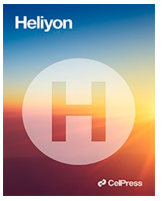
Heliyon 9 (2023) e20050

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



An over-the-horizon potential safety threat vehicle identification method based on ETC big data

Guanghao Luo [a,b], Fumin Zou [a,b], Feng Guo [a,b,*], Jishun Liu [a,b], Xinjian Cai [a,b], Qiqin Cai [c], Chenxi Xia [a,b]

[a] Fujian Key Laboratory of Automotive Electronics and Electric Drive, Fujian University of Technology, FuZhou, 350108, China
[b] Renewable Energy Technology Research Institute of Fujian University of Technology, Ningde, 352101, China
[c] School of Mechanical Engineering and Automation, Huaqiao University, Xiamen, 361021, China

ARTICLE INFO

Keywords:
ETC big data
Over-the-horizon
Speed prediction
Vehicle positioning
Potential safety threat vehicle
Identification

ABSTRACT

Smart cars rely on sensors like LIDAR and high-precision map-based perception for driving environment sensing. However, they can't detect low-speed vehicles beyond visual range, affecting safety and comfort. Manual vehicles face similar challenges. Low-speed driving contributes to expressway accidents due to limited visibility, road design, and equipment performance. To enhance safety, an over-the-horizon potential safety threat vehicle identification method using ETC big data is proposed. It consists of three layers. The first layer is the vehicle section travel speed sensing layer based on the wlp-XGBoost algorithm. The second layer is the in-transit vehicle position estimation layer based on the DR-HMM algorithm. The third layer is the Multi-information fusion of potential safety threat vehicle identification layer. Dynamic real-time detection and identification of potential safety threats in expressway sections were achieved, and simulations were conducted using real-time ETC data from Quanxia section on an ETC platform. Results show accurate prediction of vehicle speed and position in different road sections and traffic situations, with over 95% accuracy and recall in identifying potential safety threat vehicles. It perceives changes in the traffic conditions of road sections in real-time based on the changing trend of potential safety threat vehicle numbers, providing a vital reference for speed planning and risk avoidance.

1. Introduction

The expressway is a vital means of transportation, serving a crucial role on a global scale. With the construction of expressways, the travel time of all counties in the United States has decreased by one-third compared with that in 1960 [1], making great contributions to local economic development. However, the World Health Organization [2] estimates that 1.35 million people worldwide suffer fatal injuries from traffic accidents each year, and more than 50 million suffer non-fatal injuries. By 2030, traffic accidents are projected to become the fifth-highest cause of mortality globally [3]. Traffic accidents rank as the second most common cause of death after cancer in certain Southeast Asian countries [4]. Traffic accidents not only cause casualties, but also lead to large economic losses, with annual traffic accident losses accounting for about 1% of GDP in developed countries, while for economically backward and developing

* Corresponding author. Fujian Key Laboratory of Automotive Electronics and Electric Drive, Fujian University of Technology, FuZhou, 350108, China.
E-mail address: mapli@fjut.edu.cn (F. Guo).

https://doi.org/10.1016/j.heliyon.2023.e20050
Received 22 March 2023; Received in revised form 8 September 2023; Accepted 9 September 2023
Available online 11 September 2023

countries, the losses are about 3%–5% of their GDP [5].

Due to the characteristics of expressways such as full closure and high-speed limit, the overall running speed of vehicles is high, and malignant traffic accidents occur from time to time [6]. Chinese highway traffic accidents have two characteristics, one is the high intensity of accidents, and the other is the concentration of accidents at night. With 2016 statistics, China's expressways caused an average of 1 death per 0.6 traffic accidents, with an accident fatality rate five times higher than that of ordinary urban roads [7]. The low visibility at night affects the vehicle owner's judgment of abnormal conditions ahead, leading to frequent traffic accidents [7]. According to statistics, expressway traffic accidents in China account for 9.43% of the total fatalities and 31.49% of direct property losses [6]. Therefore, ensuring the traffic safety of expressways is not only conducive to improving the competitiveness of expressway transportation, but also of great significance for improving economic interests and national comprehensive strength [8].

In recent years, with the gradual improvement of road design, construction standards, and management level, the overall incidence of road accidents and malignant accidents has declined steadily [9]. During this period, the characteristics and main causes of expressway traffic accidents have changed significantly. Among them, Sangbok Lee et al. [10] found that tailgating accounted for more than 50% of all accidents. In addition, by analyzing the causes of domestic traffic accidents, it was found that traffic accidents caused by driving slowly accounted for more than 35% of all traffic accidents. Therefore, driving slowly is increasingly becoming a significant factor in frequent traffic accidents on the expressway. And the main reason is that expressway vehicles travel fast, normal driving vehicles can only judge the environment within the owner's field of vision, and can not sense the low-speed vehicles ahead in time, resulting in their approach to low-speed vehicles can not control the distance, thus leading to accidents.

For low-speed driving behavior on expressways, countries around the world have introduced corresponding laws and regulations to restrict it. For example, the German Road Traffic Regulation (StVO) [11] stipulates that motor vehicles driving on expressway 330.1 and expressway 331.1 need to exceed 60 km/h. In France, the speed needs to reach 60 km/h or more to drive on the expressway, and the minimum speed limit in the leftmost fast lane is 80 km/h. Low-speed driving causing Traffic jams can be punished with a maximum fine of 4500 Euros, 6 points, 2 years imprisonment, and 3 years suspension of driving license. The Road Traffic Safety Law of the People's Republic of China stipulates that driving a motor vehicle on an expressway at a speed that is more than 20% below the prescribed minimum speed under normal circumstances is subject to a fine of 200 yuan and 3 points.

Although many countries have relevant laws and regulations according to the road design speed standards and other factors to limit low-speed driving behavior, to some extent, the number of low-speed vehicles on the expressway has been reduced. However, due to the different driving habits of vehicle owners, there are differences in vehicle performance. The speed fluctuation between vehicles on the expressway is large, the relative speed difference between vehicles easily leads to tailgating and other traffic accidents, which is a major pain point that it can not solve.

The rapid development of intelligent driving technology in recent years has also led to the rapid development of related technologies. In the area of over-the-horizon perception, over-the-horizon perception solutions based on high-precision maps have been developed. Although high-precision maps can provide over-the-horizon lane planning for autonomous vehicles, their perception of the over-the-horizon environment is limited to the perception of static data such as lanes and road signs. For dynamic environment perception, it still relies on various sensors, so it cannot perceive over-the-horizon relatively low-speed vehicles in time.

The ETC gantry is spread throughout the highway network and can real-time record the driving data of various vehicles on the expressway [12]. By mining and analyzing real-time data, it can sense the driving status of vehicles at any location on the expressway [13,14], including vehicle speed [15], vehicle coordinates, and other key information, so that it can effectively identify vehicles with potential safety threats relative to them in the case of over the horizon for different vehicles, and provide a reference for vehicle owners to plan their speed, which is conducive to improving driving safety and comfort and providing a guarantee for creating a good expressway driving environment.

Based on a large number of ETC toll systems that have been built nationwide, this paper proposes an over-the-horizon potential safety threat vehicle identification method based on ETC big data. Firstly, based on wavelet packet transform and the XGBoost model, a vehicle section speed prediction algorithm is designed. On this basis, the DR algorithm combined with the Hidden Markov Model is used to design the vehicle section fusion positioning model. Finally, the potential safety threat identification zone and the potential safety threat speed discrimination method are designed to identify the target vehicle by utilizing the potential intersection time of the vehicle, real-time traffic volume in the section, and the type of vehicle.

2. Related work

Over-the-horizon potential safety threat vehicle identification requires a real-time perception of potential safety threat vehicles beyond visual range. It is essential to judge whether the relatively low-speed vehicle ahead has a potential safety threat to the over-the-horizon vehicle according to the current traffic situation and the-over-horizon vehicle type on the premise of obtaining the real-time vehicle speed and distance. Therefore, the current research status of the vehicle speed prediction algorithm, vehicle localization algorithm, and over-the-horizon modeling algorithm is investigated in the pre-writing stage.

2.1. Research status of vehicle speed prediction algorithm

Accurately obtaining the speed of vehicles is critical in identifying vehicles with potential safety threats beyond the visual range. Therefore, the first part of this paper's related work is to research speed prediction methods. The prediction of vehicle speed in various environments has been a growing area of study among scholars both domestically and internationally in recent years. Jaewook Shin [16] suggested a speed prediction method that uses a Markov chain with constraints on the speed. The speed constraint is determined

by the geometric shape of the road and the speed limit. Through repeated driving of the same route, the measured data based on the curve coordinates are obtained for the design model to overcome the shortcomings of Markov memorylessness. Xinyou Lin [17] proposed a speed prediction model based on driving pattern recognition and Hidden Markov Chain. Firstly, sample driving cycles were constructed using three typical driving periods. These constructed driving period segments are then clustered using the K-means algorithm. In real-time, the driving mode is identified using a Learning Vector Quantization (LVQ) neural network (NN). Subsequently, the vehicle speed is predicted based on the Markov Stochastic Matrix (MTM) corresponding to the three clustered driving modes. Bogaerts et al. [18] employed a hybrid approach combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) to develop a spatiotemporal recurrent convolutional neural network. This model effectively captures both temporal features and spatial feature variations of traffic speed, resulting in accurate traffic speed prediction. De Qichen et al. [19] put forward a new short-term traffic speed prediction framework by combining a multi-scale grid model with SBiGRU, which greatly reduced the computational complexity of spatial feature extraction. The method exhibits greater accuracy than other models in predicting speed over multiple time steps, as evidenced by the experimental results.

2.2. Research status of vehicle location algorithm

It is impossible to determine whether the vehicle has a potential safety threat only depending on the speed, and it is also necessary to make a comprehensive judgment based on the distance between vehicles. As one of the important indicators for identifying potential safety threats beyond visual range, the crucial aspect is to precisely determine the current location of the vehicle, that is, to accurately locate the vehicle. The vehicle location method is one of the research hotspots at home and abroad. Among them, Vincent Havyarimana et al. [20]proposed a fusion framework utilizing sparse Gaussian-Wigner prediction. The method assumes that the measurement noise is non-Gaussian distribution, and uses the distribution of generalized errors as the approximation of non-Gaussian density. By combining the benefits of random matrix theory and sparse characteristics, the approach delivers improved vehicle positioning capabilities. M. Tsai et al. [21] put forward a cooperative localization method, which uses the DR algorithm to improve the GPS position error in VANETs. The DR algorithm employed in this study uses travel history records to filter out erroneous GPS locations. Muhammad Tahir Abbas [22] proposed an integration Utilizing an extended Kalman filter and interacting multiple models to predict the vehicle's location on the road. Five different vehicle driving models are designed, including the constant bump model, constant acceleration model, constant speed model, constant position model, and vehicle turning model. The most suitable model is selected for position prediction based on GPS data. Finally, a geographic information system is added to the model in the prediction process to further reduce the position prediction error. Zhang et al. [23] proposed a high-precision vehicle positioning method based on neural network and roadside unit (RSU) fingerprint. The positioning area is evenly divided into several grid areas, and the received signal strength of different RSUs is collected in each grid area to construct the RSU fingerprint database. In the positioning stage, the rough coordinates of the vehicle to be located are estimated by the Back Propagation Neural Network (BPNN), and the fingerprint matching area is constructed with the above position coordinates as the center of the circle and the maximum prediction error value of BPNN as the radius, and then the precise coordinates of the vehicle to be located are calculated by local fingerprint positioning.

2.3. Research status of over-the-horizon modeling method

After obtaining the vehicle speed and distance, the potential safety threat vehicle discrimination method in the over-the-horizon situation needs to be modeled in conjunction with the real-time traffic dynamics. The modeling characteristics of over-the-horizon models vary in different domains.

In the military field, Ma et al. [24] estimated the advantages between two groups of drones through the distance and altitude difference between the two groups of drones, the firing range of the missile, and the optimal firing altitude, solving the optimization problem of the advantages between the two groups of relative drones in the over-the-horizon radar combat scenario. Hu Zhaohui et al. [25] proposed a novel approach for evaluating target threats for over-the-horizon air combat, which comprehensively considers capability threat, situation threat, and air combat intention threat, and more accurately describes the threat factors in over-the-horizon air combat. Xu An et al. [26] proposed an over-the-horizon air combat situation evaluation approach that integrates the tactical attack zone of air-to-air missiles. Based on the tactical attack area of the air-to-air missile, this method discusses the shortcomings of angle, relative distance, and altitude advantage functions in the air combat evaluation model, and improves the advantage function accordingly, to improve the over-the-horizon air combat situation evaluation model.

In the field of civil vehicles, V an Eenennaam [27,28] proposed a new method for constructing an over-the-horizon perspective of vehicle-to-vehicle communication. Through vehicle collaboration, a traffic map is constructed to provide over-the-horizon perception, and TrafficFilter is used to select vehicles within an appropriate range to forward traffic information to upstream vehicles to avoid broadcasting storms and invalid information between vehicles. Xu et al. [29] proposed a multi-vehicle cooperative over-the-horizon cognitive method based on an attention mechanism and proposed a multi-vehicle cooperative data adaptive filtering algorithm for cognitive differences caused by different environments of multi-sensor sources to reduce data interference caused by the poor performance of sensors in specific scenarios.

In summary, the related technologies in the field of over-the-horizon have achieved good milestones and laid the foundation for the study of the expressway over-the-horizon potential safety threat vehicle identification method. Before researching the over-the-horizon potential safety threat vehicle identification method, the related concepts are first defined.

3. Related definitions

Three core problems must be solved to achieve expressway over-the-horizon potential safety threat vehicle identification.

Question 1. How to use ETC transaction data to predict the real-time speed of a vehicle;

Question 2. How to use ETC transaction data to locate vehicles in real-time;

Question 3. How to determine if a vehicle is a potential safety threat to an over-the-horizon vehicle?

The following definitions are provided to address the aforementioned three issues.

Definition 1. Over-the-horizon vehicles: Over-the-horizon vehicles are vehicles that have access to the over-the-horizon potential safety threat identification function of this method.

Definition 2 ET C transaction data (EData): As the vehicle passes through the ETC gantry, the on-board unit (OBU) in the vehicle will communicate with the roadside unit (RSU) located on the gantry to exchange information. The RSU can acquire vehicle ID, gantry ID, information transaction time, vehicle expressway entrance, and other information. Subsequently, the RSU uploads this information to the ETC system. These uploaded pieces of information make up the ETC data.

Definition 3. Expressway section (QD): Expressway gantries, toll station entrance, and exit (including inter-provincial import and export) are collectively referred to as Node, two adjacent Nodes constitute an expressway section *QD*, referred to as section, as shown in equation (1):

$$QD = < Node_1, Node_2 > \tag{1}$$

where $Node_1$ is the starting point of the section and $Node_2$ is the end point of the section. For example, the starting point of section 1 is the front gantry of section 1, and the end point is the rear gantry of section 1

Definition 4. Traj: The sequence of gantries that a vehicle passes through while driving on an expressway is called *Traj*, as shown in equation (2):

$$Traj = < Node_1, \ldots, Node_n > \tag{2}$$

Among them, $Node_1$ is referred to as the starting point of the trajectory, and $Node_n$ is referred to as the endpoint.

Definition5. The time the vehicle has traveled in the current section (Δt_{drive}): The time that the vehicle has already traveled in the current section is called the time that the vehicle has traveled in the current section Δt_{drive}, and Δt_{drive} satisfies the following relation shown in equation (3):

$$0 \le \Delta t_{drive} \le T_{tradetime2} - T_{tradetime1} \tag{3}$$

Among them, $T_{tradetime1}$ is the time when the vehicle interacts with the front gantry of the current section, and $T_{tradetime2}$ is the time when the vehicle interacts with the rear gantry of the current section. The definition of the front and rear gantries in the section is shown in Fig. 1.

