Research on Road Traffic Safety Risk Assessment Based on the Data of Radar Video Integrated Sensors

Xiaoyu CAI[1], Zimu LI[2], Wufeng QIAO[3], Xiling CHENG[4], Bo PENG[5], Dong ZHANG[6]

Original Scientific Paper
Submitted: 19 June 2024
Accepted: 20 Sep 2024

[1] caixiaoyu@cqjtu.edu.cn, Chongqing Jiaotong University, College of Smart City
[2] 611220110005@mails.cqjtu.edu.cn, Chongqing Jiaotong University, College of Traffic Transportation
[3] Corresponding author, hustqiao@126.com, China Construction Seventh Engineering Division Corp.; T. Y. Lin International Engineering Consulting (China) Corp. Ltd; Chongqing Jiaotong University
[4] cxlll0809@mails.cqjtu.edu.cn, Chongqing Jiaotong University, College of Traffic Transportation
[5] pengbo351@126.com, Chongqing Jiaotong University, College of Smart City
[6] 1468457143@qq.com, Chongqing Jiaotong University, College of Traffic Transportation

Publisher:
Faculty of Transport and Traffic Sciences, University of Zagreb

ABSTRACT

To accurately prevent and warn of traffic accidents, this article proposes a method for predicting urban road traffic safety risks based on vehicle driving behaviour data and information entropy theory. This method uses data from radar video-integrated sensors to calibrate the thresholds for identifying unsafe driving behaviour, introduces recognition principles and algorithms, and analyses spatiotemporal distribution patterns. By incorporating entropy theory, an evaluation system with traffic safety entropy as the primary indicator and the unsafe driving behaviour rate as the secondary indicator is established. Clustering algorithms determine the classification number and threshold of traffic safety entropy, constructing a tunnel traffic safety risk assessment model, which is validated with road accident data. Using 13 days of data from the left lane of Qingdao Jiaozhou Bay Tunnel, the model divides traffic operation risk into high and low categories based on K-means clustering results of accident and safety entropy data. The study finds that when the safety entropy classification threshold is 0.0507, the classification accuracy is the highest at 92%. These results provide technical support for identifying road traffic safety risk points and preventing accidents.

KEYWORDS

traffic engineering; risk estimate; information entropy; unsafe driving behaviour; entropy weight method.

1. INTRODUCTION

Road traffic safety has always been the focus of attention in various countries, but scholars in urban transport have always been the hot spot of research. In the 1970s, western nations began to realise the significant impact of traffic safety on the national economy and social life. They carried out relevant research on people, vehicles, roads, the environment and so on. In the past few decades, China has completed the road construction goal of developed countries for half a century. Along with the high-speed development of the automobile industry, the problems of traffic congestion and pedestrian-vehicle mixing have brought about an extremely high risk to road safety, and the government and related organisations have been paying more and closer attention to the road safety problem.

In 1973, the Institute of Highway Science of the Ministry of Transport established the first all-round consulting service type unit in the field of road traffic safety in China, including research, design, standard specification development, safety evaluation and other work. Over the decades, China has formed a more mature system of road traffic evaluation [1], and the evaluation of road traffic safety has gradually changed from passive evaluation based on accident data to active evaluation based on driving behaviour data. Driving behaviour data is a record of vehicle driving under the interaction of human-vehicle-road-legal-environment and other factors, which is of great significance in the study of road traffic safety evaluation.

Currently, data sources for studying driving behaviour include in-vehicle OBD data, intelligent device sensors, floating vehicle GPS data, roadside video data, millimetre-wave radar and lidar. Each of these data sources has limitations. For example, OBD data, intelligent device sensors and GPS data often suffer from delayed transmission, making them less suitable for detailed micro-level driving behaviour studies. Roadside video data, while useful for extracting vehicle licence plates, speed and position, relies heavily on computer software for object detection and tracking, and its accuracy is influenced by factors such as camera placement, lens angle, traffic flow conditions and weather. Millimetre-wave radar and lidar provide high-precision trajectory data for large traffic areas but do not capture vehicle image information, limiting their application.

Despite advancements in characterising unsafe driving behaviour, existing data sources generally suffer from low precision and high latency, hindering the accurate identification of driving behaviours. Therefore, there is a need for high-precision vehicle trajectory data to study bad driving behaviours and assess potential road traffic safety risks from a micro perspective. This study addresses these needs by processing radar-video integrated sensor data to focus on the micro-level identification of unsafe driving behaviours and the assessment of road traffic safety risks based on these behaviours. Utilising high-precision vehicle trajectory data for traffic safety risk evaluation holds both theoretical and practical significance.

1) Theoretical Significance

- By extracting driving behaviour information from high-precision trajectories, it is possible to identify the characteristics of unsafe driving behaviours from a micro perspective, achieving accurate recognition of typical unsafe driving behaviours and understanding their spatiotemporal distribution patterns.
- Introducing entropy theory for the micro-level identification of unsafe driving behaviours and constructing an indicator system for traffic safety risks can enrich the road traffic safety assessment methodologies.

2) Practical Significance

- Understanding the occurrence patterns of unsafe driving behaviours can enhance the efficiency of road traffic management and ensure the smooth operation of road traffic.
- Feedback from unsafe driving behaviour data on road safety conditions helps in perceiving traffic safety risks, providing theoretical support for traffic control measures and ensuring safe travel for road users.

This paper is structured as follows. Section 2 reviews and summarises existing research related to unsafe driving behaviours and road traffic safety risk assessment. Section 3 focuses on the identification of unsafe driving behaviours. Section 4 explains the evaluation of road safety based on unsafe driving behaviours. Section 5 validates the proposed model and algorithms using real-world scenarios. Section 6 presents the conclusions of the study, discusses its limitations and outlines future research directions.

2. LITERATURE REVIEW

At present, there is much research on driving behaviour. This paper reviews the current research from two aspects: identification of unsafe driving behaviour and road traffic risk assessment.

2.1 Identification of unsafe driving behaviour

There exist many works about unsafe driving behaviour recognition, which can be divided into threshold rules and machine learning or deep learning algorithms.

For the research method of detecting unsafe driving behaviour based on threshold rules, empirical formulas or analytical derivations are used to specify the threshold of vehicle kinematic parameters, to identify and analyse driving behaviour. Common parameters include speed [2], acceleration and travel time [3], yaw rate [4], etc. Lu et al. [5] used a large-sample statistical distribution method to determine the characteristic value thresholds for unsafe driving behaviours. The unsafe driving behaviour spectrum was established. Ma et al. [6] set judgment thresholds for abnormal behaviour in speed, acceleration, and braking settings, achieving real-time detection of aggressive driving behaviour by balancing driving safety and driver/passenger comfort. In addition, researchers also prefer to use thresholds such as lateral acceleration change amount, GPS direction change amount and driving distance for unsafe driving behaviour recognition. Although these threshold methods have interpretability and computational efficiency, these rules are too simple to produce satisfactory detection results.

Another category of research methods for detecting unsafe driving behaviour is based on machine learning or deep learning algorithms to construct more complex rules. Liu et al. [7] proposed an unsafe driving behaviour recognition method based on the covariance manifold and binary classification idea of the multi-class LogitBoost classifier. Shahverdy et al. [8] learned a two-dimensional convolutional neural network (CNN) based on the recursive graph technique, which is constructed on images constructed from driving signals using the recursive graph technique. The experimental results show that the proposed method can efficiently detect driver behaviour. Hu et al. [9] characterised and simulated three typical unsafe driving behaviours: fatigue/drunkenness, recklessness and using mobile phones while driving, proposed abnormal indicators and applied them to quantitatively evaluate abnormalities. Hu et al. [10] established a novel abnormal driving detection model based on deep learning, using autoencoders as building blocks to represent the driving features of abnormal driving detection. Huang et al. [11] propose a long short-term memory (LSTM) neural networks (NN) based car-following (CF) model to capture realistic traffic flow characteristics by incorporating the driving memory. The results indicate that the LSTM-NN model can effectively capture asymmetric driving behaviour.

Liu et al. [12] proposed a one-dimensional convolutional neural network model for ADBD, and experimental results showed that the proposed one-dimensional CNN model efficiently achieved multi-classification of unsafe driving behaviour, with an average accuracy of 97%, significantly better than traditional k-nearest neighbour and support vector machine algorithms. Chen et al. [13] proposed an effective method for identifying unsafe driving behaviour based on convolutional neural networks (CNN) and transfer learning. The results show that transfer learning can effectively increase the convergence speed and recognition accuracy Xiao et al. [14] proposed the fuzzy deep attention network (FDAN) method to improve driver behaviour recognition ability. FDAN integrates fuzzy logic and attention mechanisms into deep neural networks, enhancing the model's representational power and reducing data uncertainty. Darsono et al. [15] proposed a deep learning model based on LSTM, which uses OBD-II data to classify driving behaviour. The results indicate that the proposed model exhibits a natural ability to preserve and utilise temporal information in input data, surpassing traditional machine learning methods.

2.2 Road traffic risk assessment based on driving behaviour data

Many studies have established road safety evaluation models by considering various factors of unsafe driving behaviour. Considering that the current traffic safety evaluation index system is relatively weak in terms of driver behaviour characteristics and vehicle operation characteristics, Eren et al. [16] analysed drivers' safety and unsafe behaviour through optimal path detection algorithms and Bayesian classification statistics. Chen et al. [17] conducted a road safety risk assessment process from a holistic perspective using an improved entropy TOPSIS-RSR method based on the comprehensive road safety risk index (RSRI). Yan et al. [18] constructed a data-driven driving safety assessment method based on actual driving behaviour data in tunnels, providing an effective and generally acceptable method for identifying driving risk criteria that can also be applied to traffic management and safety countermeasures with a view to possible implementation in continuous tunnels. Chen et al. [19] constructed a new nonnegative constraint focus loss classifier for predicting driving behaviour under different safety risk levels. The results showed that this method can effectively find the optimal window size, reduce data volume and reconstruction errors, and extract more significant features. Cai et al. [20] applied information entropy theory to traffic safety evaluation based on vehicle OBD data, and constructed a road traffic safety risk evaluation system using driving behaviour data as the evaluation index, achieving the discrimination of potential road traffic safety risks. Wang et al. [21] proposed a risk assessment method based on driver improper driving behaviour and abnormal vehicle status warning data, which uses entropy weight method to determine the risk responsibility weight of each warning type, and then determines the risk classification threshold based on Gaussian mixture model algorithm. Yang et al. [22] innovatively developed an indicator called "Traffic Dynamic Operational Risk (TDOR)" based on aggressive driving behaviours (ADBs) and traffic flow data for traffic safety assessment.

