


## Article

# Data Validity Analysis Based on Reinforcement Learning for Mixed Types of Anomalies Coexistence in Intelligent Connected Vehicle (ICV)

Jiahao Gao <sup>1</sup> , Chuangye Hu <sup>1</sup>, Luyao Wang <sup>1</sup> and Nan Ding <sup>1,2,\*</sup>

<sup>1</sup> College of Computer Science and Technology, Xinjiang Normal University, Urumqi 830054, China; gh0404929@gmail.com (J.G.)

<sup>2</sup> Key Laboratory of Intelligent Control and Optimization for Industrial Equipment, Dalian University of Technology, Dalian 116024, China

\* Correspondence: dingnan@dlut.edu.cn

**Abstract:** Compared with traditional anomaly analysis, intelligent connected vehicle (ICV) data validity analysis is faced with a variety of data anomalies, including sensor anomalies, driving behavior anomalies, malicious tampering, and so on, which eventually leads to anomalies in the data. How to integrate the vehicle moving characteristics, driving style, and traffic flow conditions to provide an effective data detection method has become a new problem in the field of intelligent networked vehicles. Based on ICV data, a particle swarm optimization data validity detection algorithm (TE-PSO-SVM) was proposed by combining driving style and traffic flow theory to realize the effective detection of driving data. In addition, aiming at the problem of mixed types of anomalies in complex scenes, a model pool is constructed, and a model selection algorithm based on reinforcement learning (RLBMS) is proposed. Experiments on the real data set HighD show that RLBMS has a better detection effect in complex scenes of mixed types of anomalies.

**Keywords:** intelligent connected vehicle; data validity analysis; traffic flow theory; driving style; particle swarm optimization algorithm; reinforcement learning



**Citation:** Gao, J.; Hu, C.; Wang, L.; Ding, N. Data Validity Analysis Based on Reinforcement Learning for Mixed Types of Anomalies Coexistence in Intelligent Connected Vehicle (ICV). *Electronics* **2024**, *13*, 444. <https://doi.org/10.3390/electronics13020444>

Academic Editor: Sergio Busquets-Monge

Received: 29 November 2023

Revised: 13 January 2024

Accepted: 17 January 2024

Published: 21 January 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Intelligent connected vehicle (ICV) refers to the organic combination of the Internet of vehicles and intelligent vehicles. ICV is equipped with advanced onboard sensors, controllers, actuators, and other equipment. It integrates modern communication and network technology to realize the exchange and sharing of intelligent information between vehicles and people, other vehicles, roads, backstage, etc., with safety, comfort, energy-saving abilities, high efficiency, and other characteristics [1]. As ICV technology continues to advance, more and more vehicles will be on the road, generating huge amounts of data. The data are expected to be used to improve traffic safety, optimize road transport, and improve urban planning. However, the validity analysis of ICV data is challenged by various data anomalies, including sensor anomalies, driving behavior changes, and potentially malicious tampering, resulting in anomalies in the data. Therefore, it is urgent to test the validity of ICV data. Intelligent connected vehicle (ICV) has the following three characteristics: an open wireless transmission medium; high-speed mobility of vehicle nodes; and a susceptibility to environmental influence and man-made information interference [2,3]. It is these three characteristics that make it possible for sensors or transmission lines to fail, causing the data to be tampered with, distorted, or lost. If the data are mixed with other false and messy data in the transmission process, it may cause traffic jams and even threaten the life of the driver. Ensuring the validity of ICV driving data is the key to enhancing vehicle safety, and how to ensure the validity of the data has become an extremely challenging task.

In recent years, vehicle security incidents caused by vehicle vulnerabilities have occurred many times. The increased connectivity and complexity of ICVs create a broader attack surface for security threats. Some examples of these attack surfaces include vehicle-to-everything (V2X) communication, telematics services, Bluetooth connectivity, and onboard diagnostics (OBD) ports [4]. Security researchers at one company say they have hacked Tesla's keyless system to open a car door in less than 10 s, successfully testing it on both the Model 3 and the Model Y [5]. In 2022, a 19-year-old hacker publicly said on the Internet that he had successfully found a vulnerability in the third-party software of a brand of new energy vehicles, and, through the vulnerability, he remotely invaded more than 25 vehicles of the brand in 13 countries. After the intrusion, it can remotely open and close doors, sound horns, control windows, and even achieve keyless driving [6]. These vehicle safety cases highlight the potential for outsiders to tamper with vehicle data, posing a serious threat to the data security and reliability of ICVs.

With the rapid development of autonomous driving and sensor technology (including radar, 3D LiDAR, etc.), a large amount of urban traffic data has been collected through fixed sensors or vehicle sensors (mobile) to monitor the basic state and the dynamic state of the traffic network [7]. However, since the spatial range of fixed sensors is limited, and mobile vehicle sensors may face instability during data collection, the data collection process is often accompanied by problems such as sensor failure or transmission distortion, affecting the effective communication between vehicles. According to the Texas Transportation Research report, the anomaly and loss of traffic data usually range from 16% to 93% [8]. At the 17th World Conference on Intelligent Transportation Systems, Chinese researchers pointed out in their report that abnormal and lost traffic data often occur in Beijing [9]. Therefore, low-quality problems such as missing data and abnormal data have been a serious obstacle to the development and application of ICV.

The research content of this paper is mainly applied to the field of assisted driving and automatic driving in order to solve the problem of traffic congestion and traffic accidents caused by vehicle data quality, so as to ensure that intelligent connected vehicles achieve the goal of safe, comfortable, energy-saving, and efficient driving. According to the above objectives, this paper proposes the following two methods: This paper combines driving style with traffic flow theory and designs a data validity checking algorithm based on particle swarm optimization. In real-world environments, data anomalies may appear in many forms. To solve this problem, this paper proposes an anomaly detection model based on reinforcement learning, which can flexibly adapt to different forms of anomalies, thereby improving the robustness and adaptability of data validity detection. The main contributions of this study are as follows:

- (1) Combined with the acceleration and traffic flow theory, the driving style recognition coefficient is defined, and the driving style quantitative model is designed to realize the quantification of the driver's driving style. The traffic flow model is established, and the vehicle state data are fused with the driving style and traffic flow theory to predict the vehicle speed through the LSTM network.
- (2) Based on the driving style, traffic flow theory, and vehicle driving state information, a data validity detection algorithm (TE-PSO-SVM) is designed by using a particle swarm optimization support vector machine model.
- (3) Due to the diversity of ICV data, the detection accuracy of a single model is still limited in the scenario of the mixed coexistence of multiple types of anomalies. Therefore, by combining the vehicle data, traffic flow model, and driving style, different basic models are used to construct a model pool for the diversity of anomalies. According to the characteristics of the different algorithms in the model pool, the concepts of distance threshold confidence (D\_T\_C) and predictive consensus confidence (P\_C\_C) are introduced, and then reinforcement learning is used for model selection. A model selection algorithm based on reinforcement learning (RLBMS) is designed to further improve the accuracy of data validity detection.

The rest of this paper is organized as follows: In Section 2, many similar works are presented. Section 3 introduces the method used for quantifying driving style and the method of speed prediction based on traffic flow theory. In Section 4, a data validity detection algorithm (TE-PSO-SVM) is proposed based on driving style, traffic flow theory, and vehicle driving state information. In Section 5, a model selection algorithm based on reinforcement learning (RLBMS) is proposed and introduced by taking advantage of different basic models. In Section 6, the quantification process of driving style is presented, the speed prediction based on traffic flow theory is tested, the data validity algorithm proposed in the first two sections is tested, and the conclusion is drawn. Section 7 summarizes this paper and proposes future work.

## 2. Related Work

In order to solve the above problems, various scholars have applied different methods and conducted a lot of experiments. Y. Wang et al. developed a novel and comprehensive framework that combines an Adaptive Extended Kalman Filter (AEKF) with a car-following motion model and uses a data-driven fault detector for data analysis [10]. N. Ding et al. proposed an anomaly detection algorithm based on drivers' emotions (EAD). For collaborative ICV, a method based on the driver's emotional state has been used to realize the real-time detection of safe-driving-related data through the GMM model [11]. Youcef et al. proposed a k-nearest neighbors (k-NN) anomaly detection algorithm that takes into account both the spatial and temporal information data of the traffic flow [12]. Peng et al. proposed a new method for intrusion detection (AMAEID) using an attention mechanism and an automatic encoder. The attention mechanism and autoencoder for intrusion detection (AMAEID) model utilizes a multi-layer denoising autoencoder model and a dropout network layer to encode and decode message data, obtaining a deeper representation of the potential features behind the message data. Finally, the above hidden feature representation is used to infer whether the data are an abnormal message [13]. Sharma et al. used the concept of ensemble learning to establish an ensemble detection model by integrating classifiers such as k-nearest neighbor, decision tree, AdaBoost, and random forest to perform anomaly detection on Internet of Vehicle (IoV) data [14]. Malith Ranaweera et al. integrated the traffic flow phenomena into abnormal data detection techniques to improve the assessment of vehicular network threats. This method determines whether the data are abnormal (i.e., speed and spatial spacing) by analyzing microscopic parameters derived from traffic flow theory [15]. Khodayari et al. combined neural networks to build a car-following model. On the basis of the existing car-following model, the reaction time was added in order to determine whether the data were abnormal through the neural network car-following model, and the effectiveness of the model was verified through real traffic data [16]. Tianjia He et al. detected anomalous sensor measurements by exploiting the inherent redundancy among heterogeneous sensors, which is the simultaneous response of multiple sensors to the same physical phenomenon in a correlated manner. For example, pressing on the accelerator increases the engine RPM and vehicle speed. An anomaly detector was proposed by embedding redundancy into a deep autoencoder [17]. Traditional data validity analysis methods generally focus on the data of a single vehicle or part of the vehicle and need to further consider the driver's status and traffic flow status. How to combine the vehicle's own data characteristics, driving style, and traffic flow characteristics to provide effective data detection methods has become a new problem in intelligent connected vehicles.