Definition 6. The vehicle section travel speed (V_{car}): The average speed of vehicles passing through a section is called the section speed of vehicles in the section. For example, if the vehicle passes through Section 1 for Δt hours and Section 1 is L km long, the speed of the vehicle in Section 1 is shown in equation (4):

$$V_{car} = \frac{L}{\Delta t} \tag{4}$$

where L represents the actual length of the section, and Δt represents the time taken by the vehicle from entering a section to leaving a

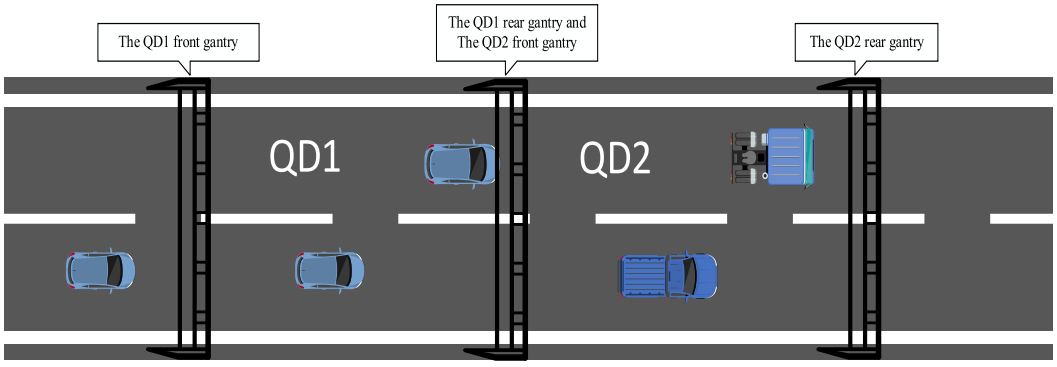


Fig. 1. Schematic diagram of the section.

section.

Definition 7. Potential safety threat vehicle: A vehicle traveling at a speed below a certain threshold of over-the-horizon speed within the potential safety threat vehicle identification zone is called a potential safety threat vehicle. Therefore, a potential safety threat vehicle can only appear in front of an over-the-horizon vehicle and travel at a speed lower than the over-the-horizon vehicle's speed.

Definition 8. Over-the-horizon potential interaction time domain, that is, the time range in which the over-the-horizon vehicle may interact with vehicles entering the same area in a certain period. Since the purpose of this paper is to identify potential safety threat vehicles, as defined in definition 7, the over-the-horizon potential interaction time domain must be located in the slice of time before the over-the-horizon vehicle enters the current segment. If the vehicle enters the current section on the same day [00:00,06:00), the potential interaction time domain of the vehicle in this section is from 18:00 of the previous day to the time node $T_{tradetime1}$, when the vehicle enters the current section; If the vehicle enters the current section on the same day [06:00,24:00), the potential interaction time domain of the vehicle in this section ranges from 00:00 of the same day to the time node $T_{tradetime1}$, when the vehicle enters the current section, as shown in equation (5):

$$T_{interactive} = \begin{cases} [T_1, T_{tradetime1}], T_{tradetime1} \in [00:00, 06:00) \\ [T_2, T_{tradetime1}], T_{tradetime1} \in [06:00, 24:00) \end{cases} \tag{5}$$

where $T_{tradetime1}$ is the time when the over-the-horizon vehicle enters the target section and interacts with the front gantry. T_1 is 18:00

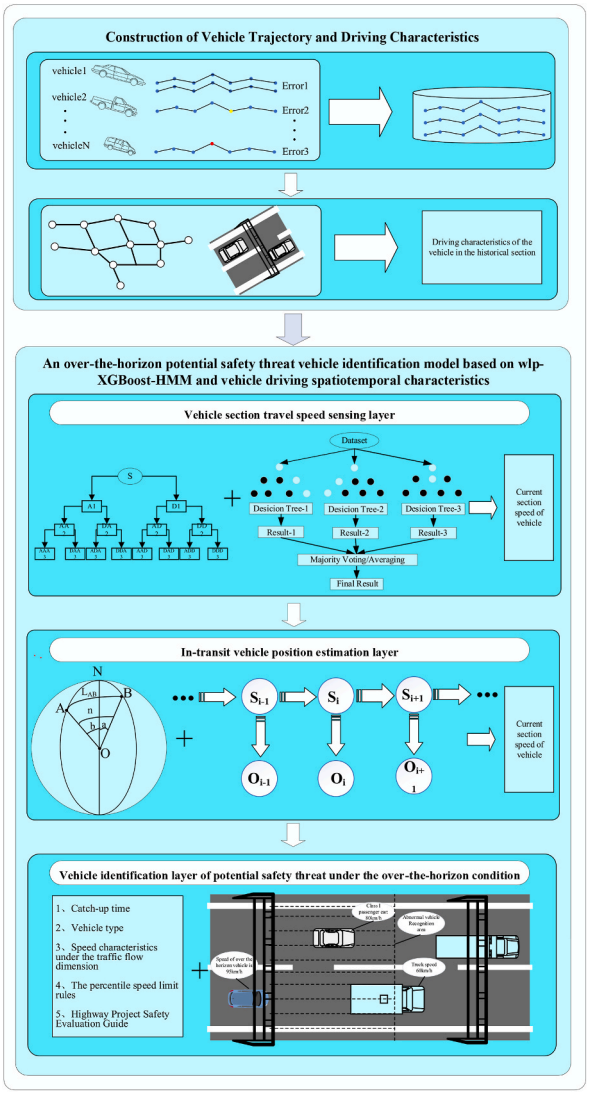


Fig. 2. Overall framework diagram.

the day before the vehicle enters the section. T_2 is 00 : 00 on the day the vehicle enters the section.

Definition 9 ET C (Electronic Toll Collection): ETC is a non-stop electronic toll collection system, also known as an automatic road payment system, which is a road toll collection method specially used for Toll roads. It is usually found on expressways, bridges, or tunnels implementing toll policies, and some sections of the city center. This article focuses on the study of expressway ETC. Because vehicles are settled uniformly at the exit toll station, ETC data record key information for cost settlement, such as the vehicle's driving trajectory, passage time, and vehicle type. This information lay the data foundation for the method proposed in this paper.

4. Modeling of over-the-horizon potential safety threat vehicle identification method for expressway

4.1. Summary of the general framework

Firstly, This paper uses the expressway ETC simulation platform to generate ETC transaction data by combining vehicle driving characteristics and gantry operation features; secondly, it deeply mines the ETC transaction data, constructs vehicle travel trajectories, and establishes vehicle historical section passage speed feature sets and traffic flow features during passage by combining gantry topology and the expressway gantry distance generation method based on Gaode Map. Finally, based on the wlp-XGBoost-HMM and vehicle driving spatiotemporal features, an over-the-horizon potential safety threat vehicle identification model is constructed to detect and identify over-the-horizon potential safety threat vehicles. The model consists of three layers in total. The first layer is the perception layer of vehicle section speed. After noise reduction of vehicle running data by wavelet packet transform, the XGBoost algorithm is used to predict the current vehicle section speed. The second layer is the position estimation layer of the vehicle in transit. After estimating the potential position of the vehicle by the DR algorithm, the accurate position of the vehicle is estimated by the

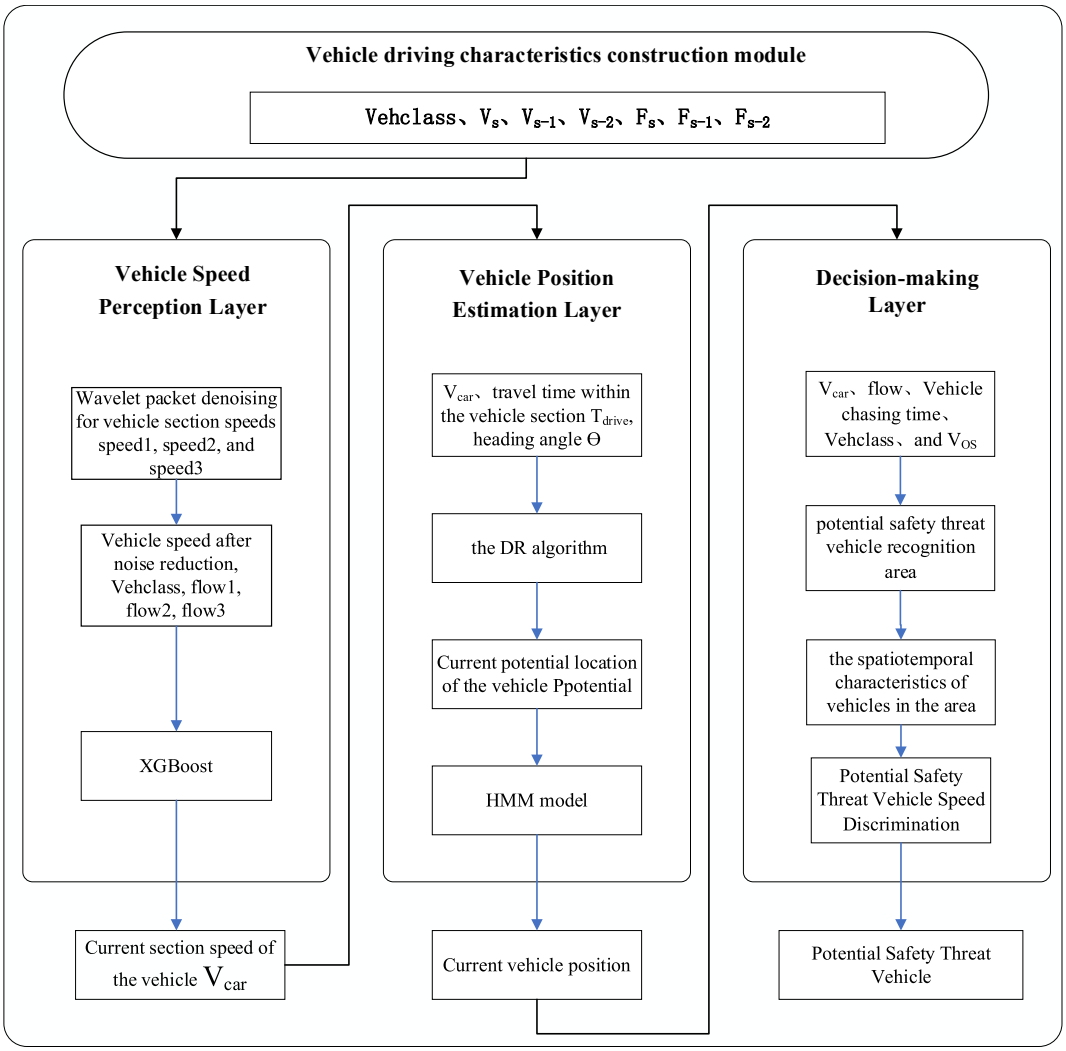


Fig. 3. Potential safety threat vehicle identification model under the over-the-horizon condition.

Hidden Markov algorithm. The third layer is the potential safety threat vehicle discrimination layer under the over-the-horizon situation with multiple information fusion, which firstly combines the potential vehicle potential catch-up time, speed characteristics under the traffic flow dimension, and vehicle type to delineate the multi-level potential safety threat vehicle identification zone, and then dynamically determines the potential safety threat vehicle speed according to the driving characteristics of each type of vehicle, the percentile speed limit rules and the "Highway Project Safety Evaluation Guide". The overall framework of the method is shown in Fig. 2.

4.2. Construction of vehicle trajectory and driving characteristics

The expressway ETC simulation platform [30] can generate ETC transaction data that matches the actual road conditions based on real-time traffic flow, road speed, vehicle type proportion, and other characteristics of the highway. Since the original simulation data generated by ETC only recorded shallow information such as vehicle and gantry transaction time, vehicle type, license plate, etc., the depth of information was insufficient. To accurately predict the section speed of vehicles and locate vehicles, the original data is deeply mined to build the historical travel path and driving characteristics of vehicles.

(1) Acquisition of vehicle's historical speed

After generating ETC transaction data through the simulation platform, construct the driving trajectory for each vehicle based on the time series. Based on the constructed driving trajectories, the historical driving characteristics of each vehicle are analyzed. Firstly, each vehicle's driving trajectory is traversed to obtain the passing time nodes of adjacent gantries in the vehicle trajectory, and the time spent by the vehicle passing through the section composed of adjacent gantries is calculated. Secondly, the coordinates of adjacent gantries are obtained, and the distance between the two adjacent gantries is calculated using the Gaode API. Then, the driving speed of the vehicle in the historical section is calculated according to the kinematic formula. To meet the model requirements in section 4.3.2, this article obtained the traffic speeds of the target section and the first two sections of the target section of the vehicle.

(2) Acquisition of segment flow and vehicle type characteristics

In order to preserve the authenticity of traffic flow to the greatest extent and avoid the distortion of calculated traffic flow caused by too long a time lag, this article only calculates the number of vehicles passing through the front and rear gantries of the section from the first 10 min of vehicle entry to the time when the vehicle trades with the front gantry of the section, as the corresponding traffic flow for each vehicle during the passage period. To match the speed of the vehicle section in the previous step, this article also calculates the traffic flow between the current section and the first two sections of the current section. The ETC data contains the vehicle type, so vehicle type characteristics are directly extracted from the data. The constructed vehicle driving characteristics are shown in the vehicle driving characteristics construction module in Fig. 3, which includes Vehclass representing the vehicle type, the current section speed V_s of the vehicle, the speed V_{s-1} of the previous section of the vehicle in the current section, the speed V_{s-2} of the previous two sections of the vehicle in the current section, the traffic flow F_s of the vehicle in the current section, the traffic flow F_{s-1} of the previous section of the vehicle in the current section, and the traffic flow F_{s-2} of the previous two sections of the vehicle in the current section

4.3. An over-the-horizon potential safety threat vehicle identification model based on wlp-XGBoost-HMM and vehicle driving spatiotemporal characteristics

To determine whether the vehicles ahead pose a potential safety threat to over-the-horizon vehicles, it is necessary to obtain the real-time traffic speed and position of the vehicles and to identify potential safety threats based on the real-time driving situation on the expressway. Therefore, the model consists of three layers: the first layer is for in-transit vehicle speed perception, the second layer is for vehicle position estimation, and the third layer is for identifying potential safety threats to over-the-horizon vehicles, as shown in Fig. 3.

4.3.1. Vehicle section travel speed sensing layer

Based on the vehicle driving characteristics mined in 4.2, the wlp-XGBoost algorithm is used to construct a vehicle section travel speed perception layer. Firstly, wavelet packet transform is used to denoise the traffic speed in historical sections of vehicles; Secondly, based on the historical section travel speed, vehicle type, and section traffic flow, predict the current section travel speed of vehicles.

4.3.1.1. Wavelet packet transform data denoising. During the process of vehicles driving on expressways, various subjective reasons cause sudden speed changes, and this sudden change in speed can cause the model to be affected by noise when directly conducting regression analysis on the data, seriously affecting the model's fit to the data. To reduce the impact of noise, it is necessary to denoise the collected data. Wavelet analysis [31] has both time-domain and frequency-domain analysis capabilities and has variable time-frequency resolution, making it suitable for analyzing sudden changes in signals. However, wavelet transform only further decomposes the low-frequency part of the signal and no longer decomposes the high-frequency part. Wavelet packet transform can not only decompose the low-frequency part of the temporal data but also the high-frequency part of the data. Therefore, this decomposition method is a more precise temporal data decomposition method, which improves the time-frequency resolution of the temporal

data [32].

The wavelet packet transform used in this paper can be described by a Binary tree, as shown in Fig. 4. Each node in the figure represents a signal frequency band, and the first digit represents the decomposition level of the wavelet packet. After wavelet packet decomposition, not only can all original signal frequency bands be covered, but also the signal can be decomposed at any multi-scale to obtain comprehensive local information [33].

Generally, a speed signal model with noise can represented by equation (6):

$$V(k) = \dot{V}(k) + \theta e(k) \tag{6}$$

where *V(k)* is the vehicle speed containing noise; $\dot{V}(k)$ is a useful signal; *e(k)* is noise; θ is noise energy. In engineering applications, valuable signals are often characterized by low-frequency components or relatively stable signals, while noise signals often manifest as high-frequency signals. Denoising processing is to remove these high-frequency signals to the greatest extent.

During this process, one important step in using wavelet packet denoising is determining the threshold parameter. The threshold parameter should be selected according to specific circumstances, and common methods include fixed threshold, empirical threshold, and adaptive threshold. In this article, an adaptive threshold method is used. Adaptive thresholding is a commonly used method for determining thresholds, which is built on the statistical characteristics of the signal. It can adaptively determine the threshold grounded upon the local characteristics of the signal.

This article employs wavelet packet transform for denoising the vehicle speed signal, and the process can be divided into four main steps.

1) Select a wavelet and determine the level of decomposition, and then perform wavelet packet decomposition on the speed characteristics. In this paper, the db4 wavelet is used for 3-level decomposition.

