

Article

Dynamic Identification Method for Potential Threat Vehicles beyond Line of Sight in Expressway Scenarios

Fumin Zou ^{1,2}, Chenxi Xia ^{1,2,*}, Feng Guo ^{1,2}, Xinjian Cai ², Qiqin Cai ³ , Guanghao Luo ^{1,2} and Ting Ye ^{1,2}

¹ Fujian Key Laboratory of Automotive Electronics and Electric Drive, Fujian University of Technology, Fuzhou 350118, China; fmzou@fjut.edu.cn (F.Z.); mapli@fjut.edu.cn (F.G.); 2211905072@smail.fjut.edu.cn (G.L.); 2211905021@smail.fjut.edu.cn (T.Y.)

² Renewable Energy Technology Research Institute, Fujian University of Technology, Fuzhou 350118, China; cxj2001@fjut.edu.cn

³ School of Mechanical Engineering and Automation, Huaqiao University, Xiamen 362021, China; 20011080002@stu.hqu.edu.cn

* Correspondence: 2211308044@smail.fjut.edu.cn; Tel.: +86-182-5259-2967

Abstract: Due to the challenge of limited line of sight in the perception system of intelligent driving vehicles (cameras, radar, body sensors, etc.), which can only perceive threats within a limited range, potential threats outside the line of sight cannot be fed back to the driver. Therefore, this article proposes a safety perception detection method for beyond the line of sight for intelligent driving. This method can improve driving safety, enabling drivers to perceive potential threats to vehicles in the rear areas beyond the line of sight earlier and make decisions in advance. Firstly, the electronic toll collection (ETC) transaction data are preprocessed to construct the vehicle trajectory speed dataset; then, wavelet transform (WT) is used to decompose and reconstruct the speed dataset, and lightweight gradient boosting machine learning (LightGBM) is adopted to train and learn the features of the vehicle section speed. On this basis, we also consider the features of vehicle type, traffic flow, and other characteristics, and construct a quantitative method to identify potential threat vehicles (PTVs) based on a fuzzy set to realize the dynamic safety assessment of vehicles, so as to effectively detect PTVs within the over-the-horizon range behind the driver. We simulated an expressway scenario using an ETC simulation platform to evaluate the detection of over-the-horizon PTVs. The simulation results indicate that the method can accurately detect PTVs of different types and under different road scenarios with an identification accuracy of 97.66%, which verifies the effectiveness of the method in this study. This result provides important theoretical and practical support for intelligent driving safety assistance in vehicle–road collaboration scenarios.

Keywords: expressway; ETC data; over-the-horizon; vehicle detection of potential threats; intelligent driving



Citation: Zou, F.; Xia, C.; Guo, F.; Cai, X.; Cai, Q.; Luo, G.; Ye, T. Dynamic Identification Method for Potential Threat Vehicles beyond Line of Sight in Expressway Scenarios. *Appl. Sci.* **2023**, *13*, 12899. <https://doi.org/10.3390/app132312899>

Academic Editors: Xinguo Jiang, Chuanyun Fu and Chuan Xu

Received: 19 October 2023

Revised: 27 November 2023

Accepted: 29 November 2023

Published: 1 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

At the end of 2022, the total mileage of China's expressways ranked first in the world, exceeding 170,000 km. Due to expressway fast travel speed and good road conditions, driving has become the first choice of transportation for people. However, expressways have brought people fast and efficient transportation [1], but also brought new challenges to expressway management department, such as serious traffic accidents, which have caused huge economic losses and casualties. According to the annual report of traffic accident statistics, the road traffic accidents that occurred nationwide in 2022 resulted in 83,085 unfortunate deaths, which is enough to raise our attention toward traffic safety. However, what is even more thought-provoking is that the number of fatalities in accidents on the expressway reached 6235, accounting for 7.5% of the total number of road traffic accidents, which is relatively high [2]. This further demonstrates the greater risks and more complex challenges inherent in expressway accidents and serves as a warning of the

harm of expressway accidents. Compared to general road accidents, expressway accidents have the characteristics of heavy casualties and large losses, which can affect the normal driving of surrounding vehicles, cause traffic congestion and accidents, seriously affect the safety of expressway, and cause huge losses to society. Therefore, it is necessary to explore possible prevention and improvement measures in order to attract deeper social attention to highway traffic safety and enhance the safety of expressway [3]. This is not only related to preserving lives, but also closely related to social stability and economic development. Through in-depth research and effective safety management, we are expected to reduce the occurrence of expressway traffic accidents and ultimately make a positive contribution to building a safe and efficient transportation system [4]. On the expressway, the overall speed of vehicles is usually high, and if the gap between the vehicle speed and the surrounding vehicle speed is too large, it may have a certain impact on the driving of the surrounding vehicles, which may lead to traffic accidents [5]. When driving on expressways, drivers can only observe the situation of vehicles behind them through their rear-view mirrors and cannot perceive danger beyond their sight range, resulting in the vehicle being unable to react in a timely manner and causing accidents. However, the construction cost of new expressway infrastructure is high, and there are many difficulties and challenges with regard to technology, capital, system, mode, and so on. Therefore, it is important to enhance the intelligent construction of expressways [6], utilizing advanced information and intelligent technologies to comprehensively, quickly, and accurately collect and process expressway information, transmit it to corresponding drivers, reduce traffic accidents, and improve the operational efficiency of expressways [7].

At present, electronic toll collection (ETC) is one of the effective methods to improve the service level of expressways [8]. With the large-scale popularization of ETC terminals and the massive deployment and application of ETC gantry systems, as of the end of June 2022, a total of 28,318 ETC gantries and 85,277 toll lanes have been built in 29 provinces across the country, including 37,947 ETC-dedicated lanes. The number of ETC users has risen to over 260 million, providing strong data support for expressway over-the-horizon safety detection. At present, many scholars have developed research on intelligent driving based on ETC data, including dynamic speed limit recognition on expressways [9], traffic flow prediction on expressways [10,11], and vehicle speed prediction [12], among other related research [13]. Through the exchange of information between vehicles, roads, and the cloud [14], the ETC gantry of highways is used to collect vehicle traffic information, and the intervention of cloud platforms can process and analyze large-scale vehicle data and road condition information, thereby detecting potential threat factors outside the driver's line-of-sight range and transmitting this warning information to corresponding vehicles subscribing to beyond-line-of-sight services. This enables drivers to comprehensively assess the surrounding environment, perceive potential hazards in advance, and minimize the potential risk of traffic accidents [15]. For example, if the vehicle behind approaches the intersection quickly, the system can alert drivers who have subscribed to the beyond-line-of-sight service in advance to avoid potential intersection accidents. This not only helps to improve the decision-making level of drivers, but also lays a solid foundation for the development of future intelligent driving assistance systems.

The structure of this article is as follows. The second section summarizes the current domestic and international research status. The third section introduces the description of the problem and provides relevant definitions. The fourth section constructs an expressway over-the-horizon PTV detection algorithm. The fifth section provides the results and analysis of the experiments. The sixth section summarizes the entire text.

2. Related Work

When driving on an expressway, it is crucial to perceive the road conditions behind you for driving safety. PTV detection in the over-the-horizon range not only plays a significant role in driving safety, but also in avoiding traffic accidents. At present, vehicle safety detection for the over-the-horizon range is a challenging emerging research field,

and research related to the over-the-horizon range mainly focuses on the air and sea. For example, Xu et al. [16] proposed an over-the-horizon air combat situation assessment method that combines air-to-air missile tactical attack areas, reflecting the true situation characteristics of over-the-horizon air combat; Merz et al. [17] proposed a perception and guidance system based on LiDAR, which enables unmanned helicopters to independently complete obstacle detection and avoidance, terrain tracking, and close-range inspection tasks beyond the line of sight; Guo et al. [18] proposed a clutter cancellation method based on FFT phase analysis using over-the-horizon radar to detect ships under the condition of short accumulation. However, there is relatively little research on the detection of threatening vehicles on expressways.

At present, there are many related studies on abnormal driving detection, which mainly use video surveillance data, which is data collected by sensors and communication devices at vehicle terminals. For example, Wang et al. [19] used a lightweight gradient elevator algorithm based on video taken by UAV to detect abnormal driving. However, in the case of imperfect video monitoring or bad weather on the expressway, video data is difficult to accurately detect driving, and detection of PTVs on the expressway is proportional to hardware requirements. Therefore, it is difficult to achieve the expected detection outcome. Zhou et al. [20] used the kernel method and an extreme learning machine to detect the abnormal driving of vehicles by acquiring acceleration sensor data, gyroscope sensor data, and magnetic field sensor data in the vehicle's associated smart phone. However, it only detects abnormal driving of individual vehicles and does not upload information to the cloud, making it impossible to transmit information to other vehicles around the road network and achieve information sharing. Hui et al. [21] used the extended neural network detection model of bidirectional long short-term memory network (BiLSTM) and fully connected neural network (FC) to detect abnormal driving using vehicle driving data collected from vehicle terminals. However, the operation of sensors and communication devices will be affected by various factors, which leads to various problems in the quality of data. Moreover, there are also disadvantages such as incomplete sample data and high sparsity in this method. ETC data, on the other hand, has the advantages of full samples, accuracy, time series, and good data quality. The use of ETC data to detect PTVs is of great significance for future expressway research.

In order to enhance the precision of PTV detection and provide more accurate traffic information, it is important to predict the speed of vehicles on the road in real time. At present, many domestic and international researchers have conducted research in the field of traffic prediction. The main traffic prediction methods include statistical models, deep learning models, and traditional machine learning models. Statistical learning models predict future values based on historical data by analyzing historical time series. Liu et al. [22] used the autoregressive integrated moving average model to predict railway passenger flow. The statistical model is suitable for simple and stable data, while the changes in expressway traffic conditions are significant and require too much calculation. In addition, it can only capture linear relationships, while traffic speed data has nonlinear characteristics. The deep learning model demonstrates its high prediction accuracy and extensive prediction ability. The deep learning model has strong learning ability and can automatically extract features and capture data correlations. Zhao et al. [23] proposed a temporal graph convolutional network (T-GCN) for traffic prediction, in which GCN and GRU are combined to capture the spatiotemporal characteristics of traffic flow, which conforms to the spatiotemporal correlation of expressway networks and the time series of ETC transaction data. Huang et al. [24] proposed long-term and short-term graph convolution network (LSGCN), which can simultaneously meet long-term and short-term prediction tasks. This framework proposed a new graph attention network, called cosAtt, and integrated cosAtt and GCN into a spatial gating block. Through spatial gated blocks and gated linear unit (GLU) convolution, LSGCN can effectively capture complicated spatiotemporal characteristics and obtain steady forecast results. Han et al. [25] propose a dynamic graph construction method to learn the temporal and spatial dependencies of

road segments. Then, a dynamic graph convolution module is proposed, which aggregates the hidden states of adjacent nodes to the focus node through message passing on the dynamic adjacency matrix. In addition, multiple fusion modules are provided to combine the auxiliary hidden states learned from traffic volume with the main hidden states learned from traffic speed. The experiments have shown relatively advanced effects. Although deep learning models have been widely applied in the field of traffic prediction, they still face some limitations, such as a large number of hyper-parameters leading to complex models, high computational costs, and the need for a large number of datasets during the training process. Compared to machine learning, it has better explanatory power, but it is not effective when dealing with large amount of data. Tong et al. [26] proposed a model of particle swarm optimization (PSO) combined with support vector regression (SVR) to predict traffic flow. The SVM model has high requirements for data pre-processing and parameter adjustment and cannot perform well for prediction tasks with complex regularity and complex factors. Yang et al. [27] used the k-nearest neighbor (KNN) algorithm to predict short-term traffic flow, which has a large computational load for datasets with large sample sizes. At the same time, when the samples are imbalanced, the prediction deviation is relatively large, and other algorithms need to be used to balance the dataset.

However, Microsoft Research has proposed a framework, LightGBM [28], to implement the gradient lifting decision tree algorithm, which is fast in operation, low in memory consumption, and high in model accuracy; supports parallel training; can process massive data; and is widely used in traffic prediction. Xia et al. [29] proposed a new traffic prediction model based on an integrated framework of bagging and LightGBM. Wang et al. [30] proposed a short-term driving time prediction model based on LightGBM. The experiment showed good advantages in prediction accuracy and training speed compared with a KNN model and gradient boosting decision tree (GBDT) model.

The use of speed prediction models to predict the speed of vehicles in transit as a feature of PTV detection has been incorporated into PTV evaluation algorithms. In driving threat assessment, some researchers have begun to apply improved mathematical models in this field. Representative analysis methods include Gaussian mixture model (GMM), analytic hierarchy process (AHP), and principal component analysis (PCA). Yang et al. [31] uses the steering wheel angle and deviation from the lane centerline distance in the normal state to construct a Gaussian hybrid model, find the threshold of the maximum likelihood function in the normal state, and divide it into normal driving and fatigue driving, which can have a high recognition rate for fatigue state. He et al. [32] combined expert consultation method and AHP to establish an evaluation model of expressway risky driving elements, and systematically analyzed risky expressway driving behavior. Zhang et al. [33] used the entropy weight principal component analysis method to establish a driving safety evaluation model, and the experimental results have good reference value for expressway safety management.

3. Problem Description

There are four core questions that must be addressed to achieve over-the-horizon PTV detection on the expressway:

- Question 1: How do we process the transaction data generated by the ETC gantry?

An ETC gantry can generate a large amount of data in a short period of time. However, due to equipment failures, network issues, and other reasons, the generated data may have errors. How to effectively process this data to ensure its accuracy and reliability is one of the challenges we currently face. In addition, we also need to address the challenge of converting ETC transaction data into data suitable for PTV detection. This conversion process involves multiple technical challenges such as data format, accuracy, and timeliness, which require in-depth research and resolution.

- Question 2: How do we obtain the speed of vehicles traveling in transit?

In the ETC system, vehicles only generate transaction records when passing through the ETC gantry, so we cannot calculate the vehicle's speed between the entrance gantry of a certain section and an exit gantry they did not pass. However, as one of the core features of PTV detection, vehicle speed is crucial for accurately evaluating the safety of vehicles. Therefore, we need to study how to obtain speed information of vehicles traveling on the road to elevate the precision of the evaluation.

- Question 3: How do we effectively evaluate the positions of vehicles in transit?

Due to the deployment intervals of ETC gantry, ranging from a few kilometers to more than ten kilometers, and the fact that these gantries only collect data when the vehicle passes through them, they cannot continuously track the vehicle's position throughout the entire driving process. Therefore, in some cases, it is difficult to accurately evaluate the position of the vehicle.

- Question 4: How do we effectively detect PTVs beyond the line-of-sight range?

Effectively detecting PTVs beyond visual range helps drivers take timely preventive measures to avoid traffic accidents. However, when vehicles are traveling at high speeds, it is difficult to accurately detect them in a short period of time. In particular, complex road environments such as bends and obstacles may also interfere with the accuracy of the detection system.

4. Expressway PTV Detection Algorithm for Over-the-Horizon

4.1. Overview of the Overall Framework of the Algorithm

This article uses the transaction data collected by ETC gantries for PTV detection beyond the line of sight in the rear. The algorithm is mainly divided into four modules: data preprocessing module, vehicle speed prediction module, vehicle positioning module, and PTV evaluation module. Figure 1 shows the overall framework structure of over-the-horizon detection.

