, showing that the proposed model can extract both the temporal features of macro traffic flow status and the non-aggregated features of micro risky driving behaviors to improve the crash risk evaluation performance.

2) Compared to models using traffic flow data, the model based on risky driving behavior demonstrates superior recall and AUC performance. This indicates that the utilizing risky driving behavior data can more precisely capture the turbulence in traffic flow prior to a crash.
3) For macro traffic flow variables, the lower the road speed, the greater the traffic flow and speed fluctuations, the higher the crash risk. For micro risky driving behaviors, the more frequency, the greater duration and acceleration of risky braking, the more surrounding vehicles would be affected, consequently heightening the crash risk.

In realistic application, roadside detectors always provide solid traffic flow data. However, the risky driving behavior data from phones or connected vehicles may be less due to their low penetration. Therefore, the model robustness to low penetration of risky driving behavior data is investigated based on a set of ablation experiments. Specifically, different ratio (aka., missing rate) of risky driving behaviors were randomly deleted from 0, 10%, 20%, to 50% in each sample to simulate several data missing scenarios. Considering the randomness in the deletion operations, five experiments were repeated with different random seeds at each missing rate. Fig. 14 is the results of the ablation experiments, which shows the model AUC and accuracy across different missing rates. With the substantial increase in the missing rate, the AUC decreases from 0.785 to 0.754, the accuracy reduces from 0.777 to 0.701. However, these model metrics are still higher than 0.7, showing that the model performance can still keep relatively high even if half of risky driving behavior data is missing. Therefore, although there will be some accuracy loss, the proposed LSTM + Transformer model is still robust to the data missing situation to keep reliable crash risk evaluation.

The results proved the advantage of using risky driving behaviors to expand the crash risk evaluation performance. With the proposed crash risk evaluation model, the study can be used to implement the safety potential prediction components in PTM. Meanwhile, several proactive safety countermeasures at both active traffic management and CV driving assistant and control can be proposed for real-world traffic management to prevent crashes:

1) For active traffic management, our method finds that a low road average speed and high speed standard deviation are more likely to cause crashes in the freeway. Therefore, speed harmonization control such as variable speed limits and ramp metering can be implemented on the high crash risk roads, which have been proven to effectively reduce temporal and spatial variations of traffic speed, therefore significantly reduce the crash risk and improve road safety [55], [56], [57].
2) For CV driving assistant and control, our methods reveals that the roads with frequent risky braking and long-duration braking/accelerations have high crash risks. Therefore, in-vehicle decision support systems like crash warning can be delivered to the approaching vehicles to extend their reaction times and reduce their risky driving behaviors [58], [59], [60]. Moreover, with the development of connected and autonomous vehicle (CAV) technology, we can develop CAV control strategies to avoid risky braking and keep smooth driving interventions during their decision-making processes. Existing studies have shown that such driving support and control strategies would be more effectively to improve the drivers' behaviors and prevent crashes [61], [62], [63].

Nonetheless, there are still a few limitations in the current study. Firstly, the used risky driving behavior data have only a penetration rate of 5-10%. A higher penetration could be considered in future research to improve model performance. Secondly, this study mainly established the relationship between traffic flow status, risky driving behaviors and crashes, other factors such as weather, lighting, and traffic composition that might have an impact on the crash were not included. Thirdly, more data can be collected to test the model transferability to other types of roadways. For instance, the ensemble weights for those two types of data may also change with different road environment, which need to be further investigated with more available data. And the temporal-spatial matching rules of risky driving behaviors need to be refined according to real-world data collection scenarios.

References

[1] (2023). *Global Status Report on Road Safety 2023*. Accessed: Dec. 21, 2023. [Online]. Available: https://www.who.int/teams/social-determinants-of-health/safety-and-mobility/global-status-report-on-road-safety-2023

[2] M. Abdel-Aty, A. Pande, and L. Hsia, "The concept of proactive traffic management for enhancing freeway safety and operation," *ITE J.*, vol. 80, no. 4, p. 34, 2010.