Let $V_j^n(t) \in U_j^n$, then $V_j^n(t)$ can be expressed as $V_j^n(t) = \sum_l d_l^{j,n} u_n(2^j t - l)$.The wavelet packet decomposition algorithm is to find $\{d_l^{j,2n}\}$ and $\{d_l^{j,2n+1}\}$ from $\{d_l^{j+1,n}\}$, as shown in equations (7) and (8):

$$\left\{d_l^{j,2n}\right\} = \sum_k a_{k-2l} d_k^{j+1,n} \tag{7}$$

$$\left\{d_l^{j,2n+1}\right\} = \sum_k b_{k-2l} d_k^{j+1,n} \tag{8}$$

2) After selecting the wavelet and deciding the level of decomposition, the next step is to calculate the optimal wavelet packet basis and tree, based on a specified entropy standard.
3) In the process, the wavelet packet decomposition coefficients are subject to threshold quantization. This involves the selection of an appropriate threshold for each coefficient, followed by quantization.
4) Wavelet packet reconstruction is carried out based on the lowest layer wavelet packet decomposition coefficient and the corresponding quantized coefficient.

Wavelet packet reconstruction is to find$\{d_l^{j+1,n}\}$by$\{d_l^{j,2n}\}$and$\{d_l^{j,2n+1}\}$, as shown in equation (9):

$$\left\{d_l^{j+1,n}\right\} = \sum_k \left[h_{l-2k} d_k^{j,2n} + v_{l-2k} d_k^{j,n+1}\right] \tag{9}$$

4.3.1.2. Speed prediction model based on XGBoost. The XGBoost algorithm is a powerful tool for fitting, evaluating, and predicting spatial sequence problems using a series of classification regression trees [34,35]. With this approach, it is possible to reconstruct data sets from a series of spatial sequence data. In the case of predicting expressway vehicle speeds, the training set data consisting of the spatial sequence is used as an input with a fixed step size, referred to as the historical spatial sequence. The output variable is the spatial sequence of the next step size, which is called the current spatial sequence. Together, these two sequences form a time window, and the length of this window is set to 3 in this model. Specifically, the speed of the next section is predicted based on the average speed of the

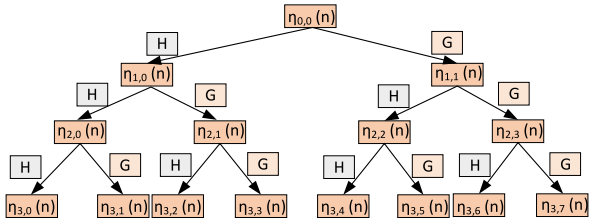


Fig. 4. Schematic diagram of wavelet packet decomposition.

first two sections of the vehicle, the vehicle type, and the corresponding traffic flow. The window slides forward one spatial step each time until it reaches the first spatial window. Fig. 5 provides a schematic representation of this process. Sliding the window transforms the spatial sequence into a new dataset consisting of several spatial windows.

The original sequence of data $S = [(p, x)_i]$, which was originally arranged in the order of expressway sections, was transformed into a supervised learning sequence $S_p = [(x_{i-2}, x_{i-1}), x_i]$ consisting of input and output variables, creating a new dataset for a supervised learning model for predicting vehicle speed on expressways. Among them, p represents the order of section positions, and x represents the speed value. Furthermore, the expressway vehicle speed prediction model is based on historical data to predict future data. Therefore, During the model evaluation phase, the forward verification method is employed. As described in section 4.3.3.1, a supervised learning sequence was constructed, where 80% of the data after wavelet packet denoising was applied for training, and the other 20% was designated as the test set. The iterated expressway speed prediction model [35] is shown in equation (10)~(11).

$$OBJ^t = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) + C \tag{10}$$

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \mathscr{W}_j^2 \tag{11}$$

In the equation, i is the sample number. $\sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i))$ which represents the error dimension between the factual and forecasted speed values and reflects the fitting degree of the model on the training set is the training loss function. At the t-th iteration, f_t refers to the function of weight assigned to the leaf nodes in the tree structure. x_i represents the value sequence of the target vehicle speed that serves as the model input. y_i refers to the real value of speed. $\hat{y}_i^{(t-1)}$ denotes the forecasted speed value obtained after the (t-1)-th iteration. $\Omega(f_t)$ represents the regularization term. C denotes the constant parameter. γ refers to the complexity added by adding one more leaf node, and γ can restrict the expansion of the tree. T represents the count of terminal nodes in the decision tree. λ is a regularization parameter that can be used to penalize the complexity of the model and help prevent overfitting by smoothing the leaf nodes. $\sum_{j=1}^{T} \mathscr{W}_j^2$ represents the squared sum of the weights assigned to the leaf nodes in the model. The complexity of the model is primarily managed by the regularization term. When the value of the regularization term is low, the model becomes simpler and less complex, which can help reduce model variance, and make the trained model more stable and less prone to overfitting [36]. The objective function is obtained through Taylor expansion, as shown in equations (12)–(14).

$$OBJ^{(t)} \simeq \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)}\right) + \sum_{i=1}^{n}\left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)\right] + \Omega(f_t) + C \tag{12}$$

$$OBJ^{(t)} \simeq \sum_{i=1}^{n}\left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)\right] + \Omega(f_t) = \sum_{i=1}^{n}\left[g_i \mathscr{W}_{q(x_i)} + \frac{1}{2} h_i \mathscr{W}_{q(x_i)}^2\right] + \gamma T + \lambda \frac{1}{2} \sum_{j=1}^{T} \mathscr{W}_j^2 = \sum_{j=1}^{T}\left[\left(\sum_{i \in I_j} g_i\right) \mathscr{W}_j + \frac{1}{2}\left(\sum_{i \in I_j} h_i + \lambda\right) \mathscr{W}_j^2\right] + \gamma T \tag{13}$$

$$OBJ^{(t)} = \sum_{j=1}^{T}\left[\left(\sum_{i \in I_j} g_i\right) \mathscr{W}_j + \frac{1}{2}\left(\sum_{i \in I_j} h_i + \lambda\right) \mathscr{W}_j^2\right] + \gamma T =$$

$$\sum_{j=1}^{T}\left[G_j \mathscr{W}_j + \frac{1}{2}\left(H_j + \lambda\right) \mathscr{W}_j^2\right] + \gamma T \tag{14}$$

By continuing the derivation of $\mathscr{W}_j$, we can obtain the target value for the predicted velocity and the optimal initial solution for the weights $\mathscr{W}_j^*$, as shown in equation (15).

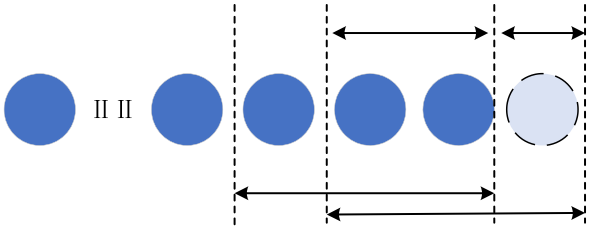


Fig. 5. Space sequence window.

$$\mathscr{W}_j^* = -\frac{G_j}{H_j+\lambda}, OBJ = -\frac{1}{2}\sum_{j=1}^{T}\frac{G_j^2}{H_j+\lambda} + \gamma T \tag{15}$$

The best split node is obtained using a greedy algorithm, and the difference between the node before and after splitting is compared. Equation (16) demonstrates the calculation of the gain value, Gain. Select the node with the highest gain value to split. Provided that either the gain value is greater than 0 or the maximum depth of the tree has not been reached, the process of splitting leaf nodes continues. However, If the gain value is less than 0 or the tree has reached its maximum depth, the splitting of the leaf nodes stops, and the optimal solution for weights and the optimal value for predicting speed are obtained by the function. At this point, The prediction model prevents overfitting, enhances the accuracy of prediction, and achieves optimal fitting results.

$$Gain = \frac{1}{2}\left[\frac{G_L^2}{H_L+\lambda} + \frac{G_R^2}{H_R+\lambda}\right] - \frac{(G_L+G_R)^2}{2(H_L+H_R+\lambda)} - \gamma \tag{16}$$

Finally, the parameters of the model are based on the Hyperopt module of Python software to achieve Bayesian optimization. Multiple parameters are optimized over a large range, ultimately resulting in the parameter combination with the best overall performance.

4.3.2. *In-transit vehicle position estimation layer*

After predicting the travel speed of vehicles in the potential interaction time domain of over-the-horizon vehicles using the vehicle passage speed sensing layer, it is necessary to further calculate the real-time position of vehicles in transit in the potential interaction time domain. Therefore, a vehicle position estimation layer is designed in the second layer of this model. Firstly, using the vehicle section passage speed V_{car}, the passage time within the vehicle section Δt_{drive} and the vehicle heading angle Θ derived from the vehicle passage speed sensing layer, the DR algorithm is used to derive the current potential vehicle position $P_{potential}$. However, the DR algorithm assumes the motion state between two points as a uniform linear motion state, it does not take into account the change of vehicle motion state and the change of road condition. And the DR algorithm deduces the new position from the original position, which will lead to the accumulation of errors, so the potential position of the vehicle obtained by the DR algorithm only and the actual vehicle position, there is a large error [37]. Therefore, based on the DR algorithm, the hidden Markov algorithm is further used to correct the vehicle potential position to obtain a more accurate vehicle position.

4.3.2.1. Vehicle potential location estimation algorithm. The main principle of the vehicle potential position estimation algorithm is based on the current vehicle position, vehicle section passage speed, heading angle θ_i and passage time using the DR algorithm to project the vehicle position at the next moment [38]. Knowing the initial position of the vehicle at the moment t_0 and its heading angle, according to the distance and passage time of the vehicle, the position of the vehicle at the next point can be projected, the position of the vehicle at the moment t_k can be expressed as equation (17):

$$\begin{cases} x_k = x_0 + \sum_{i=0}^{k-1} s_i \cos\theta_i \\ y_k = y_0 + \sum_{i=0}^{k-1} s_i \cos\theta_i \end{cases} \tag{17}$$

As shown in Fig. 6, firstly, the gantry position of the current section is used as the initial position of the vehicle, i.e., coordinates (x_0, y_0), and the directional angles of the two gantries before and after the current section are calculated as the heading angle θ_i of the vehicle. The orientation angle of the gantry composition remains constant in the same zone, so that θ_i is equal at any moment. Secondly, the vehicle's mileage s_i can be derived from the time and speed that the vehicle has passed in this section. Finally, the potential position of the vehicle (x_k, y_k) is calculated by combining the initial position, heading angle, and passing mileage of the vehicle.

4.3.2.2. Vehicle position correction method based on the Hidden Markov mode. After the potential positions of the vehicles have been obtained using the DR algorithm, the vehicle potential positions x_k and y_k are corrected using a Hidden Markov Model. The first step is

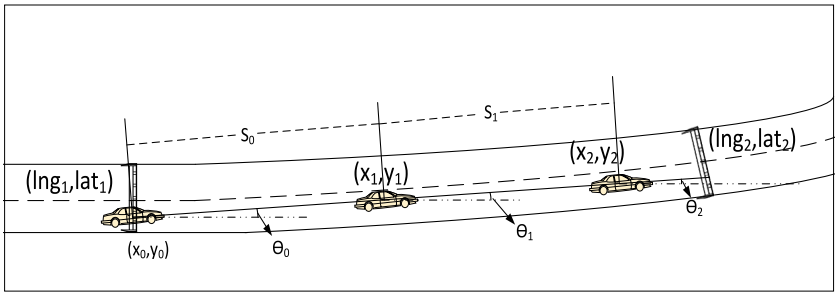


Fig. 6. Schematic diagram of potential location estimation of expressway vehicles.

to filter out candidate road segments. Each candidate road segment is considered a hidden state in the Markov chain, and the corresponding observation cost is calculated to observe whether the current trajectory matches the candidate road segment. If the current trajectory is very close to the candidate road segment, the appropriate road segment is selected. The second step is to calculate the transfer cost of each adjacent hidden state edge on the Markov chain and find the path with the minimum observation cost and minimum transfer cost on the Markov chain. Finally, the Viterbi algorithm is used to solve the road structure network, that is, the dynamic programming method is used to quickly find the optimal path with the minimum total cost and the closest to the actual path.

As shown in Fig. 7, the invisible variable z in the figure is a hidden state in the Hidden Markov Model, and the number of states is denoted by N; y is the variable affected by the variable z and can be observed, and the number of states is recorded as M; the transition probability matrix between the implicit states z is $A = [A_{ij}]_{N\times N}$, where $A_{ij} = P(q_{t+1} = q_j \mid q_t = q_i), 1 \le i,j \le N$; the observation probability matrix from the hidden state Z to the observation state y is denoted as $B = [B_j(k)]_{N\times M}$, where $B_j(k) = P(y = y_k \mid q_t = q_j), 1 \le j \le N$, $1 \le k \le M$; when $t = 1$ is the initial state, the initial state is $\lambda = (\lambda_i)$, where $\lambda_i = P(q_1 = Z_i), 1 \le i \le N$. In general, a Hidden Markov Model can be succinctly represented by a triple $\gamma = (A, B, \lambda)$ [39].

4.3.2.2.1. Acquisition of observation probability. The observation probability is the probability that a state y_i is observed from a point Z_i on the Hidden Markov Chain at time t. As shown in Fig. 8, there are two observation points y_1 and y_2 on the road r_1 and r_2, and the implicit matching points near the point y_1 have $z_1 = (r_1, p_1)$ and $z_2 = (r_1, p_2)$ on the road r_1 and r_2 respectively. The distance from the implicit matching point z_1 and z_2 to the observation point y_1 is not equal. Assuming that the great circle distance between the potential position of the vehicle and the candidate positioning point is $\|y_1 - z_1\|_{great\_circle}$, the observation probability can be regarded as a Gaussian process [40], centered on the hidden sequence z_1. Therefore, the values of the elements in the observation probability matrix are shown in equation (18), where σ is the standard deviation of the predicted quantity values of the DR algorithm.

$$B_1(1) = p(y_1 | p_1 = z_1) = \frac{1}{\sqrt{2\pi}\sigma} e\left(\frac{\|y_1 - z_1\|_{great\_circle}}{\sqrt{2}\sigma}\right)^2 \tag{18}$$

4.3.2.2.2. Acquisition of state transition probability and initial state probability. The state transfer probability is the transfer probability between hidden states on a Markov chain i.e. the probability that the state is z_j at time t+1 under the condition that the state is z_i at time t. The relationship between the state transfer probability and the hidden matching points is shown in equation (19).

$$A_{12} = p(p_2 = z_2 | p_1 = z_1) \propto e^{-ad_{12}} \tag{19}$$

Where: d_{12} denotes the distance between candidate implicit matching points z_1 and z_2, a is the parameter affecting the shortest distance element between z_1 and z_2.

The initial state probability represents the probability that a hidden state is an initial state, as shown in equation (20).

$$\lambda_i = P(q_1 = z_i) \tag{20}$$

4.3.2.2.3. Model solution. Based on the established expressway driving trajectory correction model, the Viterbi algorithm is used to address the issue, which can calculate the cost of multiple routes on multiple routes, and finally find the path with the smallest total cost and closest to the actual driving route of the vehicle. The schematic diagram of the Viterbi algorithm is shown in Fig. 9.

To implement the Viterbi algorithm, it is usually necessary to calculate the local state [41] and give its recursive formula. In the Hidden Markov Model, two local states for recursion are first given.

(1) For the first local state, the recursive expressions of the maximum probability σ_t and σ in the transition paths $z_1, z_2, \ldots z_t$ between all possible hidden states with hidden state i at time t can be calculated by equation (21) and (22):

$$\sigma_t(i) = \max_{z_1, z_2, \ldots, z_{t-1}} P(z_t = z, z_1, z_2, \ldots, z_{t-1}, y_t, y_{t-1}, \ldots, y_1 | \lambda) \ , \ i = 1, 2, \ldots, N \tag{21}$$

$$\sigma_{t+1}(i) = \max_{1\le j\le N} [\sigma_t(j) a_{ji}] b_i(y_{t+1}) \tag{22}$$

(2) The second local state can be calculated by recursive formula. For example, to calculate the hidden state $\Phi_t(i)$ at time t-1 with the maximum transition probability in the transition path $(z_1, z_2, \ldots, z_{t-1}, z_t)$ between all independent states at the current time t, see equation (23) :

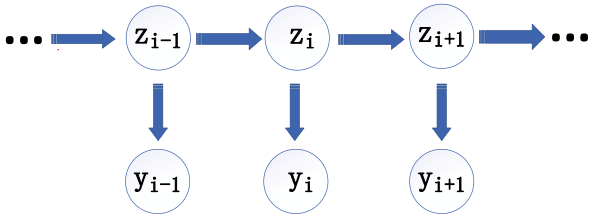


Fig. 7. Schematic diagram of the Hidden Markov Model.