1. Data preprocessing module: First, clean the abnormal data, and then check the integrity of the data. Construct the driving trajectory of each vehicle, and then use equivalent conversion to calculate the traffic flow of each section at a certain time.
2. Vehicle speed prediction module: Due to the inability to calculate the current driving speed of the vehicle, we use a machine model to predict the driving speed of the vehicle in the current section. The algorithm first applies WT to the speed data of the vehicle's historical section for noise reduction. Then, the LightGBM model is used to train and learn the speed characteristics of historical vehicle sections, and the final output is the predicted speed of the vehicle in the current section.
3. Vehicle positioning module: Set the gantry spacing through the ETC simulation system and use this spacing as the distance for over-line-of-sight detection. Through a comprehensive analysis of the characteristics of roads and vehicles, the rationality of the distribution of gantry structures is demonstrated, the detection range of over-the-line-of-sight is divided, and the position of vehicles in transit is refreshed through the transaction data time and location collected by ETC gantry structures.
4. PTV evaluation module: By extracting vehicle and road features from ETC data, the evaluation factors of the PTV are constructed. Finally, fuzzy set theory is used to detect the threat of a single vehicle driving, improving the accuracy of model judgment.

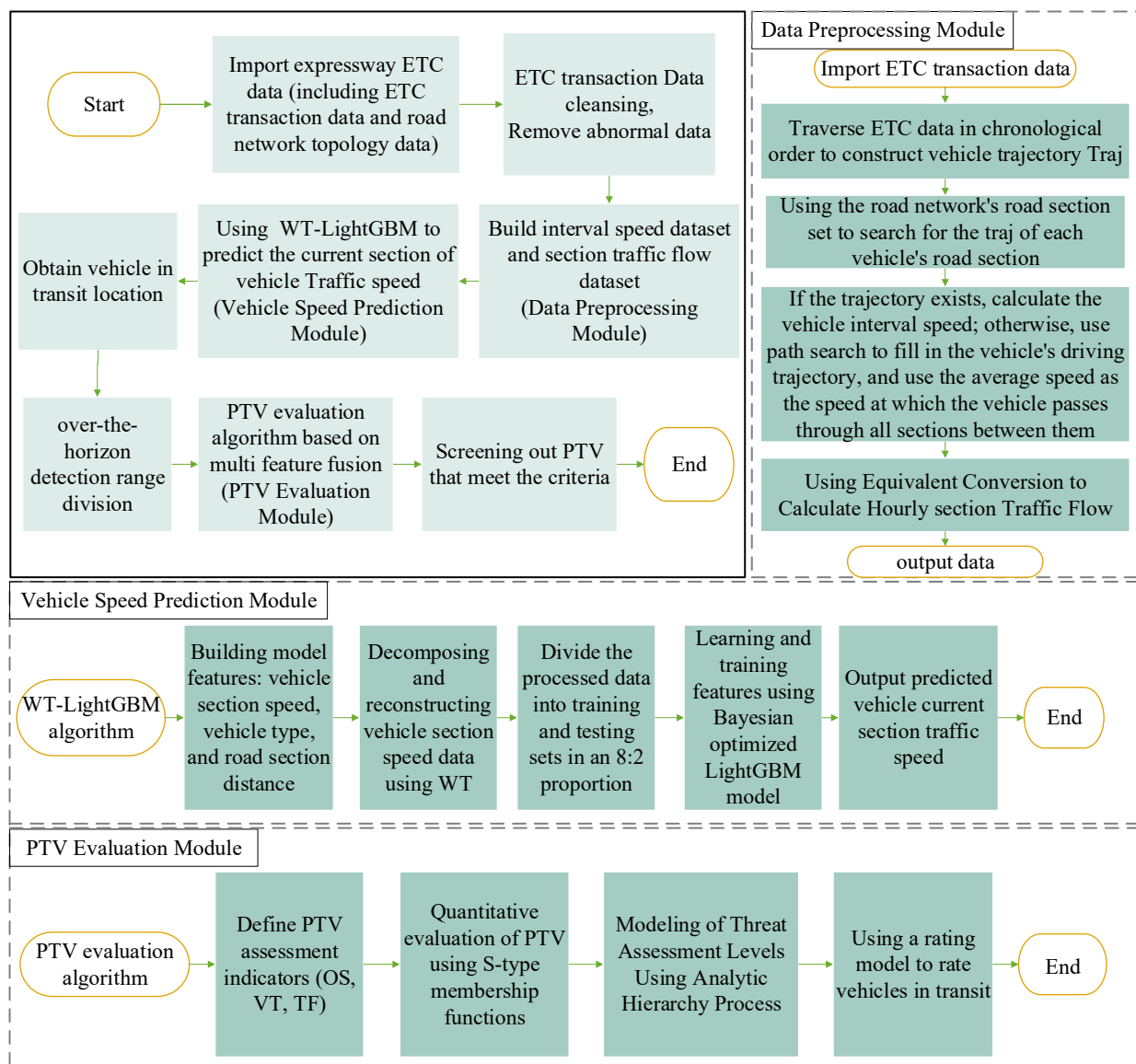


Figure 1. Overall composition of elements.

4.2. Related Definitions

To complete our method, the relevant definitions are given below:

1. ETC transaction data (EData): The EData of the expressway mainly comprises three fields: SID, VID, and TTime, which refer to toll station or gantry ID, vehicle ID, and transaction time, respectively. According to the SID field, the longitude (LNG) and latitude (LAT) of ETC toll station or gantry can be further determined. Based on the VID field, the type of the vehicle and MAC address of the vehicle's OBU device in the ETC transaction data can be further determined. In this paper, the toll station is regarded as a particular portal.
2. Expressway network (LW): All sections within the scope of this expressway research constitute the expressway network, abbreviated as the road network. The expressway network is shown in Figure 2, where blue indicates the expressway toll station and red indicates the expressway ETC gantry, and the blue line indicates the expressway network.
3. Expressway section (QD): Each gantry and toll station entrance and exit (including inter-provincial entrances and exits) of the expressway is called Node, and two contiguous Nodes constitute a section QD, referred to as a section: $QD = \langle \text{Node}_1, \text{Node}_2 \rangle$,

where Node₁ represents the beginning of the section and Node₂ represents the end of the section.

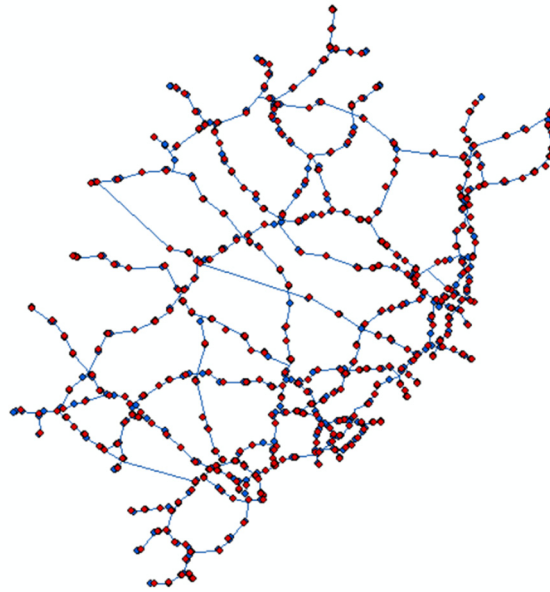


Figure 2. Schematic diagram of the expressway network.

A schematic diagram of an expressway section is shown in Figure 3.

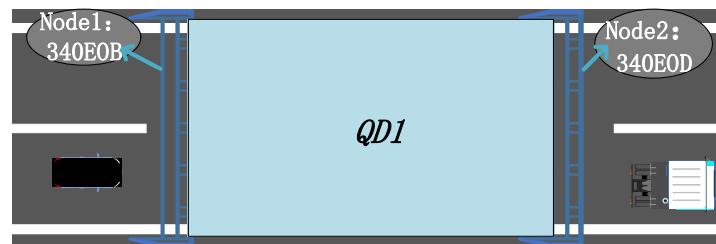


Figure 3. Schematic diagram of the section.

4. Travelling trajectory (Traj): The sequence of nodes formed by a particular vehicle on an expressway gantry is called Traj: Traj = <Node₁, ..., Node_n>, where Node₁ is called the starting point of the trajectory and Node_n is called the end point.
5. Vehicle speed (V_{car}): The average speed of a single vehicle passing through sections.

$$V_{car} = \frac{L}{\Delta t} \quad (1)$$

where L represents the actual length of the section and Δt represents the time difference between vehicles passing through a certain section.

4.3. Data Preprocessing

4.3.1. Cleaning of Abnormal Data

When a vehicle passes through the ETC gantry, a transaction record will be generated. Therefore, the ETC gantry system can generate a large amount of transaction data in a short period of time. However, due to various reasons, such as equipment failures, network issues, and so on, the transaction data may appear abnormal, which affects the accuracy of data analysis and decision-making results. In order to reduce the impact of abnormal data, the following situations need to be addressed:

1. Missing data: Vehicle transaction data is not captured effectively. For example, fields such as entry and exit station, transaction time, and vehicle model are missing.
2. Data redundancy: Multiple sets of data duplicate each other. For example, there are multiple sets of records of the same vehicle passing through the same ETC gantry.
3. Data Error: Data records that do not match the normal rules of the road, such as ETC gantries in different directions of travel, while capturing the crossing records of the same vehicle.

Therefore, in order to mine the effective information implied in the transaction data, it is necessary to carry out data cleansing, solve the outlier in the data, ensure the accuracy of the data, and support subsequent decision making and PTV detection.

4.3.2. Vehicle Speed Dataset Construction

After the preliminary cleaning of the vehicle transaction data collected by the ETC gantry system, construct the driving trajectory for each vehicle according to the chronological order; use the road section collection of the expressway network to conduct a road section search for the driving trajectory of each vehicle, traverse the two adjacent ETC gantries in the driving trajectory one by one, check whether the two adjacent gantries exist in the road section collection of the expressway network, and if they exist, calculate the vehicle travelling time Δt through the two gantries by the difference between the collected transaction times.

Then, based on the distance between the two neighboring gantries, calculate the speed of the vehicle passing through the road section V_{car} ; if it does not exist, perform a path search based on these two ETC gantries to fill in the vehicle's travelling track, calculate the average speed of the vehicle through the path based on the results of the search path, and take this average speed as the speed of the vehicle through all the sections between the two ETC gantries, thus obtaining the speed of the sections of the expressway network that each vehicle passes through in the process of travelling. In turn, the interval speed dataset of all vehicles passing through the road section is obtained. The specific construction method is shown in Algorithm 1.

Algorithm 1: Construction of the vehicle speed dataset

Input: Real-time vehicle trajectory data: Traj; expressway network topology dataset: LW.

Output: Vehicle travel speed

```

1. Traj = {Node1, . . . . . , Noden}, LW = {QD1, . . . . . , QDn}, QD = {Q, Distance};
2. FOR i = 0 to i = n - 1
3.   Nodei, Timei, Nodei+1, Timei+1 #Extract information from adjacent nodes;
4.   Δti = Timei+1 - Timei #Calculate the time difference between adjacent nodes;
5.   Q = (Nodei, Timei, Nodei+1, Timei+1) #Save the information of adjacent nodes;
6.   Δt = (Q, Δti) # Save vehicle travel time data;
7.   IF Nodei and Nodei+1 in LW # If adjacent nodes are in the topology data;
8.     Δt = Δt(Q, Δti) #Vehicle travelling time t data remains unchanged;
9.   ELSE Nodei and Nodei+1 not in LW;
10.    distance = { };
11.    {Nodei, . . . . . , Noden} ← shortest path LW #Search for the shortest path;
12.    {distancei, . . . . . distancen} ← LW # Extract the shortest distance between
    gantries;
13.    v = Distance/v # Calculate the passing speed of the front and rear gantries;
14.    {Nodei, . . . . . , Noden} = v #Add speed attribute to adjacent gantries;
15.    Δtj = distance/v #Calculate the time difference between adjacent nodes;
16.    Nodej, Nodej+1, Timej, Timej+1, Δtj #Extract information about adjacent nodes in
    shortest paths;
17.    Δt = (Q, Δtj) #Replace the original passage time and generate a new passage time;
18.  END IF
19.  Vcar = Speed acquisition(Q, Δtj); #Calculate vehicle passing speed
20. END FOR
21. RETURN Vcar

```

4.3.3. Construction of Section Traffic Flow Dataset

Section traffic flow refers to the number of vehicles passing through the same road section (in the same direction) within a certain period of time. In the statistics of traffic flow in the expressway sections, we use the converted coefficient traffic flow for subsequent dynamic PTV detection. Due to significant differences in the performance and space occupied by vehicles of different models on the road, their level of danger may also vary. When analyzing and comparing complex traffic flows, it is necessary to perform equivalent conversions for different vehicle models. The introduction of the equivalent conversion coefficient can enable vehicles of different vehicle types to be compared and counted in the same “unit”, making it easier to calculate traffic flow in sections.

According to the latest standard vehicle classification for expressway tolls in 2019, vehicles can be divided into three categories: passenger car, truck, and special operation vehicle, of which passenger cars can be divided into 4 subcategories, and trucks and special operation vehicles can each be divided into 6 subcategories, for a total of 16 models, which correspond one-to-one with the code names of the charging models in the ETC transaction flow data. Among these, codes 1–4 indicate four categories of passenger cars, 11–16 indicate six categories of trucks, and 21–26 indicate six categories of special operation vehicles. In order to carry out the conversion, it is necessary to determine the conversion coefficients for other vehicles based on the JTG B01-2014 Technical Standard of Expressway Engineering and the data provided by Fujian Expressway Information Technology Co., Ltd. (Fuzhou, China), with small passenger cars as the standard cars [34]. It should be noted that the calculation method of the equivalent conversion factor may be slightly different due to factors such as region and industry, which need to be based on local standards and the actual situation. Table 1 shows the conversion factors for different vehicle types for comparison and analysis.

Table 1. Expressway vehicle conversion factors table.

Fee Code	Vehicle Description	Category	Subcategory	Conversion Factor
1	H ≤ 9 persons, L < 6000 mm passenger car	passenger car	Class I	1
2	H = 10–19 persons and L < 6000 mm passenger car		Class II	1
3	H ≤ 39 persons and L ≥ 6000 mm passenger car		Class III	1.5
4	H ≥ 40 persons and L ≥ 6000 mm passenger car		Class IV	1.5
11	Z = 2, L < 6000 mm and M < 4.5 t	truck	Class I	1
12	Z = 2, L ≥ 6000 mm or M ≥ 4.5 t		Class II	1.5
13	Z = 3		Class III	3
14	Z = 4		Class IV	3
15	Z = 5		Class V	4
16	Z = 6		Class VI	4
21	Z = 2, L < 6000 mm and M < 4.5 t	Special operation vehicle	Class I	1
22	Z = 2, L ≥ 6000 mm or M ≥ 4.5 t		Class II	1.5
23	Z = 3		Class III	3
24	Z = 4		Class IV	3
25	Z = 5		Class V	4
26	Z = 6		Class VI	4

H represents the authorized number of passengers, L represents the length of the vehicle, M represents the maximum allowable total mass, and Z represents the total number of axles.

4.4. Section Vehicle Speed Prediction Algorithm

In order to more accurately evaluate the safety of vehicles in transit and improve the traffic safety of travelers, it is necessary to predict the traffic speed of vehicles. Due to the use of a formula to calculate the vehicle speed, it is necessary for the vehicle to pass through the front and rear gantries of the section. However, when the vehicle passes through QD and does not arrive, the platform cannot calculate the speed of the vehicle section. Therefore, it is necessary to predict the interval speed of all vehicles on the expressway. This article uses a lightweight gradient boosting machine (LightGBM) model based on WT to denoise the historical traffic speed of vehicles, and then combines it with the LightGBM model to predict the speed of the current section of the vehicle.