Driving style and traffic flow data have become crucial due to the uniqueness of driverless cars. In this context, Milardo et al. conducted data collection on multiple vehicle collision events. Through an in-depth analysis of a large amount of collision data, they concluded that there is an inseparable correlation between drivers' driving style and vehicle data [18]. This correlation not only highlights the impact of driving behavior on vehicle performance, but also provides useful insights for a deeper understanding of intelligent connected vehicle systems. Therefore, the driver's style can be quantified by the model to better judge the validity of the data in ICV. Traffic flow theory provides a theoretical

basis for data validity detection through the study of traffic flow, density, speed, and other indicators [19], which helps researchers to have a more comprehensive understanding of road traffic conditions and make a more scientific assessment of data validity detection.

### 3. Driving Style and Traffic Flow Theory

In order to reduce the number of accidents and, thus, improve road safety, Nesrine Kadri et al. designed a new recurrent neural network structure based on stacked long short-term memory (LSTM) for classifying driving behaviors. By applying the Dempster-Shafer (DS) belief function theory, the uncertainty of data was successfully overcome, thus, significantly improving the accurate classification results of driving states [20]. Amina Turki et al. proposed a hybrid real-time system based on the eye-closing ratio and mouth-opening ratio, which has two processes: offline and online. The offline process uses a pre-trained convolutional neural network (CNN) to perform a classification module to detect the driver's drowsiness. The online procedure uses Chebyshev distance to calculate the percentage of the driver's online eye-closing and yawning frequency from live video. With the help of pre-training CNN based on the ensemble learning paradigm, the driver's sleepiness state can be accurately evaluated [21]. Driving behavior affects a driver's speed on the road. Drivers in highly nervous or excited situations may present abnormal driving behavior, resulting in abnormal data. Therefore, quantifying the driving style can help us to judge the validity of the data.

Based on urban traffic flow theory, Lin, X. et al. combined the ARIMA model and the Garch model to obtain corresponding fluctuation characteristics and proposed a short-term high-speed traffic flow prediction method based on the ARMI-GARCH-M model to realize the prediction of traffic flow data [22]. Yacong Gao et al. developed an LSTM short-term traffic prediction model that can more fully capture recent spatio-temporal features and, thus, proposed a short-term traffic speed prediction method based on multi-temporal traffic flow states before and after. This method can accurately reflect the running state of the road in real time [23]. Traffic flow theory provides a theoretical basis for data anomaly detection through the study of traffic mobility, density, speed, and other indicators, which can help researchers to understand the traffic flow on the road and to make a more accurate judgment on data validity detection.

In summary, there is an important relationship between driving style and vehicle data, and traffic flow theory provides a theoretical basis to help researchers to more scientifically detect and evaluate data quality on the road. This helps to improve road traffic efficiency and safety and is of great significance for traffic management and safety research.

#### 3.1. Driving Style

When quantifying a driver's driving style, it is more accurate to consider multiple factors, such as vehicle speed, acceleration, and traffic density, which is helpful to better understand and describe the complexity of the driving behavior.

According to the driver's driving behavior scale and the driver's driving safety scale [24], the driver's driving style is divided into the following three types: cautious type, normal type, and radical type. Some researchers only consider the vehicle's data to quantify the driver's driving style [25,26]. For example, Murphey et al. [26] proposed a very classic driver style recognition coefficient algorithm,  $R_{\text{driver}}$ , through vehicle speed or acceleration. Few scholars have considered the impact of traffic density on driver style, therefore, there is a certain deviation from the actual situation, which eventually leads to inaccurate quantitative results. (1) When the traffic density is small, the vehicle belongs to free driving, and the frequent change of acceleration is normal. However, when only considering the acceleration to judge the driving style, the driver's style is likely to be judged as radical. (2) Under the same traffic density, whether the acceleration is stable or with frequent changes, the driver's driving style must be different. Only considering the density cannot distinguish the driving style.

Based on the driver style recognition coefficient,  $R_{driver}$ , proposed by Murphey et al., [27], a new driving style recognition coefficient,  $E$ , is proposed by combining traffic density and acceleration. Firstly, the density ratio ( $D$ ) is defined by the density around the vehicle, and then the driver's style recognition coefficient,  $R_{driver}$ , is calculated by the speed. Finally, the driving style recognition coefficient,  $E$ , is expressed as Equations (1)–(3), as follows:

$$J = \frac{d^2v(t)}{d^2t} \tag{1}$$

$$R_{driver} = \frac{W_J}{\bar{J}} \tag{2}$$

$$E = D * R_{driver} = \frac{k}{k_m} * \frac{W_J}{\bar{J}} \tag{3}$$

where  $k$  is the density around the target vehicle and  $k_m$  is the density corresponding to the maximum traffic flow of the road section, which is obtained according to the Greenshield traffic flow model [28], as shown in Figure 1;  $J$  is the degree of vehicle impact, where the physical meaning is the rate of change of acceleration;  $\bar{J}$  is the average impact degree obtained by normal drivers in the same traffic environment; and  $W_J$  is the standard deviation of the impact in the calculated window when the time window is within  $\omega$ . Through a large number of experiments and studies, Murphey et al. proved that, when the time window  $\omega$  is set to 3~9 s, it has a higher recognition of the driver's style.

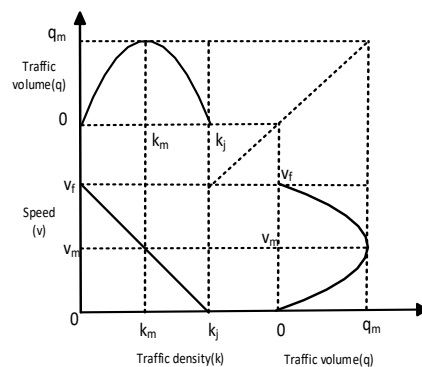


Figure 1. Greenshield traffic flow model.

Finally, the following two different thresholds are set: Norm-threshold and Agg-threshold. The calculated driving style recognition coefficient is compared with the two thresholds in Table 1 to get the final driving style category.

Table 1. Driving style classification table.

Driving Style	Driving Style Division	Driving Style Parameters
Cautious type	$E < \text{Norm-threshold}$	1
Normal type	$\text{Norm-threshold} < E < \text{Agg-threshold}$	2
Radical type	$E > \text{Agg-threshold}$	3

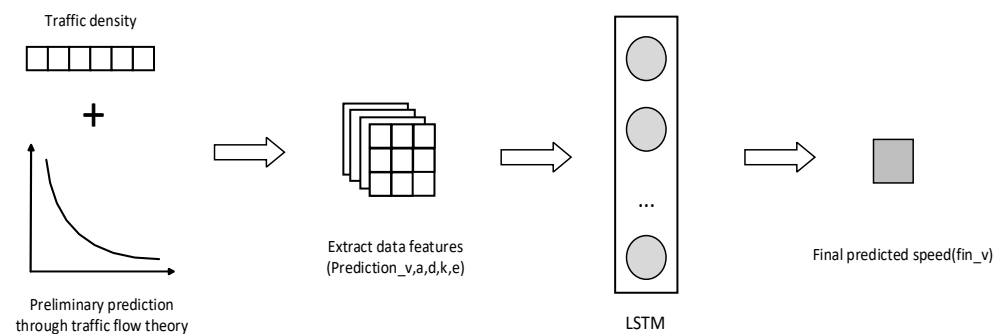
### 3.2. Speed Prediction Based on Traffic Flow Theory

The Greenshield model is one of the basic models used to describe traffic flow behavior, which was proposed by Canadian traffic engineer Bruce D. Greenshields [28]. The model provides insight into the dynamic characteristics of traffic flow by studying the relationship between vehicle density, speed, and flow [29].  $k_j$  is the jam density,  $k_m$  is the optimal density,  $v_f$  is the free-flow speed,  $v_m$  is the optimal speed, and  $q_m$  is the maximum traffic flow. The density–speed–flow relationship diagram is shown in Figure 1.