2.3 Research gap

Research on identifying unsafe driving behaviour has become relatively mature, and data-driven traffic safety risk assessment based on driving behaviour has gradually become a research hotspot. However, further research is still needed in the following two aspects:

1) Coordination between existing methods and new traffic data. The reference of new data can further optimise the parameter judgment threshold for unsafe driving behaviour, which is conducive to mining more microscopic unsafe driving behaviour. Compared with traditional abnormal driving recognition data, the data of radar-video integrated sensors has higher accuracy (centimetre level) and faster return speed (millisecond level). In addition, based on the existing methods for identifying unsafe driving behaviour, more microscopic constraint rules are expected to be incorporated.
2) Exploration of new data-driven traffic safety risk assessment methods. Introducing new data can transform passive driving, mainly based on accident data in existing road traffic safety assessment research, into active measurement guided by driving behaviour data. With the advent of the digital era of transportation, high-precision and low-latency driving behaviour data are becoming increasingly easy to obtain. Active assessment of traffic safety risks will be the future development trend of intelligent transportation. Therefore, further exploration of road traffic safety risk assessment methods driven by new data is necessary.

This article presents an objective evaluation approach that utilises data from radar-video integrated sensors and entropy theory to address safety risk assessment challenges in complex urban road traffic systems. It focuses on recognising and analysing unsafe driving behaviour using Thunder Vision all-in-one machine data, calibrating its characteristic indicator thresholds, and proposing a recognition principle and algorithm process. The method introduces an entropy theory-based evaluation system with traffic safety entropy as the primary indicator and unsafe driving behaviour rate as the secondary indicator. A clustering algorithm determines the classification number and threshold of traffic safety entropy, leading to a tunnel traffic safety risk assessment model. The feasibility of the evaluation results is verified using road accident data.

3. IDENTIFICATION OF UNSAFE DRIVING BEHAVIOUR

Unsafe driving behaviour specifically refers to speeding, abnormally low speed, unstable speed, rapid acceleration, rapid deceleration and abnormal car-following. When a driver engages in unsafe driving behaviour, it is difficult for them to make the correct response in a short period in the event of an emergency, especially when driving on urban roads with complex driving environments and relatively saturated traffic flow. The safety hazards are extremely high and can easily trigger traffic accidents.

3.1 Data analysis and preprocessing

Data preprocessing is the first work of data analysis and mining, abnormal data is a common situation in the process of data analysis. To accurately extract vehicle trajectory information, data should be screened and processed at the beginning of the study.

Data introduction

Radar video integrated sensors are multimodal intelligent traffic sensors based on millimetre wave radar and video detection. They can accurately detect and identify parameters such as the position, speed, direction and type of traffic participants in complex road environments. At the same time, deep learning algorithms are used to achieve detection and early warning of various traffic events. The data in this article is sourced from the Jiaozhou Bay Tunnel Video Traffic Flow Detection and Application System, which collects high-precision traffic trajectory data in real-time by deploying radar video integrated sensors in the field.

The data collection frequency of the Thunder Vision equipment is 10Hz, which means that data are transmitted back every 100 ms. For example, the data from the radar video integrated sensors in the Jiaozhou Bay Tunnel in Qingdao reached 2.69 million in one day. *Figure 1* shows the parsed raw data returned from the server, and *Table 1* shows the data fields and their meanings.

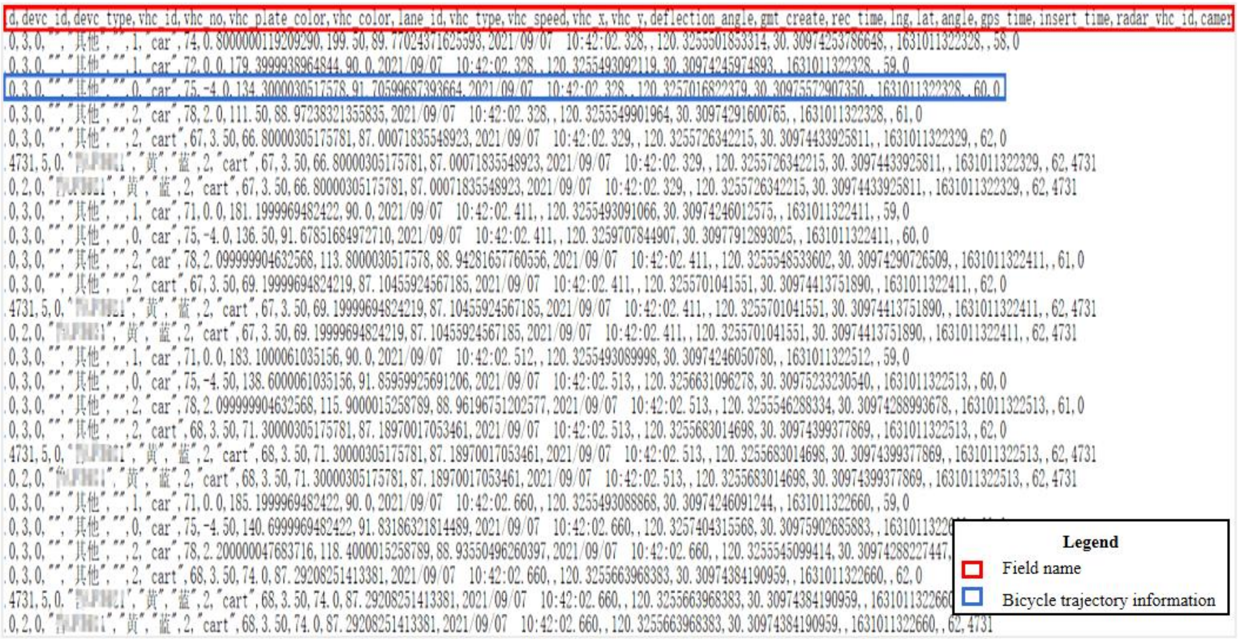


Figure 1 – The parsed raw data of the radar video integrated sensors

Table 1 – Data fields and their meanings for radar video integrated sensors

Number	Field name	Field meaning
1	devc_id	Camera device IP
2	vhc_id	Vehicle ID
3	vhc_no	Licence plate number
4	vhc_plate	Licence plate colour
5	vhc_color	Vehicle colour
6	lane_id	Lane number
7	vhc_type	Object type detected: vehicle, person
8	vhc_speed	Target speed – longitudinal speed unit, km/h
9	vhc_x	The x-coordinate of the vehicle relative to the equipment
10	vhc_y	The y-coordinate of the vehicle relative to the equipment
11	lng	Longitude
12	lat	Latitude
13	radar_vhc_id	Vehicle identification number for coordinate detection
14	gmt_create	Vehicle coordinate detection time, 100 ms level
15	rec_time	Receiving time, in seconds
16	gps_time	Detection time, converted from gmt_create
17	devc_type	3 radar data, 2 video data, 5 fused data

The original data contains irrelevant data, as well as issues like data duplication, redundancy, and errors. Therefore, it is necessary to perform data extraction, cleaning, and association on the original data.

Data preprocessing

1) Single-device data extraction

The raw data records millisecond-level data of 17 fields, including camera IP, vehicle ID, licence plate number, vehicle colour, licence plate colour, lane number, longitude and latitude, etc. In identifying abnormal driving behaviour, field data such as vehicle colour and licence plate colour are of no value to this study. To improve the efficiency of big data processing in the later stage and efficiently carry out data analysis and mining work, it is necessary to extract field data related to the research content from the original data and remove irrelevant data.

2) Multi-device data association

Data association is the process of integrating data from different sources or databases into one database through shared fields or variables to achieve data usage, analysis and cross-database queries between different databases. At present, the mainstream brands of Leishi all-in-one machines in the market are Dahua and Hikvision, and the effective detection range of their products' vehicle coordinates is around 150 m. To obtain long-distance vehicle driving data, matching and associating the extracted single-device data with feature fields is necessary.

3) Data processing of equipment splicing section

When adjacent devices collect vehicle information, there may be duplicate vehicle information detection due to overlapping collection ranges, as shown in *Figure 2*. To address this issue, the data collection interval of adjacent devices is divided to eliminate duplicate data in the splicing segment and achieve vehicle coordinate splicing of neighbouring devices.

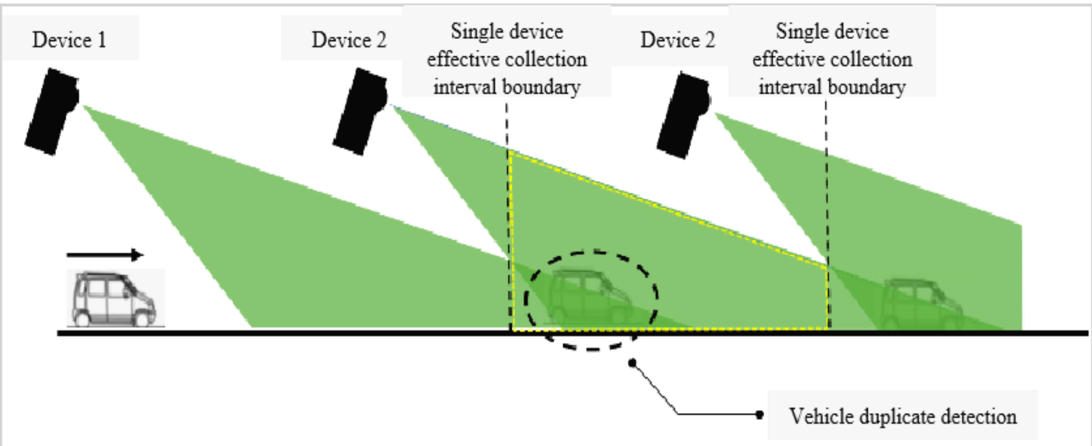


Figure 2 – Schematic diagram of vehicle coordinate stitching between adjacent devices

Abnormal data analysis and processing

Abnormal data is a common situation in data analysis, and abnormal data detection is a prerequisite for analysing the causes of data anomalies. Currently, there are many methods for detecting data anomalies, such as simple descriptive statistics and 3σ Principles, quartile tests, density-based clustering, etc. [27], these methods mainly target numerical data, and their main principle is to strip out abnormal data by analysing the overall characteristics of the data. Based on the Qingdao Jiaozhou Bay Tunnel Digital Twin Application Project, considering the data characteristics of the original data, the detection of abnormal data can be achieved by using the Numpy module in Python and rough retrieval of duplicate and NAN values using the duplicated function and is null function. On the other hand, since the data of radar-video integrated sensors meet the accuracy requirements for digital twinning, it can be combined with digital platforms to use the original trajectory data for twinning, thereby intuitively observing anomalies in the original data.