[3] A. Pande and M. Abdel-Aty, "A freeway safety strategy for advanced proactive traffic management," *J. Intell. Transp. Syst.*, vol. 9, no. 3, pp. 145–158, Jul. 2005, doi: 10.1080/15472450500183789.

[4] R. Yu, L. Han, and H. Zhang, "Trajectory data based freeway high-risk events prediction and its influencing factors analyses," *Accident Anal. Prevention*, vol. 154, May 2021, Art. no. 106085, doi: 10.1016/j.aap.2021.106085.

[5] M. Hossain, M. Abdel-Aty, M. A. Quddus, Y. Muromachi, and S. N. Sadeek, "Real-time crash prediction models: State-of-the-art, design pathways and ubiquitous requirements," *Accident Anal. Prevention*, vol. 124, pp. 66–84, Mar. 2019, doi: 10.1016/j.aap.2018.12.022.

[6] R. Yu, L. Han, M. Abdel-Aty, L. Wang, and Z. Zou, "Improving model robustness of traffic crash risk evaluation via adversarial mix-up under traffic flow fundamental diagram," *Accident Anal. Prevention*, vol. 194, Jan. 2024, Art. no. 107360, doi: 10.1016/j.aap.2023.107360.

[7] Y. Ma, J. Zhang, J. Lu, S. Chen, G. Xing, and R. Feng, "Prediction and analysis of likelihood of freeway crash occurrence considering risky driving behavior," *Accident Anal. Prevention*, vol. 192, Nov. 2023, Art. no. 107244, doi: 10.1016/j.aap.2023.107244.

[8] M. Guo et al., "A study of freeway crash risk prediction and interpretation based on risky driving behavior and traffic flow data," *Accident Anal. Prevention*, vol. 160, Sep. 2021, Art. no. 106328, doi: 10.1016/j.aap.2021.106328.

[9] S. Zhang and M. Abdel-Aty, "Real-time crash potential prediction on freeways using connected vehicle data," *Analytic Methods Accident Res.*, vol. 36, Dec. 2022, Art. no. 100239, doi: 10.1016/j.amar.2022.100239.

[10] L. Han, R. Yu, C. Wang, and M. Abdel-Aty, "Transformer-based modeling of abnormal driving events for freeway crash risk evaluation," *Transp. Res. C, Emerg. Technol.*, vol. 165, Aug. 2024, Art. no. 104727, doi: 10.1016/j.trc.2024.104727.

[11] M. Abdel-Aty, N. Uddin, A. Pande, M. F. Abdalla, and L. Hsia, "Predicting freeway crashes from loop detector data by matched case-control logistic regression," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1897, no. 1, pp. 88–95, Jan. 2004, doi: 10.3141/1897-12.

[12] R. Yu, Y. Wang, Z. Zou, and L. Wang, "Convolutional neural networks with refined loss functions for the real-time crash risk analysis," *Transp. Res. C, Emerg. Technol.*, vol. 119, Oct. 2020, Art. no. 102740, doi: 10.1016/j.trc.2020.102740.

[13] L. Wang, M. Abdel-Aty, Q. Shi, and J. Park, "Real-time crash prediction for expressway weaving segments," *Transp. Res. C, Emerg. Technol.*, vol. 61, pp. 1–10, Dec. 2015, doi: 10.1016/j.trc.2015.10.008.

[14] M. Abdel-Aty and L. Wang, "Implementation of variable speed limits to improve safety of congested expressway weaving segments in microsimulation," *Transp. Res. Proc.*, vol. 27, pp. 577–584, Jan. 2017, doi: 10.1016/j.trpro.2017.12.061.

[15] C. Xu, P. Liu, W. Wang, and Z. Li, "Identification of freeway crash-prone traffic conditions for traffic flow at different levels of service," *Transp. Res. A, Policy Pract.*, vol. 69, pp. 58–70, Nov. 2014, doi: 10.1016/j.tra.2014.08.011.

