
**International Journal of Vehicle Design**

ISSN online: 1741-5314 - ISSN print: 0143-3369
https://www.inderscience.com/ijvd

# An obstacle avoidance path selection for autonomous vehicles based on multi-dimensional data mining

Aiju Wang, Yao Yao, Zhanlei Shang

**DOI:** 10.1504/IJVD.2025.10071859

**Article History:**

| | |
|---|---|
| Received: | 03 December 2024 |
| Last revised: | 30 April 2025 |
| Accepted: | 02 May 2025 |
| Published online: | 10 July 2025 |

# An obstacle avoidance path selection for autonomous vehicles based on multi-dimensional data mining

Aiju Wang and Yao Yao*

College of Information Engineering,
Zhengzhou University of Technology,
Zhengzhou, 450044, China
Email: juzi81@163.com
Email: zzutyy@163.com
*Corresponding author

Zhanlei Shang

Engineering Training Center,
Zhengzhou University of Light Industry,
Zhengzhou, 450001, China
Email: shangzl@zzuli.edu.cn

**Abstract:** In order to overcome the problems of poor obstacle avoidance path selection, low success rate, and long time in traditional methods, a new obstacle avoidance path selection method for autonomous vehicles based on multi-dimensional data mining is proposed. The method employs the K-means algorithm to process multi-sensor data (including visual cameras, LightLaser Detection and Ranging (LiDAR), global positioning system (GPS), and traffic flow) for environmental data collection in autonomous vehicles. Based on the collected data and constraints, a target function for obstacle avoidance path selection of unmanned vehicles is constructed. The optimisation function is solved using the whale optimisation algorithm (WOA), and the optimal solution obtained is the obstacle avoidance path selection scheme for unmanned vehicles. Experimental results show that the proposed method for autonomous vehicle lane changing has a relatively large angle and short path, without collision problems. The maximum success rate of obstacle avoidance path selection is 98.56%, and the minimum time is 0.44 s.

**Keywords:** multi-dimensional data mining; autonomous vehicles; obstacle avoidance path selection; K-means algorithm; objective function; killer whale hunting algorithm.

**Reference** to this paper should be made as follows: Wang, A., Yao, Y. and Shang, Z. (2025) 'An obstacle avoidance path selection for autonomous vehicles based on multi-dimensional data mining', *Int. J. Vehicle Design*, Vol. 97, No. 5, pp.1–21.

**Biographical notes:** Aiju Wang received her Master's degree from Harbin Engineering University, China, and published extensively in top-tier conferences and journals. She has contributed significantly to the field by introducing novel models and methodologies for complex data analysis. Now, she works in Zhengzhou University of Technology. Her research interests include computer software, information security and big data analysis.

Yao Yao received her Master's degree from Zhengzhou University, China, and published extensively in top-tier conferences and journals. She has contributed significantly to the field by introducing new models and methods for Web mining. Now, she is an Associate Professor at Zhengzhou University of Technology, specialising in artificial intelligence, high performance computing and big data analysis.

Zhanlei Shang, Master's degree, Professor, Engineering Training Center in Zhengzhou University of Light Industry. He instructed such courses as Fundamentals of Computer, Programming, Data Structure and Principle and Application of Database. Meanwhile, he published academic papers more than 20, 5 of them were retrieved by EI. About 20 computer textbooks were issued under my general editorship. His research interests include Database and network.

## 1 Introduction

With the rapid development of artificial intelligence and automation technology, autonomous vehicles have gradually become a hot topic in the automotive industry and intelligent transportation system research (Qin et al., 2023). One of the core technologies of autonomous vehicles is autonomous navigation, in which obstacle avoidance path selection is a crucial part of the autonomous driving process of autonomous vehicles. It requires the vehicle to detect and identify obstacles ahead in real time, including static obstacles (such as road construction, parked vehicles, etc.) and dynamic obstacles (such as pedestrians, other vehicles, etc.), and based on this information, quickly plan a safe and feasible driving path (Mamak and Glanc, 2022; Zhou et al., 2022). The selection of obstacle avoidance paths is not only related to the safety and efficiency of autonomous vehicles, but also the key to their stable operation in complex and changing environments. The research on obstacle avoidance path selection can significantly improve the safety of autonomous vehicles. By perceiving and predicting obstacles in real-time, vehicles can make obstacle avoidance decisions in advance, avoiding collisions with obstacles and ensuring the safety of passengers and other traffic participants (Wang et al., 2024). Not only that, but it can also reduce the waiting time and delays of unmanned vehicles during driving, and improve traffic efficiency. Especially during peak hours and on congested roads, autonomous vehicles can utilise more intelligent obstacle avoidance strategies to find more efficient driving paths, alleviate traffic pressure, promote the overall development of intelligent transportation systems, and achieve more intelligent, efficient, and sustainable modes of transportation. Therefore, researching obstacle avoidance path selection methods for autonomous vehicles is of great significance. Through intelligent path selection and obstacle avoidance planning, autonomous vehicles can more effectively utilise road resources, reduce traffic congestion, and improve overall travel efficiency. Obstacle avoidance path selection is one of the core components of autonomous driving technology, and related research can promote the continuous progress and improvement of autonomous driving technology, laying the foundation for future intelligent transportation systems.

Autonomous vehicles require real-time and accurate perception of surrounding environmental information, including road structure, obstacle location, traffic flow, etc.

This requires advanced sensor technology and efficient data processing algorithms to support. In order to overcome the limitations of traditional methods, a series of new obstacle avoidance path selection algorithms have emerged in recent years. Huo and Wang (2024) proposed an obstacle avoidance path selection method for autonomous vehicles based on the cat swarm algorithm. Build a system error model that can be used to simultaneously search for trajectories and target positions, and design an obstacle avoidance path selection strategy based on this model. This strategy ensures the automatic update rate of online estimation while minimising the length of the path and avoiding collisions. By using a specific allocation strategy, the goal of minimising the maximum distance travelled is achieved, thereby ensuring the quality of obstacle avoidance path selection. However, in actual testing, it was found that due to the limited and incomplete data collected, this method has the problem of poor obstacle avoidance effect and a large gap between the expected goals. Zhang and Zhang (2022) proposed an obstacle avoidance path selection method for autonomous vehicles based on an improved ant colony algorithm. Use the grid method to establish a map model of the target vehicle's operating environment. In this map model, the target vehicle is equipped with a laser rangefinder, which uses the trilateration method to locate the current position of the vehicle body. The A* algorithm is used to guide the search process of the basic ant colony algorithm, thereby improving it. After determining the position of the target vehicle, this position is set as the starting point. Then, the improved ant colony algorithm is used to search for the optimal obstacle avoidance path to the endpoint in the previously constructed environmental map, completing the obstacle avoidance path selection. However, after testing, it was found that due to the relatively single data collected, it is difficult to determine the true traffic status, resulting in a low success rate of obstacle avoidance path selection for this method, and the actual application effect is not good. Li (2024) proposed an obstacle avoidance path selection method for autonomous vehicles based on improved convolutional networks. Construct an accurate kinematic model of a vehicle and use it to describe the kinematic characteristics of autonomous vehicles during operation. The system detects the target position of obstacles on the road, continuously obtains various information of dynamic obstacle targets on the road, and makes predictions on the future position and speed of these obstacles. We use an improved convolutional network to calculate the loss function of the dynamic obstacle avoidance path for autonomous vehicles, which measures the gap between the planned obstacle avoidance path and the actual target path. Moreover, the artificial potential field method is used to assist in finding the minimum potential energy and obtaining the final obstacle avoidance path. Due to the relatively single and poor overall data collected, it is difficult to provide important data support for subsequent analysis. And the implementation process of this method is relatively complex, resulting in an increase in obstacle avoidance path selection time.