Fig. 8. Schematic diagram of vehicle track observation probability on an expressway.

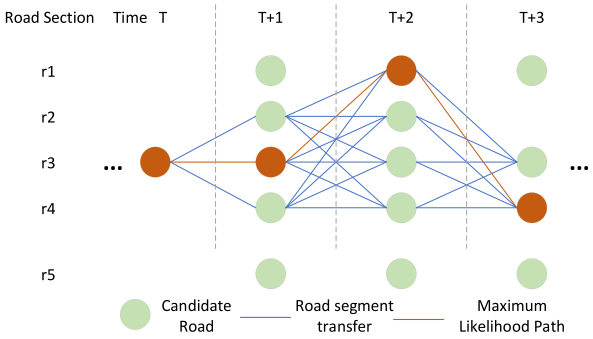


Fig. 9. Road network matching process based on Viterbi.

$$\Phi_t(i) = arg\max_{1\le j\le N}\left[\sigma_{t-1}(j)a_{ji}\right] \tag{23}$$

According to formulas (21) and (23), we can find the hidden state of T time from the departure time and then search along the return route through the preorder state until we get the hidden state sequence that can better explain the observation sequence.

4.3.3. Vehicle identification layer of potential safety threat under the over-the-horizon condition

After obtaining the vehicle speed and position from the vehicle section travel speed sensing perception layer and the in-transit vehicle position estimation layer, potential safety threat vehicles can be identified. As shown in Definition 7, a potential safety threat vehicle is defined as a vehicle located within the over-the-horizon identification zone and having a lower speed than the over-the-horizon vehicle. During the identification process, if the identification zone is too large, potential safety threat vehicles may be identified too early, affecting the driver's focus and even reducing overall highway efficiency. On the other hand, if the identification zone is too small, potential safety threat vehicles may be identified too late, leaving the driver with insufficient time to take action. Therefore, it is crucial to appropriately divide the over-the-horizon identification zone. Whether one vehicle poses a safety threat to another vehicle depends not only on the speed and distance between them but also on the real-time driving environment and the mechanical performance of the vehicles [42–45].

The potential safety threat vehicle identification layer first divides the over-the-horizon identification zone into multiple levels based on the potential catch-up time of vehicles, speed characteristics in traffic volume dimension, and vehicle type characteristics. Once the identification zone is defined, the speeds of vehicles within it are evaluated using percentile speed limits and referring to the "Highway Project Safety Evaluation Guidelines" to determine if they fall within the range of potential safety threat speeds. Finally, vehicles located within the over-the-horizon identification zone and meeting the criteria for potential safety threat speeds are classified as potential safety threat vehicles under the over-the-horizon conditions.

4.3.3.1. Selection and division of potential safety threat vehicle identification zone based on spatiotemporal characteristics of the vehicle under the over-the-horizon conditions. To correctly evaluate the driving environment of vehicles on expressways, this part analyzes the time required for over-the-horizon vehicles to catch up with potential safety threats under different speed differences and vehicle distances, and the impact of expressway section characteristics on the over-the-horizon identification, including traffic volume characteristics, speed characteristics, and vehicle type characteristics. According to the conclusions obtained, the actual requirements

of potential safety threat vehicle identification are integrated to delimit the potential safety threat vehicle identification zone and improve the identification efficiency.

(1) Vehicle potential catch-up time analysis

The motion state of the over-the-horizon vehicle and the potential safety threat vehicle in its identification zone can be regarded as a chase motion. By analyzing the time required for the over-the-horizon vehicle to catch up with the potential safety threat vehicle under different speed differences and vehicle distances, it provides a reference for the division of the over-the-horizon identification zone.

According to the speed statistics of section 1, the maximum speed of vehicles appears in passenger cars, reaching 154 km/h, while the average speed of the whole day in the section is only 100 km/h, with a difference of 54 km/h. Therefore, the maximum speed difference is set as 60 km/h in this experiment, and the difference decreases by 20 and 10, respectively, to calculate the catch-up time under different speed and distance combinations. As can be seen from Table 1, different vehicle distances and speed differences require different catch-up times. Therefore, in order to avoid prematurely identifying potential safety threat vehicles and distracting the owner's attention, too late identification will make the owner too late to react, so in the process of demarcating the identification zone should consider the possible rendezvous time between the over-the-horizon vehicle and the potential security threat vehicle, that is, the time required to catch up.

(2) Analysis of speed characteristics in traffic volume dimension

The probability of a collision occurring in congested traffic is approximately six times higher than in free-flowing traffic conditions. In transitional traffic conditions, the probability of collision is about 1.6 times higher than in free-flowing conditions. As vehicle speed increases, the probability of traffic collisions decreases, primarily due to improvements in traffic conditions [46]. Therefore, the magnitude of traffic flow affects the driving environment for vehicles, constrains the operational of the over-the-horizon vehicle when encountering potential safety threat vehicles, and also influences the safety of the expressway driving environment.

From Fig. 10, it can be observed that after a traffic flow volume of over 1272, there is a significant increase in the number of abnormal vehicles in this section, and most of the abnormal vehicles are traveling at low speeds. When the traffic flow volume decreases from its maximum value to 1467, the number of abnormal vehicles gradually decreases significantly. Additionally, from Fig. 11, it can be observed that when the traffic flow volume is below 900, the number of abnormal vehicle speeds caused by low speeds within the section is minimal and can be almost negligible. However, when the traffic flow volume exceeds 900, there is a slight increase in the number of low-speed abnormal vehicles. As long as the maximum traffic flow volume does not exceed 1325, the number of low-speed abnormal vehicles will not increase significantly. By combining the information from Figs. 10 and 11, it can be observed that traffic flow volume has a certain influence on vehicle speed. Moreover, by perceiving the changes in the number of abnormal vehicles, it is possible to infer the variations in traffic conditions on different road sections.

By considering the impact of traffic flow on vehicle speed, the section status can be categorized into three states: smooth-flowing, moderate congestion, and heavy congestion. Correspondingly, adjustments will be made to the identification zones for different traffic flow volumes. When the traffic flow is less than 900, the road section is considered smooth-flowing. The speed of the over-the-horizon vehicle is less limited by the traffic flow, and the distance between the over-the-horizon vehicle and the potential safety threat vehicle will also accelerate and become smaller. In order to perceive the difference between the current road section and the front road section due to the irregularity of the spatial distribution of traffic flow in advance, the identification zone should be increased by one level under the same conditions. When the traffic flow is greater than 1370, the road section is heavy congestion, and the speed of over-the-horizon vehicles is limited. The distance between over-the-horizon vehicles and potential safety threat vehicles also shrinks at a slower speed, which makes it meaningless to identify potential safety threat vehicles located at a longer distance. In addition, the operation space of over-the-horizon vehicles decreases when the road section is heavy congestion, so more attention should be paid to vehicles at a closer distance, and expanding the identification zone will only distract the owner's attention. Therefore, the identification zone

Table 1
Catch-up schedule.

Maximum speed difference	Vehicle distance	Catch-up time
60 km/h	6 km	6min
	4 km	4min
	2 km	2min
40 km/h	6 km	9min
	4 km	6min
	2 km	3min
20 km/h	6 km	18min
	4 km	12min
	2 km	6min
10 km/h	6 km	36min
	4 km	24min
	2 km	12min

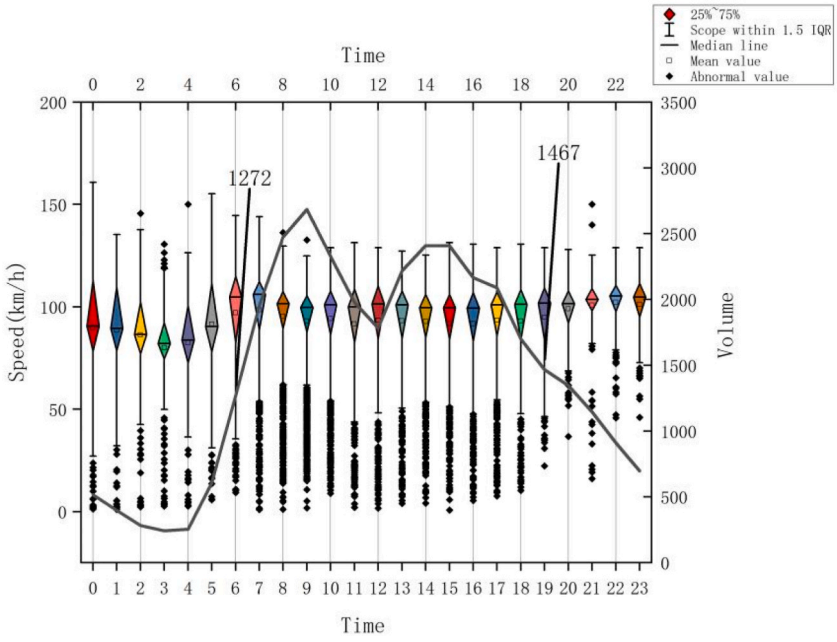


Fig. 10. Traffic volume-velocity characteristics of section 1.

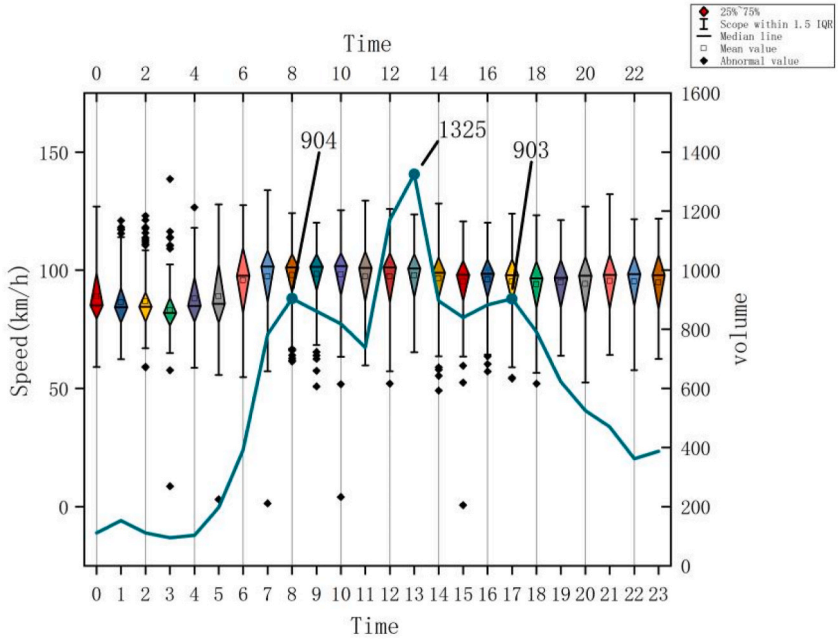


Fig. 11. Traffic volume-velocity characteristics of section 2.

should be appropriately reduced to a smaller level under the same conditions. When the traffic flow is maintained at the [900,1370] interval, it is considered that the road section is in a moderate congestion state at this time, and the identification zone does not need special adjustment.

(3) Analysis of vehicle type characteristics

On the actual road, the speed and mechanical properties of passenger cars and freight cars are quite different, and the degree of danger caused by them is different. When there are more heavy vehicles in the lane, the number of lane changes per lane also increase

[47], resulting in an increased likelihood of an accident [48]. Therefore, the proportion of different types of vehicles has a certain impact on the safety of the expressway driving environment.

Therefore, this experiment first counts the proportion of different vehicles in different sections. As shown in Table 2, in Section 1, Class 1 passenger cars have the highest number, accounting for 57.55% of the total vehicles. The second highest is Class 6 freight cars, comprising 16.64% of the total vehicle count. Class 1 freight cars rank third, accounting for 8.91% of the total vehicle count. In Section 2, the total vehicle flow is lower than in Section 1, but the distribution proportions of different vehicle types are similar to those in Section 1. Class 1 passenger cars remain the highest, accounting for 55.94% of the total vehicles. Class 6 freight cars come next, comprising 17.71% of the total vehicle count. Class 1 freight cars rank third, accounting for 10.00% of the total vehicle count.

Based on the above analysis, it can be observed that regardless of the traffic flow within each section, the distribution proportions of vehicles in each section are similar. Therefore, when defining the identification zone of potential safety threat vehicles, this paper considers that the proportions of different types of vehicles in each section are roughly equal, and the proportion of vehicle types in each section is not considered.

Secondly, analysis of the section travel speeds of different types of vehicles found that the section traffic speed of Class I passenger cars is significantly higher than the section traffic speed of other types of passenger cars, as shown in Fig. 12. Therefore, based on the vehicle classification, this method reclassifies the vehicles into three categories. All types of passenger cars are divided into two classes: Class I vehicles and Class II vehicles. Class I passenger cars are categorized as Class I vehicles, while the remaining types of passenger cars are classified as Class II vehicles. Additionally, due to the similar travel speeds and vehicle characteristics between trucks and special-purpose vehicles. Therefore, freight cars and special operation cars are categorized into Class III vehicles.

Fig. 13 shows that the average travel speeds of different types of vehicles in each section are different. The average speed of Class I vehicles is greater than that of Class II vehicles, and the speed difference between them is about 18%. Therefore, under the same conditions, Class II vehicles will identify fewer potential safety threat vehicles than Class I vehicles. However, it is important to note that Class II vehicles have slower section travel speeds, resulting in smaller speed differentials between them and potential safety-threat vehicles. For a distance of 4–6 km, it may take more than 25 min to catch up, making the identification zone beyond 4 km impractical for Class II vehicles.

On the other hand, for Class I vehicles, which have higher speeds and better mechanical performance, it may only take 5 min to catch up for a distance of 4–6 km. Therefore, the identification zone for Class I vehicles can be larger compared to Class II vehicles. Class III vehicles have the slowest speed and lower vehicle handling performance compared to Class I and Class II vehicles. The identification zone beyond 4 km is meaningless for Class III vehicles, and a 2 km zone is more suitable. However, in order to perceive changes in the traffic situation along the road section, the identification zone for Class III vehicles can be expanded to 4 km when necessary.

In summary, Class I vehicles have higher speeds and larger speed differentials with potential safety threat vehicles. Therefore, their identification zone can be selected within 2 km, 4 km, or 6 km. Class II vehicles have slower speeds compared to Class I vehicles, and the speed differential with potential safety-threat vehicles is not too large. If the identification zone is too large, early warnings may become meaningless. Therefore, their identification zone can be selected within 2 km or 4 km. As for the slowest Class III vehicles, expanding the identification zone at certain times can help detect sudden spatial changes in traffic flow along the road section. Therefore, their identification zone can also be selected within 2 km or 4 km.

Table 2
of each type of vehicle in each section.

Vehicle type	Number of vehicles in section 1	The proportion of various vehicles	Number of vehicles in section 2	The proportion of various vehicles
Class I passenger car	123,178	57.55%	20,837	55.94%
Class II passenger car	357	0.17%	106	0.28%
Class III passenger car	477	0.22%	69	0.19%
Class IV passenger car	1292	0.60%	142	0.38%
Class, I freight car	19,067	8.91%	3726	10.00%
Class II freight car	13,441	6.28%	2023	5.43%
Class III freight car	6396	2.99%	1198	3.22%
Class IV freight car	8392	3.92%	1373	3.69%
Class V freight car	5288	2.47%	1027	2.76%
Class VI freight car	35,623	16.64%	6598	17.71%
Class I special operation car	213	0.10%	16	0.04%
Class II special operation car	196	0.09%	77	0.21%
Class III special operation car	16	0.007%	0	0%
Class IV special operation car	0	0%	0	0%
Class V special operation car	0	0%	0	0%
Class VI special operation car	0	0%	56	0.15%

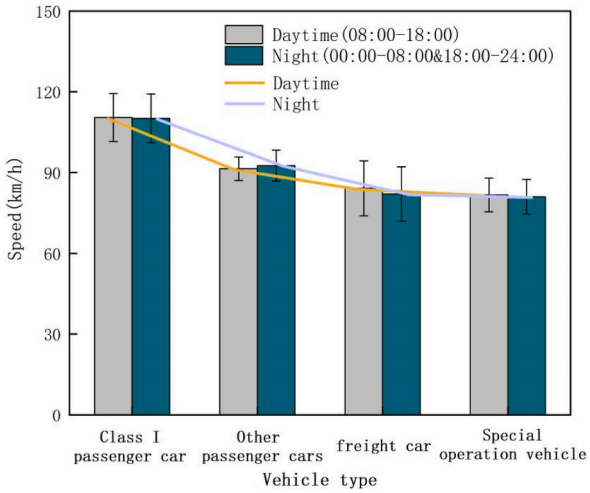


Fig. 12. Comparison of the average speed of each type of vehicle passage.