4.4.1. Wavelet Transform

The application of the ETC gantry on expressways provides convenience for predicting vehicle speed. However, due to various environmental and device-specific reasons, the collected data often contain a large amount of noise, which poses a great threat to speed prediction. In the process of analyzing real traffic flow, it was found that real speed signals often appear as low-frequency or stationary speed signals, while noise signals are more reflected in the high-frequency part [35]. Therefore, we use WT to filter the noise signal in the original data, so as to obtain more accurate section speed data. In order to separate the low-frequency and high-frequency parts from the original signal, Mallat proposed the concept of multi-resolution analysis in 1989, and based on this, he proposed a fast algorithm of discrete wavelet transform. The principle of the Mallat algorithm for wavelet decomposition and synthesis at a certain scale is shown in Figure 4.

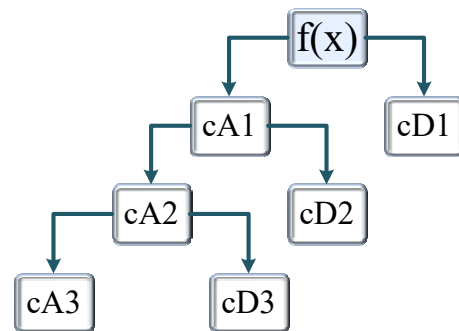


Figure 4. Schematic diagram of wavelet transform.

This algorithm can more accurately separate high-frequency noise and low-frequency signal components from the signal by analyzing it at different scale levels. By utilizing this algorithm, we can effectively filter out high-frequency noise signals while retaining low-frequency signals and obtain relatively accurate section speeds. The decomposition expression is as follows:

$$cA_{j,k} = \sum_n h(n - 2k) cA_{j-1,n} \quad (2)$$

$$cD_{j,k} = \sum_n g(n - 2k) cA_{j-1,n} \quad (3)$$

where $cA_{j,k}$ is the wavelet coefficient of the low-frequency part of signal $f(t)$ in layer j ; $cD_{j,k}$ is the wavelet coefficient of the high-frequency part of signal $f(t)$ in layer j ; and h and g are both filter coefficients, usually low-pass and high-pass filters. The original signal is decomposed into a series of wavelet bands. This process is accomplished by multiple layers of low-pass and high-pass filters, where the low-pass filter is used to extract the low-frequency components of the signal and the high-pass filter is used to extract the high-frequency components of the signal. In each layer of the decomposition process, the low-frequency portion is progressed and the high-frequency portion is used as the input for

the next layer of decomposition. This results in a series of wavelet coefficients in different frequency bands, representing local features of the original signal.

Reconstruction of the expression is seen as:

$$cA_{j-1,k} = \sum_n h(k-2n)cA_{j,n} + \sum_n g(k-2n)cD_{j,n} \quad (4)$$

Reconstruct the obtained wavelet coefficients into the original signal. This process is completed by multiple inverse filters and inverse upsampling operations. The inverse filter is used to transform the wavelet coefficients, while the inverse upsampling operation restores the splitting rate to its original size. In this way, similar results to the original signal can be reconstructed from the decomposed wavelet coefficients.

In the process of expressway speed prediction, we need to process a set of raw section speed sequence data with non-stationary characteristics. In this context, we chose the sym5 wavelet as the basis function mainly to balance the selection of data decomposition levels, as the decomposition levels are crucial for grasping the changes and trends of data. Specifically, an excessive number of decomposition layers may disrupt the inherent dynamic changes and development trends of the section speed sequence, while an insufficient number of decomposition layers may prevent the effective separation of different frequency features in the original section speed signal. Therefore, based on the current research results of wavelet transform sequence denoising [36], we set the number of decomposition layers to 3 to achieve the best signal extraction effect. Then, we apply a threshold function to process the high-frequency signal portion of each decomposition layer to reduce its impact on speed prediction. Finally, we reconstruct the low-frequency velocity signals from the bottom layer (third layer) and the high-frequency velocity signals from each layer after threshold processing to obtain the denoised section speed data.

Please refer to Algorithm 2 for specific implementation steps.

Algorithm 2: WT for noise reduction processing

Input: Raw section speed sequence data(input_data), decomposition levels, base Function (here 'sym5')

Output: Section speed sequence data after noise reduction(output_Data)

1: wavelet = Wavelet(base_function) # Initialize WT

2: decomposition = wavedec(input_data, wavelet, level = levels) # Using wavelet to perform levels wavelet decomposition on input_data and obtain the decomposition result

3: for i in range(1, len(decomposition)):

4: decomposition[i] = threshold_function(decomposition[i]) # Traverse the decomposition and apply threshold functions to the high-frequency parts of each layer

5: output_data = waverec(decomposition, wavelet) # Reconstruct the processed wavelet decomposition results into a signal (output_data)

4.4.2. Lightweight Gradient Boosting Machine Learning Model (LightGBM)

LightGBM is a GBDT algorithm developed by Microsoft, which is a variant of GBDT and is mainly used to solve problems encountered by GBDT when processing large amounts of data. LightGBM adopts a leaf-wise leaf growth strategy with depth constraints, which has low computational cost. By controlling the depth of the tree and the minimum amount of data for each leaf node, overfitting is avoided. As shown in Figure 5, select the leaf with the highest gain from all the current leaves each time for splitting until the gain is less than the given threshold or reaches the depth limit. This strategy can focus more on areas with large errors and improve the accuracy of the model but may lead to overfitting. In order to prevent overfitting, LightGBM has added depth restrictions to the leaf-wise growth strategy, allowing it to better prevent overfitting while maintaining the accuracy of the model.

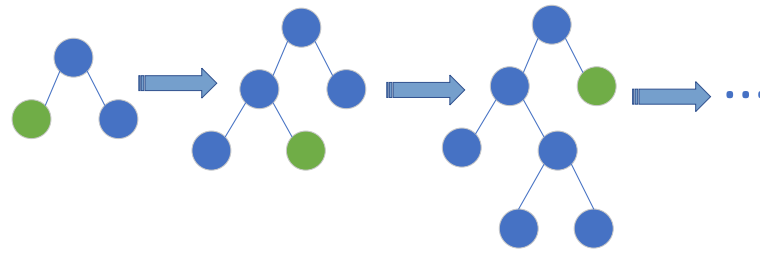


Figure 5. Decision tree strategy by leaf growth.

At the same time, LightGBM has chosen a decision tree algorithm based on histogram, which can reduce storage and computing costs. The decision tree algorithm based on histogram is an effective method to reduce data storage and accelerate model calculations. That is, the continuous floating-point eigenvalues are discretization into k integers and a histogram with a width of k . When traversing feature data, the discretization value is used as an index and as the cumulative statistics of the histogram. After traversing once, the histogram can accumulate corresponding statistics, and then search for the optimal splitting point based on the histogram. Through feature discretization, LightGBM can not only reduce storage consumption, but also reduce computing costs, and speed up training through cache hit rate optimization.

Therefore, for ETC with a large amount of data and many features, LightGBM has good processing ability. It can be well applied to predict continuous target variables, such as the predicted section vehicle speed in this article. When dealing with this problem, we take the vehicle speed and type in the historical section as the features, and the vehicle speed in the current section as the target variable, and then use the LightGBM model for training and prediction. In addition, the “vehicle type” and section features in the data belong to category features, and since LightGBM can directly process these features, this processing ability may bring significant performance improvements on ETC data. The advantage of LightGBM is particularly evident in cases where the category features in the ETC dataset are rich and have a significant impact on model performance. This not only simplifies the workflow of data preprocessing, but also improves the accuracy and efficiency of the model to a certain extent.

LightGBM uses a tree-based learning algorithm to form a powerful model by integrating the results of multiple weak learners. These weak learners are usually decision trees. In each iterative learning process, the model seeks a decision tree structure that minimizes the objective function. The objective function is composed of a loss function and regularization term. The loss function measures the gap between the predicted value and the true value of the model. For regression problems, the loss function is usually selected as mean squared error, and the formula is as follows:

$$L(y, y_v) = (y_i - y_{v_i})^2 \quad (5)$$

where y is the true value and y_v is the predicted value. The regularization term prevents the model from overfitting. For a tree model, the number of leaf nodes and the square sum of leaf node values are often used. The formula is as follows:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2 \quad (6)$$

where f is the tree model; T is the number of leaf nodes; ω is the value of leaf nodes; and γ and λ are the regularization parameters, which can affect the fitting degree of the model. Therefore, the objective function can be written as:

$$\text{Obj}(f) = \sum_i L(y_i, y_{v_i}) + \sum_k \Omega(f_k) \quad (7)$$

Finally, in order to optimize the performance of the model, we used a Python-based Hyperopt module to perform Bayesian optimization. This method gradually adjusts and narrows the parameter space based on the feedback information of the objective function, thereby efficiently searching and optimizing multiple parameters globally. In this way, we can determine a set of optimization parameters that can maximize overall performance. This method elevates the optimization process of traditional grid search and random search to a new level, making parameter adjustment more intelligent and efficient, thus enabling the optimal model configuration to be found in complex parameter spaces. The overall section vehicle speed prediction algorithm is implemented as follows (Algorithm 3):

Algorithm 3: Speed prediction algorithm for section vehicles

Input: vehicle history section travel time: $T []$, vehicle travel speed: $Vcar []$, vehicle type: $veh []$, section distance: $Dis []$;

Output: Vehicles in transit section travelling speed: V_{pre} ;

```

1. data = load_data(filepath) //import data;
2. denoised_data = wavelet_denoise(data) #Noise the data using the WT;
3. X_train, X_test, y_train, y_test = train_test_split(denoised_data) #Split the test set and the
   training set for the denoised data;
4. PARA= hyperopt(parameters) #Bayesian optimisation of parameters using hyperopt
   module;
5. model = lgb.LGBMRegressor() # Logic to build the LightGBM model;
6. train_model(model, X_train, y_train) # train the model;
7. predictions = predict(model, X_test) # predict data;
8. print(predictions) # Output predictions

```

The main objective of the algorithm is to predict the traffic speed of vehicles within a section, based on information including the traffic speed of vehicles on the historical road section, vehicle type, and road distance. First, in order to accurately capture the change of vehicle speed and eliminate the impact of noise, we use WT to denoise the original data. Next, we divided the processed data into a training set and a test set in an 8:2 ratio, and the training set data was used to train the LightGBM model. During the model training process, in order to find the optimal model parameter configuration, we adopted a Bayesian optimization strategy based on the hyperopt module. Through this adaptive parameter search strategy, the model can find the optimal parameter combination in a multi-dimensional parameter space, thereby improving the prediction accuracy of the model. Finally, the optimized model is used to predict the traffic speed of vehicles in the test set within the section. Our method realizes accurate prediction of vehicle speed by comprehensively using historical driving data, vehicle types, and road distance information, and provides an important decision-making basis for vehicle scheduling and management in intelligent transportation systems.

4.5. PTV Quantitative Evaluation Algorithm

In this section, we first define PTV, and then propose a PTV quantification method based on a fuzzy set to find PTVs from massive ETC data and rate the threat of each PTV.

4.5.1. Definition of PTV

The preprocessed ETC data can show information about the vehicles travelling in expressways. Therefore, in this work, the PTV is extracted from ETC data. In this paper, four typical metrics are defined based on ETC data.

Definition 1. (*Over Speed (OS)*) OS is one of the key factors leading to traffic accidents, indicating that if the current vehicle speed exceeds the speed of over-the-horizon vehicles, that is, if the speed is greater than V (for example, $V = 100$ km/h), it is considered as OS.

Definition 2. (Vehicle Type (VT)) VT represents the type of vehicle in transit. Different types of vehicles driving at the same speed on the expressway have different levels of danger. Therefore, it is necessary to determine the type of vehicle and conduct dynamic hazard ratings based on different vehicle types.

Definition 3. (Longtime Driving (LD)) LD represents the cumulative driving time of a vehicle on an expressway section. If the cumulative driving time of a vehicle is too long, it will lead to sleep-deprived driving of the driver, which will lead to traffic accidents. If the cumulative driving time exceeds a certain threshold, it is considered as LD. According to Chinese traffic regulations, this threshold is usually set to 4 h.

Definition 4. (Traffic Flow (TF)) TF represents the number of vehicles passing through the same road section (in the same direction) within a certain period of time. When the traffic flow in a section exceeds a certain threshold, it is considered more dangerous, such as $TF > 1500$.

Based on the above four indicators, if a vehicle has a large traffic volume, is of a type that is prone to accidents, has traveled for a long time on that road section, and is fast, then the vehicle may be considered a PTV. Of course, this also requires a comprehensive judgment based on other actual situations and data and cannot rely solely on a single indicator for evaluation. Although more indicators may improve the accuracy of detection, this will increase the difficulty of data collection and the cost of deploying and maintaining new equipment. For example, if a truck or specialized work vehicle is driving in a fast lane or frequently changing lanes, it will be detected as a PTV, but it is difficult to extract vehicle lane data from ETC data.

4.5.2. Determination of Membership Function

The evaluation of PTVs is subjective and ambiguous and cannot objectively measure whether vehicles are dangerous while in transit. In this paper, a quantitative method of PTV evaluation based on a fuzzy set is designed to accurately evaluate the vehicle hazard. The core idea of the evaluation model is to use fuzzy mathematics to make an overall evaluation of things or objects constrained by multiple factors. Fuzzy set theory was put forward as a classical concept of set theory [37] as early as 1965. The fuzzy set is an abstract concept, whose elements are uncertain and can only be determined by the membership function.

Firstly, the traffic state in the section is defined as a fuzzy set $\mu_A(x) \rightarrow [0,1]$, and the representation of the degree of affiliation adopts the vector representation. Then, determine the membership function and use the assignment method to determine the membership function. There is a positive correlation between OS, VT, LD, TF, and expressway accident rate. Based on the consensus that the faster the speed, the larger the vehicle type, the longer the cumulative driving time, the greater the traffic flow, and the higher the degree of threat, an S-type membership function was selected for the four indicators. Its calculation is shown in Formula 8, and its function diagram is shown in Figure 6.

S-type membership function:

$$\mu_A(x) = \begin{cases} 0, & x \leq a \\ 2\left(\frac{x-a}{b-a}\right)^2, & a < x \leq m \\ 1 - 2\left(\frac{x-b}{b-a}\right)^2, & m < x \leq b \\ 1, & x > b \end{cases} \quad (8)$$

where x is a measurement parameter of element A, such as vehicle traffic speed, section traffic flow, and so on, and element A represents OS, VT, LD, and TF. The closer the value of $\mu_A(x)$ is to 1, the higher the affiliation of x to the fuzzy set A, and the higher the vehicle danger. On the contrary, the closer the value of $\mu_A(x)$ is to 0, the lower the affiliation of x to

the fuzzy set A, and the lower the vehicle danger. If $\mu_A(x) = 1$, it means that x belongs to A completely, and if $\mu_A(x) = 0$, it means that x does not belong to A at all.

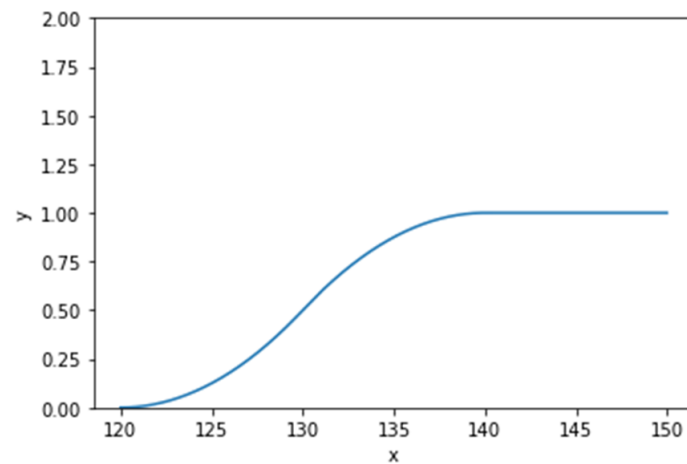


Figure 6. Schematic diagram of S-type membership function.