Through the traffic flow model, the relationship between speed and density can be obtained, but it is a macroscopic traffic flow model and cannot accurately predict the vehicle speed. Therefore, this paper combines the vehicle state data with the traffic flow model and driving style and predicts the vehicle speed through the long short-term memory (LSTM) network to further improve the prediction of the vehicle speed data [30].

Long short-term memory (LSTM) is a variant of recurrent neural networks that are used to handle the modeling and forecasting of sequential data and time series data. Its memory unit consists of a state vector and three gating vectors, namely the input gate, the forget gate, and the output gate. These gating vectors control the flow and retention of information through a sigmoid function. LSTM is designed such that it can effectively handle long-term dependencies and has a strong ability for modeling and predicting sequence data. The driving style and traffic flow data capture the driver's state and the surrounding environmental factors more fully, which is beneficial for improving the accuracy of speed prediction. In this experiment,  $n$ -dimensional data  $x$ , such as traffic flow prediction speed, driving style, and car-following distance, are selected, and then  $X = [x_1, x_2, \dots, x_t]$  is used as the input of LSTM to obtain the speed prediction value at time  $t + 1$ . The experiments show that, when  $t = 8$ , the predicted results are better.

Firstly, the number of surrounding vehicles is obtained by ICV, and the traffic density is calculated. Secondly, the speed–density relationship diagram is used to predict the speed and obtain the macroscopic speed of the vehicle. Then, the preliminary predicted vehicle speed, vehicle status, and driving style data are input into the LSTM network as features. The final predicted vehicle speed (fin\_v) is obtained as the basis for subsequent ICV data validity detection. The specific prediction process is shown in Figure 2.



**Figure 2.** Predicting vehicle speed based on traffic flow and driving style.

#### 4. Data Validity Detection Algorithm Based on Driving Style and Traffic Flow Theory

##### 4.1. Support Vector Machine

A support vector machine (SVM) is a machine learning algorithm that is widely used in supervised learning, mainly for classification and regression tasks [31]. The basic idea is to find a hyperplane that separates the different classes of data and maximizes the distance from the nearest data point to the hyperplane. The commonly used kernel functions include linear kernel function, polynomial kernel function, and radial basis function (RBF) kernel function. In the SVM classification model with RBF as the kernel function, the penalty factor ( $C$ ) and the kernel function coefficient ( $\sigma$ ) are important parameters that affect the classification accuracy. In order to improve the accuracy of anomaly detection, the values of  $C$  and  $\sigma$  must be optimized, and the particle swarm optimization algorithm is used to optimize the parameters.

##### 4.2. Particle Swarm Optimization

Particle swarm optimization (PSO) is a swarm intelligence algorithm inspired by the foraging behavior of birds [32]. PSO simulates cooperation and information sharing among individuals in the bird flock to find the optimal solution to the problem. This method can adaptively control the evolution of individuals according to the change of

population state and balance the global and local search capabilities. In PSO, the solution space is represented as the positions of a swarm of particles, where each particle represents a solution. The particle searches in the solution space and adjusts its moving direction and speed on its own historical experience and information exchange with its neighboring particles. Through continuous iterative updates, the particle swarm gradually converges to the optimal solution. The PSO algorithm is simple, easy to implement, and has good global search ability. After finding the global and local optimal values for the  $i$  particle of the  $t + 1$  generation, the update rules of its velocity and position are shown by Equations (4) and (5), as follows:

$$v_i(t + 1) = \rho v_i(t) + \lambda_1 \cdot R_1 \cdot [pbest_i - x_i(t)] + \lambda_2 \cdot R_2 \cdot [gbest - x_i(t)] \quad (4)$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (5)$$

where  $v_i(t + 1)$  is the velocity of the  $i$ th particle of the  $t + 1$  generation;  $\rho$  is the inertia weight, which is used to control the influence degree of particle historical velocity in the update;  $v_i(t)$  is the velocity of the  $i$  particle of the  $t$  generation;  $\lambda_1$  and  $\lambda_2$  are the learning factors, which represent the influence degree of particle individual experience and group experience in updating respectively;  $R_1$  and  $R_2$  are random numbers, which are used to introduce randomness and increase the diversity of the search;  $Pbest_i$  is the individual optimal solution of  $i$ th particle, that is, the best position that the particle has experienced;  $x_i(t)$  is the position of the  $i$  particle of the  $t$  generation;  $gbest$  is the group optimal solution, that is, the best position among all particles. The specific structure of the PSO-optimized SVM algorithm is shown in Algorithm 1.

---

**Algorithm 1** PSO-SVM algorithm.

---

**Input:** S, Search space; T, Maximum iterations; F, Features in the dataset.

**Output:** Optimal penalty factor (C) and kernel function coefficients( $\sigma$ ).

(1) Initialize parameters Number of particles N, inertia weight  $\rho$ , Learning factor  $\lambda_1$  and  $\lambda_2$ . Random numbers  $R_1$  and  $R_2$ .

(2) The initial position and velocity of N particles are randomly generated.

(3)  $j = 1$ .

While ( $j < T$ )

For each particle  $i$ , the velocity and position information are calculated according to Equations (4) and (5).

Its fitness value is calculated for each particle, and  $pbest_i$  and  $gbest$  are updated.

End

---

#### 4.3. TE-PSO-SVM Algorithm

By combining driving style with traffic flow theory, we can obtain the vehicle status, driver style, and road information more comprehensively and provide a more accurate data validity detection scheme. Algorithm 2 gives the overall description of the TE-PSO-SVM algorithm.

---

**Algorithm 2** TE-PSO-SVM algorithm.

---

**Input:** Vehicle location and speed information.

**Output:** Data detection result.

(1) Obtain vehicle location information and calculate the density around the target vehicle.

(2) The driving style is determined by Equation (3) and the driving style quantization table.

(3) The vehicle speed was predicted by the model in Figure 2.

(4) Combined with driving style and traffic flow data, the detection results were output by the PSO-SVM model.

---

## 5. Model Selection Algorithm Based on Reinforcement Learning

Due to the complexity of real-world data, a single model may not be sufficient to fully capture the data characteristics, resulting in insufficient data accuracy [33]. In order to make use of the advantages of different basic models and solve the problem of insufficient detection accuracy of a single model in the scene where multiple types of abnormal mixtures coexist, the following four different structural algorithm models are selected as the basic algorithms in the model pool: (1) Logistic Regression [34]: It is a linear model for classification tasks where the output layer uses a logistic function to map the input to a probability value between 0 and 1; (2) Multilayer Perceptron (MLP) [35]: It is a model based on the neural network; (3) Decision Tree [36]: It is an algorithm model based on a tree structure, layer by layer reasoning to achieve the final classification; (4) PSO-SVM [37]: It is a sample model that divides different categories by the optimal hyperplane. Finally, a model selection algorithm based on reinforcement learning is proposed in combination with the Advantage Actor–Critic algorithm.

Advantage Actor–Critic (A2C) is a reinforcement learning algorithm that combines the policy gradient (Actor) and value function approach (Critic). It aims to improve the performance of a reinforcement learning agent in an environment by simultaneously optimizing a policy and a value function [38]. A2C interacts with the environment at each step and uses the experience of these interactions to update the policy and the value function. Reinforcement learning is different from other types of machine learning methods. The training data mainly comes from various interactions with the environment. Its advantage lies in its reward and punishment mechanism, which can respond quickly to environmental information [39]. The value network scores  $s_t$  and  $s_{t+1}$  via Equations (6) and (7), respectively. TD target and TD error are shown in Equations (8) and (9). The value network and policy network updates are shown in Equations (10) and (11), as follows:

$$\hat{v}_t = v(s_t; w) \quad (6)$$

$$\hat{v}_{t+1} = v(s_{t+1}; w) \quad (7)$$

$$\hat{y}_t = r_t + \gamma \cdot \hat{v}_{t+1} \quad (8)$$

$$\delta_t = \hat{v}_t - \hat{y}_t \quad (9)$$

$$w_{new} \leftarrow w_{now} - \alpha \cdot \delta_t \cdot \nabla_w v(s_t; w_{now}) \quad (10)$$

$$\theta_{new} \leftarrow \theta_{now} - \beta \cdot \delta_t \cdot \nabla_\theta \ln \pi(a_t | s_t; \theta_{now}) \quad (11)$$

where,  $\hat{v}_t$  and  $\hat{v}_{t+1}$  are the value network score values,  $w$  is the value network parameter used to evaluate the quality of the state,  $\theta$  is the policy network parameter used to control the selected action,  $r_t$  is the reward value,  $s_t$  and  $s_{t+1}$  are the state, and  $\alpha$  and  $\beta$  are the value loss coefficient and policy loss coefficient, respectively.

### 5.1. Overall Workflow

Firstly, each candidate anomaly detector is pre-trained on the training set. Secondly, all of the detectors are run on the test data, and each detector calculates a series of abnormal scores for each test instance. According to the prediction scores and thresholds, all basic detectors can generate a set of prediction labels for the test data. Then, using the prediction score, threshold, and prediction label obtained in the previous step, two additional confidence scores are defined. Finally, these two confidence scores are integrated into the state variables of the reinforcement learning model along with the prediction scores, thresholds, and prediction labels. The specific structure is shown in Figure 3.