3.2 Determination of indicators for unsafe driving behaviour

By analysing the indicators of unsafe driving behaviour in traditional data, combined with the characteristics and content of the data, the indicators of unsafe driving behaviour and its characterisation were determined. This study identified and determined the threshold values of each characterisation indicator for various unsafe driving behaviours through an extensive literature review and research experience. The results are shown in *Table 2*.

Table 2 – Threshold of characterisation indicators of unsafe driving behaviour

Unsafe driving behaviour	Characterisation indicator	Threshold
Speeding	Speed	≥80km/h-
	Duration	≥3s
Abnormal low speed	Speed	≥80km/h
	Following Distance	≥150m
	Duration	≥2s
Unstable speed	Speed standard deviation	≥10.684119
Rapid acceleration	Acceleration	≥3m/s^2
	Duration	≥2s
Rapid deceleration	Acceleration	≥-3m/s^2
	Duration	≥2s
Abnormal car-following	TTC	≥3s
	Following distance	≥150m
	Duration	≥2s

TTC: It is an abbreviation for Time to Collision, representing collision time.

3.3 Identification of unsafe driving behaviour

Based on the threshold of each unsafe driving behaviour characterisation indicator, it can reduce the difficulty of identifying unsafe driving behaviour and improve the accuracy of model recognition. This study determined the threshold values of each characterisation indicator through extensive literature review and research experience and identified four types of unsafe driving behaviours: speeding, abnormal low speed, unstable speed and abnormal car-following. The specific process will be elaborated on elsewhere. This article designs a recognition process for the two behaviours of rapid acceleration and deceleration and writes recognition code suitable for massive data in Python. Due to space limitations, this article only introduces the identification of rapid acceleration behaviour.

Rapid acceleration

There are two indicators of rapid acceleration behaviour: acceleration and duration. Due to the lack of acceleration data provided by radar video integrated sensors, the acceleration of the vehicle needs to be calculated before identifying rapid acceleration and deceleration, as shown in *Equation 2*:

$$a_{(i+1)} = \frac{V_{(i+1)} - V_{(i)}}{\left(t_{(i+1)} - t_{(i)}\right) \cdot 3.6} \tag{1}$$

where $V_{(i)}$ is the speed at time *i*; $V_{(i+1)}$ is the speed at time *i* + 1.

To determine whether the vehicle has undergone rapid acceleration behaviour, the acceleration threshold A for rapid acceleration behaviour is introduced, and combined with previous research and industry experience, a value of 3 m/s^2 is taken.

According to relevant research, there is currently no unified standard for the duration of rapid acceleration behaviour. This study selected data from the Jiaozhou Bay Tunnel from 20 October 2022 to 24 October 2022, for five days. To ensure the accuracy of calibration parameters, the entire day was divided into four time periods based on the traffic flow characteristics of the Jiaozhou Bay Tunnel, and two hours of data were selected as representative data for each period. Among them, 0–2 refers to the period from 00:00 to 02:00 in the morning, representing the nighttime peak data of the Jiaozhou Bay Tunnel. Therefore, this article analyses the recognition results with durations of 2 s and 3 s through sample data, as shown in *Table 3*.

Table 3 – Recognition results of rapid acceleration behaviour under different durations

Time period	Date	Data count	Duration≥2s		Duration≥3s	
			Behaviour count	Behaviour rate	Behaviour count	Behaviour rate
0-2	10/20	377	43	11.41%	1	0.27%
	10/21	418	58	13.88%	1	0.24%
	10/22	523	55	10.52%	0	0.00%
	10/23	411	59	14.36%	1	0.24%
	10/24	357	51	14.29%	2	0.56%
7-9	10/20	6871	1341	19.52%	41	0.60%
	10/21	6699	1309	19.54%	61	0.91%
	10/22	5150	909	17.65%	22	0.43%
	10/23	4375	717	16.39%	17	0.39%
	10/24	6871	1314	19.12%	39	0.57%
10-11 13-14	10/20	4564	794	17.40%	13	0.28%
	10/21	4682	855	18.26%	19	0.41%
	10/22	5274	1028	19.49%	28	0.53%
	10/23	5251	900	17.14%	17	0.32%
	10/24	4564	723	15.84%	49	1.07%
16-18	10/20	6595	1460	22.14%	47	0.71%
	10/21	7272	1631	22.43%	92	1.27%
	10/22	6867	1524	22.19%	57	0.83%
	10/23	7076	1394	19.70%	45	0.64%
	10/24	6595	1119	16.97%	34	0.52%
Average behaviour rate	-	-	17.58%	-	0.55%	

From the table, it can be seen that the behaviour rate at a duration threshold of 3 seconds for rapid acceleration behaviour is much lower than that at a duration threshold of 2 seconds. In the case of a duration threshold of 3 seconds, there were 0 rapid acceleration behaviours during the 0–2 period, and 55 were identified with the same duration threshold of 2 seconds. Setting the duration to 2 seconds is more reasonable, so this article sets the duration threshold for rapid acceleration behaviour to 2 seconds.

Recognition algorithm for rapid acceleration behaviour

The detailed steps of the algorithm for identifying rapid acceleration behaviour are as follows:

Step I: Analyse and read the raw data of radar video integrated sensors, and use PyCharm programming software to convert the raw data item by item into a list for easy traversal and indexing in the future.

Step II: Group the converted data based on licence plate numbers, convert the timestamp format to time format, and sort the grouped data according to the time rules from small to large and latitude from small to large.

Step III: Starting from the second data point in each group, calculate the vehicle acceleration α at each moment according to *Equation 1* and set the initial duration T to 0.

Step IV: Determine if the acceleration exceeds the set acceleration threshold. If it does, retain the time information of the data and proceed to step 6. If not satisfied, proceed to step 5.

Step V: Determine whether it is the last data entry. If the result is yes, then there is no rapid acceleration behaviour, and the process is over. If the result is negative, reset the duration T to 0, read the data for the next time step, and proceed to step 4.

Step VI: Determine whether the duration T of the behaviour that meets the acceleration threshold for rapid acceleration is greater than or equal to 2 seconds. If the judgment result is yes, it indicates that rapid acceleration behaviour has occurred. If the judgment result is negative, it indicates that there has been no rapid acceleration behaviour. Go to step 7.

Step VII: Determine whether the current data are the last data of the vehicle. If so, the algorithm ends. If not, read the data for the next time step and proceed to step 4.

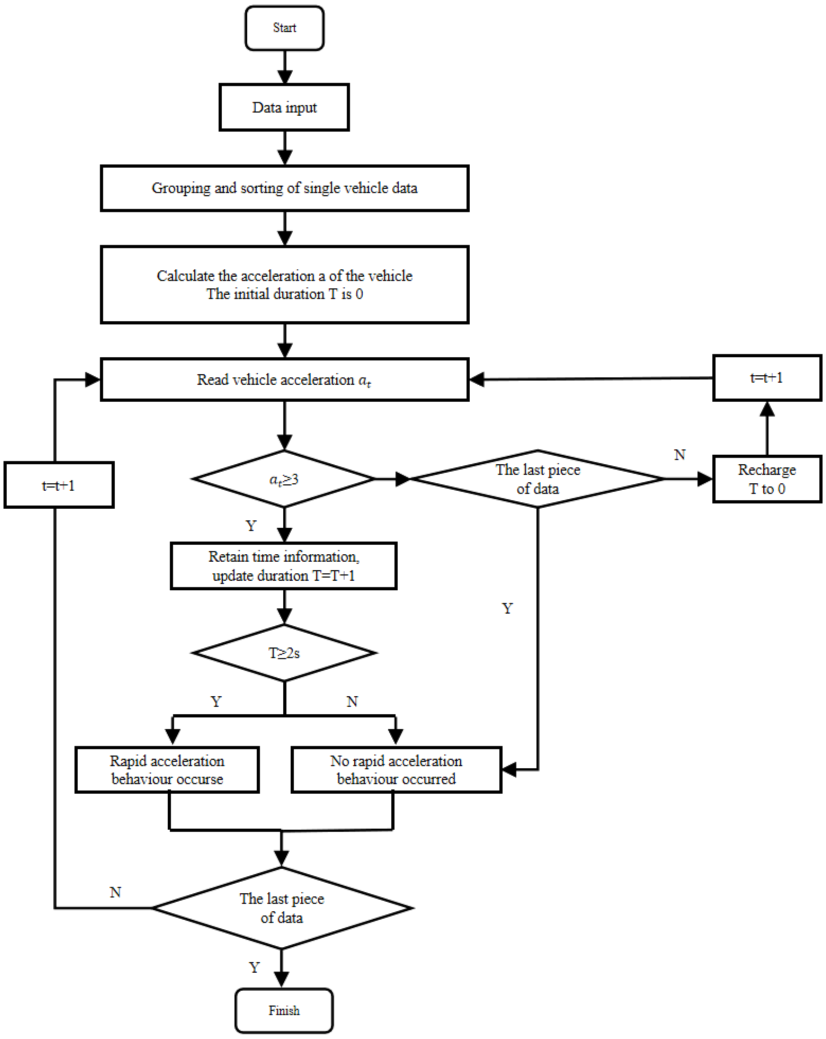


Figure 3 – Process diagram for rapid acceleration behaviour recognition

4. ROAD TRAFFIC SAFETY RISK ASSESSMENT

The entropy method can determine the index weight objectively by analysing the variation degree of each index value, avoiding the deviation caused by human factors, and has a strong mathematical theoretical basis. However, the traditional entropy method tends to produce a large deviation in the treatment of extreme values in this study, so it is improved in this paper to effectively ensure the description of objective facts.

4.1 Evaluation indicator system

This article takes the entropy value of road traffic operation under the influence of unsafe driving behaviour as the primary indicator for estimating road traffic safety risks and selects the six indicators shown in *Table 4* as the secondary indicators for road traffic safety risk assessment.