4.5.3. Construction of PTV Evaluation Algorithm

In order to better evaluate the PTV of expressway sections, a comprehensive comparative analysis was conducted on PTVs, and their severity ratings were evaluated. It is necessary to further determine the weight impact of different indicators on PTV threat assessment, that is, the degree to which different indicators have an impact on the results. The core of AHP is to apply a hierarchy to the problem to be analyzed. Based on the nature of the problem and the overall goal to be achieved, the problem is decomposed into different factors, and the factors are aggregated and combined at different levels according to their correlation, influence, and membership relationships, forming a multi-level analytical structural model. This method is applied to the multi-objective comprehensive evaluation method for hierarchical weight decision analysis.

In the field of intelligent driving, PTV threat assessment is regarded as a multi-objective comprehensive evaluation, and it is difficult to quantitatively describe what threat is. AHP is very suitable for this type of problem. Therefore, this article uses the structure of AHP to describe the modeling of threat assessment levels. Figure 7 is the analytic hierarchy process structure diagram.

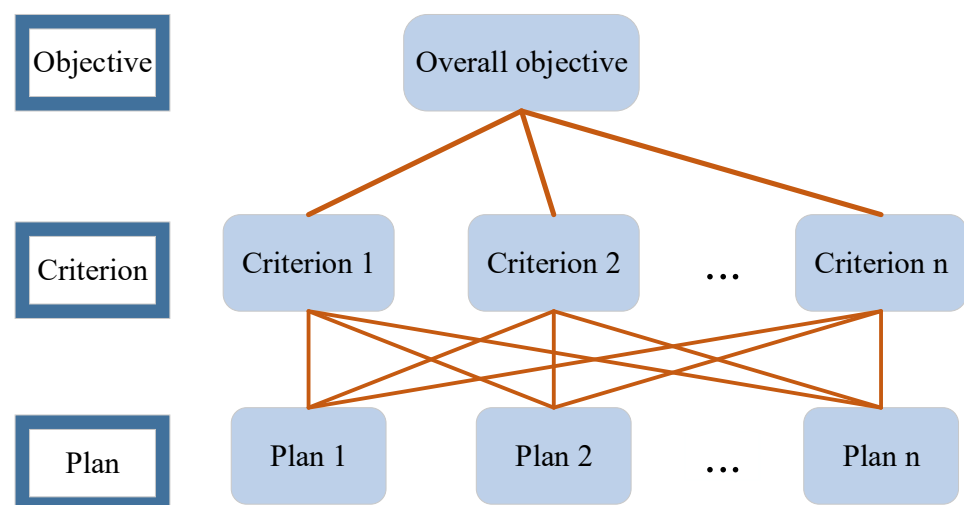


Figure 7. Analytic hierarchy process structure.

Due to the consideration of multiple characteristic indicators in PTV evaluation, it is not possible to accurately calculate the degree of influence of each factor. When assigning weights to obtain weight vectors, subjective factors account for a large proportion and can select the factors that have the greatest impact on danger. Corresponding weights can be given based on the importance of the indicators. Therefore, it is necessary to establish a judgment matrix to compare the importance of any two factors, avoiding the computational difficulty of comparing all factors together. Table 2 represents the quantitative criteria for the importance between elements, which divides the importance scale of elements into 1–9. The larger the importance scale, the more important the element is relative to another element.

Table 2. Proportional scale table.

Scale	Meaning
1	Equally important
3	One factor is slightly more important than the other
5	One factor is obviously more important than the other
7	One factor is strongly more important than the other
9	One factor is extremely more important than the other
2, 4, 6, 8	Intermediate value of two adjacent judgments
Reciprocal	A is compared with B; if the scale is 3, then B compared with A is 1/3

The judgement matrix $R = r_{ij}(n \times n)$ is constructed by filling in the values comparing the importance of two elements, as shown in Table 3.

Table 3. Judgment matrix.

Element	f_1	f_2	f_3	...	f_n
f_1	r_1/r_1	r_1/r_2	r_1/r_3	...	r_1/r_n
f_2	r_2/r_1	r_2/r_2	r_2/r_3	...	r_2/r_n
f_3	r_3/r_1	r_3/r_2	r_3/r_3	...	r_3/r_n
...
f_n	r_n/r_1	r_n/r_2	r_n/r_3	...	r_n/r_n

Note: $f_1, f_2, f_3, \dots, f_n$ represent the various types of indicators for evaluating PTVs after classification, and the elements r_{ij} of the matrix values are determined by the scale of the proportionality table.

After the construction of the judgement matrix is completed, the matrix needs to be tested for consistency, which refers to the permissible range of inconsistency determined by pairwise comparison of the matrices. The final weights can be obtained only after passing the consistency test of the judgement matrix. Calculate the consistency index (CI) with Formula (9), as shown below:

$$CI = \frac{\lambda_{\max} - n}{n - 1} \quad (9)$$

where λ_{\max} is the maximum eigenvalue of the matrix. When $CI = 0$, there is complete consistency; when CI approaches 0, there is satisfactory consistency. The larger the CI , the more severe the inconsistency. In order to measure the size of CI , a random consistency indicator RI was introduced, and the statistical data of RI is shown in Table 4. Finally, the consistency ratio CR is obtained and calculated as shown in Formula (10).

$$CR = \frac{CI}{RI} \quad (10)$$

Table 4. Random consistency index RI.

Matrix Order n	1	2	3	4	5	6	7	8	9
RI Value	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45

If the result of CR calculation is less than 0.1, the consistency verification is passed; if the contrary is true, consistency verification is not passed.

After determining the elements of the judgment matrix, in order to further determine the indicator weights of PTVs and facilitate calculation, the column vectors of matrix R are normalized to obtain the normalized matrix \hat{R} . The results are shown in Table 5. Finally, calculate the mean of the row vector in matrix \hat{R} to obtain the row vector ω , as shown in Formula (11), as the weight vector of the element.

$$\omega = \left[\frac{\sum_{i=1}^n \hat{R}_{1i}}{n}, \frac{\sum_{i=1}^n \hat{R}_{2i}}{n}, \dots, \frac{\sum_{i=1}^n \hat{R}_{ni}}{n} \right] \quad (11)$$

Table 5. Normalized judgment matrix.

Element	f_1	f_2	f_3	...	f_n
f_1	$\frac{R_{11}}{\sum_{i=1}^n R_{i1}}$	$\frac{R_{12}}{\sum_{i=1}^n R_{i2}}$	$\frac{R_{13}}{\sum_{i=1}^n R_{i3}}$...	$\frac{R_{1n}}{\sum_{i=1}^n R_{in}}$
f_2	$\frac{R_{21}}{\sum_{i=1}^n R_{i1}}$	$\frac{R_{22}}{\sum_{i=1}^n R_{i2}}$	$\frac{R_{23}}{\sum_{i=1}^n R_{i3}}$...	$\frac{R_{2n}}{\sum_{i=1}^n R_{in}}$
f_3	$\frac{R_{31}}{\sum_{i=1}^n R_{i1}}$	$\frac{R_{32}}{\sum_{i=1}^n R_{i2}}$	$\frac{R_{33}}{\sum_{i=1}^n R_{i3}}$...	$\frac{R_{3n}}{\sum_{i=1}^n R_{in}}$
...
f_n	$\frac{R_{n1}}{\sum_{i=1}^n R_{i1}}$	$\frac{R_{n2}}{\sum_{i=1}^n R_{i2}}$	$\frac{R_{n3}}{\sum_{i=1}^n R_{i3}}$...	$\frac{R_{nn}}{\sum_{i=1}^n R_{in}}$

Finally, after determining the weights of PTV evaluation indicators, it is necessary to calculate the threat of a single vehicle based on the obtained weights. This paper proposes a scoring model to evaluate the driving threat score (DTS) of vehicles driving on the road, as shown in Formula (12).

$$DTS = \left(\sum_{i \in A} \mu(i) \omega_i \right) \times 100 \quad (12)$$

where A is the PTV indicator of each item, $\mu(i)$ is the membership degree of the current indicator, and ω_i is the exercise weight of the indicator in Formula (11).

4.6. Over-the-Horizon PTV Perception Algorithm Construction

PTV information can be obtained by evaluating the threat of vehicles in transit through DTS. In the field of intelligent driving, for over-the-horizon vehicles, it is necessary to timely obtain PTV information within the rear over-the-horizon detection range. Therefore, it is necessary to update the position of vehicles in transit in real time and provide the information to the driver in a timely manner. However, due to the limitations of data collection conditions, we are unable to accurately locate the vehicle position in seconds. Only when the vehicle passes through the ETC gantry and generates transaction data can we obtain the current vehicle position information through the longitude and latitude of the gantry.

When the vehicle and the ETC gantry generate a transaction record, the transaction time of the vehicle can be obtained; then, the approximate travelling time of the vehicle in this section and the average speed of the vehicle can be calculated, so as to calculate the

distance ΔS between the over-the-horizon vehicle and the PTV, and the specific location of the vehicle can be further estimated. The calculation is shown in Formula (13).

$$\Delta S = |T_2 - T_1| \times |V_2 - V_1| \quad (13)$$

where T_1 and V_1 are the transaction time and average section speed of the PTV, while T_2 and V_2 are the transaction time and average section speed of the over-the-horizon vehicle. Providing PTV information within the over-the-horizon range to the vehicle ahead can enable the driver to make decisions in advance. The division of the perception of the over-the-horizon range will be elaborated in detail in the experimental section. Algorithm 4 shows the process of the over-the-horizon PTV perception algorithm.

Algorithm 4: Over-the-horizon PTV perception algorithm

Inputs: information on all vehicles in transit and DTS, constraints C , section transaction time DATA;

output: PTV in over-the-horizon range;

1. Known beyond line of sight detection range BVR (e.g., 2 km) # Define PTV evaluation indicators
 2. OS_car = [...] #Section travel speed of vehicles in transit;
 3. VT_car = [...] #Vehicle type of the vehicle in transit;
 4. LD_car = [...] #Accumulated travelling time for vehicles in transit;
 5. TF_car = [...] #Section traffic flow at the current moment
 6. PTVInfo[] #initialise the PTV list
 7. EvaluationInfo[] = DTS(OS_car, VT_car, LD_car, TF_car) #Evaluate the vehicle in transit, return the vehicle evaluation score
 8. FOR EACH car IN EvaluationInfo[] #Get PTV information
 9. IF car do not meet the constraints C then
 10. Do nothing
 11. ELSE
 12. PTVInfo.add(car)
 13. END IF
 14. END FOR
 15. RETURN PTVInfo[]
 16. Def AbIdentification(car[], PTVInfo[]){ #Identify PTVs in over-the-horizon range;
 17. FOR EACH car IN car[]
 18. AbCars[] = getEachCarBVR(car, PTVInfo[]) #Get PTV within over-the-horizon detection range
 19. AbnormalInfo.add(car, makeAlertInfo(T1,T2))
 20. END FOR
 21. FOR each ti in EvaluationInfo[]
 22. car = ti.getCar #Get information about the vehicle in transit;
 23. alertInfo = ti.getAlertInfo #Get information about the PTV message
 24. SendAlertInfoToCar(car, alertInfo) #send alert info to a vehicle in transit
 25. END FOR
 26. RETURN AbnormalInfo
-

5. Experimental Results and Analysis

The experimental platform is the Centos Linux 7.9.2009 (Core) operating system based on Intel(R) Core (TM) i9-10900K CPU @ 3.70 GHz and 64 GB RAM. All experiments are implemented on open-source web applications using the Python 3.8.8 version of Jupyter Notebook.

5.1. Data Source and Preprocessing

The data for this experiment are mainly divided into two categories. Firstly, they were collected by an expressway information technology company in a certain province through the ETC gantry system, with a total of 1581 and 2289 pieces of ETC transaction

data on 1 May and 1 June 2021, respectively. The original transaction data table contains 103 fields, recording information about various parts of the vehicle and gantry, including multiple information fields such as license plate number, gantry ID, transaction data, time, and gantry longitude and latitude. The parts are shown in Table 6. The * in the table was desensitized due to the privacy concerns of the data used in our experiment.

Table 6. Attribute table of ETC transaction data.

No.	Field Name	Description	Example
1	TRADEID	Vehicle identification	S0***1
2	TRADETIME	Transaction time	1 May 2021 20:00:01
3	FLAGID	Gantry ID	3502**
4	OBUID	Device MAC	66AD40**
5	ENTIME	Entrance time	3 September 2020 7:48:39
6	ENSTATION	Entrance toll station	46**
7	VEHCLASS	Vehicle type	1

Secondly, the Gaode Map API was used to crawl the distance of sections and generate topology relationship data of expressway ETC gantries, which included ETC gantries of different sections and the actual section distance. Some attributes are shown in Table 7.

Table 7. Gantry Distance Data attributes.

No.	Attribute Name	Description	Example
1	EnNodeID	Previous gantry no.	G001535*****10010
2	ExNodeID	Next gantry no.	G001535*****10011
3	Distance	Distance between gantries	5236 m
4	LNG	Longitude	118.56**
5	LAT	Latitude	24.85***

5.1.1. Vehicle Speed Dataset

After preliminary cleaning of the original ETC transaction data, the abnormal data caused by external factors is removed, and then the ETC data and topology data are combined to calculate the average speed of each vehicle segment, thereby constructing a expressway vehicle speed dataset. As shown in Figure 8, the dataset for the section of Fu-xia on 1 May 2021 is presented. Taking Figure 8a as an example, it represents the 24 h section speed data of QD_{340257–340259}, where the horizontal axis corresponds to every hour, the vertical axis represents the size of the interval speed, each box represents the overall distribution of vehicle speed within QD_{340257–340259} an hour, and the small black dots represent outlier speed data.

5.1.2. Section Traffic Flow Dataset

Calculate the section traffic flow based on the ETC data after initial cleaning, dividing time slices into 1 h intervals. One time slice includes the entire time process from start time to end time. Calculate the number of vehicles passing through the ETC gantry, which is the original traffic flow. Use the vehicle conversion coefficient to calculate the converted number of vehicles and record it as the converted traffic flow. The comparison of traffic flow before and after conversion is shown in Figure 9, where QD1–4 refers to different sections in the Fu-xia section of the expressway.