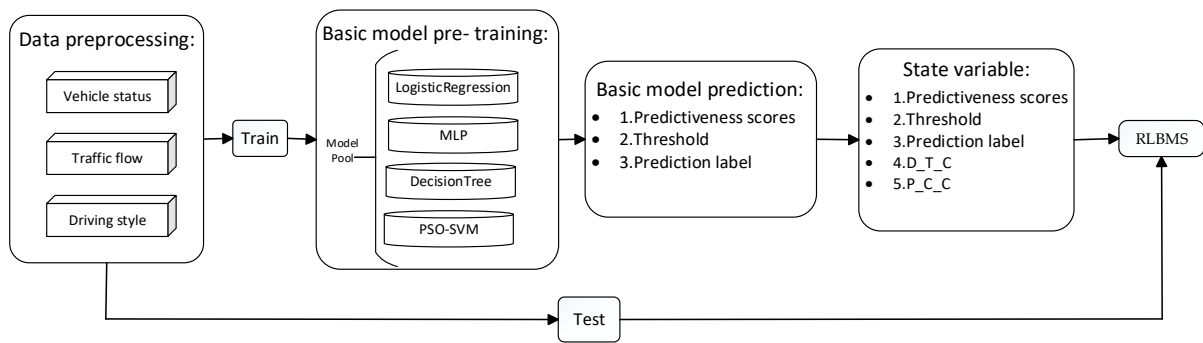


Figure 3. Algorithm flow chart.

In this experiment, the following two scores are proposed to be added to the state variables of the reinforcement learning model to improve the accuracy and robustness of the model selection algorithm based on reinforcement learning:

$$D_{T_C} = \frac{|pro - Threshold|}{Pro_{max} - Pro_{min}} \tag{12}$$

Distance threshold confidence ( $D_{T_C}$ ): The higher the prediction probability is, the higher the confidence of the corresponding instance in the current model prediction. The probability and threshold detected by the model are used to describe the rationality of the model prediction. Here, *pro* indicates the prediction probability, *Threshold* indicates the threshold, and *Pro<sub>max</sub>* and *Pro<sub>min</sub>* indicate the maximum and minimum of the prediction probability, respectively.

$$P_{C_C} = \frac{m}{M} \tag{13}$$

Predictive consensus confidence ( $P_{C_C}$ ): This is a metric that is inferred by the majority voting idea in ensemble learning. The higher the confidence level, the more consistent the prediction results of multiple models, and the more credible the prediction ability of the model. Here, *m* is the number of the same votes cast in the model and *M* is the total number of models in the model pool.

### 5.2. Introduction of Model Selection Algorithm Environment Based on Reinforcement Learning

**Actor:** The task of the actor is to learn a policy, that is, how to select actions from the state space. Policies can be deterministic (output an action directly) or probabilistic (output a probability distribution over actions). **Critic:** The task of the critic is to learn a value function that is used to estimate the long-term value of the state. The value function represents the expected reward that can be obtained after executing the current policy in a given state. The actor and the critic collaborate with each other. The actor uses the information of the value function provided by the critic to improve their policy, and the critic in turn updates the value function by interacting with the actor’s policy. This collaborative learning makes the algorithm more stable and efficient. According to the research content of this paper, the selection process of the detector is shown in Figure 4.

**States:** State refers to the information observed in the environment that is used to describe the current situation. In this paper, each state is regarded as a state variable obtained from the prediction results of the selected anomaly detector, where the state variables include an anomaly score, anomaly threshold, prediction label, distance threshold confidence, and prediction consensus confidence.

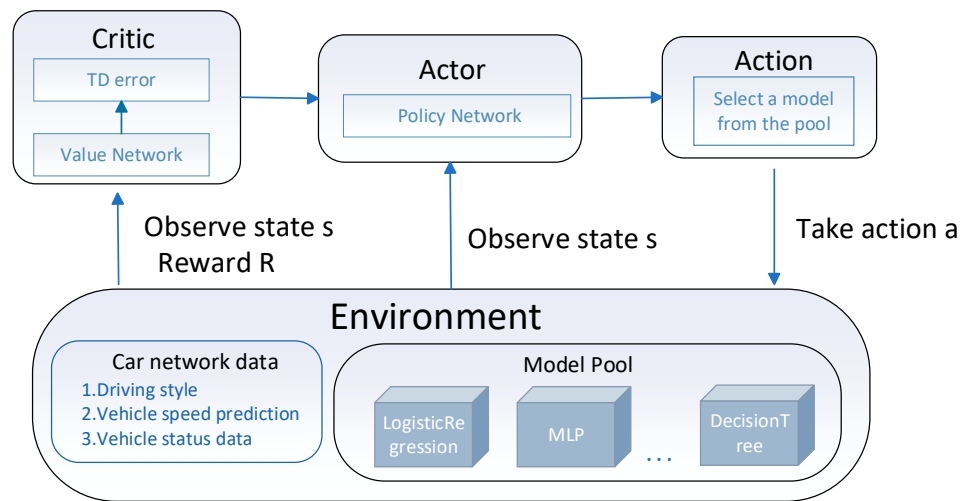


Figure 4. Dynamic selection anomaly detector.

**Action:** Action space refers to the set of actions that can be selected in each state. The size of the action space in this paper is the number of algorithms in the model pool, and the index of the candidate detector is selected from the model pool as an action.

**Reward:** The reward function defines the reward signal obtained in each state transition. The reward function in this paper is determined by the real label value and the predicted label value, as shown in Equation (14).

$$\text{Reward} = \begin{cases} Y1, \text{ True Positive}(TP) \\ Y2, \text{ True Negative}(TN) \\ Y3, \text{ False Positive}(FP) \\ Y4, \text{ False Negative}(FN) \end{cases} \quad (14)$$

### 5.3. Model Selection Algorithm Based on Reinforcement Learning

In order to take advantage of the advantages of different models, based on the PSO-SVM model, three other algorithms with different structures are added to the model pool to capture data features. According to the state information in the reinforcement learning algorithm, the corresponding action is obtained, and the algorithm that is suitable for the current state is selected from the model pool. For example, if the calculated distance threshold confidence and prediction consensus confidence are very low, it shows that the algorithm is not sensitive to such anomalies, resulting in inaccurate judgment results. It is necessary to switch other algorithms from the model pool to improve the accuracy of anomaly detection. Algorithm 3 gives the overall description of the model selection algorithm based on reinforcement learning.

---

**Algorithm 3** Model selection algorithm based on reinforcement learning.

---

**Input:** Vehicle location and speed information.

**Output:** Data detection result.

- (1) Obtain vehicle location information and calculate the density around the target vehicle.
  - (2) The driving style is determined by Equation (3) and the driving style quantization table; predict vehicle speed through the model in Figure 2.
  - (3) Train the model in the model pool, and obtain prediction scores, thresholds, and prediction labels for each algorithm in the model pool.
  - (4) Calculate  $D\_T\_C$  and  $P\_C\_C$  from the prediction score, threshold, and prediction label using Equations (12) and (13).
  - (5) Select appropriate base models from the model pool using the A2C algorithm based on state information.
-

## 6. Experiments and Results Analysis

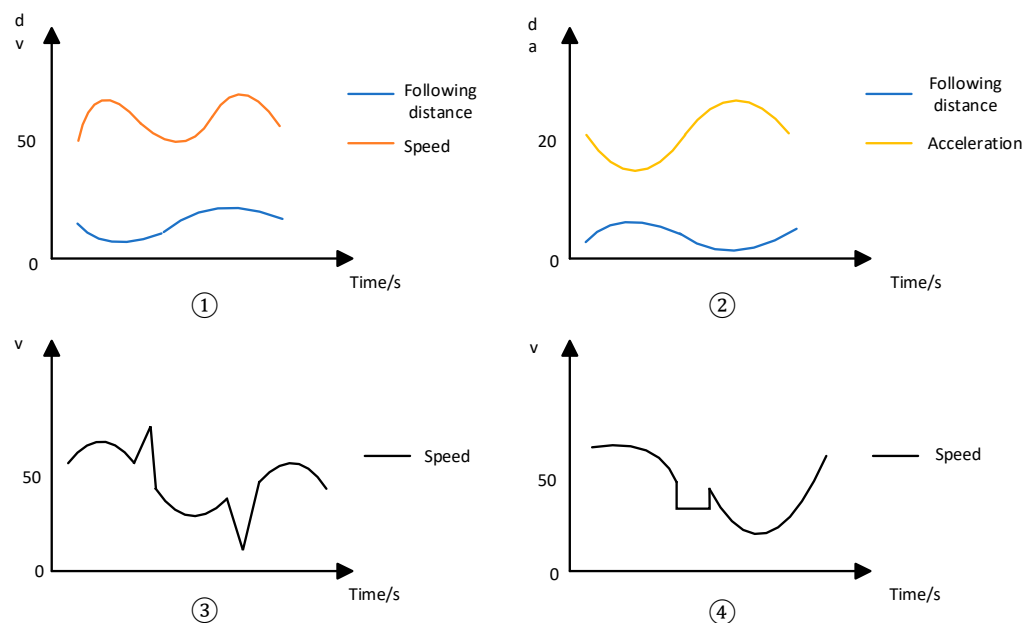
### 6.1. Data Set and Environment Description

#### (1) Selected datasets.

Based on driving style and traffic flow theory, this paper uses the HighD data set to detect the validity of the ICV data. It is part of the highway scene in the LevelX data set of the University of Aachen, Germany. A total of six locations are sampled, all of which are on the highway near Aachen. Collected by drones, each section is sampled by about 110,000 vehicles, is 410 m long and 16.5 h long, and has a sampling frequency of 25 Hz. The positioning error is usually less than 10 cm [40]. Five vehicles in the HighD dataset were randomly selected as the data set, and the data set was divided into training and testing sets in a 7:3 ratio.