Table 4 – Evaluation indicators system

First level indicator	Second level indicators
Road traffic safety entropy	Speeding rate, Abnormal low-speed rate, Unstable speed rate, Rapid acceleration rate, Rapid deceleration rate, Abnormal car-following rate

4.2 Road traffic safety risk assessment based on improved entropy method

The calculation steps for the road traffic safety risk assessment based on the improved entropy method are as follows:

Step I: Based on the results of identifying unsafe driving behaviours, calculate the probability of each unsafe driving behaviour occurring during time period *m* on road segment *i*.

$$P_{ij}^{m}=\frac{N_{ij}^{m}}{Q_{i}^{m}} \tag{2}$$

where *i* represents the number of road sections, *i* = 1, 2,..., *q*; *j* is the number of indicators, *j* = 1, 2,..., *k*; *m* is the time period, *m* = 1, 2,..., *p*. The number of vehicles that engaged in the j^{th} type of unsafe driving behaviour during time period *m* on section *i*; Q_i^m is the total number of vehicles detected in time period m for section *i*; P_{ij}^m is the incidence rate of the j^{th} type of unsafe driving behaviour in section *i* during period *m*.

Step II: Use the MAX-MIN method to standardise data and unify data dimensions.

$$x_{ij}^{m}=\frac{p_{ij}^{m}-p_{ij}^{min}}{p_{ij}^{max}-p_{ij}^{min}} \tag{3}$$

$$x_{ij}^{m}=\frac{p_{ij}^{max}-p_{ij}^{m}}{p_{ij}^{max}-p_{ij}^{min}} \tag{4}$$

where x_{ij}^m is the j^{th} indicator value of section *i* during time period *m*; p_{ij}^{min} is the minimum occurrence rate of the j^{th} type of unsafe driving behaviour on section *i*; p_{ij}^{max} is the maximum occurrence rate of the j^{th} type of unsafe driving behaviour on section *i*.

Step III: Calculate the proportion of the j^{th} indicator under each indicator value in the i^{th} section.

$$\lambda_{ij}^{m}=\frac{x_{ij}^{m}}{\sum_{i=1}^{q}x_{ij}^{m}} \tag{5}$$

where *q* represents the total number of road sections; λ_{ij}^m represents the proportion of the j^{th} indicator under each indicator value in the i^{th} section during time period *m*.

Step IV: Calculate the entropy value of each indicator.

$$E_j=\frac{1}{lnq}\sum_{i=1}^{q}\sum_{m=1}^{p}\lambda_{ij}^{m}\cdot ln\lambda_{ij}^{m} \tag{6}$$

where E_j is the entropy value of the j^{th} indicator, $E_j \geq 0$; If it exists $\lambda_{ij}^m = 0$, let it be $\lambda_{ij}^m = \lambda_{ij}^m + 0.00001$.

Step V: Calculate the weight of each indicator.

$$\begin{cases} w_{j\_2}=(1-\bar{E}^{35.35})\cdot w_{j\_0}+\bar{E}^{35.35}w_{j\_1} & E_j<1 \\ 0 & E_j=1 \end{cases} \tag{7}$$

$$w_{j\_0}=\frac{1-E_j}{\sum_{j=1}^{k}(1-E_j)} \tag{8}$$

$$w_{j\_1} = \frac{1 + \bar{E} - E_j}{\sum_{j=1}^{k}(1 + \bar{E} - E_j)} \tag{9}$$

where w_{j_2} is the entropy value of the j^{th} indicator, $\bar{E}$ is the average entropy value for all indicators with entropy values not equal to 1.

Step-VI: Calculate the safety entropy value of the i^{th} segment unit of the road during the m^{th} time period.

$$S_i^m = \sum_{j=1}^{k} w_j \cdot (-p_{ij}^m) \cdot lnp_{ij}^m \tag{10}$$

S_i^m is the safety entropy value of the j^{th} indicator within time period m.

4.3 Classification of traffic safety levels

When classifying road traffic safety levels based on correlated accident data, the abnormal data are first stripped off. Then the accident data and road safety entropy data are divided using clustering thinking. The group with the highest contour coefficient is selected as the basis for dividing road safety levels.

K-means clustering is a typical unsupervised classification algorithm that aims to automatically classify similar samples by maximising the similarity between categories and minimising the similarity between different categories. Its principle is simple, and the clustering effect is good. The algorithm flow of K-means clustering is as follows:

Step I: Enter the K value of the number of data categories to be divided.

Step II: Randomly select K points as the initial cluster centres for each category, then calculate the distance from each point to the cluster centre and assign the point to the nearest cluster centre, thus forming K clusters.

Step III: Recalculate the centroid of each cluster.

Step IV: Repeat Steps II and III until the centroid position of each cluster no longer changes or reaches the set number of iterations.

Step V: Clustering ends, generating the final category of each data and the distance from that data to the cluster centre of the category.

4.4 Determination of the optimal grading threshold

The optimal classification threshold can ensure the accuracy of dividing safety entropy values. To find the optimal classification threshold for each level, the concept of classification accuracy is introduced, and the calculation method for classification accuracy is as follows:

$$A = 1 - \frac{n_1 + n_2}{N} \cdot 100\% \tag{11}$$

where A represents the classification accuracy; N is the total number of data in adjacent clusters; $n_1$is the number of K-means clustering results in the first category, and the classification threshold is divided into the second category of data; n_2 is the K-means clustering result, and the classification threshold is divided into the number of data in the first category.

5. VERIFICATION AND RESULT ANALYSIS

5.1 Data source and processing

Data sources

The data used for the instance verification are the digital twin data of the Qingdao Jiaozhou Bay Smart Tunnel Digital Twin Traffic Operation and Control System, and Qingdao Guoxin Jiaozhou Bay Transportation Co., Ltd, which provides the accident data. To ensure the scientific and objective verification of the example, 13 days of data were selected from the long downhill section of the Jiaozhou Bay Tunnel from Huangdao to Qingdao direction to identify unsafe driving behaviour and calculate road traffic safety entropy. The data period was from 25 November 2022 to 7 December 2022, with pile numbers ZK6+900 to ZK8+300. To accurately characterise the traffic safety risks of each section unit of the Jiaozhou Bay Tunnel, the road was divided into 50/m as a section unit for road traffic safety risk calculation. *Figure 4* shows the basic road conditions of the selected data.

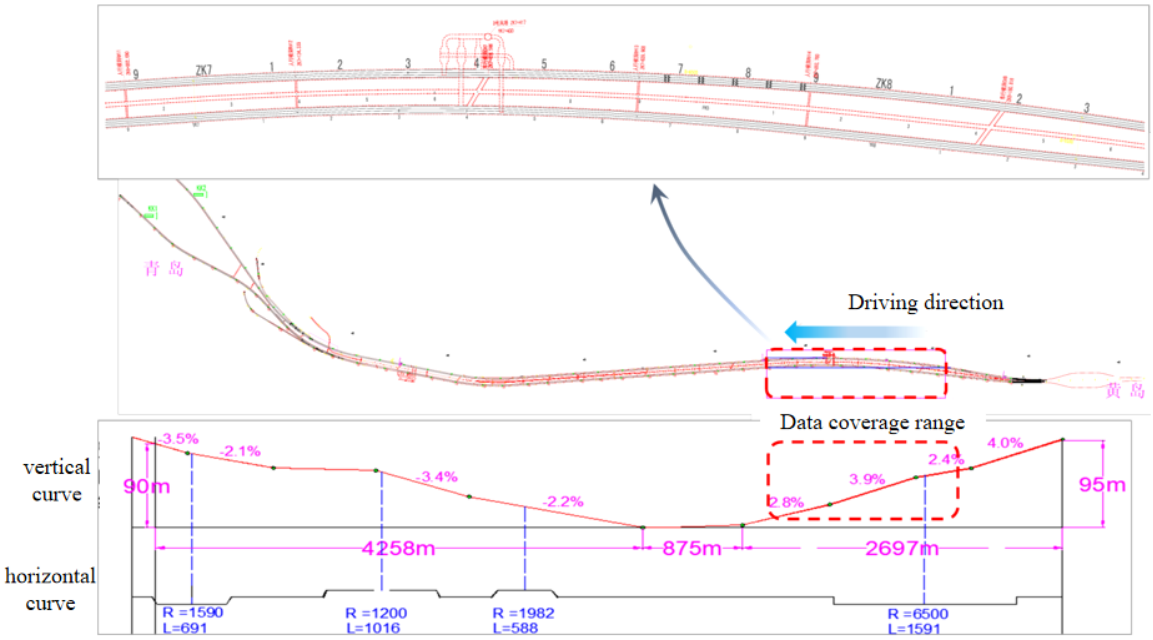


Figure 4 – Basic conditions of the selected road

Data sample size analysis

1) Introduction to data types

The raw data of Leishi stores all data in a "CSV" document at intervals of every 100 ms. The accident data records detailed information such as the station number, lane, cause and type of accident based on the date of the accident.

2) Data sample size analysis

Perform lane-by-lane statistics on traffic flow using 13-day data at 1-hour intervals and draw a time-varying line graph of traffic flow, as shown in *Figures 5–7*. By observing the time-varying patterns of traffic flow in each lane of the 13-day data, it can be found that the trend of traffic flow over time is relatively stable. The traffic flow is highest during the morning and evening peak hours, followed by the daytime off-peak period, and the nighttime off-peak period has the lowest traffic flow. The changes in data can reflect the travel demand characteristics during the morning and evening rush hour in cities. In addition, the traffic flow in the left and middle lanes is relatively close at different times, while the traffic flow in the right lane is lower. During peak hours, the hourly traffic flow in the left middle lane can reach over 1,800 vehicles, and the maximum traffic flow in the right lane is 989 vehicles.