[16] M. M. Ahmed and M. A. Abdel-Aty, "The viability of using automatic vehicle identification data for real-time crash prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 459–468, Jun. 2012, doi: 10.1109/TITS.2011.2171052.

[17] C. Wang, C. Xu, and Y. Dai, "A crash prediction method based on bivariate extreme value theory and video-based vehicle trajectory data," *Accident Anal. Prevention*, vol. 123, pp. 365–373, Feb. 2019, doi: 10.1016/j.aap.2018.12.013.

[18] P. Li, M. Abdel-Aty, Q. Cai, and C. Yuan, "This paper has been handled by associate editor tony sze. The application of novel connected vehicles emulated data on real-time crash potential prediction for arterials," *Accident Anal. Prevention*, vol. 144, Sep. 2020, Art. no. 105658, doi: 10.1016/j.aap.2020.105658.

[19] S. Roshandel, Z. Zheng, and S. Washington, "Impact of real-time traffic characteristics on freeway crash occurrence: Systematic review and meta-analysis," *Accident Anal. Prevention*, vol. 79, pp. 198–211, Jun. 2015, doi: 10.1016/j.aap.2015.03.013.

[20] Y. Lian, G. Zhang, J. Lee, and H. Huang, "Review on big data applications in safety research of intelligent transportation systems and connected/automated vehicles," *Accident Anal. Prevention*, vol. 146, Oct. 2020, Art. no. 105711, doi: 10.1016/j.aap.2020.105711.

[21] M. Guo, X. Zhao, Y. Yao, C. Bi, and Y. Su, "Application of risky driving behavior in crash detection and analysis," *Phys. A, Stat. Mech. Appl.*, vol. 591, Apr. 2022, Art. no. 126808, doi: 10.1016/j.physa.2021.126808.

[22] B. M. T. H. Anik, Z. Islam, and M. Abdel-Aty, "InTformer: A time-embedded attention-based transformer for crash likelihood prediction at intersections using connected vehicle data," 2023, *arXiv:2307.03854*.

[23] A. Pirdavani et al., "Application of a rule-based approach in real-time crash risk prediction model development using loop detector data," *Traffic Injury Prevention*, vol. 16, no. 8, pp. 786–791, Nov. 2015, doi: 10.1080/15389588.2015.1017572.

[24] M. A. Abdel-Aty, H. M. Hassan, M. Ahmed, and A. S. Al-Ghamdi, "Real-time prediction of visibility related crashes," *Transp. Res. C, Emerg. Technol.*, vol. 24, pp. 288–298, Oct. 2012, doi: 10.1016/j.trc.2012.04.001.

[25] J. Yuan, M. Abdel-Aty, L. Wang, J. Lee, R. Yu, and X. Wang, "Utilizing Bluetooth and adaptive signal control data for real-time safety analysis on urban arterials," *Transp. Res. C, Emerg. Technol.*, vol. 97, pp. 114–127, Dec. 2018, doi: 10.1016/j.trc.2018.10.009.

[26] L. Wang, M. Abdel-Aty, and J. Lee, "Safety analytics for integrating crash frequency and real-time risk modeling for expressways," *Accident Anal. Prevention*, vol. 104, pp. 58–64, Jul. 2017, doi: 10.1016/j.aap.2017.04.009.

[27] Z. Zhang, Q. Nie, J. Liu, A. Hainen, N. Islam, and C. Yang, "Machine learning based real-time prediction of freeway crash risk using crowdsourced probe vehicle data," *J. Intell. Transp. Syst.*, vol. 28, no. 1, pp. 84–102, Jan. 2024, doi: 10.1080/15472450.2022.2106564.