According to The Automatic Driving White Paper [41], the vehicle speed is the data of vehicle decision making and control, and its importance is high. Therefore, the speed is changed in the experiment to simulate the abnormal data. To simulate the error caused by data acquisition, road bumps, etc., Gaussian Noise is added to the speed and distance of the vehicle. According to the actual situation investigation, as shown in Figure 5, the following four kinds of abnormal data are defined [12,42]:

- ① When the car-following distance is small, the speed is too large;
- ② When the car-following distance is small, the acceleration is too large;
- ③ Speed mutation;
- ④ The speed appears as a constant value.



**Figure 5.** Data anomaly definition.

#### (2) Environment Description.

The experimental system is Windows 11; the CPU is the 13th generation Intel (R) Core (TM) i9-13900HX 2.20 GHz; the memory is 16 GB; Our experiment was conducted using Python 3.9 in PyCharm IDE version 2023.3.3.

#### (3) Table 2 shows the description of the relevant parameters.

**Table 2.** Explanation of relevant experimental parameters.

Parameter	Numerical Value	Description
Norm-threshold	0.4	Normal type threshold
Agg-threshold	1.1	Aggressive threshold
$k_m$	0.08	Optimal density
$\omega$	30	Time window
$\rho$	0.8	Inertia weight
$\lambda_1, \lambda_2$	0.5, 0.5	Learning factor
T	10	Number of iterations
Gaussian Noise	(0, 0.3), (0, 0.4)	(mean, variance)
(Y1, Y2, Y3, Y4)	(1.5, 0.5, -0.5, -1)	Reward value
$\alpha$	0.5	Value Loss Coefficient
$\beta$	0.1	Policy Loss Coefficient
lr	$7 \times 10^{-4}$	Learning rate
$\gamma$	0.9	Discount factor

### 6.2. Evaluating Indicator

When evaluating the experimental results, this paper selects Precision (P), Recall (R), and F1-Score (F1), which are commonly used in the field of data anomaly detection as evaluation criteria. Precision is the proportion of all of the samples judged by the model to be positive categories that are actually positive categories. Recall is the proportion of samples that are actually positive that are correctly classified as positive by the model. The F1-Score is the harmonic mean of Precision and Recall, which can be used to measure the balanced performance of the model on normal samples and abnormal samples. The calculation methods of Precision (P), Recall (R), and F1-Score (F1) are shown in Equations (15)–(17).

$$Precision = \frac{TP}{TP + FP} \quad (15)$$

$$Recall = \frac{TP}{TP + FN} \quad (16)$$

$$F1 - Score = \frac{2 \times P \times R}{P + R} \quad (17)$$

The anomaly is regarded as ‘positive,’ and the normal data point is ‘negative.’ According to the definition, True Positive (TP) is the anomaly correctly predicted, True Negative (TN) is the normal data correctly predicted, False Positive (FP) is the normal data point wrongly predicted as abnormal, and False Negative (FN) is the abnormal data point wrongly predicted as normal.

### 6.3. Driving Style Quantification

In order to clearly show the steps needed to quantify the driving style, this paper presents the specific process diagram of the driving style quantification of car ID = 974 through the driving style quantification method proposed in Section 3.

The traffic density while the vehicle is driving, the standard deviation of the impact degree within the window, and the final quantization table of the driver style are shown in Figures 6–8, respectively.

Traffic density refers to the number of vehicles passing on a unit road section. A higher traffic density means more vehicles on the road, which makes the traffic environment more complex.

The standard deviation of impact within the window is a statistic of the change rate of the acceleration derivative over a period of time. This indicator can reflect the driver’s acceleration and deceleration behavior. The standard deviation of the larger impact indicates that the driver is more inclined to accelerate and decelerate frequently and perform aggressive driving behavior.

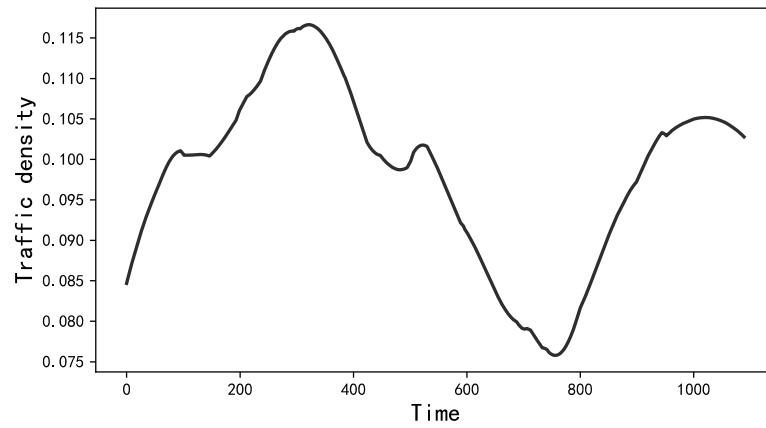


Figure 6. Traffic density.

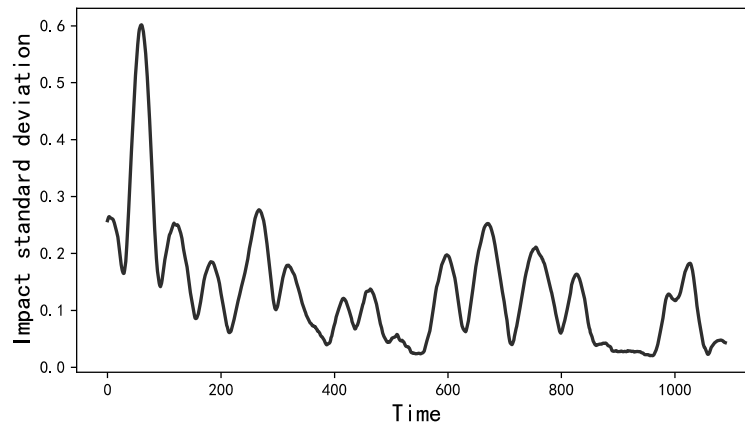


Figure 7. The standard deviation of impact within the window.

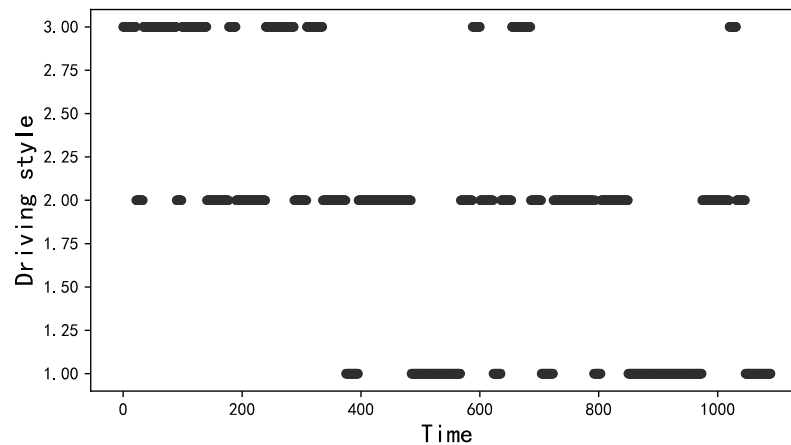


Figure 8. Quantification of driving style.

If the standard deviation of the impact degree in the window of a driver is large in an environment with high traffic density, it indicates that their driving style is more aggressive and there may be safety risks. Under different traffic characteristics, not only will vehicle sensors convey different data characteristics, but drivers will also show different driving styles. By quantifying the driving style, researchers can better understand the driving preferences and behaviors of drivers, which is helpful for data validity detection.

### 6.4. Speed Prediction Based on Traffic Flow Theory

To further investigate the relationship between traffic density, speed, and flow, the HighD data set of the highway scene of Aachen University in Germany is used to calculate the traffic density, speed, and flow of the data set, and the least square method [43] is used to obtain the relationship between density and speed in the data set, as shown in Figure 9.