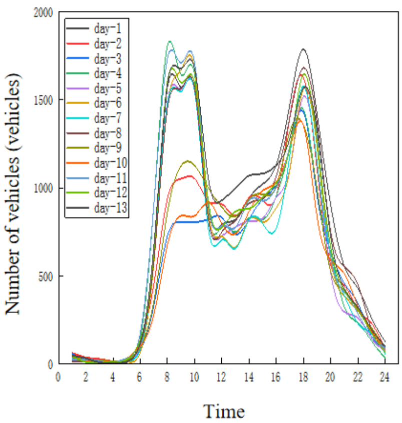


Figure 5 – Time-varying diagram of hourly traffic flow in the left lane

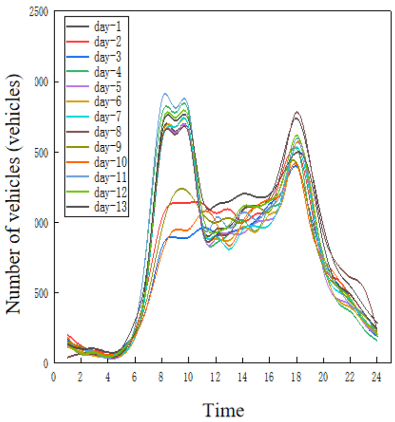


Figure 6 – Time-varying diagram of hourly traffic flow in the middle lane

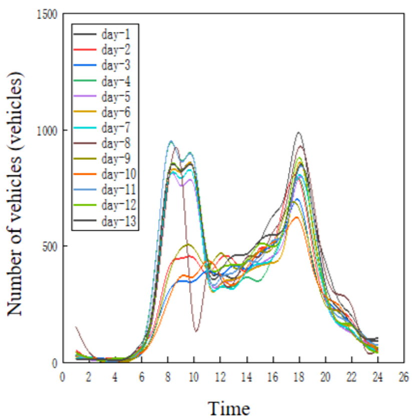


Figure 7 – Hourly traffic variation diagram of the right lane

To ensure the accuracy of road safety entropy classification results, accident data from 2018 to 2021 were statistically analysed, and the statistical results are shown in *Figure 8*. According to the statistical results, the middle lane has the highest number of traffic accidents, the left lane has slightly lower traffic accidents than the middle lane, and the right lane has the lowest number of traffic accidents. There are over 400 accident data points over the past four years, providing assurance for the accuracy of the entropy classification results of road traffic safety.

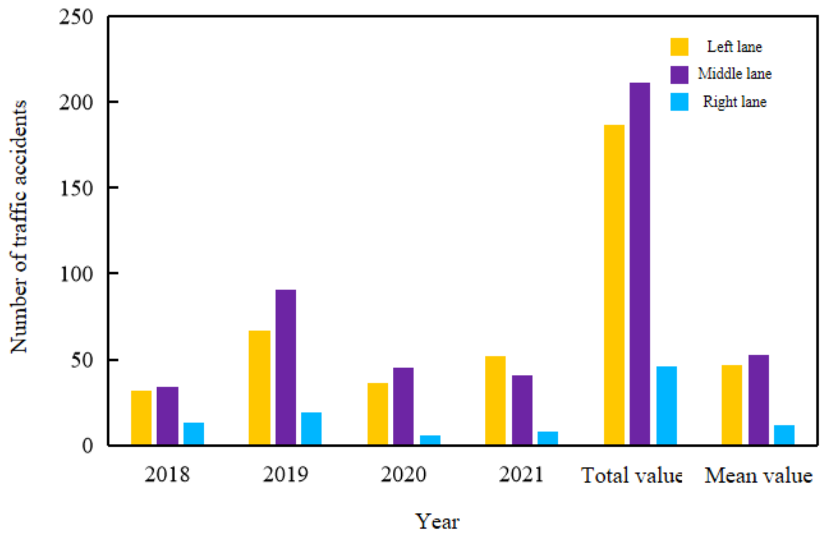


Figure 8 – Number of lane separation accidents from 2018 to 2021

In summary, the data used for example verification in this article have the characteristics of rich content, sufficient data volume and reliable data, which can effectively support the identification of unsafe driving behaviour and the estimation of road traffic safety risks based on unsafe driving behaviour data in this article.

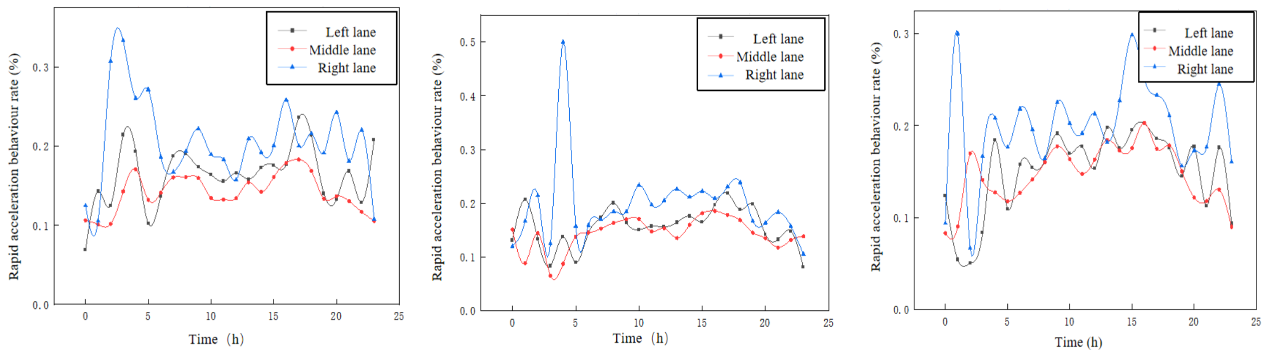
5.2 Identification of unsafe driving behaviour

Identifying unsafe driving behaviour is preliminary work for road traffic safety risk assessment, and the implementation of the identification program is mainly based on four modules: NumPy, pandas, collections and data time. After the original data is processed by code, a new table will be generated, which includes the acceleration data, speed standard deviation, TTC data, front licence plate data, front distance and identification result data of unsafe driving behaviour of the vehicle at each moment. The identified data results are shown in *Figure 9*.

	vhcNo	laneId	vhcSpeed	lat	mTime	加速度	是否急加速	是否急减速	是否超速	标准差
0		2	63	36.007507	2022-11-25 06:27:28	-9999.000000	否	否	否	5.50906
1		2	63	36.007732	2022-11-25 06:27:29	0.000000	否	否	否	5.50906
2		2	65	36.007956	2022-11-25 06:27:30	0.555556	否	否	否	5.50906
3		2	67	36.008181	2022-11-25 06:27:31	0.555556	否	否	否	5.50906
4		2	72	36.008406	2022-11-25 06:27:32	1.388889	否	否	否	5.50906
...		...	...	...	...	...	...	...	...	...
95		2	79	36.013740	2022-11-25 06:27:07	-0.555556	否	否	否	5.26836
96		2	81	36.013877	2022-11-25 06:27:08	0.555556	否	否	否	5.26836
97		2	64	36.014099	2022-11-25 06:27:09	-4.722222	否	否	否	5.26836
98		2	70	36.014324	2022-11-25 06:27:10	1.666667	否	否	否	5.26836
99		2	79	36.014522	2022-11-25 06:27:11	2.500000	否	否	否	5.26836

Figure 9 – Identification results

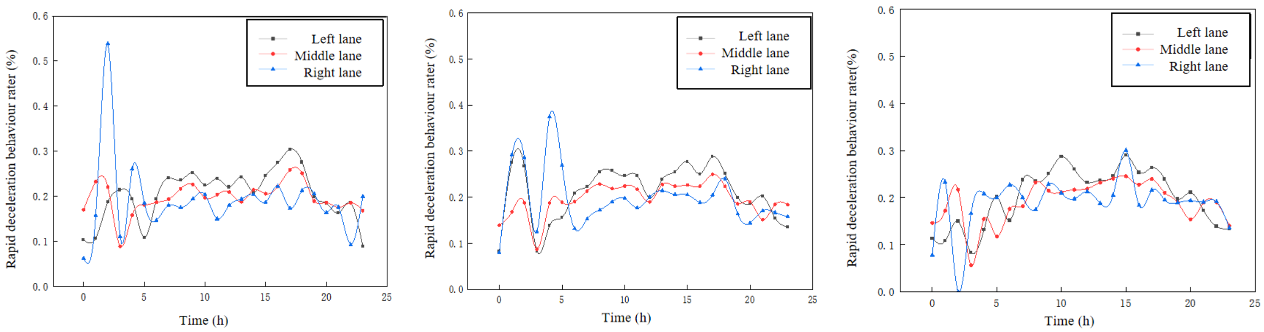
Based on the previously determined characterisation index threshold and recognition algorithm process, the recognition results of two types of unsafe driving behaviour rates, namely rapid acceleration and rapid deceleration, on the long downhill section from Huangdao to Qingdao in Jiaozhou Bay Tunnel from 25 November 2022 to 27 November 2022, are shown in *Figure 10 and Figures 11*.



(a) *(b)* *(c)*

Figure 10 – Rate of rapid acceleration behaviour in lane division: a) Rapid acceleration behaviour rate-11.25; b) Rapid acceleration behaviour rate-11.26; c) Rapid acceleration behaviour rate-11.27

Analysing the recognition results of rapid acceleration behaviour in each lane in *Figure 10*, from the lane dimension, the rate of rapid acceleration behaviour in the right lane is the highest and the difference between it and the other two lanes is more obvious. From a temporal perspective, the peak of the rapid acceleration behaviour rate of the three lanes appears more frequently during periods of low traffic volume, and the overall rapid acceleration behaviour rate of the three lanes shows the same trend over time. The analysis of its causes may be related to the occurrence rate of rapid acceleration behaviour and road traffic flow. The larger the traffic flow, the closer the distance between vehicles and vehicles, and thus the lower the occurrence rate of rapid acceleration behaviour.



(a) *(b)* *(c)*

Figure 11 – Rate of rapid deceleration behaviour in lane division: a) Rapid deceleration behaviour rate-11.25; b) Rapid deceleration behaviour rate-11.26; c) Rapid deceleration behaviour rate-11.27

Analysing the recognition results of sudden deceleration behaviour in each lane in *Figure 11*, from the lane dimension, the sudden deceleration behaviour rate of the left lane is slightly higher than the other two lanes most of the time, while the sudden deceleration behaviour rate of the right lane is the lowest. From a temporal perspective, the peak of rapid deceleration behaviour rate in the right lane occurs during periods of low traffic volume, while the peak of rapid deceleration behaviour rate in the left and middle lanes both occur during periods of high traffic volume. The overall rapid acceleration behaviour rate of the three lanes shows the same trend over time. From this, it can be seen that both rapid acceleration behaviour and rapid deceleration behaviour are influenced by traffic flow. The smaller the traffic flow, the higher the probability of rapid acceleration behaviour occurring, and the larger the traffic flow, the higher the probability of rapid deceleration behaviour occurring.

These observations underscore the correlation between lane-specific traffic conditions and driving behaviours. The data indicate that rapid acceleration is more likely in low-traffic density environments, facilitating higher speeds, while rapid deceleration occurs more frequently under high-traffic conditions, necessitating more frequent braking. This understanding is crucial for analysing traffic dynamics and developing targeted road safety interventions.