A new obstacle avoidance path selection method for autonomous vehicles based on multi-dimensional data mining is proposed with the expected goal of solving the problems of poor obstacle avoidance path selection performance, low success rate, and long time in the above methods. This method introduces multidimensional data mining techniques, which can extract valuable information and patterns from massive amounts of visual cameras, LightLaser Detection and Ranging (LiDAR), global positioning system (GPS), and traffic data, providing data support for obstacle avoidance path selection. The technical route of this study is as follows:

1 The method employs the K-means algorithm to process multi-sensor data (including visual cameras, LiDAR, GPS, and traffic flow) for environmental data collection in autonomous vehicles.

2 Based on the collected data and constraints, a target function for obstacle avoidance path selection of unmanned vehicles is constructed. The optimisation function is solved using the whale optimisation algorithm (WOA), and the optimal solution obtained is the obstacle avoidance path selection scheme for unmanned vehicles.

3 The effectiveness, success rate, and duration of obstacle avoidance path selection for autonomous vehicles were selected as evaluation indicators to validate the practical application of this method.

## 2 Design of obstacle avoidance path selection method for autonomous vehicles

### *2.1 Data collection of autonomous vehicle environment based on multi-dimensional data mining*

1 *Visual camera*

Industrial cameras often have characteristics such as high speed, high resolution, and high sensitivity, making them suitable for industrial automation, machine vision, autonomous vehicles in industrial applications, and other industrial inspection applications. Compared with ordinary consumer-grade cameras, industrial cameras have better durability, reliability, and stability, a wider range of operating temperatures, and performance advantages such as resistance to dust, water, and earthquakes. The technical parameters of industrial cameras are detailed in Table 1.

**Table 1** Technical parameters of visual camera

| *Technical parameter* | *Parameter value* |
|---|---|
| Part number | BFS-US–16S2C-CS |
| Resolving power | 1440 × 1080 |
| Display frame rate | 226 |
| Pixel size | 3.45 μm |
| Interface | USB 3.1 Gen 1 |
| Power requirements | 8~24 V |

2 *Laser LiDAR*

LiDAR can accurately calculate the distance between a target object and a vehicle by emitting a laser beam and measuring the time it takes for the laser to reflect back. The accuracy of this distance measurement is very high, reaching a precision of centimetres or even millimetres. For example, when measuring the distance to the vehicles in front, LiDAR can accurately determine the distance between two vehicles, providing accurate distance information for unmanned vehicles' following, overtaking, and other operations (Matsushita et al., 2022; Wang et al., 2024). Compared to other sensors such as cameras, Camera-based sensors mainly use image recognition to estimate the distance of objects. However, this method is easily affected by factors such as lighting, object shape, and texture, while the distance measurement of LiDAR is relatively independent of these factors and more stable and accurate. Select the multi harness Velodyne-VLP32 LiDAR for data acquisition, and its specific technical parameters are shown in Table 2.

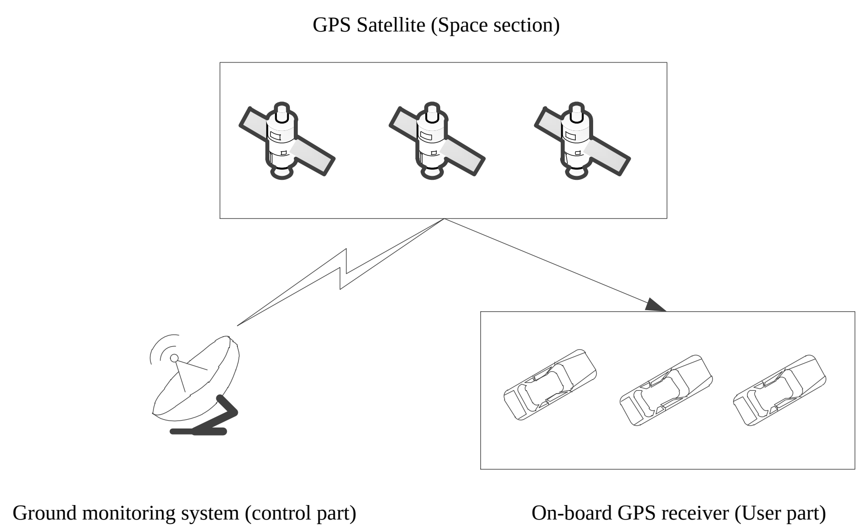
**Table 2** Technical parameters of LiDAR

| *Technical parameter* | *Parameter value* |
|---|---|
| Sensor | 32 channels |
| Measure distance | 200 m |
| Field of view angle | Vertical at 40°, water at 306° |
| Rotation rate | 5 ~ 20 Hz |
| Wavelength | 903 nm |
| Working voltage | 10.5 ~ 18 V |
| Single echo | About 600,000 points per second |
| Double echo | About 1.2 million points per second |

## 3 *GPS*

GPS can provide precise latitude and longitude coordinates for autonomous vehicles, thereby which allows for the determination of their absolute position on the Earth's surface. This is the foundation for navigation and path planning of autonomous vehicles. Compared with traditional map matching based positioning methods, GPS positioning is more direct and accurate. In areas without detailed maps or in which map updates are not timely, GPS can still determine the location of vehicles and provide basic location references for them (Jia, 2022). The working principle of the onboard GPS for autonomous vehicles is shown in Figure 1.

**Figure 1** Working principle of vehicle GPS



The car-mounted GPS device uploads the recorded vehicle driving data to the backend data management center through the communication network at regular intervals, and finally generates the GPS trajectory data of the vehicle. Each vehicle's GPS data records the vehicle's licence plate number, latitude and longitude coordinates, recording time, elevation, and other driving information in detail, which can provide important data support for obstacle avoidance path selection for unmanned vehicles (Qian et al., 2023).

## 4 *Traffic flow data*

Assuming $D$ represents the set of traffic flow data samples, $|D| = L$ represents the number of road samples, $D_i$ represents the flow data samples $D_i = (A, V_i)$ monitored by one of the traffic roads, $A = \{a_1, a_2, \dots, a_m\}$ represents the set of $m$ historical traffic flow time series datasets, and the length of each time series dataset needs to be pre-set. $V_i = N$ represents the main feature set, and $h$ represents a certain road intersection. $D_i$ records a total of 10 time units from the start of recording zero time to the end of recording. For each time series $a_i$ of road $D_i$, there are $h$ valid time points. Each valid time point can be recorded as $t_{(ai,s)}$, satisfying $t_{(ai,s)} < t_{(ai,s+1)}$ (Ma et al., 2020).