[28] G. A. Bodvarsson, S. T. Muench, and Washington State Transportation Center. (Aug. 2010). *Effects of Loop Detector Installation on the Portland Cement Concrete Pavement Lifespan: Case Study on I-5*. Accessed: Dec. 1, 2023. [Online]. Available: https://rosap.ntl.bts.gov/view/dot/22405

[29] M. Bonera, B. Barabino, G. Yannis, and G. Maternini, "Network-wide road crash risk screening: A new framework," *Accident Anal. Prevention*, vol. 199, May 2024, Art. no. 107502, doi: 10.1016/j.aap.2024.107502.

[30] M. Bonera, B. Barabino, and G. Maternini, "Road network safety screening of county wide road network. The case of the Province of Brescia (Northern Italy)," in *Advances in Road Infrastructure and Mobility*, A. Akhnoukh, K. Kaloush, M. Elabyad, B. Halleman, N. Erian, S. Enmon II, and C. Henry, Eds., Cham, Switzerland: Springer, 2022, pp. 525–541, doi: 10.1007/978-3-030-79801-7_38.

[31] A. P. Afghari, M. M. Haque, and S. Washington, "Applying a joint model of crash count and crash severity to identify road segments with high risk of fatal and serious injury crashes," *Accident Anal. Prevention*, vol. 144, Sep. 2020, Art. no. 105615, doi: 10.1016/j.aap.2020.105615.

[32] J. Wang, T. Luo, and T. Fu, "Crash prediction based on traffic platoon characteristics using floating car trajectory data and the machine learning approach," *Accident Anal. Prevention*, vol. 133, Dec. 2019, Art. no. 105320, doi: 10.1016/j.aap.2019.105320.

[33] C. M. Day et al., "Detector-free optimization of traffic signal offsets with connected vehicle data," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2620, no. 1, pp. 54–68, Jan. 2017, doi: 10.3141/2620-06.

[34] T. Imkamon, P. Saensom, P. Tangamchit, and P. Pongpaibool, "Detection of hazardous driving behavior using fuzzy logic," in *Proc. 5th Int. Conf. Electr. Eng./Electron., Comput., Telecommun. Inf. Technol.*, May 2008, pp. 657–660, doi: 10.1109/ecticon.2008.4600519.

[35] Y. Ge, W. Qu, C. Jiang, F. Du, X. Sun, and K. Zhang, "The effect of stress and personality on dangerous driving behavior among Chinese drivers," *Accident Anal. Prevention*, vol. 73, pp. 34–40, Dec. 2014, doi: 10.1016/j.aap.2014.07.024.

[36] Y. Yao et al., "Development of urban road order index based on driving behavior and speed variation," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2673, no. 7, pp. 466–478, Jul. 2019, doi: 10.1177/0361198119853576.

[37] R. Yu and M. Abdel-Aty, "Multi-level Bayesian analyses for single- and multi-vehicle freeway crashes," *Accident Anal. Prevention*, vol. 58, pp. 97–105, Sep. 2013, doi: 10.1016/j.aap.2013.04.025.

[38] R. Yu and M. Abdel-Aty, "Utilizing support vector machine in real-time crash risk evaluation," *Accident Anal. Prevention*, vol. 51, pp. 252–259, Mar. 2013, doi: 10.1016/j.aap.2012.11.027.

[39] Y. Ali, F. Hussain, and M. M. Haque, "Advances, challenges, and future research needs in machine learning-based crash prediction models: A systematic review," *Accident Anal. Prevention*, vol. 194, Jan. 2024, Art. no. 107378, doi: 10.1016/j.aap.2023.107378.

[40] P. Li, M. Abdel-Aty, and J. Yuan, "Real-time crash risk prediction on arterials based on LSTM-CNN," *Accident Anal. Prevention*, vol. 135, Feb. 2020, Art. no. 105371, doi: 10.1016/j.aap.2019.105371.

[41] Y. Yang, Y. Yin, Y. Wang, R. Meng, and Z. Yuan, "Modeling of freeway real-time traffic crash risk based on dynamic traffic flow considering temporal effect difference," *J. Transp. Eng., Part A: Syst.*, vol. 149, no. 7, Jul. 2023, Art. no. 04023063, doi: 10.1061/jtepbs.teeng-7717.