Based on the analysis of the above recognition results, there are significant differences in the rates of various unsafe driving behaviours on different lanes in the spatial dimension. Taking rapid acceleration as an example, the right lane has the highest rate of rapid acceleration behaviour and a significant difference compared to the other two lanes. The difference in rapid acceleration behaviour rates between the left and middle lanes is relatively small, while the rapid acceleration behaviour rate in the middle lane is the lowest. In terms of the time dimension, the peak values of the rapid acceleration behaviour rate of the three lanes are more frequent during low traffic volume periods, and the overall trend of the rapid acceleration behaviour rate over time is the same. The traffic flow in the right lane of the Jiaozhou Bay Tunnel is much smaller than that in other lanes, while the traffic flow in the left middle lane is relatively close. Therefore, the drivers in the right lane have a higher rate of rapid acceleration behaviour, while the rates in the left and middle lanes are similar.

Road traffic safety evaluation needs to consider the traffic characteristics of different lanes. If the data of three lanes are mixed, the differences between indicators will be weakened, thereby affecting the accuracy of indicator weights. In addition, after analysing and studying the accident data of each lane on the road section in recent years, it was found that there has been a sudden change in the accident data of the right lane. The number of accidents in the middle lane is significantly higher than that in the left lane, and the annual accident data for the left lane is relatively stable. The sudden change in tunnel accident data may be related to traffic control measures. The data segment selected in this article has not changed the road driving environment in the past two years. Therefore, considering the stability of unsafe driving behaviour in each lane and the natural environment, this article chooses the left lane as the object to carry out the example verification work of this article. Statistical analysis was conducted on the rates of six types of unsafe driving behaviours on the left lane of the long downhill section of the Jiaozhou Bay Tunnel from Huangdao to Qingdao using the Pandas module, covering a period of 13 days from 25 November 2022 to 7 December 2022. The results are shown in *Figure 12.*

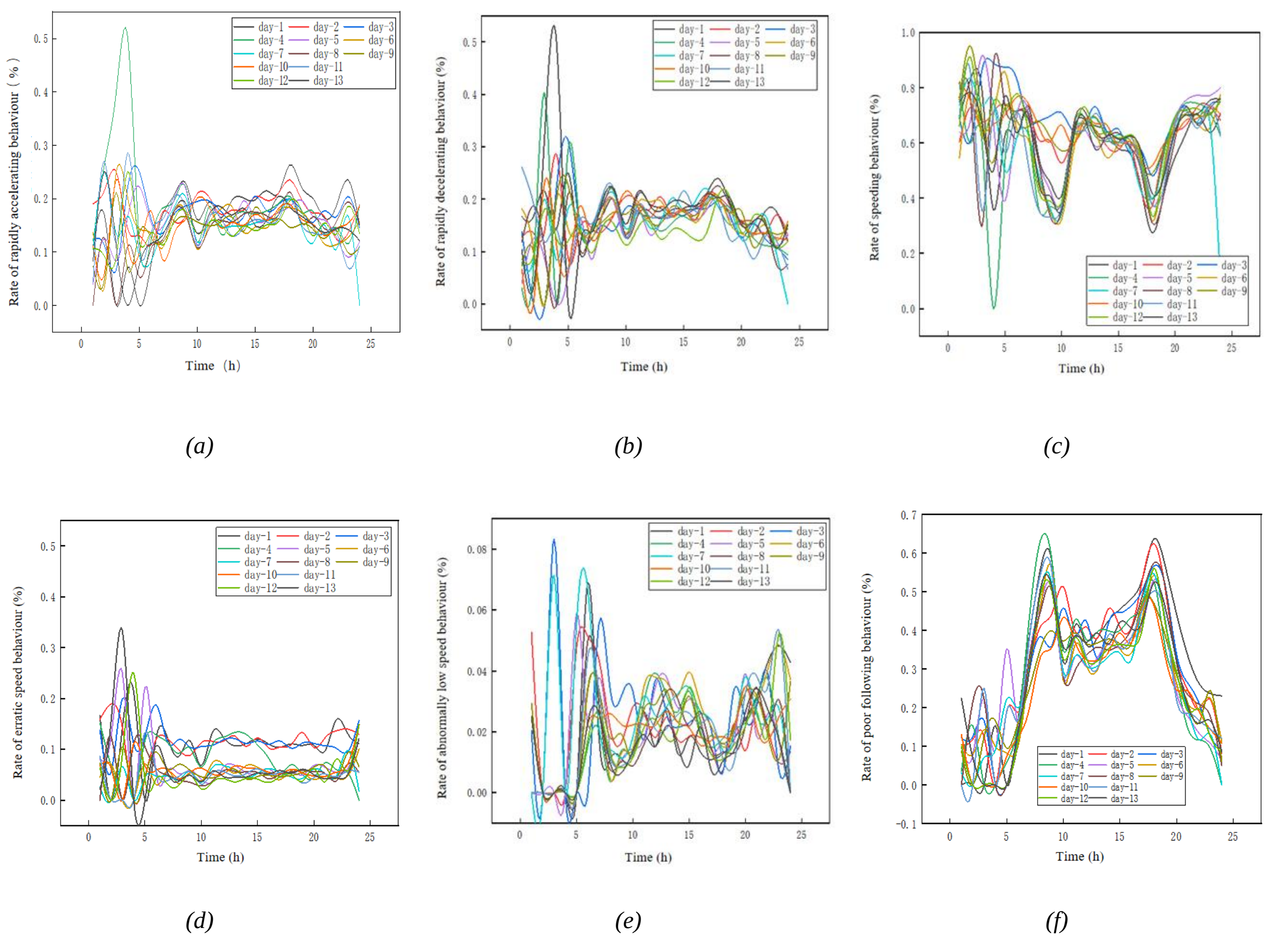


Figure 12 – Rate of unsafe driving behaviour in the left lane: a) rapid acceleration rate; b) rapid deceleration rate; c) speeding rate; d) unstable speed rate; e) abnormal low-speed rate; f) abnormal car-following rate

5.3 Calculation and analysis of entropy value for road traffic safety

The road traffic safety entropy is determined by the index value of each unsafe driving behaviour. Based on the data on unsafe driving behaviour rate obtained above, this part uses the improved entropy method to calculate and analyse the traffic safety entropy of each section.

Indicator weight calculation

To verify the stability of entropy method weighting, the weight data of each behaviour was calculated using 13 days of data, and the average weight of each behaviour for 13 days was used as the weight result for each unsafe driving behaviour. The calculation results are shown in *Table 5*.

Table 5 – Recognition results of rapid acceleration behaviour under different durations

Date	**Indicator weight**					
	Rapid acceleration	**Rapid deceleration**	**Speeding**	**Unstable speed**	**Abnormal low speed**	**Abnormal car-following**
day-1	0.23	0.15	0.18	0.10	0.12	0.22
day-2	0.14	0.09	0.34	0.15	0.12	0.16
day-3	0.17	0.13	0.21	0.11	0.13	0.25
day-4	0.08	0.09	0.33	0.20	0.10	0.19
day-5	0.21	0.25	0.14	0.11	0.10	0.19
day-6	0.10	0.29	0.15	0.19	0.11	0.16
day-7	0.12	0.26	0.20	0.12	0.11	0.19
day-8	0.21	0.24	0.16	0.11	0.11	0.17
day-9	0.13	0.18	0.08	0.27	0.13	0.22
day-10	0.13	0.20	0.13	0.19	0.08	0.26
day-11	0.14	0.16	0.19	0.16	0.10	0.25
day-12	0.13	0.19	0.19	0.13	0.11	0.25
day-13	0.13	0.08	0.29	0.16	0.07	0.27
Mean value	0.15	0.17	0.20	0.15	0.11	0.22
Standard deviation	0.04	0.07	0.07	0.04	0.02	0.04

Calculation of entropy value for road traffic safety

Based on the weights of unsafe driving behaviour given in *Table 5*, the road traffic safety risk entropy values have been calculated. Additionally, the number of accidents for each section of the road in 2020 and 2021 has been compiled. The results are summarised in *Table 6*.

Table 6 – Number of unit accidents on various road sections in 2020 and 2021

Road section number	Safety entropy value	Number of accidents			Road section number	Safety entropy value	Number of accidents		
		2020	2021	Mean value			2020	2021	Mean value
1	0.028568	0	0	0	15	0.049589	0	0	0
2	0.070766	12	20	16	16	0.018179	3	1	2
3	0.050339	0	2	1	17	0.04499	0	0	0
4	0.051534	15	14	14.5	18	0.051299	4	4	4
5	0.048705	1	0	0.5	19	0.042482	0	0	0
6	0.051198	12	20	16	20	0.049342	3	3	3
7	0.046557	0	0	0	21	0.054853	3	3	3
8	0.05178	1	0	6.5	22	0.047582	0	0	0
9	0.050669	4	9	0.5	23	0.060679	7	2	4.5
10	0.056232	5	5	5	24	0.043897	0	1	0.5
11	0.054137	7	8	7.5	25	0.048713	0	0	0
12	0.052077	1	0	0.5	26	0.046506	0	0	0
13	0.047701	0	0	0	27	0.041949	2	0	1
14	0.054327	6	6	6	28	0.044581	0	0	0

The number of accidents near the bottom of the slope (section unit numbers 2–11) was much higher than that in the non-slope bottom section (section unit numbers 19–28). Based on this, a comparative analysis is conducted on the safety entropy value and accident number between the slope bottom section and the non-slope bottom section, the results are shown in *Figures 13 and 14*.