If the unmanned vehicle environment data variable $V_{jk}$ represents the flow value of the *j*th time series attribute at the *k*th time point, its function form can be expressed by the following formula:

$$V_{jk} = f\left(t_{(a_k,1)}\right) \tag{1}$$

Based on the above analysis, the *k*th time series of $D_i$ can be represented by the following formula:

$$TS_{ik} = \left[ f\left(t_{(a_k,1)}\right) f\left(t_{(a_k,2)}\right) \cdots f\left(t_{(a_k,h)}\right) \right] \tag{2}$$

The time series corresponding to sample $D_i$ of unmanned vehicle environment data can be represented in the form of a traffic local time series data matrix, and the specific calculation formula is as follows:

$$\mathrm{TS}_i = \begin{bmatrix} TS_{i1} \\ TS_{i21} \\ \vdots \\ TS_{in} \end{bmatrix} = \begin{bmatrix} f_i\left(t_{(a_k,1)}\right) f_i\left(t_{(a_k,2)}\right) \dots f_i\left(t_{(a_k,h)}\right) \\ f_i\left(t_{(a_2,1)}\right) f_i\left(t_{(a_2,2)}\right) \cdots f_i\left(t_{(a_2,h)}\right) \\ \vdots \quad \vdots \quad \vdots \quad \vdots \\ f_i\left(t_{(a_n,1)}\right) f_i\left(t_{(a_n,2)}\right) \cdots f_i\left(t_{(a_n,h)}\right) \end{bmatrix} \tag{3}$$

## 5 *Multi-dimensional data mining*

Assumption $T$ represents the preliminary clustering set formed by visual camera data, LiDAR data, GPS data, and traffic flow data, and $k$ represents the number of clusters. Run the K-means algorithm multiple times, by using different random initial centroids each time, and then select the clustering result with the minimum sum of squared errors (SSE) as the final solution. This method involves multiple attempts to find a better initial centroid. The DB value evaluates clustering performance by calculating the ratio between the separation and compactness of each cluster. It calculates the ratio. The numerator of the ratio is the average distance from all points within each cluster to its centroid point. The denominator is the average distance between centroids of different clusters. The smaller the DB value, the better the clustering effect. Therefore, the optimal number of clusters can be determined by minimising the DB value.

The multi-dimensional data mining process for unmanned vehicle environment based on K-means algorithm is as follows:

*Step 1*: Calculate the centroid set $M = [m_1, m_2 ... m_k]$ of the cluster based on $T$.

*Step 2*: Randomly select $k$ feature vectors $x_1, x_2 ... x_k \in M$ from $M$ as the initial clustering centers.

*Step 3*: Use standardised Euclidean distance $d_{(i,j)=}\sqrt{\sum_{t=1}^{n}\left(\frac{m_{it}-x_{jt}}{s_t}\right)^2}$ to calculate the distance from the $n$ dimensional centroid $m_i$ to each initial cluster center, and cluster the centroid into the nearest cluster to that point $c_i = arg\min_j \left| d_{(i,j)} \right|$.

*Step 4*: Calculate the average coordinates of all points in each cluster and use this average as the new cluster center. The calculation formula is as follows:

$$x_j = \frac{\sum_{i=1}^{m} 1\{c_i = j\} m_i}{\sum_{i=1}^{m} 1\{c_i = j\}} \tag{4}$$

*Step 5*: Repeat steps 3 and 4 until the clustering center no longer moves widely and satisfies the convergence criterion function $J(c, x) = \sum_{i=1}^{m} \left\| m_i - x_{c_i} \right\|^2$, and output the unmanned vehicle environment dataset $T''$.

The data preprocessing step is crucial when obtaining multidimensional data from various sensors such as data from visual cameras, LiDAR, GPS, and traffic flow sensors. The following is a detailed explanation of data synchronisation and calibration, outlier removal or denoising techniques, as well as feature extraction and selection processes:

1 *Data synchronisation and calibration*

For data from different sensors, such as data from visual cameras, LiDAR, and GPS, it is necessary to ensure that their timestamps are synchronised. This can be achieved by integrating a unified time reference in the data collection system, or using post-processing software to align the data based on timestamps. In practical applications, LiDAR data contains key information such as the time when the laser is emitted and received, which can be used for time synchronisation with other sensor data. At the same time, GPS data usually contains timestamp information, which facilitates time alignment with other data sources.

2 *Data calibration*

Space calibration between sensors is another important step. Since the installation positions and angles of different sensors differ, they need to be spatially calibrated to ensure data consistency in space. This typically involves using known physical markers or reference points to determine the relative position and angle between sensors through measurement and calculation. For visual cameras and LiDAR, spatial calibration can be achieved by capturing the same scene and matching feature points. For GPS data, it may be necessary to calibrate GPS data by combining map information and satellite signals.

3 *Outlier removal*

Outliers are mainly caused by sensor failures, data transmission errors, or environmental interference. In the data preprocessing stage, distance based methods are used to identify and remove these outliers. This method can select one or more suitable combinations based on the distribution characteristics of the data and actual application scenarios.

4 *Noise reduction technology*

Noise reduction technology is used to reduce noise and interference in data, which improves the accuracy and reliability of the data. For visual camera data, image filtering algorithms can be used to reduce image noise. For LiDAR data, Kalman filtering can be used to reduce noise interference. For GPS data, differential GPS technology or fusion of other sensor data can be used to improve positioning accuracy.

5 *Feature extraction and selection*

Feature extraction is the process of extracting key information from raw data that is useful for subsequent analysis. For visual camera data, image features such as edges, corners, and textures can be extracted. Similarly, for LiDAR data, laser features such as distance, angle, and reflection intensity can be extracted. Similarly, for GPS data, location features such as longitude, latitude, and velocity can be extracted. Similarly, for traffic flow data, traffic characteristics such as traffic flow, vehicle speed, and lane occupancy can be extracted velocity can be extracted; For traffic flow data, traffic characteristics such as traffic flow, vehicle speed, and lane occupancy can be extracted.

Feature selection is the process of selecting the most useful subset of features for subsequent tasks from the extracted features. When selecting features, factors such as correlation, redundancy, interpretability, and computational efficiency need to be considered. Through feature selection, the model structure can be simplified, model performance can be improved, and the risk of overfitting can be reduced.

In summary, this method utilises the K-means algorithm to mine multidimensional data such as visual cameras, LiDAR, GPS, and traffic flow, enabling the collection of environmental data for autonomous vehicles. This provides an important data foundation for subsequent obstacle avoidance path selection.

## *2.2 Objective function for obstacle avoidance path selection of autonomous vehicles*

When we are identifying relevant features and using them to construct the objective function, the following points should be noted:

1 *Understanding business requirements*: Firstly, it is necessary to clarify the business requirements and analysis objectives in order to determine which features are relevant to the objectives.