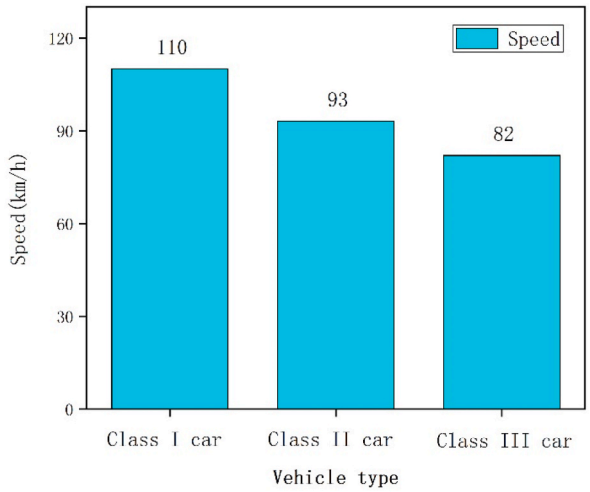


Fig. 13. Speed statistics of various vehicles after classification.

Based on the analysis of potential catch-up time, speed in the traffic dimension, and vehicle type characteristics, we can determine the spatiotemporal characteristics that vehicles face when encountering potential safety threat vehicles. From this, we can derive the division range of identification zones, as shown in Table 3.

4.3.3.2. Speed identification method of potential safety threat under the over-the-horizon condition. The Safety Evaluation Guide for

Table 3
Division of vehicle identification zone with potential safety threats.

Over-the-horizon vehicle type	Traffic volume on the road where the over-the-horizon vehicle is located	Operating space	The trend of decreasing distance	Scope of identification zone
Class I vehicles	1370≤flow	Small	Slow	2 km
	900<flow<1370	Medium	Medium	4 km
	flow≤900	Big	Quick	6 km
Class II vehicles	1370≤flow	Small	Slow	2 km
	900<flow<1370	Medium	Medium	4 km
	flow≤900	Medium	Medium	4 km
Class III vehicles	1370≤flow	Small	Slow	2 km
	900<flow<1370	Small	Slow	2 km
	flow≤900	Medium	Medium	4 km

Highway Projects [49] states that when the average difference in running speed is less than 10 km/h, the accident rate of one million vehicles is only 0.46. When the average difference in running speed is 10–20 km/h, the accident rate of one million vehicles is 1.44. When the average difference in running speed is greater than 20 km/h, it will cause poor coordination of the running speed of the expressway, and the accident rate of one million vehicles is 2.76.

When the minimum speed of small cars adopts 25% of the same type of cars, and the minimum speed of heavy vehicles adopts 30% of the same type of cars, the safety of expressway operation can be effectively improved, and the economic efficiency of expressway can be better played. In terms of the maximum speed limit, when the 85% speed of small cars is used as its maximum speed limit standard, and the 90% speed of heavy vehicles is used as its maximum speed limit standard, it is conducive to controlling the dispersion of running speed, reducing the traffic accident rate and accident severity, and ensuring a high highway traffic operation efficiency [50].

Through the analysis of the speeds of different types of vehicles on a section of the expressway in Fujian Province for a day, it was found that the 25th percentile speed of Class I vehicles is 105 km/h, and the 85th percentile speed is 118 km/h. For Class II vehicles, the 30th percentile speed is 91 km/h, and the 90th percentile speed is 96 km/h. As for Class III vehicles, the 30th percentile speed is 77 km/h, and the 90th percentile speed is 94 km/h. The 25th percentile speed of Class I vehicles is 89% of the 85th percentile speed, resulting in a speed difference of 13 km/h. For Class II and Class III vehicles, the 30th percentile speeds are 95% and 82% of the 90th percentile speeds, respectively, resulting in speed differences of 5 km/h and 17 km/h. As shown in Table 4.

Based on the percentile speed limit rules, the following conclusions can be drawn: The upper limit and lower limit of normal vehicle speeds for each vehicle type differ by 11%, 5%, and 18%, respectively. Therefore, if the speed of a Class I vehicle is lower than 11% of the normal speed for vehicles of the same type, it is considered a low-speed vehicle. Similarly, if the speed of a Class II vehicle is lower than 5% of the normal speed for vehicles of the same type, it is classified as a low-speed vehicle. For Class III vehicles, if their speed is lower than 18% of the normal speed for vehicles of the same type, they are considered low-speed vehicles.

When the speed differential between vehicles is between 10 and 20 km/h, the million-vehicle accident rate increases by 2–3 times compared to a speed differential within 10 km/h [49]. Combined with the characteristics of the over-the-horizon potential safety threat vehicle identification process that the potential safety threat vehicle is relative to the over-the-horizon vehicle itself, is to the over-the-horizon vehicle as the center, so the speed of the over-the-horizon vehicle no matter how fast or slow, its speed for its own is normal vehicle speed. Therefore, the speed of the over-the-horizon vehicle is regarded as the normal speed in the process of potential safety threat vehicle identification. On this basis, the potential safety threat vehicle speed discrimination method in the over-the-horizon situation is defined.

For over-the-horizon Class I vehicles, as shown in Fig. 13, their average speed is above 110 km/h, and an 11% speed differential exceeds 10 km/h. Therefore, according to the percentile speed limit rule, any vehicle with a speed lower than 11% of its speed should be considered a potential safety threat speed. For Class II vehicles, their speeds are lower compared to Class I vehicles. If we follow the percentile speed limit rule, the speed difference between normal vehicle speed and potential safety threat speed is within 10 km/h. Therefore, during actual driving, it does not pose a safety hazard. Therefore, according to the "Guidelines for Highway Project Safety Evaluation," which outlines the relationship between speed differential and traffic accident rates, speeds lower than 15% of its vehicle speed are considered potential safety threat speeds. For Class III vehicles, which are mostly loaded vehicles with long braking distances, the speed differentials are too large to control the following distance effectively. With an average speed ranging between 80 and 81 km/h, an 18% speed difference already exceeds 10 km/h. Therefore, according to the percentile speed limit rule, speeds lower than 18% of its vehicle speed are considered potential safety threat speeds. As shown in equation (24).

$$V_{pst}\begin{cases} \leq 89\%V_{OS}\ , & \text{Vehclasss} = 1 \\ \leq 85\%V_{OS}\ , & \text{Vehclasss} \in [2,6] \\ \leq 82\%V_{OS}\ , & \text{Vehclasss} \in [11,26] \end{cases} \tag{24}$$

Among them, V_{pst} is the speed of potential safety threats, V_{os} is the speed of the over-the-horizon vehicle, and Vehclasss is the vehicle type of the over-the-horizon vehicle.

5. Experimental results and analysis

5.1. Data sources and attributes

The test data uses the expressway ETC simulation platform to simulate the real-time transaction data of the Quanxia section. The original transaction data generated includes vehicle number, transaction time, gantry longitude, latitude, and other information, as shown in Table 5 below. The test section is the section from Quanzhou, Fujian to Xiamen, as shown in Fig. 14 in red.

Table 4
Statistical analysis of vehicle speed.

Over-the-horizon vehicle type	25%/30%vehicle speed	85%/90%vehicle speed	Speed Difference	Difference between the upper limit and lower limit of normal vehicle speed
Class I vehicles	105 km/h	118 km/h	13 km/h	11%
Class II vehicles	91 km/h	96 km/h	5 km/h	5%
Class III vehicles	77 km/h	94 km/h	17 km/h	18%

Table 5
ETC transaction data attribute table.

Serial number	Attribute Name	Examples	Attribute Name	Examples
1	TRADEID	S0***1(位)	ENTIME	2020/9/3 7:48:39
2	TRADETIME	2020/9/3 20:00:01	ENSTATION	46**
3	FLAGID	3502**	LNG	118.56**
4	FLAG TYPE	0	LAT	24.85***
5	OBUID	66AD40**	VEHCLASS	1

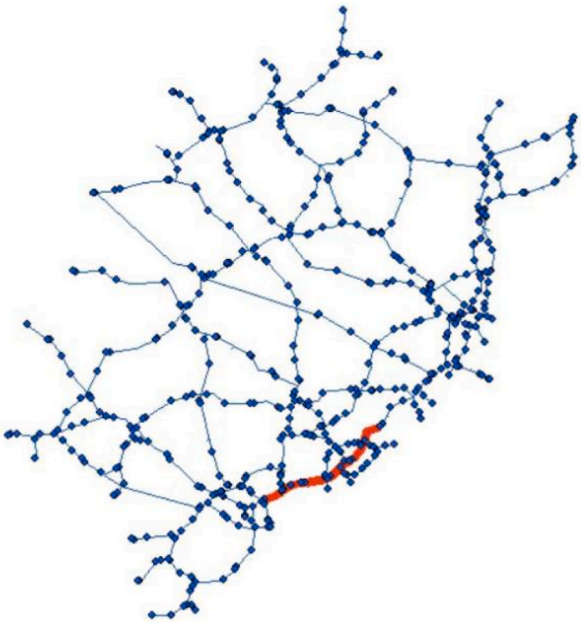


Fig. 14. Experimental road schematic diagram.

5.2. Vehicle section travel speed sensing layer experiment

5.2.1. Data preprocessing-wavelet packet noise reduction

During the experimental process, the ETC transaction data generated by the ETC simulation platform is matched with the expressway network topology to obtain the complete vehicle trajectory and extract the driving characteristics of the historical section. The historical driving speed of the vehicle is decomposed and reconstructed using wavelet packet transform to reduce noise.

Partial data after wavelet packet denoising is shown in Table 6, where the row index represents the vehicle number, speed1 denotes the passing speed of the vehicle in section 1, speed2 denotes the passing speed in section 2, and speed3 denotes the passing speed in section 3. Analysis of the mean square error and standard deviation of the data reveals that the mean square error of the original data is 595.7, while the overall mean square error of the data after wavelet packet denoising is 562.4. In terms of standard deviation, the standard deviation of the original data is 23.7, while the standard deviation of the denoised data is 21.8. It can be observed that the

Table 6
Table comparing data before and after wavelet packet denoising.

Vehicle ID	Original data			Data after wavelet packet denoising		
	Speed1	Speed2	Speed3	Speed1	Speed2	Speed3
1	104.102	93.515	92.406	104.793	93.708	92.063
2	97.690	87.748	89.192	99.428	96.083	81.288
3	98.973	104.287	55.504	96.772	102.921	55.081
4	89.561	95.763	95.150	89.229	100.935	95.390
5	86.354	88.528	84.490	87.053	88.721	84.911
6	67.801	70.885	69.399	69.727	70.783	69.511
7	57.514	73.231	61.567	58.050	73.439	57.728
8	77.078	86.603	76.318	78.844	88.782	76.393
9	77.774	77.505	75.531	78.707	84.418	69.074
10	82.118	86.982	86.339	82.845	87.467	86.579

denoised data is more stable, which is consistent with the characteristic of vehicles on expressways to maintain a stable driving speed under normal road conditions. Furthermore, the denoised data has a smaller mean square error and standard deviation, which is beneficial for achieving better experimental results in subsequent experiments.

5.2.2. Selection of evaluation indicators

The use of suitable evaluation metrics can provide an accurate and intuitive reflection of a model's predictive performance, allowing for the comparison of predicted and actual data values to quantify the model's efficacy. For this study, the mean absolute error (MAE) and root mean square error (RMSE) were selected as evaluation metrics. The MAE, being the absolute value of the average error, eliminates any positive or negative offsets and better reflects the true prediction error. On the other hand, RMSE can capture data dispersion and is calculated using equations (25) and (26).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{25}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2} \tag{26}$$

where: n represents the total amount of data in the training set.

Although MAE and RMSE provide a more accurate assessment of the model's prediction accuracy, the large fluctuations in the original vehicle speed values during data collection can result in relatively larger MAE and RMSE values. To improve the interpretability of the evaluation metrics and eliminate the impact of the size of the test set values, R^2 has been introduced as a means of assessing the predictive performance of the model, as shown in equation (27). This measure eliminates the impact of degrees of freedom on the prediction results. A higher R^2 indicates a better model fit, with the range of possible values between 0 and 1.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2}{\sum_{i=1}^{n}\left(\bar{y}_i - y_i\right)^2} \tag{27}$$

5.2.3. Analysis of experimental results

To enhance the verification of the model's robustness, this study conducts speed prediction experiments on two distinct roads, each of which encompasses three sections. Among them, there is no service zone on road 1 and one service zone on road 2. After the roads are selected, the trained wlp-XGBoost models are predicted on the test set, and the differences between the predicted and actual values are analyzed to calculate the MAE, RMSE and R^2 of the prediction models. The prediction results of the Lasso model, feedforward neural network model, and XGBoost model were also compared. Table 7 presents the evaluation of prediction performance on road 1, while Fig. 15~17 illustrate the comparison of predicted and actual values among the three models.

Through Table 7 and Figs. 15–17, it is evident that the prediction curves generated by the Lasso regression model, feedforward neural network, and the wlp-XGBoost model follow the changing trend of the actual curve and the R^2 values of the three models are 0.901, 0.882, and 0.885, respectively, indicating satisfactory fitting performance. However, The wlp-XGBoost model exhibited a lower MAE of 2.851 and RMSE of 3.872 compared to the Lasso regression model and the feedforward neural network model. Based on the comprehensive analysis of all evaluation indicators, the wlp- XGBoost model was found to have higher accuracy in predicting vehicle speed in expressway sections.

Because the closer the R^2 value is to 1, the higher the fitting degree of the model is. When predicting the speed of Road 1, the main reason why the R^2 values of the three models are higher and similar is that the road conditions on the expressway are better most of the time. The vehicle maintains a high speed, and the owner does not experience rapid acceleration and deceleration. And because there is no service zone in Road 1, the probability of noise in the data set is also reduced, so the model can fit the data well.

In addition, compared to the other two models, The wlp-XGBoost model exhibits a better performance next to the points where the speed changes sharply and can identify the vehicle speed more accurately, to avoid the missed detection of vehicles with potential safety threats due to the large difference between the predicted speed and the actual speed, or the false detection of vehicles without potential safety threats.

The above three models are closer to each other when prediction is performed on Road 1, and the prediction results are all close to the true values. This is because road 1 has no service area influence, so there is less noise in the dataset used for model training and prediction.

Table 7
Model prediction performance evaluation table in Road 1.

Model selection	MAE	RMSE	R^2
Lasso regression model	3.399	4.555	0.885
Feedforward neural network model	3.338	4.445	0.882
wlp-XGBoost model	2.851	3.872	0.901

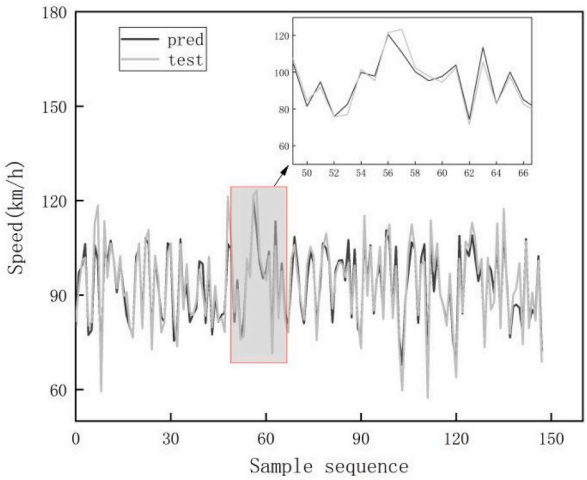


Fig. 15. Graph of Lasso regression prediction results in Road 1.

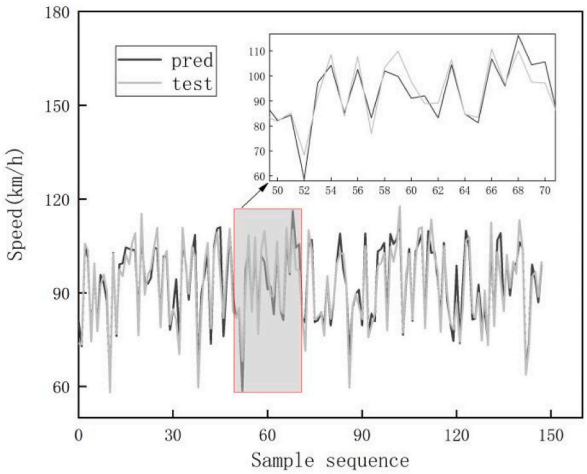


Fig. 16. Graph of Feedforward Neural Network Model prediction results in Road 1.