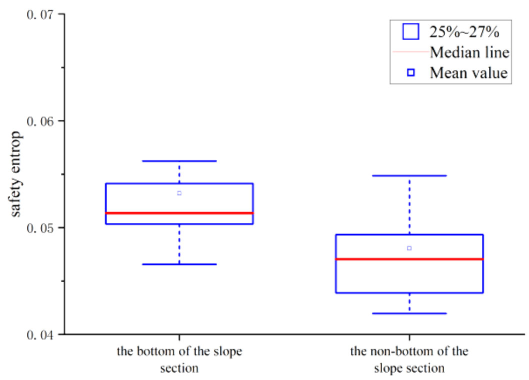


Figure 13 – Safety entropy of slope and non-slope sections

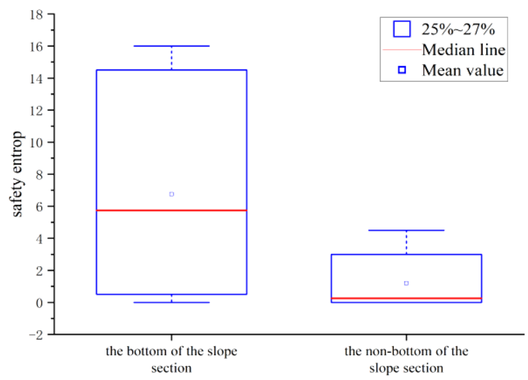


Figure 14 – Accident number of slope and non-slope sections

From these figures, it can be seen that both in terms of mean and median, the number of accidents in the bottom section of the slope is higher than that in the non-bottom section, and the safety entropy value of the bottom section is also higher than that of the non-bottom section. Although there is a significant difference in the number of accidents between the slope bottom section and the non-slope bottom section, and the difference in safety entropy values is relatively small, both reflect the overall trend of higher safety risks in the slope bottom section than in the non-slope bottom section, indicating that road traffic safety entropy can effectively characterise road traffic safety risks.

5.4 Classification of road traffic safety risk states

The reasonable division of road traffic safety entropy is the guarantee to perceive the potential risk of traffic operation. In previous studies, accident data are often used to evaluate road traffic safety. Therefore, this paper classifies road traffic safety entropy based on accident data.

Remove abnormal data

Select the traffic safety entropy values of 28 road sections and accident data from 2020 to 2021 to classify the risk levels of road traffic safety. Draw a two-dimensional scatter plot of the number of safety entropy accidents to eliminate the interference of abnormal data on the classification results, as shown in *Figure 15*.

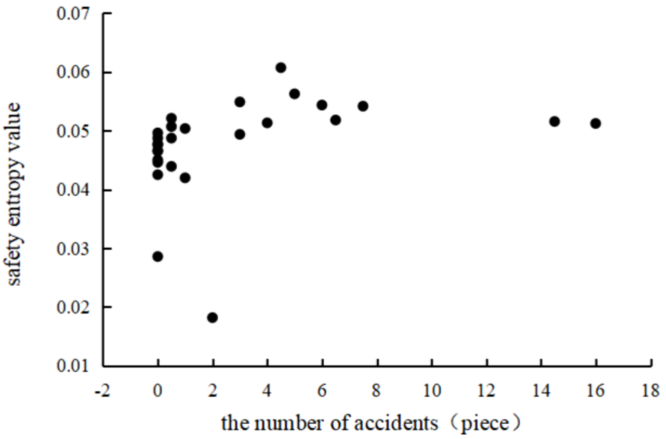


Figure 15 – Scatterplot of traffic accident-safety entropy

As shown in the above figure, the three sets of data on the right side of the scatter plot are far away from other data. These three data sets have a more significant number of accidents, and the difference in safety entropy values with different data is slight. Therefore, it can be considered that these three sets of data are isolated data. Therefore, it has been decided to remove these three sets of isolated points and use the data after removing the isolated points as input data for K-means clustering to classify road traffic safety.

Determination of optimal grading results

To determine the optimal safety entropy classification number, the K-means method was used to cluster the data after removing outliers, with clustering numbers of 2, 3 and 4 respectively. The optimal clustering effect was determined by calculating the contour coefficient of each class, thus determining the classification number of traffic safety entropy. The clustering results are shown in *Figures 16–18*. The contour coefficient calculation results corresponding to the number of clusters in each category are shown in *Table 7*.

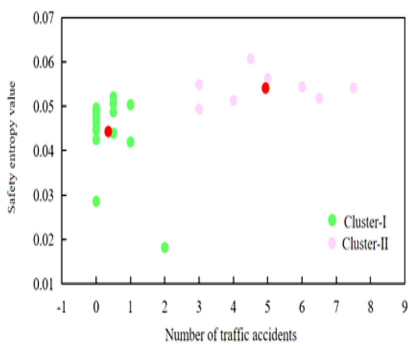


Figure 16 – Traffic accident count - safety entropy value K-means clustering into two categories

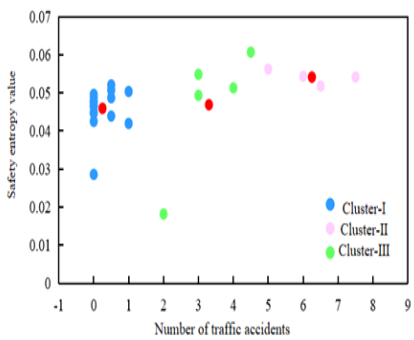


Figure 17 – Traffic accident count - safety entropy value K-means clustering into three categories

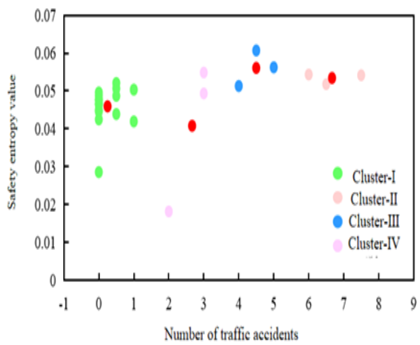


Figure 18 – Traffic accident count - safety entropy value K-means clustering into four categories

Table 7 – Profile coefficient table

Clustering numbers	Profile coefficient
2	0.757
3	0.717
4	0.732

According to the contour coefficient, it can be concluded that the effect of clustering 2 is the best. *Figure 16* shows that when clustering into two categories, there are more data samples between each category and a significant difference in cluster centres. In the results of clustering three and four categories, some categories have small sample sizes, and the clustering centres transmit unequal information to different categories in the dimensions of accident number and safety entropy. When the number of clusters is 4, the cluster centre of the number of accidents in Cluster-II is higher than that in Cluster-I, but the cluster centre of the safety entropy value is smaller than that in Cluster-I. So, the best classification results are selected into two categories: high risk and low risk.

Determination of the optimal grading threshold

Take the safety entropy data of each category with a clustering number of 2, sort the safety entropy values from small to large, and draw a scatter plot, as shown in *Figure 19*. The graph shows that the safety entropy values of different categories are not entirely separated, so it is necessary to define the optimal threshold for safety entropy values.

The safety entropy values of the two types of clustering centres with a clustering number of 2 are the upper and lower limits of the optimal classification threshold. In increments of 0.001, the K-means method classifies the sample data as Cluster-I under different classification thresholds. The classification threshold is defined as the number of false positive samples n1 for Cluster-II, and the clustering result of the K-means method is Cluster-II. The classification threshold is the number of false positive samples n2 for Cluster-I. Then, the classification accuracy under different thresholds is calculated according to *Equation 11*, and the calculation results are plotted as a line graph shown in *Figure 20*.

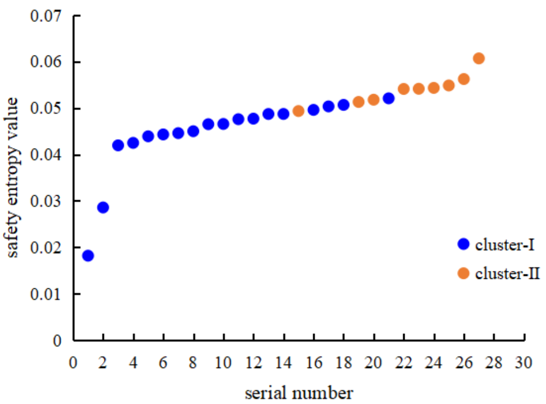


Figure 19 – Sorted scatterplot of security entropy

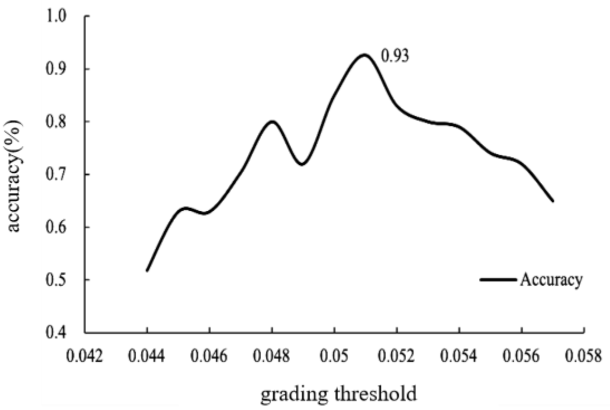


Figure 20 – Accuracy of different classification thresholds

The above figure shows that when the classification threshold for selecting the safety entropy value is 0.0507, the accuracy of defining high and low risks reaches 92%. Therefore, the optimal classification threshold is determined to be 0.0507.

5.5 Result analysis

According to the calibration results of the optimal classification threshold, the risk levels of 28 road segment units were divided, and the division results are shown in *Table 8*. According to the division results, 11 out of 28 road segment units are high-risk, while 17 are low-risk.

Table 8 – Results of risk classification of each road section unit

Road segment units	Safe entropy	Risk levels	Road segment units	Safe entropy	Risk levels
1	0.028568	low	15	0.049589	low
2	0.070766	high	16	0.018179	low
3	0.050339	low	17	0.04499	low
4	0.051534	high	18	0.051299	high
5	0.048705	low	19	0.042482	low
6	0.051198	high	20	0.049342	low
7	0.046557	low	21	0.054853	high
8	0.05178	high	22	0.047582	low
9	0.050669	low	23	0.060679	high
10	0.056232	high	24	0.043897	low
11	0.054137	high	25	0.048713	low
12	0.052077	high	26	0.046506	low
13	0.047701	low	27	0.041949	low
14	0.054327	high	28	0.044581	low

Statistics were collected on the number of accidents in high and low-risk road sections, as shown in *Figure 21*.