2 *Feature engineering*: Construct a subset of features related to the target by extracting and selecting features, and we perform appropriate feature scaling, normalisation, and other processing to improve the stability and performance of the model.

3 *Model construction*: Select appropriate machine learning algorithms or deep learning models based on business requirements and data characteristics, and use the processed features as inputs to construct the objective function.

4 *Model evaluation and optimisation*: Use cross validation, A/B testing, and other methods to evaluate and optimise the model to ensure its accuracy and reliability. Also need to pay attention to the overfitting and underfitting issues of the model and make appropriate adjustments and improvements.

In the objective function of obstacle avoidance path selection for autonomous vehicles, acceleration cost is an indicator used to measure the cost or impact of acceleration changes when the vehicle travels along a potential path. Acceleration is a physical quantity that describes how quickly an object's velocity changes. For autonomous vehicles, the magnitude and frequency of acceleration can affect various aspects such as driving comfort, safety, and energy consumption. In the path planning of autonomous vehicles, the acceleration cost needs to be included in the objective function to ensure that the planned path is the one that the vehicle can actually travel safely and stably. The calculation formula for acceleration cost is as follows:

$$C_{acc} = \omega_{acc} u^T \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} u \tag{5}$$

In the formula, $\omega_{acc}$ represents the weight coefficient of acceleration, $u^T$ represents acceleration, and $u$ represents acceleration.

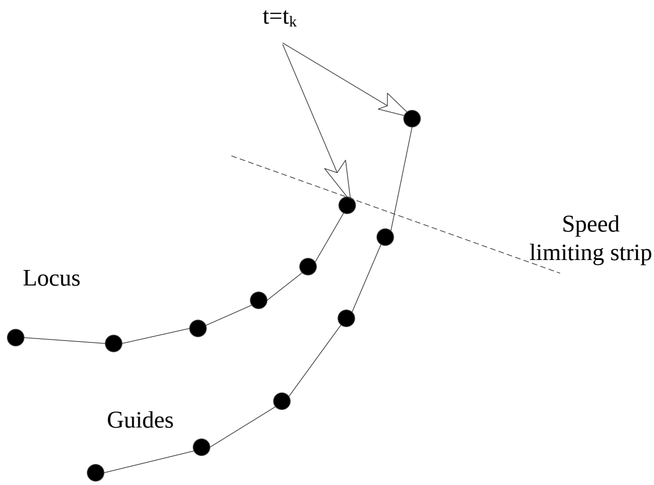
For autonomous vehicles, in the objective function of obstacle avoidance path selection, the yaw angle cost is an indicator measuring the cost or impact of yaw angle changes when the vehicle travels along a specific path. Yaw angle refers to the angle between the longitudinal axis of a vehicle and its direction of travel, which reflects how much the vehicle steers during travel. The calculation formula for the cost of yaw angle is as follows:

$$C_{yaw} = \omega_{yaw} z^T \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} z \tag{6}$$

In the formula, $\omega_{yaw}$ represents the weight coefficient of the yaw angle, $z^T$ represents the yaw angle, and $z$ represents the yaw acceleration (Hu et al., 2020).

In the path planning of autonomous vehicles, the reference line is a pre-set ideal driving route. It is usually constructed based on map information (such as high-precision maps), including the basic geometric shape of roads, lane information, road boundaries, etc. The reference line provides a rough direction and path framework for the vehicle. Further process the results obtained from the mixed A* search that account for speed limit constraints, as shown in Figure 2, by directly discarding the trajectory points after the end of the speed limit zone. At the same time, we set the reference information of the last point on the planned trajectory as the information of the speed limit zone to ensure that the planned trajectory strictly complies with traffic regulations.

**Figure 2** Schematic diagram of the reference line cost of time vs. its approach




The calculation formula for the reference line cost is as follows:

$$C_{ref} = \omega_{ref}\left(x - x_{ref}\right)^T Q_{ref}\left(x - x_{ref}\right) \tag{7}$$

The kinematic constraints of unmanned vehicles adopt a vehicle kinematic model based on the bicycle model, where the state quantity is $x = \left[p^x, p^y, v, \theta\right]^T$, and the vehicle kinematic constraints are as follows:

$$\dot{x} = \begin{bmatrix} v\cos\theta \\ v\sin\theta \\ \dot{v} \\ \dfrac{v}{L}\tan\delta \end{bmatrix} \tag{8}$$

In the formula, $\delta$ represents the front wheel steering angle, indicating that the kinematic model of autonomous vehicles is a nonlinear constraint. If the ILQR algorithm is used to linearise the state transition equation to handle this nonlinear constraint, the following relationship holds:

$$\begin{cases} f(x,u) \approx f(\hat{x},\hat{u}) + \nabla_{x,u} f(\hat{x},\hat{u})\begin{pmatrix} x-\hat{x} \\ u-\hat{u} \end{pmatrix} \\ \bar{f}(\delta x, \delta u) = F\begin{pmatrix} \delta x \\ \delta u \end{pmatrix} \end{cases} \tag{9}$$

In the formula, $x$ and $u$ respectively represent the initial state quantity and control quantity of the unmanned vehicle system, and $\hat{x}$ and $\hat{u}$ respectively represent the current state quantity and control quantity of the unmanned vehicle system (Li et al., 2023).

The obstacle avoidance constraints during the movement of autonomous vehicles are as follows:

$$\dot{h}(x) + \gamma h(x) \geq 0 \tag{10}$$

In the formula, $h(x)$ represents the geometric distance between the trajectory point and the origin (Ren and Liu, 2022), $\dot{h}(x)$ represents a continuously differentiable function, and $\gamma \in [0,1]$ represents a constant.

Based on the above analysis, the objective function for obstacle avoidance path selection of autonomous vehicles is as follows:

$$\min_{u_0,\ldots,u_{N-1}} J = \sum_{k=0}^{N-1} \left\{ \frac{\dot{x}}{2} C_{acc} + C_{yaw} + C_{ref} + \frac{Ch(x)}{2} + cu_k^T r_k \right\} \tag{11}$$

In the formula, $C$ represents the quadratic coefficient matrix, $c$ represents the quadratic coefficient matrix, $u_k^T$ represents the weight matrix of the unmanned vehicle state variables, and $r_k$ represents the weight matrix of the unmanned vehicle control variables.

In summary, by considering acceleration cost, reference line cost, vehicle kinematics constraints, obstacle avoidance constraints, etc., building an objective function for obstacle avoidance path selection of autonomous vehicles can provide an important foundation for subsequent analysis.

### *2.3 Solution of obstacle avoidance path selection objective function based on killer whale hunting algorithm*

The objective function of obstacle avoidance path selection for autonomous vehicles is subject to multiple constraints, and the innovation of the killer whale hunting algorithm lies in its ability to effectively handle these complex constraints. Compared with traditional algorithms, it can integrate these constraints into optimisation mechanisms that mimic the behaviour of killer whale swarms, which makes the process of solving the problem more in line with the actual operating environment of autonomous vehicles and improves the feasibility of path selection.