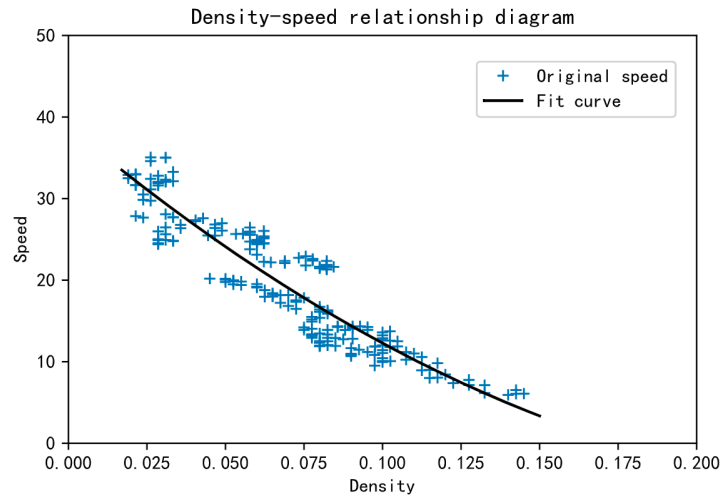


Figure 9. Density–speed relationship.

Since the above model is a macroscopic traffic flow model, it cannot accurately predict the vehicle speed. Therefore, the method of vehicle state data combined with driver style and traffic flow theory in Section 3 is used to predict the vehicle speed.

In this paper, the root mean square error (RMSE) and mean absolute error (MAE) are used to calculate and measure the performance of the prediction model, as follows:

$$RMSE = \sqrt{\left(\frac{1}{n} * \Sigma(v_i - \hat{v}_i)^2\right)} \tag{18}$$

$$MAE = \frac{1}{n} * \Sigma|v_i - \hat{v}_i| \tag{19}$$

where  $n$  is the number of samples,  $v_i$  is the actual observation value,  $\hat{v}_i$  is the predicted value, and  $\Sigma$  is the sum of all samples. In general, the smaller the root mean square error and the mean absolute error, the higher the accuracy of the model.

Five vehicles were randomly selected in this experiment to validate the proposed method. The predicted speed of the vehicle is calculated and the RMSE and MAE are used as the evaluation standard criteria. From the conclusions drawn in Tables 3 and 4, it can be concluded that the predicted speed derived from the vehicle state data combined with considering the driver style and traffic flow theory has a higher accuracy compared to the data derived from the traffic flow theory only.

Table 3. Comparison table for MAE speed prediction.

Vehicle ID	RMSE	Traffic Flow Theory	LSTM Speed Prediction Based on Driving Style and Traffic Flow
1111		4.06	0.57
869		3.34	0.60
392		5.16	0.52
224		3.21	0.55
131		1.56	0.38
AVG		3.37	0.52

**Table 4.** Comparison table for RMSE speed prediction.

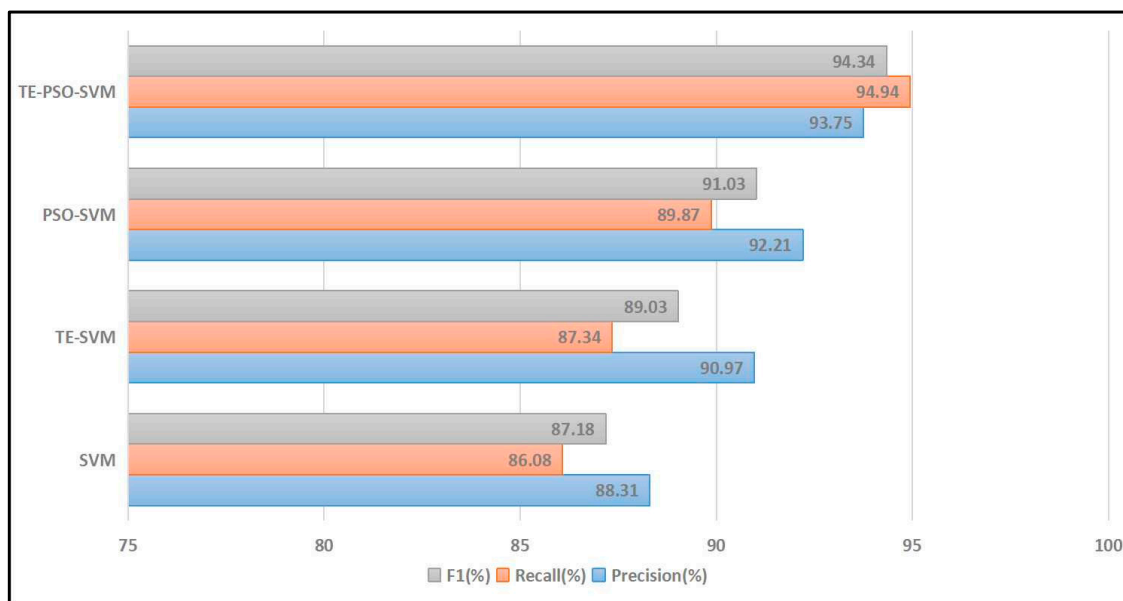
Vehicle ID \ MAE	Traffic Flow Theory	LSTM Speed Prediction Based on Driving Style and Traffic Flow
1111	3.61	0.45
859	3.10	0.51
392	4.87	0.36
224	2.77	0.40
131	1.38	0.35
AVG	3.15	0.41

The experimental results show that considering the factors that affect the vehicle speed in the prediction of vehicle speed can effectively improve the accuracy of vehicle speed prediction and provide more reliable information for driving decision making and data analysis.

6.5. Anomaly Detection Results and Discussion

6.5.1. Data Validity Detection Experiment Based on Driving Style and Traffic Flow Theory

This experiment combines the support vector machine model optimized by particle swarm optimization and compares the three-dimensional intelligent connected vehicle data (v, a, d) with the six-dimensional intelligent connected vehicle data (v, a, d, e, k, fin\_v) combined with driving style and traffic flow theory. In order to verify the performance of the algorithm in different environments, Gaussian Noise with a mean of 0 and a variance of 0.3 and Gaussian Noise with a mean of 0 and a variance of 0.4 is added to the detection data in Figures 10 and 11, respectively. The final detection results are shown in Figures 10 and 11.



**Figure 10.** Performance comparison of different algorithms in 0.3 Gaussian Noise.

Through the analysis of the experimental results in Figures 10 and 11, it can be seen that the data detection algorithm combining driver style and traffic flow theory has different degrees of improvement in the evaluation indicators such as Precision, Recall, and F1, compared with the traditional algorithm, which proves that the algorithm adding driver style and traffic flow theory has better performance in different external environments. At the same time, these experimental results also actually verify that there is a close relationship between vehicle state data, driving style, and traffic flow characteristics.

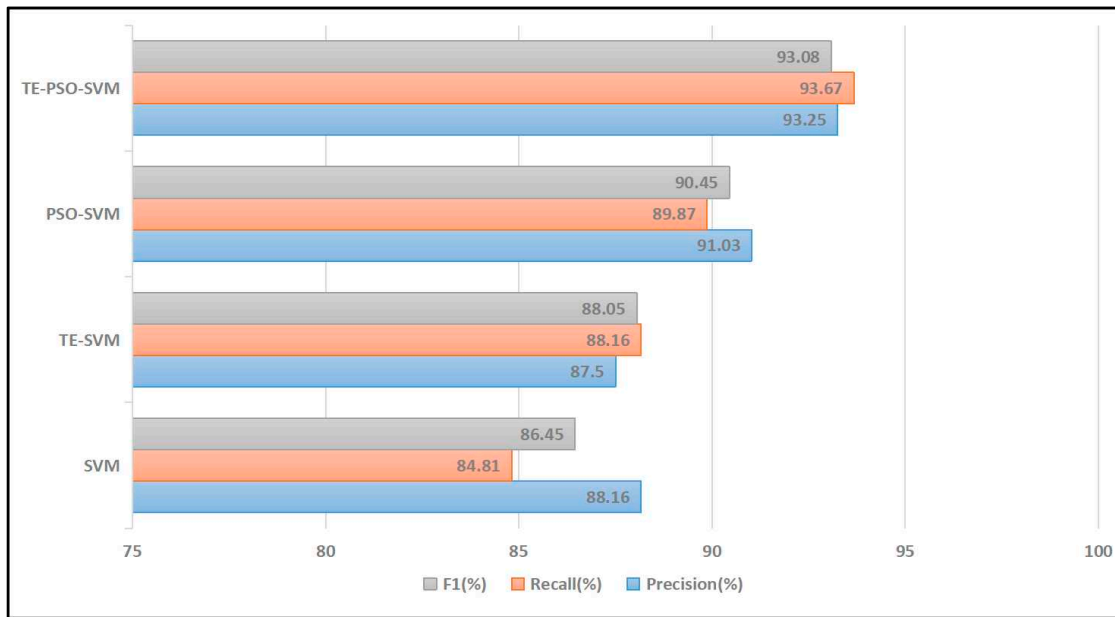


Figure 11. Performance comparison of different algorithms in 0.4 Gaussian Noise.