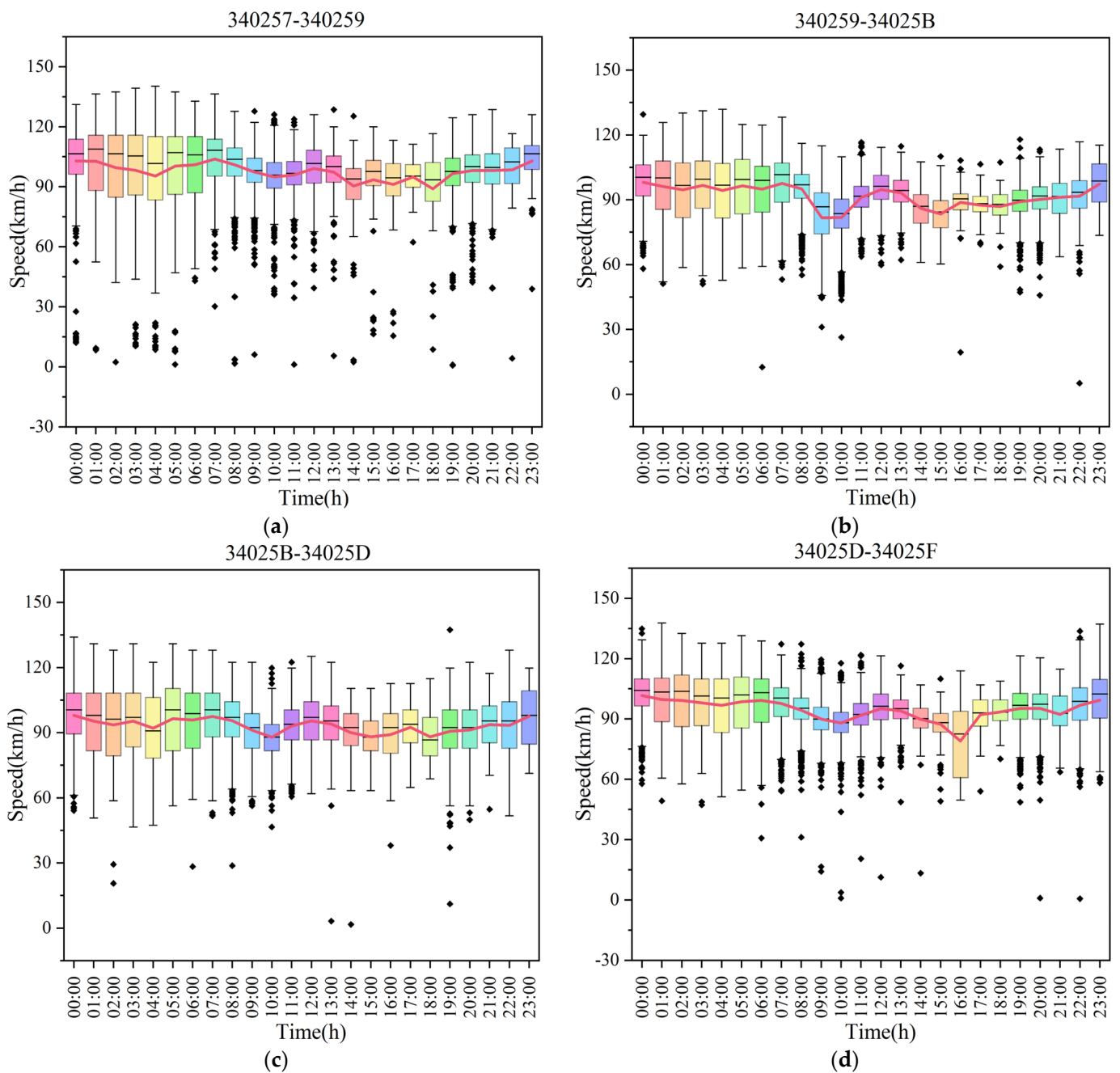


Figure 8. Distribution of interval speed status within 24 h. (a) The 24 h section vehicle speed data of QD_{340257–340259}. (b) The 24 h section vehicle speed data of QD_{340259–34025B}. (c) The 24 h section vehicle speed data of QD_{34025B–34025D}. (d) The 24 h section vehicle speed data of QD_{34025D–34025F}.

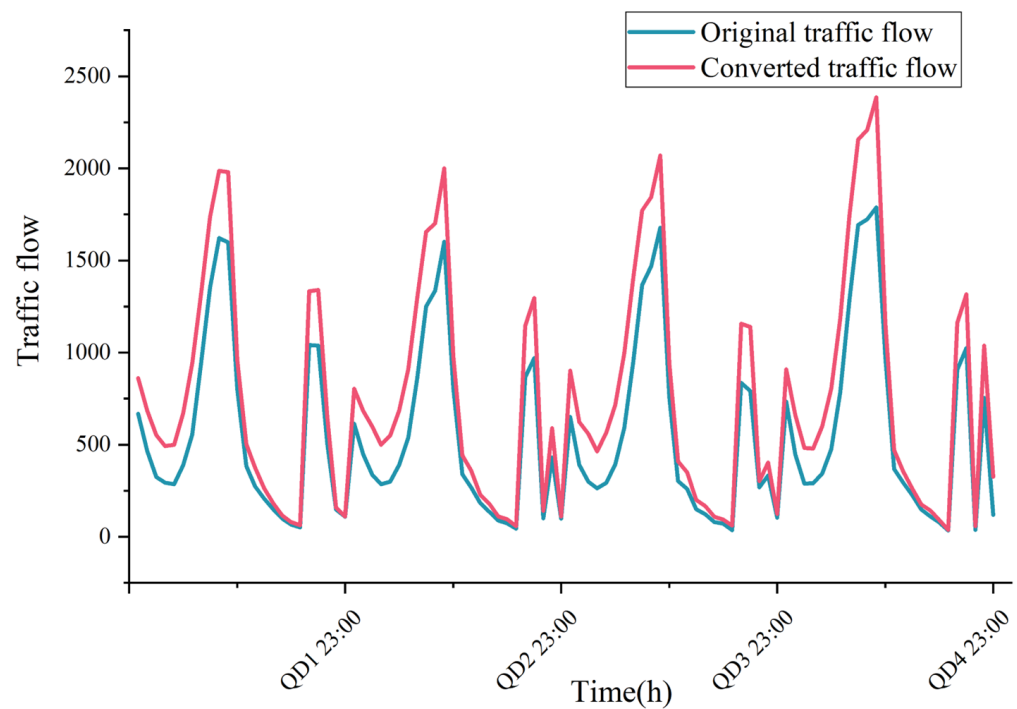


Figure 9. Comparison of traffic flow before and after conversion.

5.2. Experimental Study on Speed Prediction of Section Vehicles Based on WT-LightGBM

5.2.1. Evaluation Indicators and Parameter Settings

The experiment uses preprocessed data for in transit vehicle speed prediction. Firstly, WT is used to decompose and reconstruct the speed data of each section of the vehicle, and noise reduction is performed. Then, put it into the LightGBM model for training and learning. In order to evaluate the predictive performance of the model, the experiment uses root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2) as evaluation indicators. The metrics are defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2} \quad (14)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}| \quad (15)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (16)$$

where y_i is the actual vehicle speed, \hat{y} is the predicted vehicle speed, and N is the sample size. RMSE and MAE reflect the degree of deviation between the real value and the predicted value. The smaller the value, the better the quality of the model and the more accurate the prediction. R^2 is used to observe whether the forecast error is greater than or less than the mean reference error, and the larger the value, the better the forecast effect.

In this experiment, when using WT for data denoising processing, sym5 wavelet was used, with three decomposition layers. At the same time, the LightGBM model used the hyperopt module to perform Bayesian optimization on the parameters. The final determined model parameters are shown in Table 8.

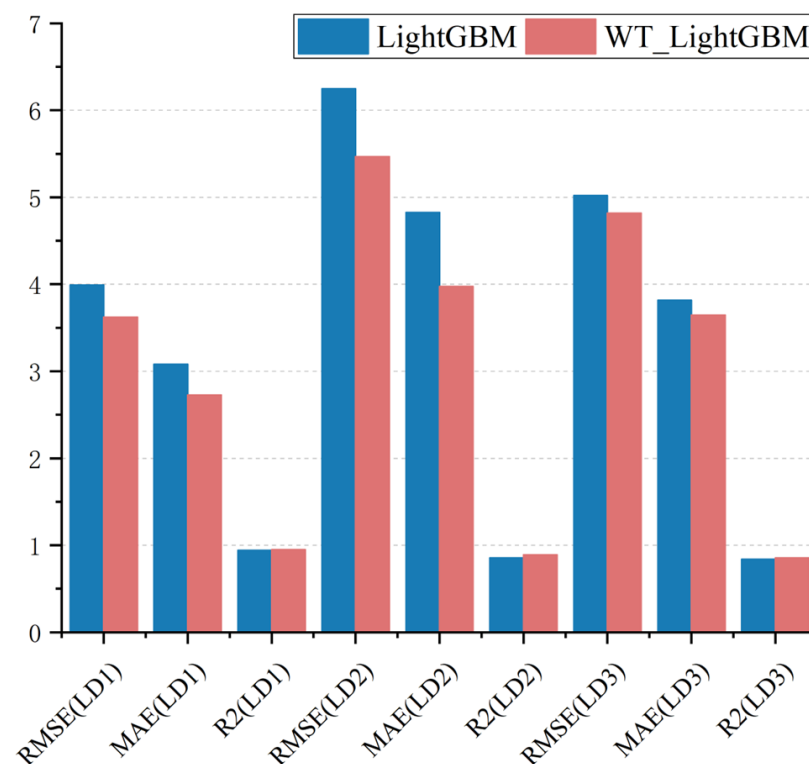
Table 8. Model parameter values.

Parameter Name	Value
learning_rate (step size per iteration)	0.1
n_estimators (number of iterations)	470
num_leaves (number of leaf nodes)	31
max_depth (the maximum depth of the tree)	7
Subsample (random sampling rate of samples per tree)	0.79
subsample_freq (sampling frequency)	8

5.2.2. Result Analysis

In order to verify the effectiveness and superiority of our proposed WT-LightGBM model in predicting the speed of vehicle sections, the experiment selected transaction data generated by ETC frames of three sections (LD) of Fujian Province's expressway on 1 May 2021 and 1 June 2021, for speed prediction experiments. Among them, LD₁ is the section with no service area for normal traffic flow on weekdays, LD₂ is the section with service area for normal traffic flow on weekdays, and LD₃ is the section with more traffic flow on holidays, and for sections with service areas, each section includes five sections.

The experimental results are shown in Figure 10. It can be clearly seen from the figure that the prediction results of the WT-LightGBM model are significantly better than those of the LightGBM model in any road segment. Among them, under the normal traffic flow and service area of LD₂, the RMSE and MAE of model WT-LightGBM are 12.46% and 17.63% lower than LightGBM, respectively, and the R² of model WT-LightGBM is 3.98% higher than LightGBM. This is because the LightGBM model did not consider the noise of the data, resulting in low prediction accuracy. It also proves that it is feasible to use WT-LightGBM to improve the accuracy of the prediction.

**Figure 10.** Comparative experiment between WT-LightGBM and LightGBM.

In order to verify the reliability of model WT-LightGBM, this paper also applies several different machine learning and deep learning models to test the data of these three road sections, including support vector regression (SVR), k-nearest neighbor (KNN), gate recurrent unit (GRU), random forest (RF), and XGBoost. Analyze the difference between the predicted value and the actual value, and calculate the RMSE, MAE, and R^2 of the prediction model. The prediction performance evaluation is shown in Tables 9–11.

Table 9. Comparison of LD₁ prediction performance evaluation.

Model	RMSE	MAE	R^2
SVR	6.1504	4.7862	0.8561
KNN	5.1107	3.842	0.9
GRU	5.2293	4.0131	0.8156
RF	4.3362	3.1638	0.9276
Xgboost	4.9316	3.9715	0.9075
WT-LightGBM	3.6233	2.7256	0.95

Table 10. Comparison of LD₂ prediction performance evaluation.

Model	RMSE	MAE	R^2
SVR	9.3619	7.423	0.6751
KNN	7.7573	5.4283	0.7769
GRU	5.8393	4.4038	0.7837
RF	6.9826	4.9885	0.8215
Xgboost	6.6867	5.2048	0.8342
WT-LightGBM	5.4668	3.9739	0.8892

Table 11. Comparison of LD₃ prediction performance evaluation.

Model	RMSE	MAE	R^2
SVR	8.2056	6.58	0.573
KNN	6.751	5.1302	0.7109
GRU	6.3946	4.4773	0.7244
RF	5.3202	4.0131	0.8092
Xgboost	6.2162	4.9936	0.7549
WT-LightGBM	4.8155	3.6436	0.8529

Observing the data in Table 9, it can be seen that model WT-LightGBM performs better than other models in the experiment, with the R^2 score of 0.95, and RMSE and MAE of 3.6233 and 2.7256, respectively. Model WT-LightGBM can effectively handle large-scale high-dimensional data and maintain high robustness in the presence of noise or large fluctuations in the data. These characteristics make the WT-LightGBM model highly practical and widely adaptable in predicting vehicle traffic speeds. The SVR model may have poor performance in processing big datasets or multi-feature datasets, while ETC data has more features. Although the KNN model is simple and easy to use, it has high computational complexity on the big dataset. Although GRU can process sequence data, there may be overfitting issues for this problem, with the R^2 score of 0.8156. RF exhibits strong generalization ability, with the R^2 score of 0.9276, and has lower RMSE and MAE. The advantage of RF is that it can handle high-dimensional data, but its training time is relatively long. XGBoost and LightGBM are both GBDT models that perform well on many problems. The R^2 score of XGBoost reached 0.9075, but in comparison, LightGBM performed even better. LightGBM has the characteristics of fast speed and efficient memory utilization, which are especially suitable for processing large-scale data. The WT-LightGBM model we propose is an optimized model for LightGBM, which uses wavelet denoising before data processing to eliminate section speed noise and preserve original data features, improving prediction accuracy and reducing errors.

In the experiment LD₂, the data environment is more complex because there are service areas and noisy data on the road section. Nevertheless, the WT-LightGBM model still performs better than other models, with the R^2 score of 0.8892, and RMSE and MAE of 5.4668 and 3.9739, respectively. This indicates that the WT-LightGBM model has excellent robustness when dealing with noisy data.

In the experiment LD₃, it was considered that there is a high traffic volume during holidays, and there is a service area, resulting in high data volatility. In this more complex situation, the WT-LightGBM model still achieved the best results, with the R^2 score of 0.8529, and RMSE and MAE of 4.8155 and 3.6436, respectively. This proves that the WT-LightGBM model can maintain high prediction accuracy when dealing with data with a large amount of volatility.

To better observe the experimental effect of the model, this article selects LD₁ and uses the WT-LightGBM model to predict the visualization of the current vehicle traffic speed. The results are shown in Figure 11, where blue represents the actual vehicle speed and yellow represents the predicted vehicle speed. From the figure, it can be seen that the predicted vehicle speed has a similar trend to the actual vehicle speed, indicating that the model can accurately predict vehicle speed.

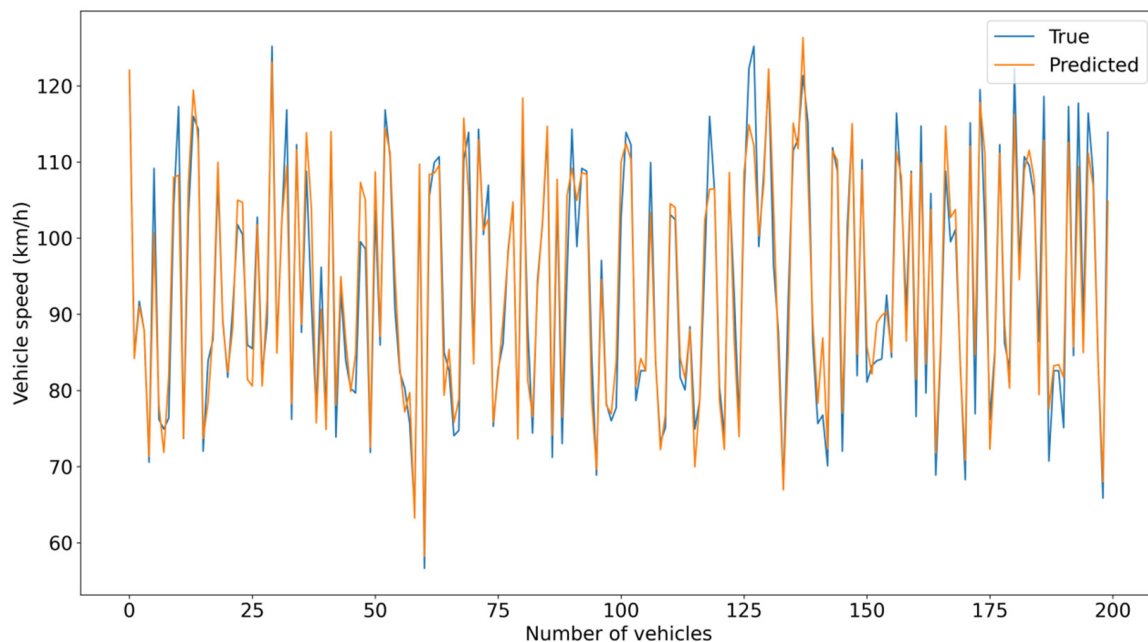


Figure 11. Visualization of vehicle traffic speed prediction.