Assuming there are $N$ killer whales in the population, they can operate in one, two, or even multiple dimensions. Based on different numbers and dimensions of killer whales, a mathematical model is then constructed as follows:

$$X = [x_1, x_2, \ldots, x_N] = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,D} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N,1} & x_{N,2} & \cdots & x_{N,D} \end{bmatrix} \tag{12}$$

In the formula, $X$ represents the killer whale population corresponding to all obstacle avoidance path selection objective function candidate solutions, $x_N$ represents the position of the *N*th killer whale individual corresponding to the candidate solution, and $x_{N,D}$ represents the position of the *N*th killer whale in the *D*th dimension.

Abstracting the behaviour of killer whales communicating and driving away schools of fish through sonar as the chase phase of OPA. The chase phase is divided into two parts: driving away the fish swarm and surrounding the fish swarm, with parameters $p_1$ and v set as random constants between [0,1]. When the random value is greater than $p_1$, it will drive away the school of fish, otherwise it will surround the school of fish. When driving away a school of fish, killer whales need to drive it to the surface of the water. There are two situations based on the size of the school of fish. In the first situation, when the school of fish is small or the trapping space is relatively simple, killer whales can easily locate the school of fish. In the second scenario, when the school of fish is large or the enclosure space is complex, the killer whale population approaches the school of fish by having individual members take the lead and controls the position of the centre of the group to avoid deviation. Set a fixed parameter of $q$, use the first method when the random number is greater than $q$, otherwise choose the second method. In the stage of driving away fish schools, the speed and position update formula of killer whales is as follows:

$$\begin{cases} v^t_{chase,1,i} = a \times \left(d \times x^t_{best} - F \times \left(b \times M^t + c \times x^t_i\right)\right) \\ v^t_{chase,2,i} = e \times x^t_{best} - x^t_i \end{cases} \tag{13}$$

$$\begin{cases} x^t_{chase,1,i} = x^t_i + v^t_{chase,1,i} \ if\ rand > q \\ x^t_{chase,2,i} = x^t_i + v^t_{chase,2,i} \ if\ rand \le q \end{cases} \tag{14}$$

In the formula, *a*, *b*, and *d* represent random values between [0,1], $e$ represents random values between [0,2], $F = 2$, $q = 0.9$, $x^t_{best}$ represents the optimal position of the killer whale, and $x^t_i$ represents the current position of the killer whale.

When herding the school of fish, individual killer whales communicate with each other to determine their next position. The formula for adjusting the position of the killer whale at this time is as follows:

$$x^t_{chase,3,i,k} = x^t_{j1,k} + u \times \left(x^t_{j2,k} - x^t_{j3,k}\right) \tag{15}$$

In the formula, $j1$, $j2$ and $j3$ represent three randomly selected different killer whales.

After successfully surrounding the school of fish, the killer whales hunt and consume the fish in an orderly manner, and then returned to their original positions. The process of exchanging positions with the next killer whale is called the attack stage. The calculation formula for the position and speed of killer whales during the attack phase is as follows:

$$v^t_{attack,1,i} = \left(x^t_{first} + x^t_{second} + x^t_{third} + x^t_{four}\right) / 4 - x^t_{chase,i} \tag{16}$$

$$x^t_{attact,i} = x^t_{chase,i} + g_1 \times v^t_{attact,1,i} + g_2 \times v^t_{attack,2,i} \tag{17}$$

In the formula, $x^t_{first}, x^t_{second}, x^t_{third}, x^t_{four}$ represents the four best positioned killer whales, $g_1$ represents a random number between [0,2], and $g_2$ represents a random number between [–2.5, 2.5]. Considering the behaviour of exchanging positions when killer whales attack, some of the weaker killer whales give up their positions to those with better positions, and there is a certain probability that the positions of other weaker killer whales will serve as a reference for determining the lower limit of a variable in the calculation.

The convergence mechanism of the killer whale hunting algorithm in solving the objective function of obstacle avoidance path selection is a complex and sophisticated process, involving the combination of global and local search, dynamic parameter adjustment, elite retention strategy, and diversity preservation mechanism. Through the joint action of these strategies, the algorithm can gradually approach and eventually converge to the optimal solution. The steps for solving the objective function of obstacle avoidance path selection based on the killer whale hunting algorithm are as follows:

*Step 1*: Population initialisation. Determine population size $N$, dimension $D$, maximum number of iterations $Max\_iter$, selection probabilities $p_1$ and $p_2$, upper and lower limits of search space $lb$ and $ub$, and initialise the killer whale population accordingly.

*Step 2*: Calculate the fitness function. Evaluate the fitness function values of each killer whale and select the best individual and their location.

*Step 3*: During the chase phase, killer whales use selection factor $p_1$ to drive or surround the school of fish to complete the chase process. During this period, they use sonar to locate the school of fish and adjust their position. The killer whale updates its position according to the formulas (13)–(14).

*Step 4*: During the attack phase, killer whales hunt by attacking schools of fish, and their position is updated using formulas (16)–(17).

*Step 5*: Population update. After the attack phase, the killer whale population updated their positions, with some of them being replaced with $lb$.

*Step 6*: Determine whether the current iteration count has reached the maximum iteration count. If it has not, terminate the loop and repeat the process of step 2 above. Otherwise, output the solution result of the objective function for obstacle avoidance path selection of unmanned vehicles.

In summary, applying the killer whale algorithm to solve the objective function of obstacle avoidance path selection aims to obtain the optimal obstacle avoidance path selection scheme, laying an important foundation for the further development of obstacle avoidance theory for autonomous vehicles. The overall process of obstacle avoidance path selection for autonomous vehicles is shown in Figure 3.

## 3 Experimental design

### *3.1 Experimental scheme*

In order to verify the effectiveness of the obstacle avoidance path selection method for autonomous vehicles based on multi-dimensional data mining designed in this paper, relevant experiments were conducted. The specific experimental plan is as follows:

1 *Experimental subjects*

A specific type of autonomous vehicle was selected as the research object, and the technical parameters of the research object are shown in Table 3.

**Figure 3** Overall process of obstacle avoidance path selection for autonomous vehicles

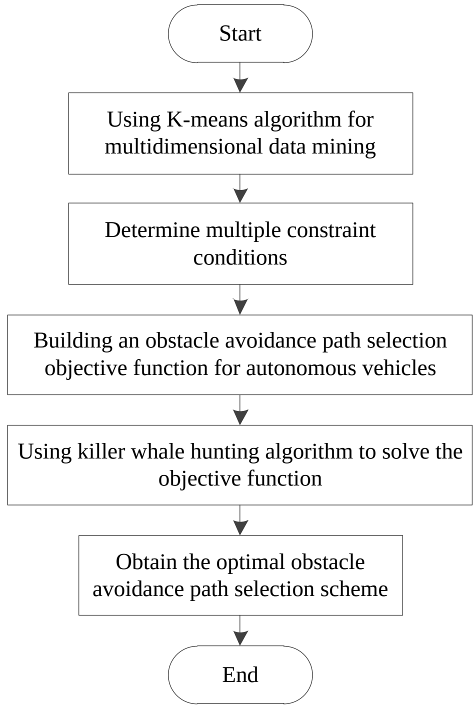



**Table 3** Technical parameters of experimental objects

| *Technical parameter* | *Parameter description* |
|---|---|
| Size | 4490 × 2014 × 1980 mm |
| Quality | 2015 kg |
| Peak power | 228 kw |
| Maximum climbing gradient | 30% slope |
| Charging performance | Capable of fast charging |
| Autonomous driving level | L4 |

The experimental scenario is shown in Figure 4.