In addition, whether in the detection of three-dimensional data or six-dimensional ICV data, the SVM algorithm optimized by the PSO algorithm shows high performance in different environments. In summary, through these experimental results, this paper concludes that the data validity algorithm combining driver style and traffic flow theory has more advantages than traditional algorithms and verifies the significant effect of the PSO algorithm in optimizing the SVM algorithm.

6.5.2. Model Selection Data Validity Detection Experiment Based on Reinforcement Learning

In this experiment, the effectiveness of six-dimensional ICV data (v, a, d, e, k, fin\_v) was verified under different Gaussian Noises. According to the parameters in Table 2, the reward function of the model selection algorithm based on reinforcement learning (RLBMS) was set, and the four algorithms in the model pool were compared with RLBMS. The experimental results are shown in Figures 12 and 13.

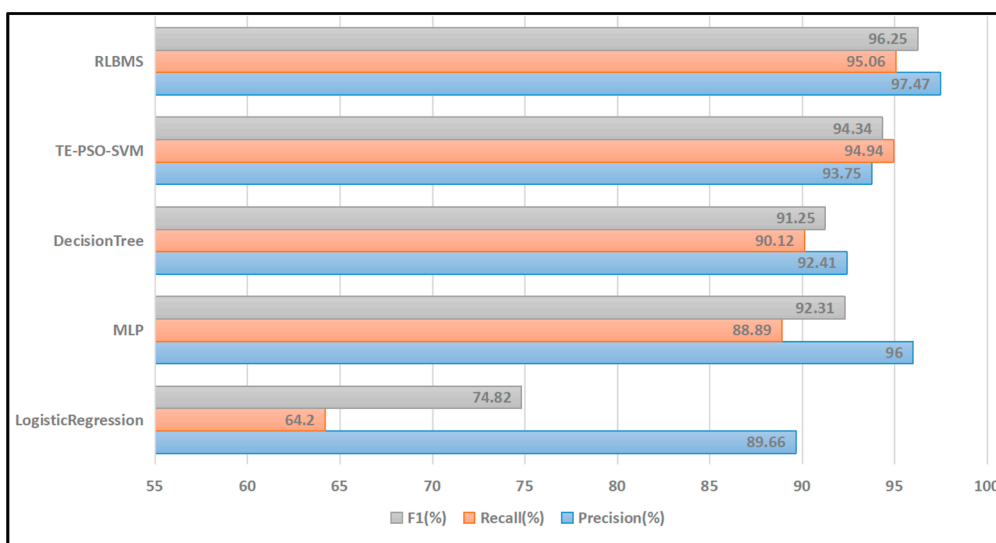
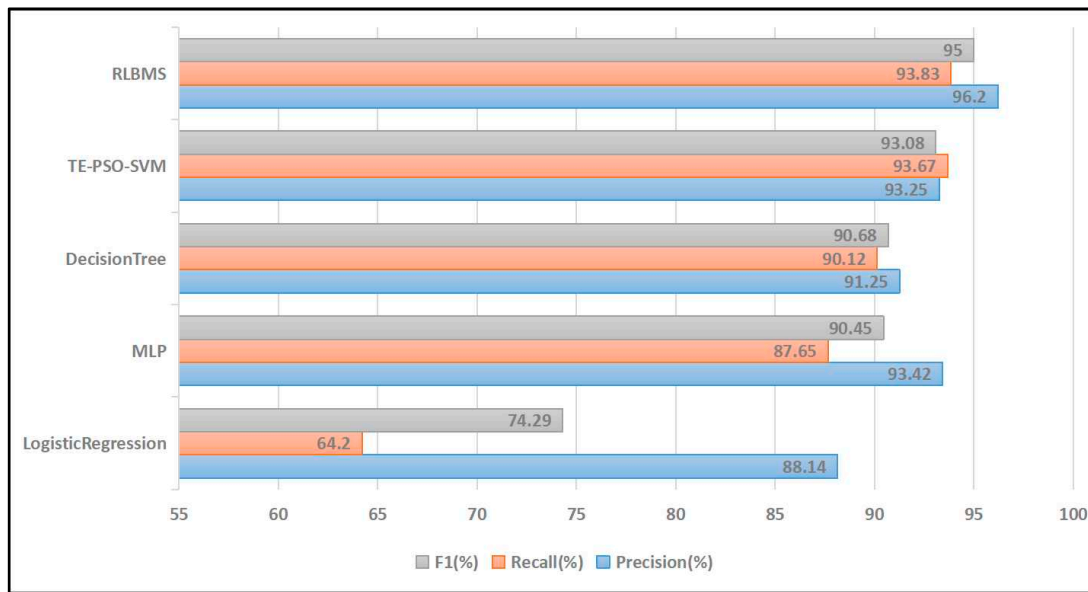


Figure 12. Performance comparison between basic model and RLBMS under 0.3 Gaussian Noise.





**Figure 13.** Performance comparison between basic model and RLBMS under 0.4 Gaussian Noise.

In order to more fully capture the characteristics of the data and take advantage of different model structures, this experiment constructs a model pool and adds four models (LogisticRegression, DecisionTree, PSO-SVM, and MPL algorithm) to the anomaly detection model pool of reinforcement learning as the basic model. Through the verification of the algorithm, the results of Figures 12 and 13 are obtained. From these two figures, it can be concluded that, under different environmental conditions, the Precision, Recall, and F1 of the proposed algorithm model are further improved compared with the basic model in the model pool.

These findings demonstrate the superiority of reinforcement-learning-based model selection algorithms in improving anomaly detection performance. The proposed algorithm can more intelligently select the best model that adapts to the current environment and task, thus further enhancing the accuracy of anomaly detection. By comprehensively utilizing different basic models, this method effectively exerts their respective advantages and makes it possible to obtain superior detection results in different environments.

In addition, in order to further explore the influence of distance threshold confidence (D\_T\_C) and predictive consensus confidence (P\_C\_C) on the classification performance of RLBMS algorithm, this paper also conducts relevant experiments and evaluates their contributions to the accuracy and robustness of anomaly detection. In this paper, the following experiments are carried out under different Gaussian Noise levels (0.3 and 0.4): the effect of adding two confidence levels (D\_T\_C, P\_C\_C) and not adding two confidence levels on the performance of the algorithm. The experimental results are shown in Figures 14 and 15.

Through the analysis of the results of Figures 14 and 15, the following conclusions can be drawn: the distance threshold confidence and the prediction consensus confidence have a positive impact on the accuracy and robustness of the algorithm and can improve the performance of the algorithm. When using these two confidence scores at the same time, the algorithm shows better performance under different noise levels.

Therefore, the comprehensive consideration of distance threshold confidence and prediction consensus confidence can significantly improve the performance of the algorithm, make it more suitable for different noise environment data, and achieve better detection results. These results are of great significance for further optimizing the RLBMS algorithm and improving its practical application value.

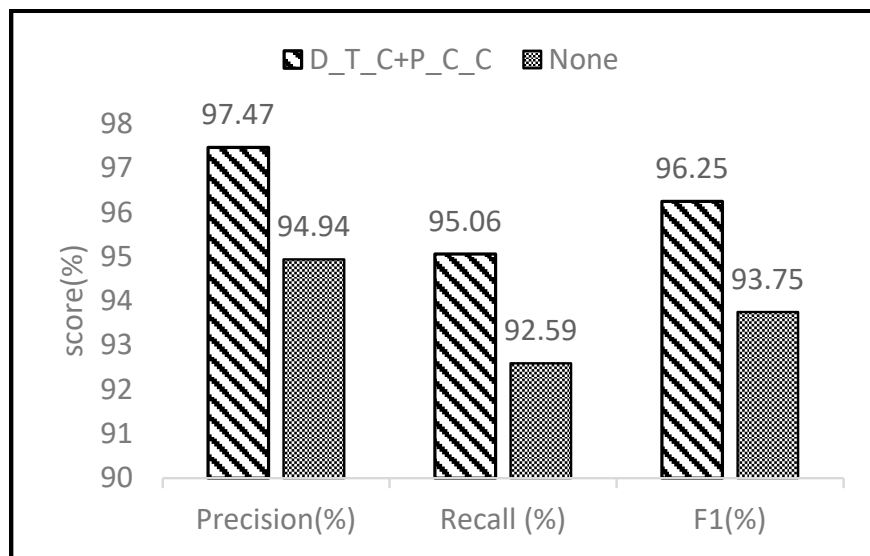


Figure 14. Under Gaussian Noise (0, 0.3), the influence of two confidence degrees on the algorithm.