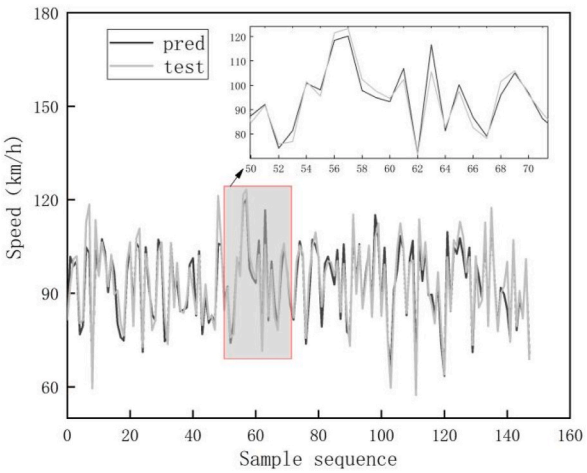


Fig. 17. Graph of wlp-XGBoost Model prediction results in Road 1.

And when predicting the speed of vehicles on road 2, the feedforward neural network model and lasso regression model were found to be significantly less effective compared to the wlp-XGBoost model. The lasso regression model and the feedforward neural network model had MAE values of 11.533 and 10.363, respectively, which were much larger than the wlp-XGBoost model of 5.423, while the R^2 coefficients representing the degree of model fit were 0.801, 0.598, and 0.563, the Lasso regression model and the feedforward neural network model have a large gap with the wlp-XGBoost model. The prediction performance evaluation within road 2 is shown in Table 8, and Figs. 18–20 display the comparison between the predicted and actual values of the three models.

Through analysis, it was found that the Lasso regression model and feedforward neural network model prediction effect fluctuate greatly. It is mainly due to the existence of a service area in the second section of Road 2, individual vehicles will enter the service area for a short rest, and the speed of its section passage will plummet, which leads to a certain noise generated by the concentration of speed features, so the speed prediction effect of Lasso regression model and feedforward neural network model at road 2 is not ideal. The wlp-XGBoost model, on the other hand, mitigates the noise influence by noise reduction of the original data and obtains more accurate results. In addition, the prediction results of the wlp-XGBoost model are consistent with the actual vehicle speed from the local enlargement of each model, so even for the noisy road sections, the wlp-XGBoost speed prediction model can achieve better results with better robustness.

In summary, in the actual operating environment of the expressway, factors that lead to vehicle speed fluctuations such as taking a rest in the service area, traffic jams, etc. Are inevitable. Using the wlp-XGBoost model to predict the speed of expressway vehicle sections, we can get more accurate section speed through fewer speed space sequences.

5.3. Vehicle position estimation layer experiment

5.3.1. Visual analysis of vehicle positioning experiments

As shown in Fig. 21(a), most of the vehicle positions calculated by the DR algorithm are distributed outside the expressway. The positioning deviation of the DR algorithm is due to the vehicle position obtained by the DR algorithm through continuous iteration, so the positioning error of the DR algorithm will continue to accumulate. The second influencing factor is that this experiment approximates the angle formed by the gantries on both sides of the expressway to the heading angle of the vehicle. This method is feasible on a straighter expressway, but the road section used in this paper has a certain range of bends, so the angle formed by the two gantries cannot be completely equivalent to the heading angle of the vehicle.

As shown in Fig. 21(b), the vehicle position calculated by the fusion positioning model is consistent with the actual position of the vehicle. The fusion positioning model uses the DR algorithm to calculate the initial position of the vehicle and then uses the Hidden Markov Model to correct the trajectory of the vehicle position. The main reason why the hidden Markov algorithm can correct the trajectory of the vehicle and make its trajectory closer to the real trajectory is that the expressway network is sparse. Therefore, the hidden Markov algorithm has fewer hidden sequences on different road sections, and the hidden sequences are basically on the same road section. Therefore, there are fewer hidden sequences with similar observation probability and transition probability between different road sections, so that the Viterbi algorithm can accurately find the best path.

5.3.2. Algorithm accuracy analysis

In this method, the expressway is considered as a road consisting of various sections, and the vehicles travel at different speeds in each section. Therefore, when conducting vehicle localization experiments, the experiments are conducted in units based on expressway segments. Firstly, the distance D_1 between the two positions is calculated using the vehicle position coordinates obtained from the vehicle fusion positioning model and the front gantry coordinates of the current section. Secondly, the actual distance D_2 traveled by the vehicle is calculated using the kinematic formula. Finally, the positioning error is obtained based on the difference between D_1 and D_2. The calculation formula is shown in equation (28).

$$\sigma = |D_1 - D_2| \tag{28}$$

To obtain the positioning accuracy of the DR algorithm and the Hidden Markov Model, error statistics are made for the DR algorithm and the fusion positioning model respectively, and the statistical results are shown in Fig. 22(a) and (b).

The positioning error range of the DR algorithm is 10∼100 m. Because this experiment selects a section of road with small curves, the DR algorithm assumes that the vehicle is in a uniform straight-line motion state, so when the test is conducted on a section with curves, there is a certain deviation between the actual vehicle state and the algorithm assumption, resulting in a large deviation between the calculation results and the actual results. When the positioning test is carried out on the road section close to the straight line, the vehicle state is consistent with the actual vehicle state, and the algorithm calculation result is close to the actual situation with a small error, so the experimental error fluctuates between 10–100 m.

The positioning error range of the fusion positioning model is 5∼50 m. Because the fusion positioning model corrects the potential

Table 8
Road 2 model prediction performance evaluation.

Model selection	MAE	RMSE	R^2
Lasso regression model	11.533	18.279	0.563
Feedforward neural network model	10.363	16.264	0.598
wlp-XGBoost model	5.423	10.257	0.801

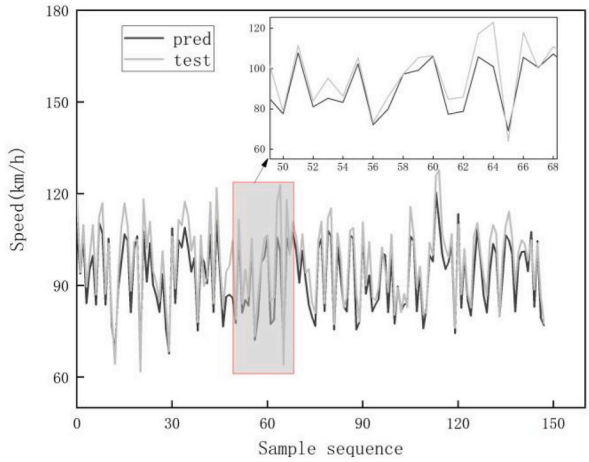


Fig. 18. Graph of Lasso regression prediction results in Road 2.

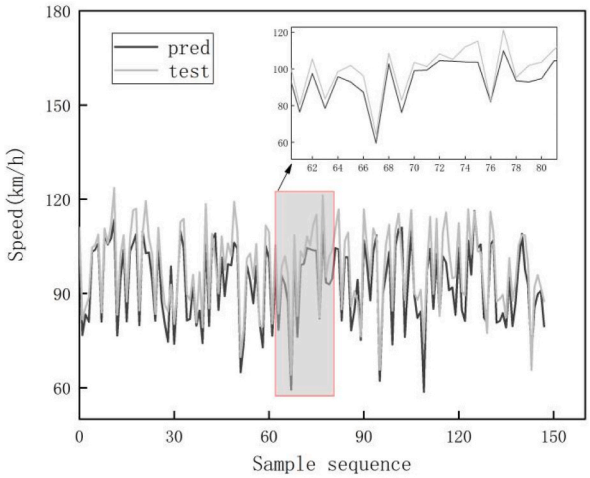


Fig. 19. Graph of feedforward neural network model prediction results in Road 2.

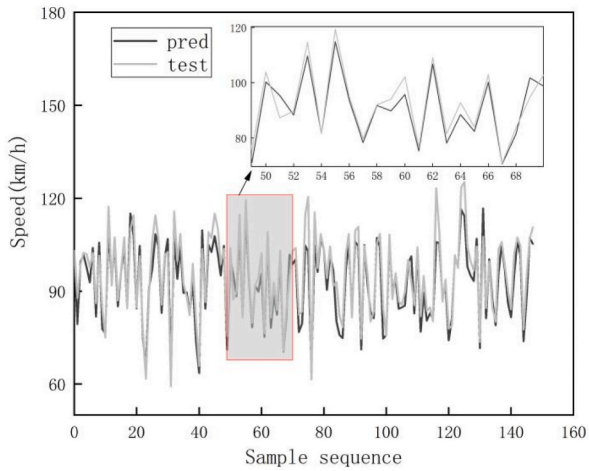


Fig. 20. Graph of wlp-XGBoost prediction results in Road 2.

(a) Experimental results of the DR algorithm

(b) Experimental results of the fusion positioning model

Fig. 21. Experimental analysis diagram of vehicle positioning.

position of vehicles, although the expressway network is relatively simple and can more accurately correct the trajectory of the drift road network to the road network, the potential position of vehicles is calculated according to linear motion, but the vehicles are not necessarily linear motion, so there is still some error between the vehicle position calculated by the fusion positioning model on the potential position and the actual vehicle position, but the error is small.

5.3.3. Analysis of algorithm robustness

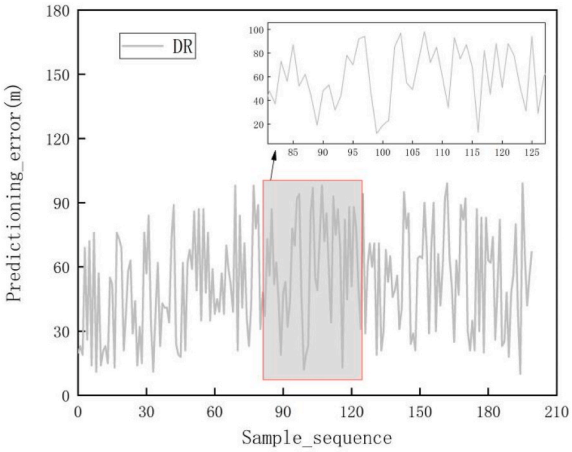
To further verify the performance of the DR algorithm and fusion positioning model in expressway vehicle location and deviation correction, the location accuracy of its experimental results is analyzed in depth. Randomly sample the vehicle position data set calculated by the DR algorithm and fusion positioning model, respectively, and calculate the average positioning error between the vehicle position calculated by the DR algorithm and fusion positioning model and the actual vehicle position. The sample size is 100~1000, the sample interval is 10, and each data volume is repeated 100 times to calculate the average value. Study the error fluctuation evolution law of each method in the vehicle positioning process. The experimental results are shown in Fig. 23.

It can be seen from Fig. 23(a) and (b) that the average positioning error of the DR algorithm fluctuates greatly when the data volume is small. With the increase in the data volume, the average positioning error of the DR algorithm tends to be stable and fluctuates around 55.0 m. The average positioning error of the fusion positioning model fluctuates around 26.6 m, and the average positioning error is significantly lower than that of the DR algorithm, which verifies the effectiveness of the fusion positioning model. Compared with the DR algorithm, the location accuracy of the fusion positioning model is improved by 51.6%, and the location error tends to be stable with the increase of the amount of sample data, which shows that this method has high accuracy and strong robustness.

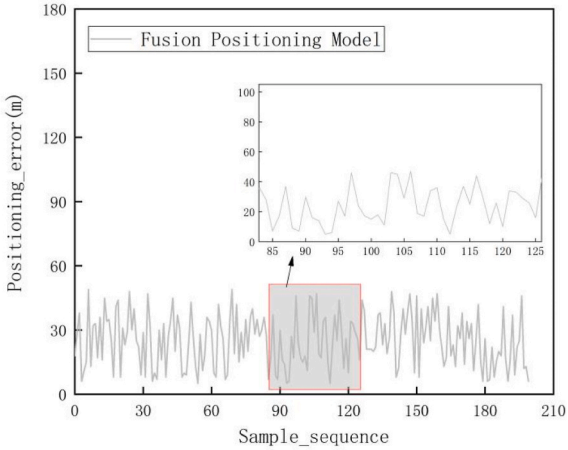
To sum up, there is an average error of about 20 m between the vehicle position calculated by the fusion positioning model and the actual vehicle position. However, this paper focuses on sensing the over-the-horizon information, and the positioning error of about 20 m will not affect the identification results and the safe driving of over-the-horizon vehicles. Therefore, the location accuracy and robustness of the fusion positioning model meet the performance requirements for the identification of vehicles with potential safety threats beyond the visual range.

5.4. Experiment of vehicle identification layer of potential safety threat under the over-the-horizon condition

After obtaining the real-time vehicle speed and vehicle position based on the vehicle section travel speed sensing layer and in-transit vehicle position estimation layer, it randomly enters the third layer, which is the vehicle identification layer of potential



(a) The DR algorithm positioning error statistics graph



(b) The fusion positioning model error statistics graph

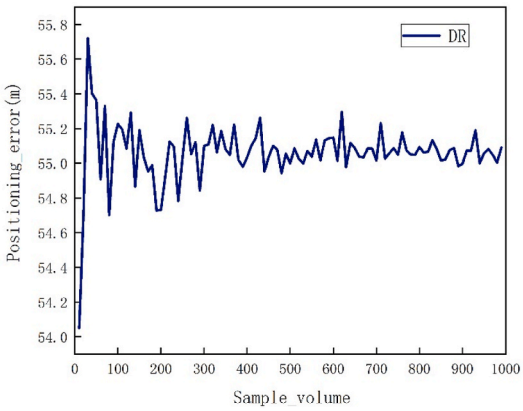
Fig. 22. Positioning error diagram.

safety threat under over-the-horizon conditions. Firstly, according to the current speed of the vehicle, the traffic volume of the section, the type of over-the-horizon vehicles, and the vehicle chasing time, different ranges of potential safety threat vehicles identification zones are divided so that the vehicles with potential safety threat to over-the-horizon vehicles can be discriminated more reasonably in different driving environments. Secondly, according to the potential safety threat speed discrimination method for all kinds of vehicles, the potential safety threat speed discrimination for the vehicles in the identification zone. Finally, potential safety threat vehicles are identified based on whether the vehicle is located in the potential safety threat identification zone and whether the vehicle speed meets the requirements for potential safety threat speed for each type of vehicle.

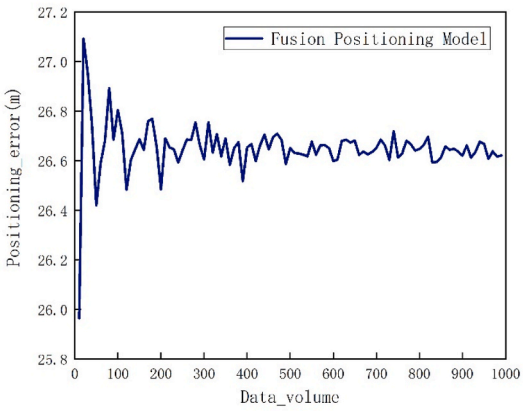
To better verify the effectiveness of the method, static and dynamic experiments were conducted. Static experiments reflect changes in the vehicles identified as potential safety threats as the identification zone expands for the same vehicle at the same moment in time; dynamic experimentation reflects the pattern of the number of potential safety threats identified at different times within a given identification zone as the scope of identification changes for the same vehicle. In the experiment, according to the proportion of each vehicle type on the expressway, we selected Class I vehicles, Class II vehicles, and Class III vehicles for the experiment, and the number ratio is 5:2:3. Among them, car1∼car5 are Class I vehicles, car6 and car7 are Class II vehicles, and car8∼car10 are Class III vehicles.

5.4.1. Analysis of experimental results of static identification of vehicles with potential safety threats beyond visual range

It can be seen from the static experiment Fig. 24 (a) to Fig. 24 (c), that with the expansion of the range of potential safety threat vehicle identification zone, the number of potential safety threat vehicles identified will also increase or remain unchanged. Among them, Car1 is the vehicle that enters this section at 5:35 a.m., and Car5 is the vehicle that enters this section at 9:10 a.m. It can be seen from the comparison of the potential safety threat vehicles in their identification zones that when the identification zone of the potential safety threat vehicles is 2 km, the number of potential safety threat vehicles identified by the two vehicles is not much different,



(a) Experimental diagram of the DR algorithm data sampling



(b) Experimental diagram of the fusion positioning model data sampling

Fig. 23. Graph of experimental results of data sampling.

but when the scope is expanded to 6 km, the number of potential safety threat vehicles identified by car5 is significantly more than that of car 1.