During the experiment, various experimental scenarios were set up to evaluate whether vehicles could travel efficiently and safely, avoiding unnecessary parking and waiting, in a low traffic density environment. In high-density traffic environments, the ability of vehicles to flexibly avoid obstacles, maintain a safe distance, and effectively respond to various emergencies is tested. We considered different types of obstacles, including static obstacles (such as buildings, trees, etc.) and dynamic obstacles (such as pedestrians, other vehicles, etc.), analysed the obstacle avoidance strategies of vehicles under different types of obstacles, and evaluated whether vehicles can effectively identify and avoid obstacles while maintaining the optimality of the driving route. While ensuring driving stability and safety, we evaluate whether the vehicle can maintain a stable driving state and accurately identify and avoid obstacles ahead on straight roads, and test whether the vehicle can smoothly turn and avoid obstacles according to the road geometry on curved roads.

**Figure 4** Experimental scene (see online version for colours)

## 2 *Experimental data*

Using laser radar, cameras, millimetre wave laser radar, etc. to obtain relevant experimental data, and performing abnormal data removal and deduplication processing on the experimental data, providing an important data foundation for subsequent experimental verification. The types of data collected are as follows:

1 *Distance data*: Distance measurement values from sensors such as LiDAR and millimetre wave radar, such as the distance between a vehicle and an obstacle, measured in metres. These data are continuous values used to determine the position relationship of obstacles.

2 Speed data includes the vehicle's own driving speed (obtained by GPS, IMU, or vehicle speed sensors), measured in metres per second or kilometres per hour, as well as the speed of the target object relative to the vehicle (obtained by millimetre wave radar).

3 Angle data, including the angle of the target object relative to the vehicle measured by millimetre-wave radar, the angle information of each point in the LiDAR point cloud data, and the steering angle of the vehicle's steering wheel, is measured in degrees or radians and is crucial for determining the vehicle's travel direction and the relative orientation of obstacles.

4 *Image data*: 2D or 3D visual image data obtained by a camera. Image data is a matrix composed of pixels, each with information such as colour (such as RGB values) and brightness.

5 *Lidar point cloud*: Three-dimensional point cloud data obtained by scanning with a LiDAR. Point cloud data is a collection of a large number of discrete three-dimensional point coordinates, each point containing $x$, $y$, z coordinate information, reflecting the spatial position of objects in the surrounding environment of the vehicle.

## 3 *Experimental indicators*

The effectiveness of obstacle avoidance path selection, the success rate of obstacle avoidance, and the time required for obstacle avoidance path selection in unmanned vehicles were used as indicators to verify the practical application effects of the method in Huo and Wang (2024), the method in Zhang and Zhang (2022), and the proposed method. When we applied different methods to the obstacle avoidance path selection process of autonomous vehicles, it was found that there were no collisions with obstacles during the obstacle avoidance process and the paths were short, which indicates that these methods have good effects on obstacle avoidance path selection for autonomous vehicles. The success rate index of obstacle avoidance for autonomous vehicles is an important parameter for measuring their ability to successfully avoid obstacles during driving. It is the ratio of the number of successful obstacle avoidances to the total number of tests under specific testing conditions. The higher this value, the higher the accuracy of obstacle avoidance path selection. When exploring the effectiveness of obstacle avoidance path selection for human-driven vehicles, time index is a crucial factor. The time index typically refers to the total time that drivers need to perceive obstacles, make decisions, and execute obstacle avoidance operations when they encounter obstacles. A shorter time indicates a higher efficiency in obstacle avoidance path selection.

## *3.2 Experimental result*

The obstacle avoidance path selection effects of unmanned vehicles under the application of three methods are shown in Figures 5–7.

**Figure 5** Obstacle avoidance path selection results of Huo and Wang (2024) method: (a) single obstacle and (b) multiple obstacles (see online version for colours)

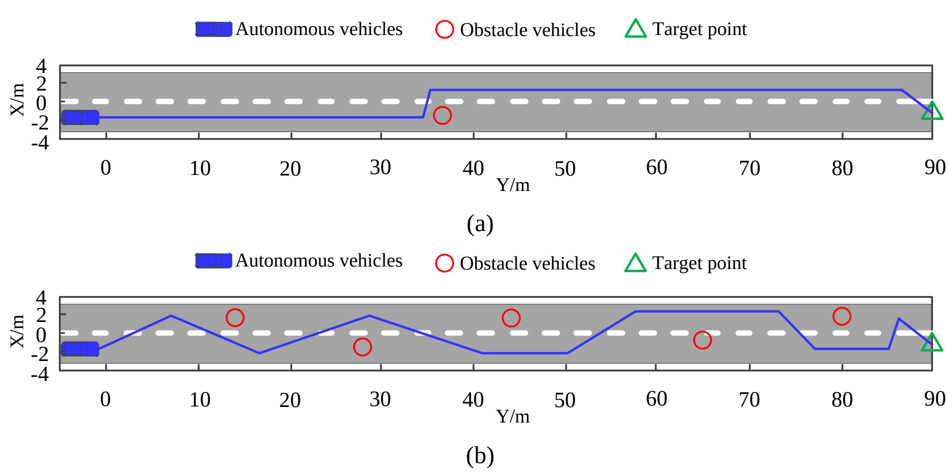
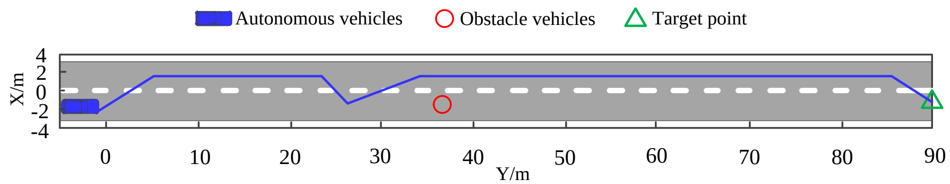
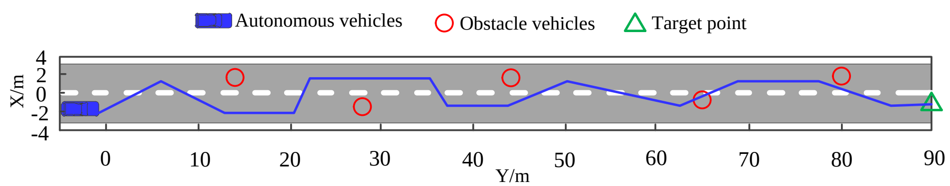


**Figure 6** Obstacle avoidance path selection results of Zhang and Zhang (2022) method: (a) single obstacle and (b) multiple obstacles (see online version for colours)