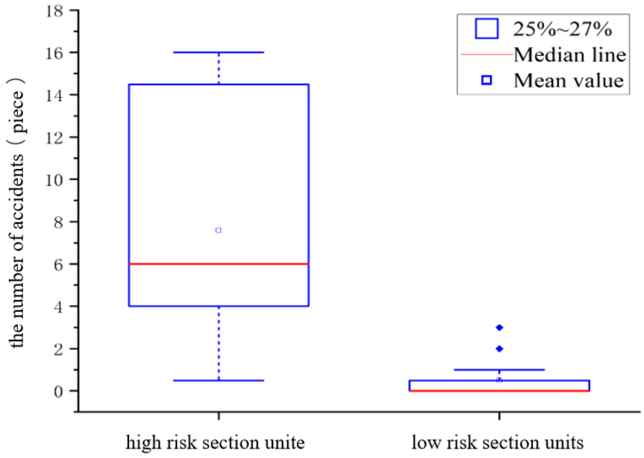


Figure 21 – Number of road section unit accidents with different risk levels

From the above figure, it can be seen that there is a significant difference between the mean and median number of accidents in the high and low-risk sections, with the number of accidents in the low-risk section being much lower than that in the high-risk section. The maximum number of accidents occurring on low-risk road sections is 3, and the minimum number of accidents occurring on high-risk road sections is also 3. The classification threshold for road traffic safety risks has a good effect on the dimension of accidents, and the determination of the dimension of accidents and determining risk levels is generally feasible. However, due to the incomplete equivalence between traffic safety risks and accidents and the model's limitations, there are also situations where the number of accidents on certain high-risk sections is relatively low, and the number of accidents on low-risk sections is relatively high.

6. CONCLUSIONS

This article uses data from the radar video integrated sensors for identifying unsafe driving behaviour, analyses the correlation between unsafe driving behaviour and traffic accidents, and establishes a road traffic safety risk estimation method based on an improved entropy method.

1) Based on high-precision and low latency data from the radar video integrated sensors, the characterisation parameters and thresholds of seven types of unsafe driving behaviours (speeding, abnormal low speed, unstable vehicle speed, rapid acceleration, rapid deceleration, abnormal car following and sudden lane changes) were measured and calibrated. A specific recognition algorithm process was constructed to achieve accurate recognition of unsafe driving behaviours, providing an important data foundation for road traffic safety risk assessment.
2) Taking the left lane of the Jiaozhou Bay Tunnel in Qingdao as an example, the model was validated using 13 days of measured data. Based on actual accident data and K-means clustering results of safety entropy values, traffic operation risks were classified into high risk and low risk, and the optimal classification threshold of safety entropy values was studied with the goal of model recognition accuracy. The results show that when the safety entropy classification threshold is 0.0507, the risk level judgement of the research section is most accurate, with an accuracy rate of 92%. This has practical reference significance for safety warning and control in the Jiaozhou Bay Tunnel in Qingdao.
3) Research has shown that the road traffic safety risk estimation method based on driving behaviour data can effectively estimate road traffic safety risks. The unsafe driving behaviour recognition method constructed in this paper can effectively identify road risk points and achieve road risk estimation through the constructed road traffic safety assessment model, thereby realising the prevention and early warning of road traffic risks.

The driving behaviour data records the interaction results between the driver and other factors in the "human vehicle road law environment" system. Mining driving behaviour data can help evaluate driving risks and improve traffic safety and efficiency. There are still some shortcomings and areas that need further research in this study, mainly reflected in the following aspects:

1) The sudden lane change behaviour is typical unsafe driving behaviour. Although this article has identified this behaviour, due to the complexity of the methods for determining the start and end times of lane change behaviour, manual verification can only be used on small sample data. With the development of future vehicle road collaboration and autonomous driving, it is imperative to develop recognition algorithms for massive data. Therefore, further research is needed on the recognition algorithms for this behaviour.
2) This article mainly focuses on the study of longitudinal unsafe driving behaviour. However, on urban roads, lateral unsafe driving behaviours such as sticking and riding on lines also pose significant safety hazards. With the continuous development of data collection technology, further exploration is needed on how to accurately characterise these behaviours based on trajectory data.
3) The ultimate goal of using driving behaviour data to assess road traffic safety risks is to scientifically improve and optimise the road driving environment. How to effectively combine security risks with operational control measures is a topic that needs further research.

FUNDING

This work was supported by Science and Technology Innovation Key R&D Program of Chongqing (CSTB2023TIAD-STX0030), Science and Technology Innovation Key R&D Program of Chongqing (No. CSTB2022TIAD-STX0003), Technology Innovation and Application Development Special Key R&D Program of Chongqing (grants no. CSTB2022TIAD-KPX0104), Chongqing Transportation Science and Technology Program (grants no. CQJT-2024CZ6-1).

REFERENCES

[1] Tang ZZ, Zhang TJ, He Y. *Road Traffic Safety Evaluation.* Beijing, China: People's Communications Press; 2008.

[2] Nguyen TH, Lu DN, Nguyen DN, Nguyen HN. Dynamic basic activity sequence matching method in abnormal driving pattern detection using smartphone sensors. *Electronics.* 2020;9(2):217. DOI: 10.3390/electronics9020217.

[3] Mcgehee DV, et al. Extending parental mentoring using an event-triggered video intervention in rural teen drivers. *Journal of Safety Research.* 2007;38(2):215–227. DOI: 10.1016/j.jsr.2007.02.009.

[4] Li F, Zhang H, Che H, Qiu X. Dangerous driving behavior detection using smartphone sensors. *IEEE 19th international conference on intelligent transportation systems (ITSC). 1-4 Nov. 2016, Rio de Janeiro, Brazil.* 2016. p. 1902-1907. DOI:10.1109/ITSC.2016.7795864.

[5] Lu J, Wang K, Jiang YM. Real-time identification method of abnormal road driving behavior based on vehicle driving trajectory. *Journal of Traffic and Transportation Engineering*. 2020;20(6):227-235. DOI: 10.19818/j.cnki.1671-1637.2020.06.020.

[6] Ma Y, et al. On-line aggressive driving identification based on in-vehicle kinematic parameters under naturalistic driving conditions. *Transportation research part C: emerging technologies*. 2020;114:554-571. DOI: 10.1016/j.trc.2020.02.028.

[7] Li C, Liu Y. Abnormal driving behavior detection based on covariance Manifold and Logitboost. *Laser & Optoelectronics Progress.* 2018;55(11):338-345. DOI: 10.3788/LOP55.111503.

[8] Shahverdy M, Fathy M, Berangi R, Sabokrou M. Driver behavior detection and classification using deep convolutional neural networks. *Expert Systems with Applications.* 2020;149:113240. DOI: 10.1016/j.eswa.2020.113240.

[9] Hu J, et al. Abnormal driving detection based on normalized driving behavior. *IEEE Transactions on Vehicular Technology*. 2017;66(8):6645-6652. DOI: 10.1109/TVT.2017.2660497.

[10] Hu J, Zhang X, Maybank S. Abnormal driving detection with normalized driving behavior data: A deep learning approach. *IEEE transactions on vehicular technology*. 2020;69(7):6943-6951. DOI:10.1109/TVT.2020.2993247.

[11] Huang X, Sun J, Sun J. A car-following model considering asymmetric driving behavior based on long short-term memory neural networks. *Transportation Research Part C: Emerging Technologies*. 2018;95:346–362. DOI: 10.1016/j.trc.2018.07.022.

[12] Liu J, et al. One-dimensional convolutional neural network model for abnormal driving behaviors detection using smartphone sensors. *International Conference on Networking Systems of AI (INSAI). 19-20 Nov. 2021, Shanghai, China.* 2021.p. 143-150. DOI:10.1109/INSAI54028.2021.00035.

[13] Chen S, et al. Vehicles driving behavior recognition based on transfer learning. *Expert Systems with Applications*. 2023;213:119254. DOI:10.1016/j.eswa.2022.119254.

[14] Xiao W, et al. FDAN: Fuzzy deep attention networks for driver behavior recognition. *Journal of Systems Architecture.* 2024;147:103063. DOI:10.1016/j.sysarc.2023.103063.

[15] Darsono AM, et al. Utilizing LSTM networks for the prediction of driver behavior. *Przeglad Elektrotechniczny*. 2024;100(04):182-185. DOI: 10.15199/48.2024.04.34.

[16] Eren H, Makinist S, Akin E, Yilmaz A. Estimating driving behavior by a smartphone. *IEEE 2012 IEEE Intelligent Vehicles Symposium (IV),3-7 Jun. 2012, Madrid, Spain. 2012.* p. 234–239. DOI: 10.1109/IVS.2012.6232298.

[17] Chen F, Wang J, Deng Y. Road safety risk evaluation by means of improved entropy TOPSIS–RSR. *Safety science.* 2015;79:39-54. DOI: 10.1016/j.ssci.2015.05.006.

[18] Yan Y, et al. Driving risk assessment using driving behavior data under continuous tunnel environment. *Traffic injury prevention.* 2019;20(8):807-812. DOI:10.1080/15389588.2019.1675154.

[19] Chen J, Wu ZC, Zhang J. Driving safety risk prediction using cost-sensitive with nonnegativity-constrained autoencoders based on imbalanced naturalistic driving data. *IEEE transactions on intelligent transportation systems*. 2019;20(12):4450-4465. DOI:10.1109/TITS.2018.2886280.

[20] Cai X, et al. Road traffic safety risk estimation based on driving behavior and information entropy. *China J. Highw. Transp.* 2020;33(06):190-201. DOI:10.1155/2020/3024101.

[21] Wang T, et al. Traffic risk assessment based on warning data. *Journal of Advanced Transportation*. 2022;2022(9):1-11. DOI:10.1155/2022/1191239.

[22] Yang H, Zhao X, Luan S, Chai S. A traffic dynamic operation risk assessment method using driving behaviors and traffic flow data: An empirical analysis. *Expert Systems with Applications*. 2024;249:123619. DOI:10.1016/j.eswa.2024.123619.

蔡晓禹，李子木，乔午锋，程茜伶，彭博，张东

基于雷视一体机数据的道路交通安全风险评估研究

摘要

为了准确预防和预警交通事故，本文提出了一种基于车辆驾驶行为数据和信息熵理论的城市道路交通安全风险预测方法。本方法使用雷视一体机的数据来校准识别不安全驾驶行为的阈值，介绍了识别原理和算法，并分析了安全驾驶行为的时空分布模式。通过引入熵理论，建立了以交通安全熵为主要指标、以不安全驾驶行为率为次要指标的评价体系。采用聚类算法确定交通安全熵的分类数和阈值，构建隧道交通安全风险评估模型，并用道路事故数据进行验证。最后，该模型利用青岛胶州湾隧道左车道上的13天驾驶行为数据进行案例分析，根据事故和安全熵数据的K-means聚类结果，将交通运行风险分为高、低两类。研究发现，当安全熵分类阈值为0.0507时，分类准确率最高，为92%。这些结果为识别道路交通安全风险点和预防事故提供了技术支持。

关键词

交通工程；风险评估；信息熵；不安全驾驶行为；熵权法