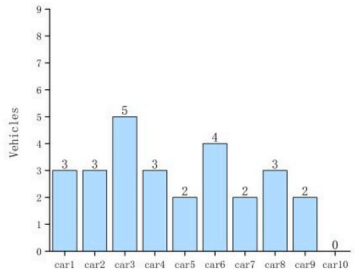
This is because the time when car1 entered this road is not within the peak traffic period, and the expansion of the identification zone does not lead to a sudden change in the number of vehicles with potential safety threats. While car5's driving time is the peak period of traffic volume. Because of the spatial characteristics of traffic volume, some sections are relatively congested. When the identification zone is 2 km and 4 km, the identification zone does not cover the area with more traffic volume. When the identification zone is expanded to 6 km, the identification zone covers the area with more traffic volume, resulting in a significant increase of potential safety threat vehicles in the car5 identification zone. Therefore, traffic volume is a key factor in the process of over-the-horizon identification.

Through the analysis of the three subgraphs in Fig. 24(a)–24(c), it is found that when the identification zone is 2 km, the average number of potential safety threat vehicles in the identification zone of various vehicles is 3.2 for class I, 3 for class II, and 1.67 for class III; When the identification zone is 4 km, there are 7 vehicles, 6.5 vehicles, and 4.33 vehicles respectively; When the identification zone is 6 km, there are 10 vehicles, 10 vehicles, and 6 vehicles respectively. The minimum value of the average identification number is always in class III vehicles, the maximum value is always in class I vehicles, and the median value is in class II vehicles. It can be seen that the identification process of vehicles with potential safety threats over the horizon is not only affected by the traffic volume on the road section but also the vehicle type is a key factor in the process of over-the-horizon identification.

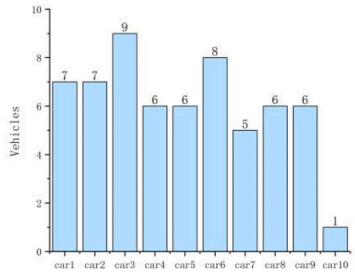
Therefore, it is reasonable to take the traffic volume and vehicle type as the identification zone division factor in the process of identifying vehicles with potential safety threats beyond visual range, and the vehicle speed is affected by the traffic volume to some extent, and should also be considered as an influencing factor.

5.4.2. Analysis of experimental results of dynamic identification of vehicles with potential safety threats beyond visual range

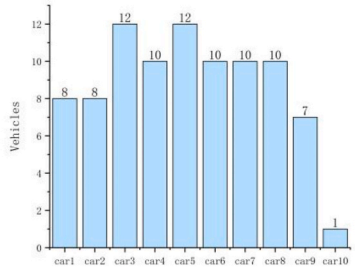
Potential safety threat vehicles in front of the over-the-horizon vehicle will be continuously overtaken by the over-the-horizon vehicle, but the potential safety threat vehicles detected by the over-the-horizon vehicle will not necessarily decrease over time, and may even keep increasing because new vehicles are constantly entering in the identification zone. To better verify the above



(a) Static state experiment diagram of 2km identification zone



(b) Static state experiment diagram of 4km identification zone



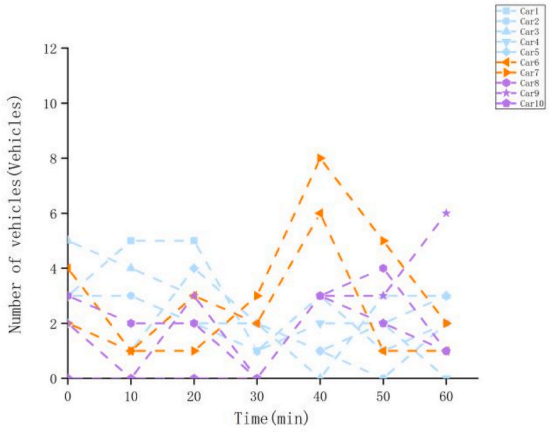
(c) Static state experiment diagram of 6km identification zone

Fig. 24. Static identification of Vehicles with Potential Safety Threats.

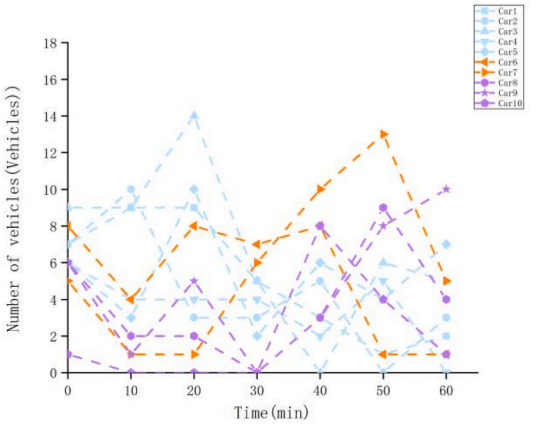
hypothesis, dynamic experiments on potential safety threat vehicle identification were conducted.

It can be seen from Fig. 25(a) to Fig. 25(c) that car10 did not detect a potential safety threat vehicle when it first entered the current roadway and the identification zone range was set to 2 km. As time goes by, when it has driven for 40 min, there are 3 potential safety threat vehicles in the identification zone; When the identification zone is extended to 4 km and 6 km, the number of potential safety threat vehicles is 8 and 10, and the number is increased significantly. By analyzing the time point of car10 entering the current road section, it was found that it entered the current road section at 7:50 a.m. when the traffic volume on the road was in a state of continuous climbing. When it drove for 40 min, the morning peak phenomenon appeared. Due to the uneven distribution of traffic volume, resulting in normal traffic volume on the road section where car10 is located, while its front 2–4 km area is more congested, it leads to a surge of potential safety threat vehicles in the corresponding identification zone. This also proves that when the traffic volume in the area where the over-the-horizon vehicle is located is not saturated, the identification zone can be adjusted one level larger, which can provide a reference for sensing the change in traffic dynamics of the road section. And whether there is congestion or an accident, it has an important impact on the car owner's driving safety, and early perception will help them to plan their speed reasonably.

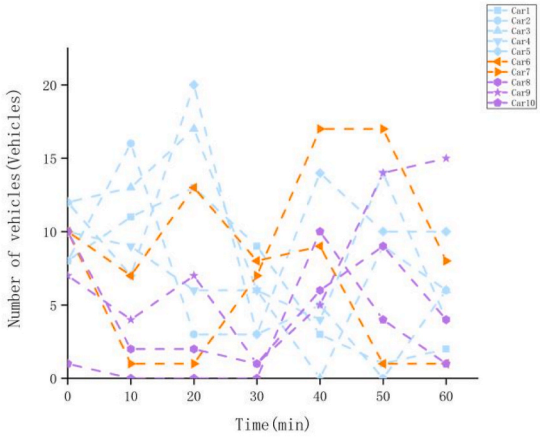
In addition, it can be seen from the experimental results of car7 that after driving for 40 min on this road, car7 itself came to a more congested area, and the number of potential safety threat vehicles in the identification zone surged. When the identification zone is 6 km, there are 17 vehicles in the identification zone, 10 vehicles in 4 km, and 8 vehicles in 2 km. Based on the analysis of the actual



(a) Dynamic experiment diagram of 2km identification zone



(b) Dynamic experiment diagram of 4km identification zone



(c) Dynamic experiment diagram of 6km identification zone

Fig. 25. Vehicle identification diagram of dynamic potential safety threats.

situation, when the vehicle itself is in a congested area, the main reason for the traffic accident is the rear-end collision and scrapes caused by the emergency stop of the vehicle in front or the emergency lane change, rather than the rear end collision caused by the lack of speed planning due to the perception of vehicles beyond visual range. Therefore, at this time, we should pay more attention to the traffic situation in the near distance. It is meaningless to identify in a wide range and interfere with the owner's judgment of the current environment. Therefore, it is reasonable to set the identification zone at 2 km.

In summary, when the area where the over-the-horizon vehicle is located is not in congestion, the identification zone should be adjusted one level larger under the same situation, and when it is in a traffic jam area, the identification zone should be adjusted one level smaller under the same conditions, verifying the rationality of the identification zone division.

Since vehicles enter the experimental section at different periods, their driving environment is quite different, so it is not practical to simply compare the number of potential safety threat vehicles in the identification zone at a certain moment, but should be analyzed based on the average value. From the average number of potential safety threat vehicles in the identification zone of each time node among the seven-time nodes in this section, when the identification zone is 2 km, 4 km, and 6 km respectively, the maximum number of potential safety threat vehicles identified is 2.86, 4.38 and 10.86 respectively, which are all generated in the class I vehicles; The median values are 2.71, 5.86 and 8.71, which are all generated in Class II vehicles; The minimum values are 0.86, 2 and 2.29, which are all generated in Class II vehicles. In the actual driving environment of the expressway, the overall speed of Class I vehicles is the fastest, followed by Class II vehicles, and Class III vehicles are the slowest. Therefore, there is a corresponding law for the number of potential safety threat vehicles corresponding to each type of vehicle. Therefore, the identification result of this method conforms to the actual driving law, which verifies the effectiveness of this method.

5.4.3. Experimental robustness analysis of over-the-horizon potential safety threat vehicle identification

Based on the identification results of the over-the-horizon static experiment and dynamic experiment, combined with the simulation experiment data generated by the ETC simulation platform, the identified potential safety threat vehicles are verified.

As shown in Table 9, it can be found by verifying the distance and speed percentages between the identified potential safety threat vehicles and the over-the-horizon vehicles in the static experiment of car1 when the identification zone is 2 km. The predicted distance between the three potential safety threat vehicles identified by Car1 and the over-the-horizon vehicle is 1.32 km, 0.98 km, and 0.65 km, respectively; the percentage of predicted vehicle speed and over-the-horizon vehicle speed are 85%, 70%, and 72%, respectively. The distance between vehicles indicates that the location of vehicles with potential safety threats is within the potential safety threat identification zone of car1, and the percentage of vehicle speed is less than 89%, which meets the requirement of potential safety threat speed for over-the-horizon class I vehicles. Therefore, there is no false identification in the static experiment with the car1 identification zone of 2 km。

From the static experiment of the 4 km identification zone of car9, as shown in Table 10, it can be seen that the potential safety threat vehicles identified by car9 are all within its identification zone. From the distance dimension, there is no false identification. However, from the perspective of vehicle speed, the predicted speed of vehicle 5 with a potential safety threat is 82% of the over-the-horizon vehicle speed, and the actual speed is 88% of the over-the-horizon vehicle speed. The actual speed does not meet the requirements of potential safety threat speed for over-the-horizon class III vehicles. Therefore, the potential safety threat vehicle 5 does not pose a potential safety threat to the over-the-horizon vehicle, which is caused by the misjudgment of this method in the speed dimension.

In addition, through the verification of simulation data, it is found that there is no missed identification event in the experiment. According to the experimental results of the wlp-XGBoost speed prediction model and the vehicle fusion positioning model, the main reasons are as follows: 1. The wlp-XGBoost speed prediction model can more accurately predict the vehicle speed, and the predicted speed mainly has a negative deviation from the actual speed, that is, the predicted speed is slightly less than the actual speed. This also leads to the method can reduce missed judgment when judging the speed of potential safety threats. 2. The vehicle fusion positioning model can estimate the vehicle position more accurately. Although there is a certain deviation between the estimated vehicle position and the actual position, its average deviation is only about 1% of the 2 km identification zone range and can be reduced to within 0.5% when the identification zone is expanded. Therefore, this also reduces the misclassification of this method in the vehicle distance dimension and ensures the robustness of this method for the combination of different types of vehicles and identification zones.

5.4.4. Experimental correctness analysis of over-the-horizon potential safety threat vehicle identification

Since there is a certain chance of analysis of identification results of a single vehicle, the overall identification correct rate of static experiment and dynamic experiment are counted separately, and the effectiveness of the method was further verified using the

Table 9
Identification Results of Static Experiment with Car1 Identification Zone of 2 km.

Vehicle ID	Predicting distance between cars	The actual distance between cars	Predicted vehicle speed to over-the-horizon vehicle speed ratio	Actual speed to over-the-horizon speed ratio
No. 1 potential safety threat vehicles	1.32 km	1.29 km	85%	86%
No. 2 potential safety threat vehicles	0.98 km	0.94 km	70%	72%
No. 3 potential safety threat vehicles	0.65 km	0.60 km	72%	74%

Table 10
Identification results of the static experiment with car9 identification zone of 4 km.

Vehicle ID	Predicting distance between cars	The actual distance between cars	Predicted vehicle speed to over-the-horizon vehicle speed ratio	Actual speed to over-the-horizon speed ratio
No. 1 potential safety threat vehicles	2.56 km	2.55 km	78%	80%
No. 2 potential safety threat vehicles	3.53 km	3.50 km	70%	72%
No. 3 potential safety threat vehicles	3.41 km	3.40 km	75%	74%
No. 4 potential safety threat vehicles	1.38 km	1.38 km	80%	81%
No. 5 potential safety threat vehicles	0.95 km	0.94 km	82%	88%
No. 6 potential safety threat vehicles	3.82 km	3.78 km	69%	73%
No. 6 potential safety threat vehicles	3.05 km	3.03 km	68%	69%

identification accuracy A and recall rate R. The calculation formula are shown in equations (29) and (30).

$$A = \frac{T_{r_{ac}}}{T_R} \tag{29}$$

$$R = \frac{T_{r_{ac}}}{T_{ac}} \tag{30}$$

Where T_{r_ac} is the actual number of vehicles with potential safety threats identified by this method, T_R is the total number of vehicles with potential safety threats identified by this method, and T_{ac} is the total number of vehicles with potential safety threats in the actual driving environment.

As shown in Table 11, the total number of vehicles with actual potential safety threats in static experiments with various combinations of identification zones and vehicle types in different environments is 180. The total number of potential safety threat vehicles identified by this method is 176, and among these 176 vehicles, the actual number of vehicles with potential safety threat is 171, with an identification accuracy of 97.1% and a recall rate of 95%. In the dynamic experiments, the total number of vehicles with actual potential safety threats in different environments with various combinations of identification zones and vehicle types is 1006, and the total number of vehicles with potential safety threats identified by this method is 980. And among these 980 vehicles, while the actual number of vehicles with potential safety threats is 963, the identification accuracy rate is 98.3% and the recall rate is 95.7%.

To sum up, the method proposed in this paper can accurately identify the vehicles that pose potential safety threats to over-the-horizon vehicles.

5.5. Real-time analysis of the response of the method

The ETC simulation platform generates real-time simulation transaction data based on ETC transaction data features, but since ETC transaction data only records information such as vehicle node passage time, vehicle unique identifier, and vehicle type. For the acquisition of vehicle speed and location information, it is necessary to dig deeper into the ETC transaction data to obtain them. The process of obtaining the speed and location information requires a certain amount of computation time, i.e., the response time delay of over-the-horizon potential safety threat vehicle identification using this method. Excessive time delay can lead to large positioning deviation of the vehicle and affect the effect of over-visual distance perception detection. Therefore, this section analyzes the response time delay and robustness of the proposed method to verify whether the method meets the real-time requirements.

5.5.1. Robustness analysis of identification method

As shown in Figs. 26 and 200 vehicles were randomly selected from the response time data of over-the-horizon potential safety threat vehicle identification, and the response time delay when using this method for over-the-horizon potential safety threat vehicle identification was counted. As can be seen from the figure, most of the time the response time delay of this method is within 250 ms, but there are a few vehicles with a response delay of more than 250 ms.

Analysis of the results revealed that vehicles with response delay greater than 250 ms detected significantly more vehicles with

Table 11
Statistical table of correct identification results.

	T_{ac}	T_R	T_{r_ac}	Accuracy	Recall Rate
Static experiments	180	176	171	97.1%	95%
Dynamic experiments	1006	980	963	98.3%	95.7%

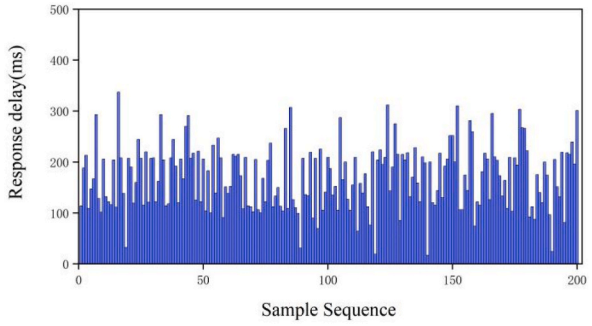


Fig. 26. Single-sample response latency statistics.

potential security threats than the remaining vehicles. In addition, combined with the analysis of its travel trajectory, the primary reason for the response time delay of more than 250 ms is that it travels in the peak period of expressway traffic, but due to the uneven spatial distribution of expressway traffic, the location of this vehicle itself is not in the core area of congestion, it can maintain normal speed, while its over-the-horizon identification zone covers the core area of congestion, the speed of vehicles in the core area of congestion is limited, therefore, the front of this vehicle is relatively Therefore, there are more relatively low-speed vehicles in front of this car relative to its speed, and the required calculation time is relatively longer. The second is the influence of the vehicle type, the speed of the class I vehicle is much larger than the class II and class III class vehicles, so the potential safety threat vehicles in the over-the-horizon identification zone will be relatively more. Finally, it is a small number of vehicles due to the driving behavior of the reason, its speed is fast, the over-the-horizon identification zone of the relatively low speed vehicles increased, and the required calculation time also becomes longer.