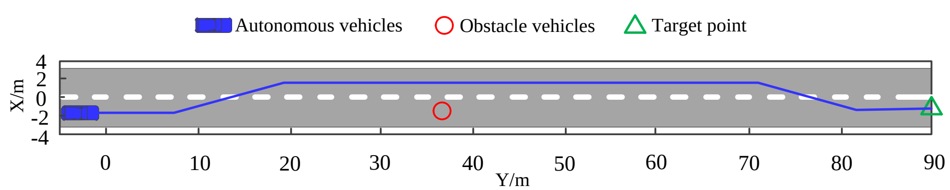



(a)

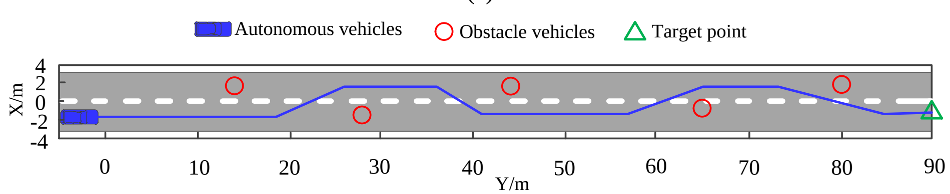



(b)

**Figure 7** The obstacle avoidance path selection results of the proposed method: (a) single obstacle and (b) multiple obstacles (see online version for colours)




(a)




(b)

According to the analysis of the results in Figure 5, the obstacle avoidance process of unmanned vehicles using the Huo and Wang (2024) method has a small lane changing angle and a long path, so the obstacle avoidance process obstacle avoidance path selection is not good. From the analysis of the results in Figure 6, the lane changing path during the obstacle avoidance process of unmanned vehicles using the Zhang and Zhang (2022) method is relatively long, and collision problems occur in multiple obstacle scenarios, resulting in unsatisfactory obstacle avoidance path selection. Compared with the Huo and Wang (2024) method and Zhang and Zhang (2022) method, the proposed method has a larger lane changing angle and shorter path for unmanned vehicles, and no collision problems occurred, indicating that this method has good obstacle avoidance path selection effect.

The success rates of obstacle avoidance path selection for autonomous vehicles under three different methods are shown in Table 4.

**Table 4** Success rate of obstacle avoidance path selection

| *Number of experiments* | *Huo and Wang (2024) method/%* | *Zhang and Zhang (2022) method/%* | *Proposed method/%* |
|---|---|---|---|
| 10 | 75.41 | 80.15 | 98.56 |
| 20 | 71.27 | 84.78 | 97.48 |
| 30 | 75.89 | 81.33 | 98.11 |
| 40 | 69.74 | 85.64 | 97.18 |
| 50 | 68.54 | 87.15 | 95.87 |
| 60 | 74.18 | 83.33 | 97.48 |
| 70 | 75.66 | 87.29 | 95.22 |
| 80 | 74.13 | 79.68 | 98.47 |
| 90 | 71.28 | 78.47 | 95.88 |
| 100 | 76.39 | 77.45 | 97.91 |

According to the results in Table 4, as the number of experiments increases, the success rates of obstacle avoidance path selection for unmanned vehicles using all three methods show a significant trend. Among these success rates, the maximum success rate of obstacle avoidance path selection for unmanned vehicles in the Huo and Wang (2024) method is 75.89%, and the Zhang and Zhang (2022) method is 87.29%. The proposed method is 98.56%, which is 22.67% and 11.27% higher than those in the Huo and Wang (2024) method and Zhang and Zhang (2022) method, respectively. The minimum success rate of obstacle avoidance path selection for autonomous vehicles using the Huo and Wang (2024) method is 68.54%, the Zhang and Zhang (2022) method is 77.45%. The proposed method is 95.22%, which is 26.68% and 19.77% higher than the Huo and Wang (2024) and Zhang and Zhang (2022) methods, respectively. After comparison, the proposed method has a higher success rate in obstacle avoidance path selection for autonomous vehicles and better practical application effects.

The obstacle avoidance path selection time of unmanned vehicles under the application of three methods is shown in Table 5.

According to the analysis of the data in Table 5, the maximum obstacle avoidance path selection time for the Huo and Wang (2024) method is 1.97 s, the Zhang and Zhang (2022) method is 2.54 s. The proposed method is 0.91 s, which is 1.06s and 1.63s lower than the Huo and Wang (2024) method and the Zhang and Zhang (2022) method, respectively. The minimum obstacle avoidance path selection time for unmanned vehicles in the Huo and Wang (2024) method is 1.06 s, the Zhang and Zhang (2022) method is 1.68 s. The proposed method is 0.44 s, which is 0.62 s and 1.24 s lower than those in the Huo and Wang (2024) method and the Zhang and Zhang (2022) method, respectively. After comparison, the proposed method has a shorter obstacle avoidance path selection time for autonomous vehicles, which can ensure that the autonomous vehicles can quickly complete the obstacle avoidance action.

**Table 5** Obstacle avoidance path selection time

| *Number of experiments* | *Huo and Wang (2024) method/s* | *Zhang and Zhang (2022) method/s* | *Proposed method/s* |
|---|---|---|---|
| 10 | 1.58 | 2.11 | 0.63 |
| 20 | 1.24 | 2.45 | 0.55 |
| 30 | 1.17 | 1.96 | 0.47 |
| 40 | 1.06 | 1.68 | 0.86 |
| 50 | 1.38 | 1.94 | 0.48 |
| 60 | 1.69 | 2.13 | 0.91 |
| 70 | 1.57 | 2.54 | 0.58 |
| 80 | 1.31 | 2.17 | 0.47 |
| 90 | 1.21 | 2.36 | 0.59 |
| 100 | 1.97 | 2.18 | 0.44 |

Based on the above analysis, the multidimensional-data-mining-based obstacle avoidance path selection method for autonomous vehicles has shown significant advantages. This method effectively mines multi-dimensional data from sources such as visual cameras, LiDAR, GPS, and traffic flow information through the K-means algorithm, enabling comprehensive environmental data collection for autonomous vehicles. On this basis, an objective function for obstacle avoidance path selection was constructed based on the collected data and constraints, and the optimal obstacle avoidance path selection scheme was obtained by solving the objective function using the killer whale algorithm, fully demonstrating the reliability and efficiency of this method in practical applications. This method pioneers the application of multidimensional data mining techniques in obstacle avoidance path selection for autonomous vehicles, achieving comprehensive collection and analysis of environmental data. At the same time, by constructing the objective function and using advanced optimisation algorithms for solving, this method can obtain more accurate obstacle avoidance path selection schemes. This innovation not only improves the safety of autonomous vehicles, but also provides strong technical support for the development of future intelligent transportation systems. The scientific significance of this method lies in its combination of multidimensional data mining and advanced optimisation algorithms, providing new ideas and methods for obstacle avoidance path selection of autonomous vehicles. By comprehensively collecting and analysing environmental data, this method can more accurately reflect the actual road conditions and make more reasonable obstacle avoidance decisions. In addition, the application of killer whale hunting algorithm makes the process of solving the problem more efficient and can obtain the optimal solution in a short period of time.