[42] Z. Gao, J. Xu, R. Yu, and L. Han, "Utilizing angle-based outlier detection method with sliding window mechanism to identify real-time crash risk," *J. Transp. Saf. Secur.*, vol. 16, no. 2, pp. 157–174, Feb. 2024, doi: 10.1080/19439962.2023.2189762.

[43] L. Wenqi, L. Dongyu, and Y. Menghua, "A model of traffic accident prediction based on convolutional neural network," in *Proc. 2nd IEEE Int. Conf. Intell. Transp. Eng. (ICITE)*, Sep. 2017, pp. 198–202, doi: 10.1109/ICITE.2017.8056908.

[44] K. Yang, X. Wang, M. Quddus, and R. Yu, "Predicting real-time crash risk on urban expressways using recurrent neural network," *Presented at the Transp. Res. Board 98th Annu. Meeting Transp. Res. Board*, 2019. [Online]. Available: https://trid.trb.org/view/1573386

[45] J. Yuan, M. Abdel-Aty, Y. Gong, and Q. Cai, "Real-time crash risk prediction using long short-term memory recurrent neural network," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2673, no. 4, pp. 314–326, Apr. 2019, doi: 10.1177/0361198119840611.

[46] Y. Zheng, L. Han, J. Yu, and R. Yu, "Driving risk assessment under the connected vehicle environment: A CNN-LSTM modeling approach," *Digit. Transp. Saf.*, vol. 2, no. 3, pp. 211–219, 2023, doi: 10.48130/dts-2023-0017.

[47] R. Zhang et al., "High-risk event prone driver identification considering driving behavior temporal covariate shift," *Accident Anal. Prevention*, vol. 199, May 2024, Art. no. 107526, doi: 10.1016/j.aap.2024.107526.

[48] S. Tipirneni and C. K. Reddy, "Self-supervised transformer for sparse and irregularly sampled multivariate clinical time-series," *ACM Trans. Knowl. Discovery Data*, vol. 16, no. 6, pp. 1–17, Dec. 2022, doi: 10.1145/3516367.

[49] K. Lee et al., "Learning missing modal electronic health records with unified multi-modal data embedding and modality-aware attention," 2023, *arXiv:2305.02504*.

[50] L.-C. Chen et al., "Self-supervised learning-based general laboratory progress pretrained model for cardiovascular event detection," *IEEE J. Transl. Eng. Health Med.*, vol. 12, pp. 43–55, 2024, doi: 10.1109/JTEHM.2023.3307794.

[51] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[52] A. Vaswani et al., "Attention is all you need," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.* Curran Associates, 2017, pp. 1–11. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[53] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, "Ensemble deep learning: A review," *Eng. Appl. Artif. Intell.*, vol. 115, Oct. 2022, Art. no. 105151, doi: 10.1016/j.engappai.2022.105151.

[54] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.* Curran Associates, 2017, pp. 1–10. [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

[55] J. Ma et al., "Freeway speed harmonization," *IEEE Trans. Intell. Vehicles*, vol. 1, no. 1, pp. 78–89, Mar. 2016, doi: 10.1109/TIV.2016.2551540.

[56] Z. Cheng, J. Lu, and Y. Li, "Freeway crash risks evaluation by variable speed limit strategy using real-world traffic flow data," *Accident Anal. Prevention*, vol. 119, pp. 176–187, Oct. 2018, doi: 10.1016/j.aap.2018.07.009.

[57] T. Hasan, M. Abdel-Aty, and N. Mahmoud, "Freeway crash prediction models with variable speed limit/variable advisory speed," *J. Transp. Eng., A, Syst.*, vol. 149, no. 3, Mar. 2023, Art. no. 04022159, doi: 10.1061/jtepbs.teeng-7349.