5.3. PTV Quantitative Evaluation Algorithm

5.3.1. Determination of Characteristic Parameters

The PTV evaluation of the expressway comprehensively considers several evaluation indexes, among which the membership function based on a fuzzy set is used to quantitatively evaluate different PTVs. To ensure the rationality of the evaluation, it is necessary to determine the segmented value interval of the membership function, which has a direct impact on the accuracy of PTV evaluation. Therefore, in order to scientifically and reasonably set these parameters, we have comprehensively considered the following two factors.

Firstly, feature analysis of evaluation indicators is crucial. The evaluation indicator characteristics of PTV are usually contained in a large amount of data, so we adopt effective data mining methods to analyze the distribution of indicators from a statistical perspective. Through in-depth mining and analysis of data, we can better understand the essential characteristics of indicators and determine the corresponding interval of segmented function values.

Secondly, we combine the limitations of traffic laws and regulations. The main purpose of traffic regulations is to regulate driving behavior, reduce economic losses caused by traffic accidents, and improve the safety level of roads. Therefore, when setting the value range of the membership function, we must comply with relevant traffic regulations and ensure that the set parameters not only accurately reflect the performance of PTV, but are also consistent with traffic regulations to ensure the legality and operability of the evaluation results.

By comprehensively considering the above two factors, we can set reasonable segmented value intervals for the membership function on a scientific basis, thereby ensuring the accuracy and reliability of expressway PTV evaluation. The adoption of this method will make the evaluation results more credible and provide a useful decision-making basis for the management and planning of expressway.

The experiment selected data from the Fu-xia section for statistical analysis, and the results are shown in Figure 12.

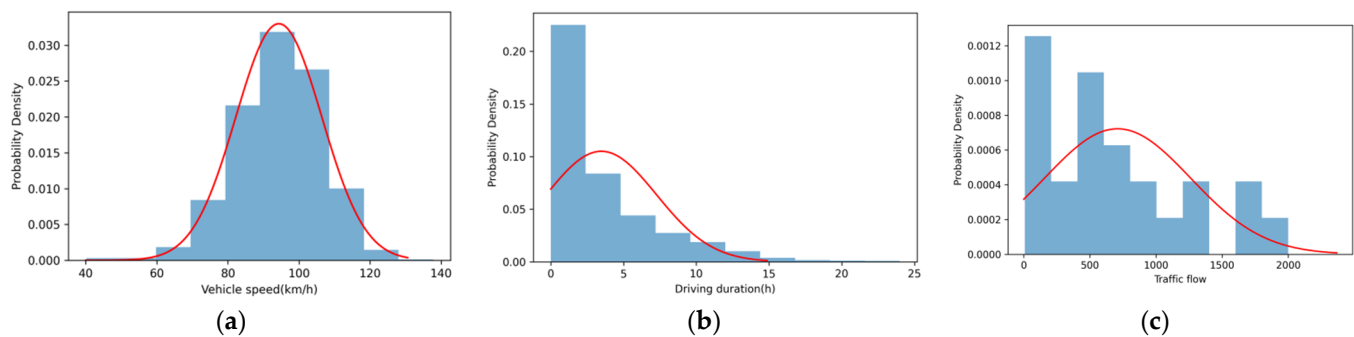


Figure 12. Statistical distribution of traffic characteristics. (a) Vehicle speed distribution. (b) Vehicle driving duration distribution. (c) Traffic flow distribution.

The probability curve of vehicle speed distribution shown in Figure 12a shows proper normal distribution characteristics and reaches the highest point at 90 km/h, providing a reference for OS feature threshold setting. Due to the dynamic nature of PTV evaluation, which is relative to over-the-horizon vehicles, the degree of danger of its OS characteristics needs to be compared with the speed of over-the-horizon vehicles. If there is a significant difference in speed between vehicles traveling on the road and over-the-horizon vehicles, it will have a certain impact on driving safety. The speed difference between vehicles can easily lead to lateral accidents and rear-end collisions. Some scholars have studied accident data, and Zhong L et al. [38] used statistical regression methods to study the relationship between the speed difference between large and small vehicles and the accident rate. They concluded that there is a positive correlation between the speed difference and the accident rate, that is, as the average speed difference increases, the accident rate gradually increases, and accidents caused by the speed difference together account for one-third of the total number of accidents. Therefore, the threshold is set for vehicle speeds greater than those of over-the-horizon vehicles.

The probability curve of vehicle driving duration distribution shown in Figure 12b shows that by referring to the definition of sleep-deprived driving duration in China (driving duration is more than 4 consecutive hours, and rest time is less than 20 min, or driving duration in a working day is more than 8 h), combined with the distribution curve, the driving duration drops sharply after more than 4 h. Therefore, it can be inferred that the driving duration of most drivers is usually less than 4 h, which is also the most suitable time parameter. To sum up, we set the time threshold of sleep-deprived driving as 4 h.

Figure 12c shows the probability curve of traffic flow distribution, with a threshold set to 750 based on the curve distribution. The size of road traffic flow directly affects the saturation of road traffic, and the saturation of roads directly affects the occurrence and development of traffic accidents.

Finally, for different types of vehicles traveling on the expressway at the same speed, the level of danger varies. It is necessary to determine the type of vehicle and conduct a hazard assessment based on different vehicle types. In order to observe the proportion of various types of vehicles driving on the expressway, this article uses data from a province's road network of 6.826061 million sections to statistically analyze the number of different types of vehicles. The results are shown in Table 12. It can be seen that Class 1 passenger cars account for a relatively large proportion, accounting for 82.93% of the total number of vehicles, followed by Class 6 trucks, accounting for 7.02% of the total number. The majority of overall vehicle types are passenger cars and trucks, accounting for 84.33% and 15.52%, respectively, while special operation vehicles only account for 0.13%. Therefore, the larger the number of vehicle types, the greater their danger.

Table 12. Statistics of the number of different types of vehicles.

Vehicle Type	Code Name	Class	Number	Percentage
Passenger car	1	Class I	5,660,545	82.93%
	2	Class II	19,434	0.28%
	3	Class III	33,503	0.49%
	4	Class IV	43,130	0.63%
Truck	11	Class I	300,813	4.40%
	12	Class II	104,710	1.53%
	13	Class III	45,950	0.67%
	14	Class IV	87,488	1.28%
	15	Class V	42,095	0.62%
	16	Class VI	479,562	7.02%
Special operation vehicle	21	Class I	4201	0.06%
	22	Class II	4175	0.06%
	23	Class III	210	0.003%
	24	Class IV	139	0.002%
	25	Class V	3	0.00004%
	26	Class VI	103	0.001%

5.3.2. Modeling of PTV Hazard Rating

Based on ETC data, we conducted in-depth mining and analysis of the PTV and proposed four membership functions to quantify it to evaluate its threat. Based on the analysis results of expressway vehicle driving and traffic characteristics mentioned above, we have determined the specific membership functions as follows. The design of these membership functions is based on the results of specific parameters and data analysis, aiming to provide an accurate and reliable numerical basis for evaluating the threat of PTV.

- Membership function of OS:

$$\mu_{OS}(v) = \begin{cases} 0, & v \leq V_{BVR} \\ 2\left(\frac{v-V_{BVR}}{1.5V_{BVR}-V_{BVR}}\right)^2, & V_{BVR} < v \leq 1.2V_{BVR} \\ 1 - 2\left(\frac{v-1.5V_{BVR}}{1.5V_{BVR}-V_{BVR}}\right)^2, & 1.2V_{BVR} < v \leq 1.5V_{BVR} \\ 1, & v > 1.5V_{BVR} \end{cases} \quad (17)$$

- Membership function of VT:

$$\mu_{VT}(m) = \begin{cases} 0, & m \leq 0 \\ 2\left(\frac{m-0}{6-0}\right)^2, & 0 < m \leq 1 \\ 1 - 2\left(\frac{m-6}{6-0}\right)^2, & 1 < m \leq 6 \\ 1, & m > 6 \end{cases} \quad (18)$$

- Membership function of LD:

$$\mu_{LD}(t) = \begin{cases} 0, & t \leq 4 \\ 2\left(\frac{t-4}{10-4}\right)^2, & 4 < t \leq 8 \\ 1 - 2\left(\frac{t-10}{10-4}\right)^2, & 8 < t \leq 10 \\ 1, & t > 10 \end{cases} \quad (19)$$

- Membership function of VF:

$$\mu_{VF}(f) = \begin{cases} 0, & f \leq 750 \\ 2\left(\frac{f-750}{1500-750}\right)^2, & 750 < f \leq 1000 \\ 1 - 2\left(\frac{f-1500}{1500-750}\right)^2, & 1000 < f \leq 1500 \\ 1, & f > 1500 \end{cases} \quad (20)$$

Formula (17): Note that V_{BVR} is the speed of over-the-horizon vehicles. According to research findings, there is a positive correlation between the speed difference between two vehicles and the accident rate. The greater the difference in vehicle speed, the lower the level of driving safety. Specifically, in the first case, when the vehicle speed does not exceed over-the-horizon vehicle speed, it is considered a normal level. In the second case, when the vehicle's speed exceeds the speed of over-the-horizon vehicles and is less than 120% of the speed of over-the-horizon vehicles, calculate the corresponding membership degree. In the third case, when the speed of vehicles exceeds 120% of the speed of over-the-horizon vehicles and is less than 150% of the speed of over-the-horizon vehicles, calculate the corresponding membership degree. When the speed of vehicles exceeds 150% of the speed of over-the-horizon vehicles in the fourth case, the membership degree is 1.

Formula (18): The degree of danger varies for different types of vehicles, and there is a positive correlation between vehicle type and accident rate. In the first case, the membership level without vehicle type is set to 0. In the second case, compared to other types of vehicles, Class I passenger cars have a smaller volume and a lower level of danger. Calculate the corresponding membership degree. In the third case, Class II passenger cars refer to vehicles with 8 or more seats but less than 19 seats, Class III passenger cars refer to vehicles with 20 or more seats but less than 39 seats, Class IV passenger cars refer to vehicles with 40 or more seats, and the risk level of vehicles with 40 or more seats gradually increases. The corresponding membership degrees are calculated separately. In the fourth case, trucks and special operation vehicles have a relatively large weight and are prone to rollover. If a traffic accident is caused, the damage is generally more severe. Therefore, these two types of vehicles are considered equally dangerous and have a membership degree of 1.

Formula (19): According to the limitations of reference traffic regulations, in the first case, when the driving time is less than or equal to 4 h, the membership degree is 0. In the second case, if the driving time is less than or equal to 4 h and less than or equal to 8 h, calculate the corresponding membership degree. In the third case, if the driving time is greater than 8 h and less than or equal to 10 h, calculate the corresponding membership degree. In the fourth case, when the driving time is greater than 10 h, the membership degree is 1.

Formula (20): This is based on the analysis of characteristic indicators of section traffic flow, as well as previous researchers' analysis of the relationship between traffic flow and accident rate. In the first case, when the traffic flow is less than or equal to 750, the membership degree is 0. In the second case, when the traffic flow is greater than 750 and less than 1000, calculate the corresponding membership degree. In the third case, when the traffic flow is greater than 1000 and less than 1500, calculate the corresponding membership degree. In the fourth paragraph, when the traffic flow exceeds 1500, the membership degree is 1.

Based on the measurement results of PTV indicators using various membership functions and corresponding weights, we use Formula (12) to calculate the comprehensive score of vehicles traveling on the road. To achieve this, we need to follow the AHP calculation criteria, compare the importance of each indicator in pairs, and construct a judgment matrix as shown below (Table 13):

Table 13. PTV indicators judgment matrix.

Element	f_1	f_2	f_3	f_4
f_1	1	3	4	3
f_2	$1/3$	1	$\frac{1}{2}$	2
f_3	$\frac{1}{4}$	2	1	2
f_4	$1/3$	$\frac{1}{2}$	$\frac{1}{2}$	1

Among them, f_1 , f_2 , f_3 , and f_4 represent OS, VT, LD, and VF, respectively. The elements in the matrix are obtained by experienced experts by pairwise comparison of their importance based on the scale table, which is the key to the analytic hierarchy process.

After constructing the judgment matrix, it is necessary to perform consistency verification on the matrix according to Formula (10), and the verification result $CR = 0.0679 < 0.1$, which passes the consistency verification. The final calculated indicator weights are as follows:

$$\omega = [0.506, 0.165, 0.214, 0.115]$$

Among them, each value of ω corresponds to the weight of PTV evaluation indicators f_1 , f_2 , f_3 , and f_4 . Through the operation of the AHP method, we will obtain the relative weights of each indicator, enabling accurate and reliable comprehensive evaluation of the PTV threat score. This comprehensive evaluation process will help to better understand and grasp the comprehensive threat level of PTVs and provide a scientific basis for relevant decision-making.

5.3.3. Result Analysis

PTV detection is a multi-dimensional evaluation process that not only considers the risk level of individual indicators, but also comprehensively considers the impact of each indicator on the overall driving process. To verify the effectiveness of the model, experimental data was obtained from 12,274 vehicles during the busy holiday season in Fuxia, Fujian Province, in 2021. The historical traffic speeds of 12,274 vehicles were calculated, and the WT-lightGBM model was used to predict the driving speed of the next section (as the driving section). Then, the PTV evaluation algorithm was used to assess the threat of the vehicles. Based on the above data, the experiment used accuracy and recall to evaluate the effectiveness of the model, and the calculation formula is as follows.

- Accuracy

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (21)$$

- Recall

$$Recall = \frac{TP}{TP + FN} \quad (22)$$

where TP indicates that the prediction is positive and correct; TN indicates that the prediction is negative and correct; FP indicates that the actual negative class is classified into the positive class; FN indicates that the actual positive class is classified into the negative class. The accuracy rate reflects the ratio of the number of samples correctly classified by

the classifier to the total number of samples for a given test dataset. The results are shown in Table 14, and it can be seen from the table that the accuracy and recall of this model are 98.03% and 99.55%, respectively. The results indicate that the proposed algorithm can effectively detect vehicles with potential hazards.

Table 14. Algorithm result display.

Model Index	Result
Accuracy	98.03%
Recall	99.55%

In addition, due to the inevitable subjectivity and uncertainty in PTV assessment, we converted the assessment results from qualitative classification to threat level classification to more effectively measure the threat degree of PTVs. To illustrate our research findings, this article proposes a potential threat index for quantitative evaluation of PTVs, which distinguishes potential PTVs based on information from a large amount of ETC data. In order to clearly express the characteristics of high-risk vehicles, we determined three reference values for describing PTVs with different threat levels by calculating the membership values of all vehicles, namely no threat, low threat, moderate threat, and high threat. The classification criteria for threat levels are based on the calculation results of membership values, which is also an innovative method proposed in this article for quantifying PTV threat levels. The specific results are shown in Table 15 below:

Table 15. PTV threat classification.