Regardless of the type of vehicles, whether the road is congested or not, and whether the vehicle driving behavior is good or not, the over-the-horizon potential safety threat vehicle identification using this method can sense and detect the potential safety threat vehicles within the over-the-horizon range on time with a response time delay of no more than 350 ms. Therefore, this method has good timeliness and meets the requirement of real-time over-the-horizon sensing and detection.

5.5.2. Identification method response time delay robustness analysis

The results of the single-sample latency analysis are subject to chance, so this subsection performs a multisample random sampling analysis on the response time data. As shown in Fig. 27, the response time data are randomly sampled from the response time data with the number of samples from 1000 to 50,000, the sample volume interval is 100, each sample volume is repeated 100 times and the average value is taken to obtain the average response latency. As can be seen from the figure, when the sample capacity is less than 14,000, the average response time delay has certain fluctuations, but the fluctuations are small, all within 1 ms. When the sample size is greater than or equal to 14,000, the average response delay tends to be stable, and it is maintained at 175.77 ms, basically not fluctuating with the change of sample size.

Since the samples are randomly sampled 1000–50,000 from all data, and each sample capacity is repeated 100 times, it is sufficient to ensure that the samples cover the expressway all-environment passage characteristics, i.e., the samples cover various combinations of road conditions in the actual operation of the expressway such as vehicle all-weather passage time, all-vehicle type, good and bad vehicle driving behavior, and road passage under each traffic volume. It can be seen that the average response delay of this method is small and stable under the full environment of the expressway, which has good robustness and meets the over-the-horizon perception detection requirements of the expressway.

In summary, the response delay of this method is small and the response delay remains stable in the whole environment of the expressway. The response latency of single sample sampling does not exceed 350 ms, and the average response latency is kept at about 175 ms in the diverse sampling experiments, which does not affect the experimental effect of this method, and the cost function-based multitasking scheduling algorithm MTS used in the simulation platform can guarantee the platform response time stability and small

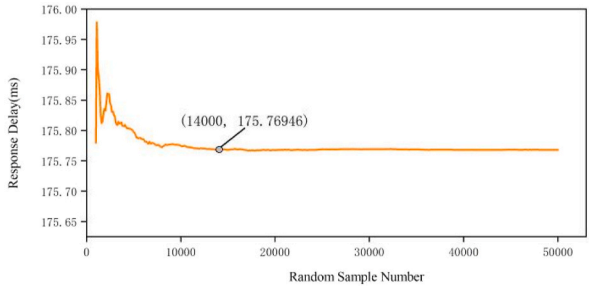


Fig. 27. Multi-sample average response delay statistics.

fluctuations. Therefore, the response delay in this paper is negligible.

6. Conclusion and future work outlook

The present study proposes a vehicle-road cooperative over-the-horizon potential safety threat vehicle identification method based on ETC big data for the current limitation that vehicles on expressways cannot perceive information beyond visual range. The main innovation points of this paper are as follows.

Firstly, in the vehicle section travel speed sensing layer, the wlp-XGBoost model is proposed. The wavelet packet transform is utilized for the decomposition and reconstruction of the speed set of the vehicle section to reduce noise in the data, and the noise reduction data is used for training to reduce the impact of noise and make the model fitting data more accurate; The XGBoost model was applied to solve the expressway spatial sequence regression problem, and good results were obtained by the wlp-XGBoost model.

Secondly, in the in-transit vehicle position estimation layer, the DR-HMM vehicle fusion model is proposed. Two gantries coordinate angles are used as the vehicle heading angle, and the DR algorithm is employed to calculate the potential position of the vehicle, considering that the expressway section tends to be straight and there is no significant curve; On this basis, taking into consideration the characteristic of fewer forks on the expressway, the Hidden Markov Model is used to rectify the potential position of vehicles, to achieve the accurate location of vehicle positions.

Thirdly, in the vehicle identification layer of potential safety threat under over-the-horizon condition, the potential safety threat vehicle identification zone is firstly determined based on the spatio-temporal characteristics of the over-the-horizon situation derived from the over-the-horizon vehicle type, the speed characteristics under the traffic volume dimension, and the vehicle catch-up time. Secondly, on the basis of considering the speed coordination and accident risk rate, the corresponding potential safety threat vehicle speed is set respectively by combining the type of over-the-horizon vehicles and relevant laws and regulations. Finally, according to the speed and distance between vehicles, it is judged whether the vehicles passing in the over-the-horizon potential interaction time domain pose a potential safety threat to over-the-horizon vehicles.

Last but not least, the most important innovation of this method lies in its longer sensing distance and wider vehicle coverage compared to LIDAR, millimeter wave radar, GPS, and other driving environment sensing solutions. Compared with the over-the-horizon perception scheme based on a high-precision map, its perception of over-the-horizon relatively low-speed vehicles is not possible by the over-the-horizon perception scheme based on a high-precision map.

This paper also has some shortcomings, that is, it does not take into account the potential safety threats that affect the vehicle identification results (such as gantry health, and vehicle access to service areas). In future work, we will further integrate the vehicle characteristics with potential safety threats, ETC gantry health, and other factors, to improve the identification method of vehicles with potential safety threats beyond visual range.

Author contribution statement

Guanghao Luo; Fumin Zou; Feng Guo: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper Xinjian Cai; Jishun Liu: Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper Qiqin Cai; Chenxi Xia: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

Data availability statement

The authors do not have permission to share data.

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.

References

[1] J. Weber, Route change on the American freeway system, J. Transport Geogr. 67 (2018) 12–23.
[2] WHO (World Health Organization), Global Status Report on Road Safety 2018, 2018 [Online]. Available: https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/. (Accessed 5 December 2019).
[3] F.L. Mannering, C.R. Bhat, Analytic methods in accident research: methodological frontier and future directions, Analytic methods in accident research 1 (2014) 1–22.
[4] Ministry of Public Health, Annual Epidemiological Surveillance Report, 2010 (Thailand).
[5] V. Ratanavaraha, S. Suangka, Impacts of accident severity factors and loss values of crashes on expressways in Thailand, IATSS Res. 37 (2) (2014) 130–136.
[6] Yeyao Sun, Wenjia Wan, Xinge Zhu, Yu Fengquan, Analysis and Prevention of Highway Traffic Accidents Traffic and Transportation, 2020 (S2), 112-114+117.
[7] PSM China, 2017. Compilation of Road Traffic Accident Statistical Data of the People's Republic of China, Traffic Management Bureau of the Public Security Ministry of the People's Republic of China, 2016.
[8] A. Chand, S. Jayesh, A.B. Bhasi, Road traffic accidents: an overview of data sources, analysis techniques and contributing factors, Mater. Today: Proc. 47 (2021) 5135–5141.
[9] T. Wedajo, E.T. Quezon, M. Mohammed, Analysis of road traffic accident related of geometric design parameters in Alamata-Mehoni-Hewane section, Int. J. Sci. Eng. Res. 8 (1) (2017) 874–881.

[10] S. Lee, B.Y. Jeong, Comparisons of traffic collisions between expressways and rural roads in truck drivers, Safety and health at work 7 (1) (2016) 38–42.
[11] Bundesministeriums der Justiz und f ur Verbraucherschutz,Straßenverkehrs-Ordnung, StVO, 2013.
[12] F. Guo, F. Zou, S. Luo, H. Chen, X. Yu, C. Zhang, L. Liao, Positioning method of expressway ETC gantry by multi-source traffic data, IET Intell. Transp. Syst. (2022).
[13] L. Zhao, Q. Wang, B. Jin, C. Ye, Short-term traffic flow intensity prediction based on CHS-LSTM, Arabian J. Sci. Eng. 45 (2020) 10845–10857.
[14] J.M. Chiou, H.T. Liou, W.H. Chen, Modeling time-varying variability and reliability of freeway travel time using functional principal component analysis, IEEE Trans. Intell. Transport. Syst. 22 (1) (2019) 257–266.
[15] F. Zou, Q. Ren, J. Tian, F. Guo, S. Huang, L. Liao, J. Wu, Expressway speed prediction based on electronic toll collection data, Electronics 11 (10) (2022) 1613.
[16] J. Shin, M. Sunwoo, Vehicle speed prediction using a Markov chain with speed constraints, IEEE Trans. Intell. Transport. Syst. 20 (9) (2018) 3201–3211.
[17] X. Lin, G. Zhang, S. Wei, Velocity prediction using Markov chain combined with driving pattern recognition and applied to Dual-Motor Electric Vehicle energy consumption evaluation, Appl. Soft Comput. 101 (2021), 106998.
[18] T. Bogaerts, A.D. Masegosa, J.S. Angarita-Zapata, E. Onieva, P. Hellinckx, A graph CNN-LSTM neural network for short and long-term traffic forecasting based on trajectory data, Transport. Res. C Emerg. Technol. 112 (2020) 62–77.
[19] D. Chen, X. Yan, X. Liu, S. Li, L. Wang, X. Tian, A multiscale-grid-based stacked bidirectional GRU neural network model for predicting traffic speeds of urban expressways, IEEE Access 9 (2020) 1321–1337.
[20] V. Havyarimana, Z. Xiao, A. Sibomana, D. Wu, J. Bai, A fusion framework based on sparse Gaussian–Wigner prediction for vehicle localization using GDOP of GPS satellites, IEEE Trans. Intell. Transport. Syst. 21 (2) (2019) 680–689.
[21] M.F. Tsai, P.C. Wang, C.K. Shieh, W.S. Hwang, N. Chilamkurti, S. Rho, Y.S. Lee, Improving positioning accuracy for VANET in real city environments, J. Supercomput. 71 (2015) 1975–1995.
[22] M.T. Abbas, M.A. Jibran, M. Afaq, W.C. Song, An adaptive approach to vehicle trajectory prediction using multimodel Kalman filter, Transactions on Emerging Telecommunications Technologies 31 (5) (2020), e3734.
[23] Yuexia Zhang, Liu Chong, Vehicle Location Method Based on Neural Network and Roadside Unit Fingerprint Computer Simulation, 2022, pp. 114–118, 04.
[24] Y. Ma, G. Wang, X. Hu, H. Luo, X. Lei, Cooperative occupancy decision making of Multi-UAV in Beyond-Visual-Range air combat: a game theory approach, IEEE Access 8 (2019) 11624–11634.
[25] Chaohui Hu, Yue Lv, An Xu, Threat Assessment Method of Over-the-horizon Targets Based on Situation Prediction Electro-Optical and Control, 2020, pp. 8–12+26, 03.
[26] An Xu, Chen Xing, Zhanwu Li, Hu Xiaodong, The Over-the-horizon Air Combat Situation Assessment Method Based on Tactical Attack Area Fire and Command and Control, 2020, pp. 97–102, 09.
[27] M. van Eenennaam, G. Heijenk, Providing over-the-horizon awareness to driver support systems, in: Proceedings of 4th IEEE Workshop on Vehicle to Vehicle Communications, 2008, June, pp. 19–25. V2VCOM.
[28] E.M. Eenennaam, Providing Over-the-horizon Awareness to Driver Support Systems by Means of Multi-Hop Ad Hoc Vehicle-To-Vehicle Communication (Master's Thesis, University of Twente, 2008.
[29] Fanqi Xu, Research and Simulation of Multi-Vehicle Cooperative BVR Cognitive Method (Master's Thesis, Beijing University of Posts and Telecommunications, 2021. https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD202201&filename=1021127386.nh.
[30] Zou Fumin, Guo Feng, Luo Sijie, Liao Lichao, Li Nan&Xing Yue Research and Design of Expressway ETC Simulation Platform Journal of System Simulation.
[31] S.S. Band, E. Heggy, S.M. Bateni, H. Karami, M. Rabiee, S. Samadianfard, A. Mosavi, Groundwater level prediction in arid areas using wavelet analysis and Gaussian process regression, Engineering Applications of Computational Fluid Mechanics 15 (1) (2021) 1147–1158.
[32] A.K. Karaev, O.S. Gorlova, M.L. Sedova, V.V. Ponkratov, N.S. Shmigol, S.E. Demidova, Improving the accuracy of forecasting the TSA daily budgetary fund balance based on wavelet packet transforms, Journal of Open Innovation: Technology, Market, and Complexity 8 (3) (2022) 107.
[33] P. Ding, X. Liu, H. Li, Z. Huang, K. Zhang, L. Shao, O. Abedinia, Useful life prediction based on wavelet packet decomposition and two-dimensional convolutional neural network for lithium-ion batteries, Renew. Sustain. Energy Rev. 148 (2021), 111287.
[34] O. Sagi, L. Rokach, Approximating XGBoost with an interpretable decision tree, Inf. Sci. 572 (2021) 522–542.
[35] A. Shehadeh, O. Alshboul, R.E. Al Mamlook, O. Hamedat, Machine learning models for predicting the residual value of heavy construction equipment: an evaluation of modified decision tree, LightGBM, and XGBoost regression, Autom. ConStruct. 129 (2021), 103827.
[36] A.I.A. Osman, A.N. Ahmed, M.F. Chow, Y.F. Huang, A. El-Shafie, Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia, Ain Shams Eng. J. 12 (2) (2021) 1545–1556.
[37] W. Li, R. Chen, Y. Yu, Y. Wu, H. Zhou, Pedestrian dead reckoning with novel heading estimation under magnetic interference and multiple smartphone postures, Measurement 182 (2021), 109610.
[38] B. Wu, C. Ma, S. Poslad, D.R. Selviah, An adaptive human activity-aided hand-held smartphone-based pedestrian dead reckoning positioning system, Rem. Sens. 13 (11) (2021) 2137.
[39] B. Li, W. Zhao, K. Zhou, L. Xu, G. Liu, T. Xiao, Z. Zhu, Research on location algorithm of mobile network based on hidden markov model, in: International Conference on Signal and Information Processing, Networking and Computers, Springer Nature Singapore, Singapore, 2022, September, pp. 938–946.
[40] Y. Hu, B. Lu, A hidden markov model-based map matching algorithm for low sampling rate trajectory data, IEEE Access 7 (2019) 178235–178245.
[41] G.D. Forney, The viterbi algorithm, Proc. IEEE 61 (3) (1973) 268–278.
[42] X. Ma, Q. Yu, J. Liu, Modeling urban freeway rear-end collision risk using machine learning algorithms, Sustainability 14 (19) (2022), 12047.
[43] K. Wang, X. Feng, H. Li, Y. Ren, Exploring influential factors affecting the severity of urban expressway collisions: a study based on collision data, Int. J. Environ. Res. Publ. Health 19 (14) (2022) 8362.
[44] C. Wang, F. Chen, Y. Zhang, J. Cheng, Analysis of injury severity in rear-end crashes on an expressway involving different types of vehicles using random-parameters logit models with heterogeneity in means and variances, Transportation Letters (2022) 1–12.
[45] C. Wang, M. Zhong, H. Zhang, S. Li, Impacts of real-time traffic state on urban expressway crashes by collision and vehicle type, Sustainability 14 (4) (2022) 2238.
[46] Z. Zheng, Empirical analysis on relationship between traffic conditions and crash occurrences, Procedia-Social and Behavioral Sciences 43 (2012) 302–312.
[47] S. Moridpour, E. Mazloumi, M. Mesbah, Impact of heavy vehicles on surrounding traffic characteristics, J. Adv. Transport. 49 (4) (2015) 535–552.
[48] K.K. Hyun, S.K. Mitra, K. Jeong, A. Tok, Understanding the effects of vehicle platoons on crash type and severity, Accid. Anal. Prev. 149 (2) (2021), 105858.
[49] JTG/T B05, Guidelines for Safety Evaluation of Highway Projects [S], 2004.
[50] Minwen Tang, Research on Standards and Methods of Expressway Speed Limit, Master's Thesis, Hunan University, 2008. https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD2008&filename=2008080578.nh.