[58] B. Ryder, B. Gahr, P. Egolf, A. Dahlinger, and F. Wortmann, "Preventing traffic accidents with in-vehicle decision support systems—The impact of accident hotspot warnings on driver behaviour," *Decis. Support Syst.*, vol. 99, pp. 64–74, Jul. 2017, doi: 10.1016/j.dss.2017.05.004.

[59] Y. Wu, M. Abdel-Aty, Q. Cai, J. Lee, and J. Park, "Developing an algorithm to assess the rear-end collision risk under fog conditions using real-time data," *Transp. Res. C, Emerg. Technol.*, vol. 87, pp. 11–25, Feb. 2018, doi: 10.1016/j.trc.2017.12.012.

[60] Z. Wang, O. Zheng, L. Li, M. Abdel-Aty, C. Cruz-Neira, and Z. Islam, "Towards next generation of pedestrian and connected vehicle in-the-loop research: A digital twin co-simulation framework," *IEEE Trans. Intell. Vehicles*, vol. 8, no. 4, pp. 2674–2683, Apr. 2023, doi: 10.1109/TIV.2023.3250353.

[61] L. Wang, H. Zhong, W. Ma, M. Abdel-Aty, and J. Park, "How many crashes can connected vehicle and automated vehicle technologies prevent: A meta-analysis," *Accident Anal. Prevention*, vol. 136, Mar. 2020, Art. no. 105299, doi: 10.1016/j.aap.2019.105299.

[62] H. Li, G. Zhao, L. Qin, H. Aizeke, X. Zhao, and Y. Yang, "A survey of safety warnings under connected vehicle environments," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 5, pp. 2572–2588, May 2021, doi: 10.1109/TITS.2020.3026309.

[63] M. G. Oikonomou, A. Ziakopoulos, A. Chaudhry, P. Thomas, and G. Yannis, "From conflicts to crashes: Simulating macroscopic connected and automated driving vehicle safety," *Accident Anal. Prevention*, vol. 187, Jul. 2023, Art. no. 107087, doi: 10.1016/j.aap.2023.107087.

Lei Han received the bachelor's and master's degrees from the College of Transportation Engineering, Tongji University, in 2020 and 2023, respectively. He is currently pursuing the Ph.D. degree in transportation engineering with the University of Central Florida (UCF). He is a Research Associate with UCF. His research interests include traffic safety analysis, intelligent transportation systems, and deep learning applications in transportation engineering.

Mohamed Abdel-Aty (Senior Member, IEEE) is currently a Pegasus Professor and a Trustee Chair with UCF, Orlando, FL, USA. He is leading the Future City Initiative with UCF. He is also the Director of the Smart and Safe Transportation Lab. He has managed more than 90 research projects. He has delivered more than 30 keynote speeches at conferences around the world. He has published more than 800 articles, more than 450 in journals (As of July 2024, Google Scholar citations: 33,873, H-index: 102). His main expertise and interests are in the areas of ITS, simulation, CAV, and active traffic management. He is the Editor Emeritus of Accident Analysis and Prevention. He has received the 2020 Roy Crum Distinguished Service Award from the Transportation Research Board, the National Safety Council's Distinguished Service to Safety Award, the Francis Turner Award from ASCE, and the Lifetime Achievement Safety Award and the S. S. Steinberg Award from ARTBA in 2019 and 2022, respectively. He has also received with his team multiple international awards, including the Prince Michael Road Safety Award, London, in 2019.

Rongjie Yu received the bachelor's degree from Tongji University and the master's and Ph.D. degrees in traffic engineering from UCF in 2012 and 2013, respectively. He is currently a Full Professor with the College of Transportation Engineering, Tongji University. His research interests include traffic safety, human behavior, and safety evaluation of connected and autonomous vehicles.

Chenzhu Wang (Member, IEEE) received the Ph.D. degree in transportation engineering from Southeast University in 2023. He is currently a Post-Doctoral Researcher in transportation engineering with UCF. His research interests include traffic safety analysis, statistical models, intelligent transportation systems, machine learning, and deep mining.