Range	$[0, \mu)$	$[\mu, \mu + \sigma)$	$[\mu + \sigma, \mu + 2\sigma)$	$[\mu + 2\sigma, 1)$
Threat classification	No threat	Low threat	Moderate threat	high threat

By using this novel PTV threat level quantification method, we can more accurately assess the potential threat of vehicles in transit and provide an important reference for traffic management and safety decision making, which has certain application prospects. In order to better validate the effectiveness and reliability of this method, a comparative analysis was conducted on the PTVs detected by this method. Some PTVs are shown in Table 16. The * in the table was desensitized due to the privacy concerns of the data used in our experiment.

Table 16. PTV example.

License Plate	OS	μ_{OS}	VT	μ_{VT}	LD	μ_{LD}	VF	μ_{TF}	DTS
HuangHu****7	82.12 km/h	0.39	16	1	5.42 h	0.11	1798	1	50.2
HuangWan****0	96.65 km/h	0.63	14	1	4.65 h	0.023	1798	1	60.29
LvMin****6	144.17 km/h	0.99	1	0.056	5.1 h	0.067	1761	1	63.84
LanMin****V	139.84 km/h	1	1	0.056	6.9 h	0.47	1761	1	73.01
HuangJin****3	137.75 km/h	1	15	1	4.21 h	0.0024	1673	1	78.65

5.4. Over-the-Horizon PTV Perception Algorithm

5.4.1. Over-the-Horizon Perception Range Division

In order to provide more timely road information and improve the traffic safety of travelers, it is necessary to provide timely information about road conditions and PTV driving on the road. The traffic speed of the vehicle is calculated by the vehicle information obtained from the ETC gantry, so as to determine the section where the vehicle is located. After determining the section where the vehicle is located, the PTV over-the-horizon

distance (safety distance) can be further detected. The selection of safe distance must meet the best safe time and space for warning the driver to avoid danger, so as to remind the driver to control their speed at the first opportunity and in a timely manner. Therefore, the early and late warning of PTVs beyond the sight distance is not conducive to the coordinated response of drivers, and the effectiveness of the safety warning is lost. The statistical analysis of all vehicle speeds in some sections is carried out by using a box diagram, and the summary of data results is shown in Table 17.

Table 17. Vehicle speed distribution table.

Data Distribution	Traffic Speed (km/h)
Maximum	136.21
75th percentile	102.71
Average	95.68
25th percentile	84.23
Minimum	60.63

Observing Table 17, it is found that the maximum speed of the vehicle reaches 136.21 km/h, while the minimum speed of the section is only 60.63 km/h; the speed difference between the two is 75.49 km/h. Because of this, we analyze the speed difference from 10 km/h to 80 km/h, and the time when they can meet at different vehicle distances. The results are shown in Table 18.

Table 18. Schedule of two vehicles meeting.

	1 km	2 km	3 km	4 km	5 km	6 km
10 km/h	6 m	12 m	18 m	24 m	30 m	38 m
30 km/h	2 m	4 m	6 m	8 m	10 m	12 m
50 km/h	1.2 m	2.4 m	3.6 m	4.8 m	6 m	7.2 m
60 km/h	1 m	2 m	3 m	4 m	5 m	6 m
80 km/h	0.75 m	1.5 m	2.25 m	3 m	3.75 m	4.5 m

Take the maximum speed difference as an example. When the speed difference is 80 km/h, and the distance between the two vehicles is 1 km, it only takes 45 s for the two vehicles to meet. If the warning time is too late, the driver will not be able to respond in time, and the 1 km warning distance is too short. If driving in a straight road section, the driver can observe the road conditions well and make cooperative responses. When the distance is 6 km, it takes 4.5 min for two cars to meet, which is enough time for travelers to make decisions in advance. Therefore, we select 2 km as the safety warning distance, and use 4 km and 6 km as the auxiliary safety warning distance to allow drivers to perceive the over-the-horizon road conditions in advance.

However, different vehicle types have different traffic speeds on the same road, and this paper selects the vehicle traffic data of the section from gantry 340241 to 340243 on 1 May 2021, and analyzes the average traffic speed of different types of vehicles. The results are shown in Figure 13, which shows that there are certain differences in the average traffic speed of different vehicle types. Among them, the average traffic speed of Class I passenger cars is significantly greater than other types of vehicles, reaching about 110 km/h. The average speed of other types of passenger cars is about 90 km/h. The average speed of trucks and special operation vehicles is generally about 80 km/h. Therefore, through the fine division of vehicle types, the potential dangerous vehicles can be effectively evaluated.

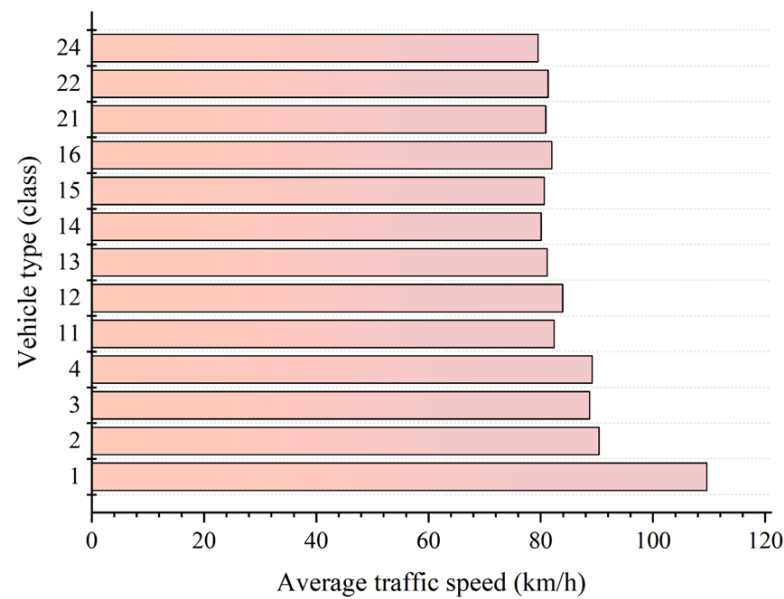


Figure 13. Comparison of the average speed of different vehicle types.

After detecting a PTV, it is necessary to update the vehicle location information in real time and provide the information to the driver in time. However, only when the vehicle passes through the ETC gantry and generates transaction data can the location information of the vehicle be obtained at the current moment. When the vehicle is passing through QD's Node₁ before reaching Node₂, the platform is unable to accurately monitor the location of the vehicle. Therefore, the distribution interval of ETC gantries directly affects the vehicle location refresh, and if the ETC gantries' distribution interval is too large, the vehicle location refresh is not timely. Therefore, it is necessary to analyze the impact of ETC gantry distribution interval on the provision of potentially dangerous vehicle information. Table 19 analyzes the refreshing time of different types of vehicle positions at different gantry intervals, and the refreshing time of vehicle positions is expressed as the time of a vehicle passing the section.

Table 19. Traffic schedule of different types of vehicles at different gantry intervals.

	Class I Passenger Car	Other Passenger Car	Truck and Special Operation Vehicle
1 km	32.723 s	40 s	45.004 s
2 km	65.445 s	80 s	90.009 s
3 km	98.167 s	120 s	135.013 s
4 km	130.89 s	160 s	180.018 s
6 km	196.335 s	240 s	270.027 s

According to the data in Table 19, if the gantry distribution interval is 1 km, the time for Class I passenger cars to pass through a section is 32.723 s, which means that the position of Class I passenger cars will be refreshed every 32.723 s. Similarly, the positions of other passenger cars, trucks, and special operation vehicles will be refreshed at 40 s and 45.004 s, respectively. According to the analyzed meeting schedule of two vehicles, it can be observed that the maximum speed difference is 80 km/h, and the safe distance is 1 km, so it takes 45 s for two vehicles to meet. Therefore, the driver has enough time to make countermeasures after obtaining the position of the PTV. Then, the gantry distribution

interval of 1 km is within the feasible range. By the same token, 2 km is also within the feasible range.

If the gantry distribution interval is 4 km, the time for Class I passenger cars to pass through a section is 130.89 s, and other passenger cars, freight cars, and special work vehicles are refreshed in 160 s and 180.018 s, respectively. According to the schedule of meeting two vehicles, the maximum speed difference is 80 km/h, and the safety distance is 4 km; then, it takes 180 s for two vehicles to meet, so the gantry distribution interval is 4 km in the feasible range. However, as for Gaode Map, the most professional Internet map maker in China, its car version displays the current real-time traffic condition information at a refresh frequency of once every 2 min. Therefore, compared with the data refresh frequency of Gaode Map, the frequency of refreshing data is 60.191 s more when the gantry distribution interval is 4 km. Therefore, the gantry distribution interval should be less than 4 km to be close to the refresh frequency of Gaode Map (13.06.1.2061).

5.4.2. Result Analysis

In order to analyze the influence of the ETC gantry distribution interval on the detection of PTVs, the experiment uses an ETC simulation platform to select a part of the expressway section test area and set different ETC gantry distribution intervals, such as 1 km, 2 km, 3 km, and so on [39]. In the ETC simulation application system function, the platform can simulate the actual situation of multiple concurrent expressways, providing micro-level multitask debugging technical support for traffic flow and single vehicles in transit supervision. The random number generator is used to simulate the driving of three types of Class I passenger cars, as well as other passenger cars and trucks in the test area. At the same time, the ETC gantries are used to collect information on the passage time and speed of the vehicles, and to record the PTVs that are monitored. It should be noted that in order to facilitate the use of data collected by ETC gantry for PTV detection, we represent the distance distributed by the gantry as the distance for over-line-of-sight detection.

After data preprocessing, the ETC transaction data of some sections are randomly extracted, the time series of vehicle traffic speed is generated, and 1000 groups of PTVs are selected for testing and evaluation. In order to analyze the rationality of gantry distribution, this article used 11 sets of data collected at different gantry intervals for comparison and analyzed the impact of gantry density in different kilometers on PTV detection. The results are shown in Figure 14.

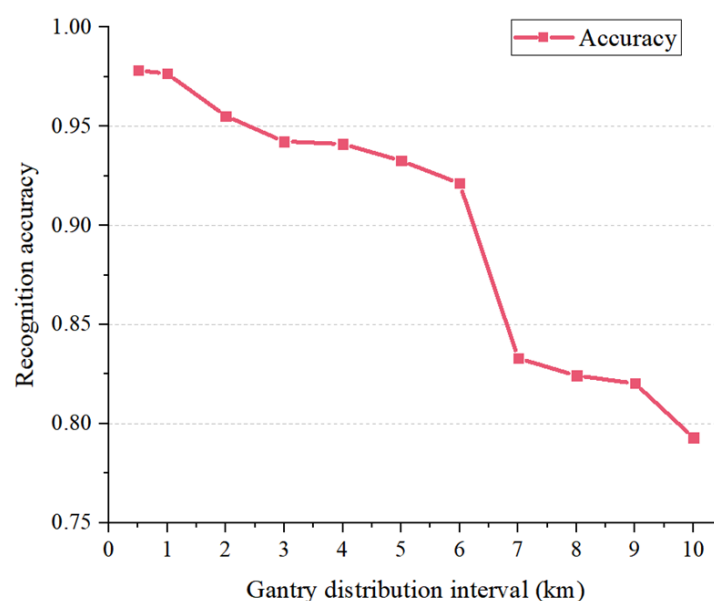


Figure 14. Comparison of different gantry intervals.

It can be seen from Figure 14 that the recognition accuracy of PTVs gradually decreases with the increase of the distribution interval of the gantry, especially at 6 km, which may be due to the fact that PTVs exceed the over-the-horizon vehicles during the refresh interval. Take the Class I passenger car and truck with the highest and lowest average speed as an example: the average speed difference is 30–40 km/h. If the safety distance is 2 km, it takes 3–4 min for the Class I passenger car to catch up with the truck. If the speed of the Class I passenger car is 100–120 km/h, the three-minute driving distance is 5–6 km. Therefore, if the gantry is not deployed within 6 km, it will largely lead to the Class I passenger car in the rear having exceeded the truck before refreshing the location information of the Class I passenger car, which will greatly reduce the accuracy of recognition. However, from an economic perspective, if the construction cost of ETC gantries is high and the distribution is too dense, this will cause pressure on the economy. If the distribution is too sparse, it will affect the real-time road condition refresh frequency, resulting in the delay of road condition information updates and misleading the driver. Therefore, according to the analysis of the gantry refresh frequency and the consideration of the actual construction cost, the gantry distribution interval of 2 km is the best.

The experiment also tracked and counted different types of over-the-horizon PTVs with 600 s (10 min) slices. A statistical analysis of PTVs at different safety distances of 2 km, 4 km, and 6 km was conducted. Due to the lack of transaction data of special operation vehicles, the experiment only analyzes two types of vehicles, such as passenger cars and trucks. The experimental results are shown in Figure 15. The model indicated in orange is trucks, and the model indicated in blue is passenger cars.

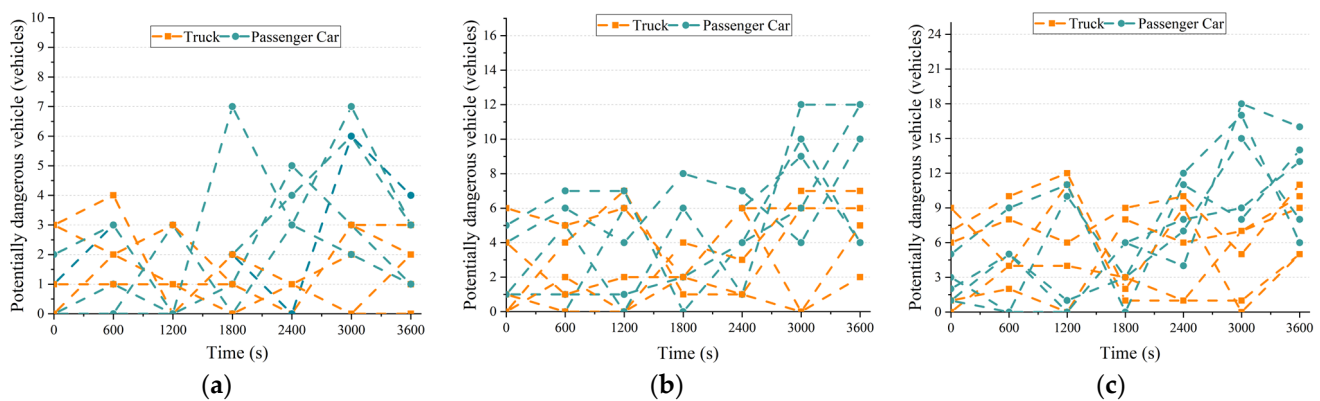


Figure 15. Statistics of PTVs behind the safety distance. (a) safety warning distance: 2 km. (b) safety warning distance: 4 km. (c) safety warning distance: 6 km.

The safety distance in Figure 15a is 2 km. Since the speed of the trucks fluctuates less throughout the whole process, the surrounding PTVs are more stable than passenger cars, generally around 2. These PTVs may be caused by other trucks in order to ensure timeliness, leading to relative speeding. At 9:00 for the passenger car, there were approximately two PTVs around, and in the next half hour, the number of PTVs within 2 km reached approximately six. Since the vehicle speed characteristics will be affected by the traffic flow in the section and the geographical environment, half an hour later, when the vehicle passes through a road section with expressway toll booths, especially the ETC channel, the owner tends to forget to reduce the speed to a reasonable range. The speed of the vehicle is within 30 km/h or 20 km/h, which is potentially threatening compared with normal driving vehicles. However, the traffic flow from the ramp to the main road is generally relatively small, and many car owners will involuntarily speed up. Therefore, on ramps and toll stations, a small number of passenger cars will appear to be relatively speeding.