The scalability of the multidimensional-data-mining-based obstacle avoidance path selection method for autonomous vehicles needs to be comprehensively considered when facing different traffic complexities and environmental challenges in the real world. In high-density traffic environments, algorithms need to swiftly and accurately process dynamic information, adapt to diverse traffic rules and driver behaviours, and cope with environmental challenges such as severe weather and complex road geometries. To reduce computational overhead, to ensure real-time algorithm response, optimisation algorithms and parallel computing techniques should be adopted. In terms of hardware, high-quality sensor configurations and high-performance computing platforms are key to enhancing environmental perception and decision-making capabilities. In addition, algorithm optimisation strategies, including the incorporation of machine learning techniques, can bolster adaptability and robustness; Hardware upgrades, encompassing redundant design and fault detection mechanisms, aim to enhance system reliability and safety. Data fusion of multiple sensor information, as well as collaborative communication with other vehicles and transportation infrastructure, further optimises obstacle avoidance path selection. In this process, it is crucial to strengthen data security and privacy protection, comply with relevant laws and regulations, and ensure the safety and compliance of autonomous vehicles. In summary, expanding the multidimensional-data-mining-based obstacle avoidance path selection method for autonomous vehicles requires comprehensive consideration of multiple factors such as algorithm optimisation, hardware upgrades, data fusion and collaboration, security and privacy protection.

## 4 Conclusion

The rapid development of autonomous vehicle technology results from technological advancements in multiple fields, including sensor technology, artificial intelligence, machine vision, control theory, etc. The integration of these technologies enables autonomous vehicles to perceive the surrounding environment in real time, understand traffic rules, predict the behaviour of other traffic participants, and make decisions based on the information gathered, achieving autonomous navigation and obstacle avoidance. Therefore, this paper proposes a new obstacle avoidance path selection method for autonomous vehicles based on multi-dimensional data mining. The experimental results show that the proposed method for autonomous vehicle lane changing has a relatively large angle and short path, without collision problems. The maximum success rate of obstacle avoidance path selection is 98.56%, and the minimum time is 0.44s. The obstacle avoidance selection results are reliable. This paper studies ways in which autonomous vehicles can better adapt to complex traffic environments, reduce the occurrence of traffic accidents, improve road traffic capacity, and offer safer, more convenient, and efficient transportation modalities to people. Although the multidimensional-data-mining-based obstacle avoidance path selection method for autonomous vehicles has shown great potential, it still faces the challenges of relying on high-quality sensor data and dynamic traffic scenarios. To overcome these limitations, future research directions should focus on incorporating more real-time adaptive functionalities, such as using advanced machine learning algorithms to augment vehicles' abilities to comprehend and anticipate complex traffic environments, and comprehensively evaluating vehicle performance through extensive testing under varying climatic conditions, diverse road types, and different traffic density levels. At the same time, improving sensor data quality and fusion capabilities is also key, including enhancing sensor design to improve their adaptability to harsh environmental conditions, and optimising data fusion algorithms to integrate information from different sensors, thereby enhancing the accuracy and reliability of environmental perception. These efforts will collectively facilitate the continuous advancement and refinement of autonomous driving technology.

## Conflicts of interest

All authors declare that they have no conflicts of interest.

## References

Hu, C., Zhao, L., Cao, L., Control, M.P. and Observer, E.S. (2020) 'Steering control based on model predictive control for obstacle avoidance of unmanned ground vehicle', *Measurement and Control*, Vol. 53, Nos. 3–4, pp.501–518.

Huo, N. and Wang, Y.Q. (2024) 'Prediction and obstacle avoidance detection algorithm based on driverless', *Computer and Digital Engineering*, Vol. 52, No. 2, pp.411–415.

Jia, Z.C. (2022) 'Vehicle obstacle avoidance path planning method based on deep data mining', *Advances in Transportation Studies*, Vol. S2, No. 1, pp.12–22.

Li, S.P. (2024) 'Dynamic obstacle avoidance path planning method for autonomous vehicles based on improved convolutional networks', *New Technology and New Products of China*, Vol. 3, No. 3, pp.15–17.

Li, W., Wang, Y., Zhu, S.X., Chen, J.P., Guo, S.J., Ren, J.L. and Wang, D.B. (2023) 'Path tracking and local obstacle avoidance for automated vehicle based on improved artificial potential field', *International Journal of Control, Automation, and Systems*, Vol. 21, No. 5, pp.1644–1658.

Ma, Y., Mao, Z., Wang, T., Qin, J. and Meng, X. (2020) 'Obstacle avoidance path planning of unmanned submarine vehicle in ocean current environment based on improved firework-ant colony algorithm', *Computers and Electrical Engineering*, Vol. 87, No. 1, pp.106773–106784.

Mamak, K. and Glanc, J. (2022) 'Problems with the prospective connected autonomous vehicles regulation: finding a fair balance versus the instinct for self-preservation', *Technology in Society*, Vol. 71, No. 1, pp.1–7.

Matsushita, T., Chida, Y. and Tanemura, M. (2022) 'Obstacle avoidance path generation and following control for a vehicle-type weeder', *Proceedings of the Japan Joint Automatic Control Conference*, Vol. 12, No. 1, pp.216–222.

Qian, Y., Sun, H. and Feng, S. (2023) 'Obstacle avoidance method of autonomous vehicle based on fusion improved A*APF algorithm', *Bulletin of the Polish Academy of Sciences. Technical Sciences*, Vol. 71, No. 2, pp.15–26.

Qin, H., Yu, B., Dun, Y. and Bai, Y. (2023) 'Effect of cognitive experiences and attitudes on commuters' travel choice behavior for autonomous vehicles', *Journal of Urban Planning and Development*, Vol. 149, No. 2, pp.1–13.

Ren, Y. and Liu, J. (2022) 'Automatic obstacle avoidance path planning method for unmanned ground vehicle based on improved bee colony algorithm', *Jordan Journal of Mechanical and Industrial Engineering*, Vol. 16, No. 1, pp.11–18.

Wang, L., Jing, H., Zhong, G., Wang, J. and Wang, T. (2024) 'Research on autonomous vehicle obstacle avoidance path planning with consideration of social ethics', *Sustainability (2071-1050)*, Vol. 16, No. 11, pp.1–10.

Wang, Y., Xing, Y. and Cong, Y. (2024) 'Motion control of autonomous vehicles after a tyre blow-out based on differential-flatness-MPC', *International Journal of Vehicle Design*, Vol. 94, Nos. 3–4, pp.252–273.

Zhang, X.Y. and Zhang, X.J. (2022) 'Research on obstacle avoidance path selection of AGV based on improved ant colony algorithm', *Automation and Instrumentation*, Vol. 21, No. 6, pp.52–56.

Zhou, D., Ma, Z., Zhang, X. and Sun, J. (2022) 'Autonomous vehicles' intended cooperative motion planning for unprotected turning at intersections', *IET Intelligent Transport Systems*, Vol. 16, No. 8, pp.1058–1073.