The safety distance in Figure 15b is 4 km. Due to the increase of the safety distance, the number of PTVs around the truck has increased, and the number of PTVs tracked and counted every 600 s is about four; while the speed of passenger cars fluctuates greatly, in

the first half hour, there were approximately four PTVs around, and after half an hour, due to the influence of the geographical environment, the number of PTVs reached nine.

The safety distance in Figure 15c is 6 km, and the PTVs around the truck have also increased due to the increase in the safety distance. The number of PTVs tracked every 600 s is approximately 8; while the bus has approximately 5 PTVs around the first half hour, half an hour later, the PTVs reach 14 due to the influence of the geographical environment.

In order to capture the changing characteristics of PTV on the expressway in the dimension of safety distance, and to analyze the distribution law of PTV, the experiment conducts statistical analysis on trucks and passenger vehicles according to different safety distances in the current time. The experimental results are shown in Figure 16. Observing the experimental results, it can be seen that the safety distance dimension of PTV around passenger cars and trucks is an overall growth curve. With the increase of the safety distance, the PTV also increases to varying degrees, showing the objective law of expressway vehicle driving.

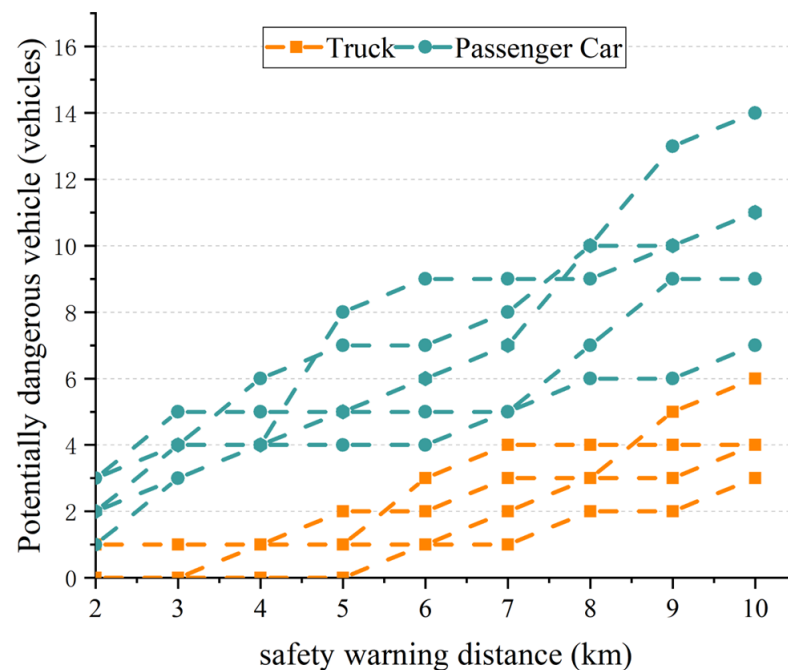


Figure 16. Distribution of PTVs within different safety distances at the current time.

6. Conclusions and Future Work Outlook

The safety issue of driving on expressway is a crucial challenge in the field of transportation. Its importance lies not only in ensuring the safety of individual travel, but also in the stability and efficiency of the entire transportation system. With the acceleration of urbanization and the continuous growth of vehicle numbers, traffic safety issues on expressway have gradually become a common focus of attention in both industry and academia. We are committed to proposing an innovative method for detecting potential threat vehicles beyond the line of sight based on ETC data by deeply mining a large amount of ETC transaction data. In this study, we adopted a series of comprehensive measures aimed at contributing to improving the supervision effect of traffic management, strengthening road safety, and improving traffic efficiency.

Firstly, we have carefully cleaned and organized the massive collection of ETC transaction data and constructed vehicle trajectories and segment traffic flow, laying a solid foundation for subsequent experiments and analysis. In addition, WT combined with LightGBM was used to predict the current driving speed of the vehicle, and its predicted RMSE, MAE, and R^2 reached 3.623, 2.725, and 0.95, respectively, which was better than traditional models. This improved the accuracy of predicting the current driving speed

of the vehicle, thereby enhancing the effectiveness of PTV detection. On this basis, we innovatively adopted the spacing of ETC gantry distribution as the range of over-the-line-of-sight detection, thereby significantly improving the accuracy of vehicle position estimation. This method not only helps to detect potential threats in a timely manner, but also provides a more comprehensive guarantee for road safety. Finally, based on the PTV features extracted from ETC data, a fuzzy method based on a membership function was constructed to quantitatively measure the threat of PTVs. The detection accuracy and recall were 98.03% and 99.55%, respectively, proving the effectiveness of our proposed method.

Although various factors need to be comprehensively considered in different application scenarios, our proposed method is an overall framework with the ability to explore and quantify the different factors that may have an impact. This framework provides feasibility for further improving the supervision effect of traffic management and improving road safety and traffic efficiency.

In the future, we will be committed to in-depth research on methods of transmitting warning information to drivers, especially in emergency situations, such as warning that there is an ambulance, fire truck, and other emergency vehicle within a few kilometers behind the driver or in front of them, so that drivers can take avoidance measures in advance. In addition, we plan to introduce more datasets to more comprehensively support PTV detection research, in order to adapt the work to more application scenarios and achieve greater application value of this research.

Author Contributions: Conceptualization, F.Z. and C.X.; methodology, F.Z.; software, C.X.; validation, F.G. and X.C.; formal analysis, F.Z.; investigation, C.X.; resources, F.G.; data curation, C.X.; writing—original draft preparation, C.X.; writing—review and editing, F.Z.; visualization, Q.C.; supervision, G.L. and T.Y.; project administration, X.C.; funding acquisition, F.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work is partially supported by the Renewable Energy Technology Research institution of Fujian University of Technology Ningde, China (funding number: KY310338); the 2020 Fujian Province “Belt and Road” Technology Innovation Platform (funding number: 2020D002); the Provincial Candidates for the Hundred, Thousand and Ten Thousand Talent of Fujian (funding number: GY-Z19113); the Patent Grant project (funding numbers: GY-Z18081, GY-Z19099, GY-Z20074); Horizontal projects (funding number: GY-H-20077); municipal-level science and technology projects (funding numbers: GY-Z-22006, GY-Z-220230); Fujian Provincial Department of Science and Technology Foreign Cooperation Project (funding number: 2023I0024); the Open Fund project (funding numbers: KF-X19002, KF-19-22001).

Data Availability Statement: Restrictions apply to the availability of these data. Data were obtained from Fujian Expressway Information Technology Co., Ltd. and are available from the authors with the permission of Fujian Expressway Information Technology Co., Ltd.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Yue, J. Research on the Impact of Expressway Development on Transport. *Transpo World* **2020**, *549*, 9–10.
2. Liu, X.; Xiao, L.; Jin, S.Y. Analysis on Characteristics and Causes of Serious Traffic Accidents on Freeways Based on NAIS. *Highw. Automot. Appl.* **2022**, *212*, 32–37.
3. Wang, S.; Zhang, Q.; Chen, G. V2V-CoVAD: A vehicle-to-vehicle cooperative video alert dissemination mechanism for Internet of Vehicles in a highway environment. *Veh. Commun.* **2022**, *33*, 100418. [[CrossRef](#)]
4. Peivandi, M.; Ardabili, S.Z.; Sheykhivand, S.; Danishvar, S. Deep Learning for Detecting Multi-Level Driver Fatigue Using Physiological Signals: A Comprehensive Approach. *Sensors* **2023**, *23*, 8171. [[CrossRef](#)]
5. Khan, Z.H.; Altamimi, A.B. A New Traffic System on Driver Sensitivity and Safe Distance Headway. *Appl. Sci.* **2023**, *13*, 11262. [[CrossRef](#)]
6. Dong, L. Research on the Development of Expressway Informatization and Intelligent Management. *Transp. Bus. China* **2021**, *631*, 104–107.

7. Wang, P.; Lu, Y.; Chen, N.; Zhang, L.; Kong, W.; Wang, Q.; Qin, G.; Mou, Z. Research on the Optimal Deployment of Expressway Roadside Units under the Fusion Perception of Intelligent Connected Vehicles. *Appl. Sci.* **2023**, *13*, 8878. [\[CrossRef\]](#)
8. Wang, Y.X.; Zeng, C.H.; Wang, L.P.; Zhang, H. Talking about the impact of ETC technology on the development of smart highways in China. *Commun. Sci. Technol.* **2021**, *44*, 185–186.
9. Zou, F.M.; Guo, F.; Tian, J.S.; Luo, S.J.; Yu, X.; Gu, Q.; Liao, L.C. The Method of Dynamic Identification of the Maximum Speed Limit of Expressway Based on Electronic Toll Collection Data. *Sci. Program.* **2021**, *2021*, 4702669. [\[CrossRef\]](#)
10. Chen, Z.Y.; Zou, F.M.; Guo, F.; Gu, Q. Short-Term Traffic Flow Prediction of Expressway Based on Seq2seq model. In Proceedings of the International Conference on Frontiers of Electronics, Information and Computation Technologies, Changsha, China, 21–23 May 2021.
11. Tian, J.S.; Zou, F.M.; Guo, F.; Gu, Q.; Ren, Q.; Xu, G. Expressway Traffic Flow Forecasting based on SF-RF Model via ETC Data. In Proceedings of the International Conference on Frontiers of Electronics, Information and Computation Technologies, Changsha, China, 21–23 May 2021.
12. Zou, F.M.; Ren, Q.; Tian, J.S.; Guo, F.; Huang, S.B.; Liao, L.C.; Wu, J.S. Expressway Speed Prediction Based on Electronic Toll Collection Data. *Electronics* **2022**, *11*, 1613. [\[CrossRef\]](#)
13. Guo, F.; Zou, F.M.; Luo, S.J.; Chen, H.B.; Yu, X.; Zhang, C.; Liao, L.C. Positioning Method of Expressway ETC Gantry by Multi-Source Traffic Data. *IET Intell. Transp. Syst.* **2022**, 1–15. [\[CrossRef\]](#)
14. Xiong, X.; Yang, X.; Liu, Z.; Zhu, X. Research on cloud-network-edge-terminal architecture and service of vehicle-road collaboration. *Appl. Electron. Tech.* **2019**, *45*, 14–18.
15. Beasley, M. Smart Vehicular Networks for Reducing Road Accidents and Traffic Congestion. GS4 Georgia Southern Student Scholars Symposium. 108. 2016. Available online: https://digitalcommons.georgiasouthern.edu/research_symposium/2016/2016/108 (accessed on 18 October 2023).
16. Xu, A.; Chen, X.; Li, Z.W.; Hu, X.D. A method of situation assessment for beyond-visual-range air combat based on tactical attack area. *Fire Control. Command. Control.* **2020**, *45*, 97–102.
17. Merz, T.; Kendoul, F. Beyond visual range obstacle avoidance and infrastructure inspection by an autonomous helicopter. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots & Systems, San Francisco, CA, USA, 25–30 September 2011.
18. Guo, X.; Ni, J.L.; Liu, G.S. The ship detection of sky wave over the horizon radar with short coherent integration time. *J. Electron. Inf. Technol.* **2004**, *26*, 613–618.
19. Xue, Q.W.; Jiang, Y.M.; Lu, J. Risky driving behavior recognition based on trajectory data. *China J. Highw. Transp.* **2020**, *33*, 84–94.
20. Zhou, H.F.; Liu, H.P.; Shi, H.X. Abnormal driving behavior detection based on the smart phone. *CAAI Trans. Intell. Syst.* **2016**, *11*, 410–417.
21. Hui, F.; Guo, J.; Jia, S.; Xing, M.H. Detection of abnormal driving behavior based on BiLSTM. *Comput. Eng. Appl.* **2020**, *56*, 116–122.
22. Liu, S.Y.; Liu, S.; Tian, Y.; Sun, Q.L.; Tang, Y.Y. Research on Forecast of Rail Traffic Flow Based on ARIMA Model. In *Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2021; Volume 1792, p. 012065.
23. Zhao, L.; Song, Y.; Zhang, C.; Liu, Y.; Wang, P.; Lin, T.; Deng, M.; Li, H. T-gcn: A temporal graph convolutional network for traffic prediction. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 3848–3858. [\[CrossRef\]](#)
24. Huang, R.; Huang, C.; Liu, Y.; Dai, G.; Kong, W. LSGCN: Long Short-Term Traffic Prediction with Graph Convolutional Networks. In Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, Yokohama, Japan, 11–17 July 2020; Volume 7, pp. 2355–2361.
25. Han, L.; Du, B.; Sun, L.; Fu, Y.; Lv, Y.; Xiong, H. Dynamic and multi-faceted spatio-temporal deep learning for traffic speed forecasting. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, Singapore, 14–18 August 2021; pp. 547–555.
26. Tong, J.; Gu, X.; Zhang, M.; Wan, J.; Wang, J. Traffic flow prediction based on improved SVR for VANET. In Proceedings of the 2021 4th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE), Online, 26–28 March 2021; pp. 402–405.
27. Yang, L.; Yang, Q.; Li, Y.; Feng, Y. K-nearest neighbor model based short-term traffic flow prediction method. In Proceedings of the 2019 18th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES), Wuhan, China, 8–10 November 2019; pp. 27–30.
28. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. Available online: <https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf> (accessed on 18 October 2023).
29. Xia, H.; Wei, X.; Gao, Y.; Lv, H. Traffic prediction based on ensemble machine learning strategies with bagging and lightgbm. In Proceedings of the 2019 IEEE International Conference on Communications Workshops (ICC Workshops), Shanghai, China, 20–24 May 2019; pp. 1–6.
30. Wang, F.; Cheng, H.; Dai, H.; Han, H. Freeway short-term travel time prediction based on lightgbm algorithm. In *IOP Conference Series: Earth and Environmental Science*; IOP Publishing: Bristol, UK, 2021; Volume 638, p. 012029.
31. Yang, P. Driving State Recognizing Based on Driving Behavior Signals. Master's Thesis, North China University of Technology, Beijing, China, 2010.

32. He, W.Y.; Lv, B.; Wang, Z.Y.; Wu, B.; Li, L.B. Analysis of Risky Driving Behavior on Freeways Based on an AHP Method. *J. Transp. Inf. Saf.* **2016**, *34*, 91–95+102.
33. Zhang, X.B.; Chen, X. Driving Behavior Analysis of Operating Vehicles Based on the Internet of Vehicles. *Sci-Tech Dev. Enterp.* **2020**, *463*, 61–63.
34. *JTG B01-2014*; Technical Standard for Highway Engineering. China Communications Press: Beijing, China, 2015.
35. Dou, C.; Zhang, S.; Zhao, L. Selection of Wavelet Thresholds in GNSS Time Series Denoising. *J. Gansu Sci.* **2021**, *33*, 6–9.
36. Zhao, R.M.; Cui, H. Improved threshold denoising method based on wavelet transform. In Proceedings of the 2015 7th International Conference on Modelling, Identification and Control (ICMIC), Sousse, Tunisia, 18–20 December 2015; IEEE: New York, NY, USA, 2015; pp. 1–4.
37. Zadeh, L.A. 'Fuzzy sets'. *Inf. Control* **1965**, *8*, 338–353. [[CrossRef](#)]
38. Zhong, L.D.; Sun, X.D.; Chen, Y.S.; Zhang, J.; Zhang, G.W. The relationship between crash rates and average speed difference between cars and large vehicles on freeway. *J. Beijing Univ. Technol.* **2007**, *33*, 185–188.
39. Zou, F.M.; Guo, F.; Luo, S.J.; Liao, L.C.; Li, N.; Xing, Y. Research and Design of ETC Simulation Platform for Expressway. *J. Syst. Simul.* **2022**, 1–17. Available online: <http://kns.cnki.net/kcms/detail/11.3092.V.20220930.1547.003.html> (accessed on 18 October 2023).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.