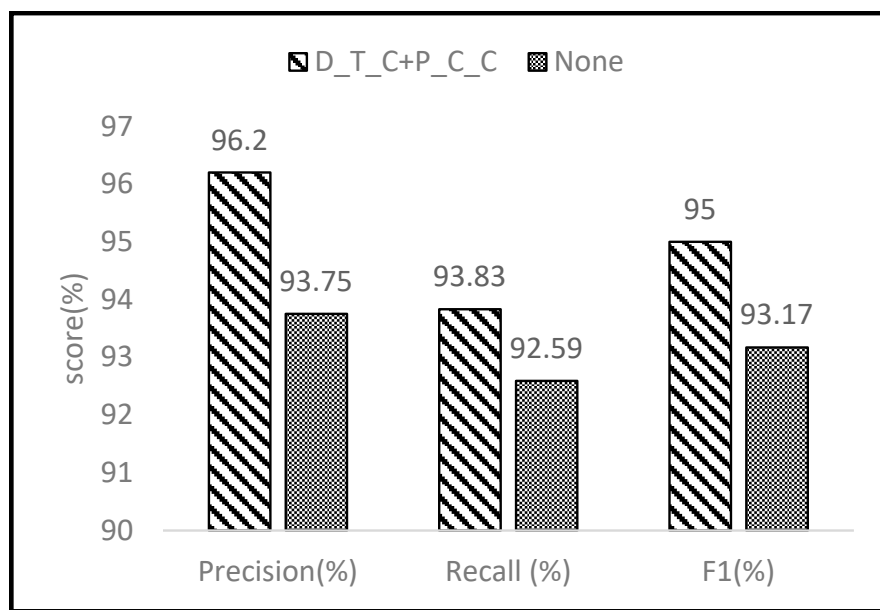


Figure 15. Under Gaussian Noise (0, 0.4), the influence of two confidence degrees on the algorithm.

### 7. Conclusions

In order to improve the accuracy of the validity detection of ICV data, this paper defines the driving style recognition coefficient and designs the quantitative model of driving style, a traffic flow model is established to predict the vehicle speed, and a data validity checking algorithm based on particle swarm optimization is proposed. Through the data validity test experiment based on driving style and traffic flow theory, it can be concluded that many factors should be considered when evaluating the validity of ICV data, such as the influence of the driver’s state and the traffic flow theory. Incorporating these key factors into data analysis methods can help to improve the accuracy and reliability of data analysis.

Due to the diversity of ICV data, the detection accuracy of a single model is still limited in scenarios where multiple types of anomalies coexist. Therefore, this paper uses multiple algorithm models with different structures to construct a model pool to capture data features. According to the characteristics of different algorithms in the model pool,

the distance threshold confidence and the prediction consensus confidence are introduced to improve the performance of the algorithm. Finally, according to the above methods, a model selection algorithm based on reinforcement learning is proposed, which can flexibly adapt to different forms of anomalies. Through the effectiveness experiment of model selection data based on reinforcement learning, this paper concludes that the model pool can more fully capture the data characteristics, facilitate the classifier to understand and analyze the data, and dynamically select the classifier, so as to make a more accurate judgment. In addition, it has also been proven that the D\_T\_C and the P\_C\_C have a positive impact on the accuracy and robustness of the algorithm, which can improve the performance of the algorithm.

In the current research, driving style and traffic flow characteristics have been combined for data analysis, but ICV systems also contain a large amount of data, such as voice data, road data, and driver image data. How to further combine multivariate data for analysis, solve the traffic congestion and traffic accidents caused by data anomalies, and ensure the effectiveness of ICV data is still a big challenge. In view of the diversity of anomalies, this study proposes a model pool to better capture data characteristics and improve the effectiveness of data analysis. In future work, we will increase the parameters of the state variables in the reinforcement learning algorithm and improve the algorithms in the model pool.

**Author Contributions:** Conceptualization, J.G. and N.D.; Methodology, J.G. and N.D.; Software, J.G. and N.D.; Supervision, C.H. and L.W.; Validation, J.G. and C.H.; Visualization, L.W.; Writing—original draft, J.G. and N.D.; Writing—review and editing, N.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported by Xinjiang Uygur Autonomous Region Natural Science Foundation Project, grant number (2021D01E20) and The National Natural Science Foundation of China, grant number (62072071) (62262066).

**Data Availability Statement:** Research data are available at: <https://www.highd-dataset.com> (accessed on 1 September 2023).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Wang, B.; Han, Y.; Wang, S.; Tian, D.; Cai, M.; Liu, M.; Wang, L. A Review of Intelligent Connected Vehicle Cooperative Driving Development. *Mathematics* **2022**, *10*, 3635. [CrossRef]
2. Dibaei, M.; Zheng, X.; Jiang, K.; Maric, S.; Abbas, R.; Liu, S.; Zhang, Y.; Deng, Y.; Wen, S.; Zhang, J. An overview of attacks and defences on intelligent connected vehicles. *arXiv* **2019**, arXiv:1907.07455.
3. Zheng, K.; Zheng, Q.; Chatzimisios, P.; Xiang, W.; Zhou, Y. Heterogeneous vehicular networking: A survey on architecture, challenges, and solutions. *IEEE Commun. Surv. Tutor.* **2015**, *17*, 2377–2396. [CrossRef]
4. Rajapaksha, S.; Kalutarage, H.; Al-Kadri, M.O.; Petrovski, A.; Madzudzo, G.; Cheah, M. Ai-based intrusion detection systems for in-vehicle networks: A survey. *ACM Comput. Surv.* **2023**, *55*, 1–40. [CrossRef]
5. Tesla's Keyless Access System Can Be Hacked and Driven Away in 10 Seconds. Available online: <https://finance.sina.com.cn/tech/2022-05-17/doc-imcwipik0280338.shtml> (accessed on 3 November 2023).
6. 'hacks' 25 Teslas in One Go! 10 Years Old Writing Code, Not Taking Classes, Starting a Company. Available online: <https://new.qq.com/rain/a/20220113A04NVG00> (accessed on 8 November 2023).
7. Laharotte, P.-A.; Billot, R.; Come, E.; Oukhellou, L.; Nantes, A.; El Faouzi, N.-E. Spatiotemporal analysis of bluetooth data: Application to a large urban network. *IEEE Trans. Intell. Transp. Syst.* **2014**, *16*, 1439–1448. [CrossRef]
8. Tan, H.; Feng, G.; Feng, J.; Wang, W.; Zhang, Y.-J.; Li, F. A tensor-based method for missing traffic data completion. *Transp. Res. Part C* **2013**, *28*, 15–27. [CrossRef]
9. Zhang, T.; Zhang, D.-G.; Yan, H.-R.; Qiu, J.-N.; Gao, J.-X. A new method of data missing estimation with FNN-based tensor heterogeneous ensemble learning for internet of vehicle. *Neurocomputing* **2021**, *420*, 98–110. [CrossRef]
10. Wang, Y.; Masoud, N.; Khojandi, A. Real-time sensor anomaly detection and recovery in connected automated vehicle sensors. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 1411–1421. [CrossRef]
11. Ding, N.; Ma, H.; Zhao, C.; Ma, Y.; Ge, H. Driver's emotional state-based data anomaly detection for vehicular ad hoc networks. In Proceedings of the 2019 IEEE International Conference on Smart Internet of Things (SmartIoT), Tianjin, China, 9–11 August 2019; pp. 121–126.

12. Djenouri, Y.; Belhadi, A.; Lin, J.C.-W.; Cano, A. Adapted k-nearest neighbors for detecting anomalies on spatio-temporal traffic flow. *IEEE Access* **2019**, *7*, 10015–10027. [[CrossRef](#)]
13. Wei, P.; Wang, B.; Dai, X.; Li, L.; He, F. A novel intrusion detection model for the CAN bus packet of in-vehicle network based on attention mechanism and autoencoder. *Digit. Commun. Netw.* **2023**, *9*, 14–21. [[CrossRef](#)]
14. Sharma, P.; Liu, H. A machine-learning-based data-centric misbehavior detection model for internet of vehicles. *IEEE Internet Things J.* **2020**, *8*, 4991–4999. [[CrossRef](#)]
15. Ranaweera, M.; Seneviratne, A.; Rey, D.; Saberi, M.; Dixit, V.V. Anomalous data detection in vehicular networks using traffic flow theory. In Proceedings of the 2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall), Honolulu, HI, USA, 22–25 September 2019; pp. 1–5.
16. Khodayari, A.; Ghaffari, A.; Kazemi, R.; Braunstingl, R. A.; Ghaffari, A.; Kazemi, R.; Braunstingl, R. A modified car-following model based on a neural network model of the human driver effects. In *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*; IEEE: Piscataway, NJ, USA, 2012; Volume 42, pp. 1440–1449.
17. He, T.; Zhang, L.; Kong, F.; Salekin, A. Exploring inherent sensor redundancy for automotive anomaly detection. In Proceedings of the 2020 57th ACM/IEEE Design Automation Conference (DAC), San Francisco, CA, USA, 20–24 July 2020; pp. 1–6.
18. Milardo, S.; Rathore, P.; Amorim, M.; Fugiglando, U.; Santi, P.; Ratti, C. Understanding drivers' stress and interactions with vehicle systems through naturalistic data analysis. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 14570–14581. [[CrossRef](#)]
19. Ni, D. *Traffic Flow Theory: Characteristics, Experimental Methods, and Numerical Techniques*; Butterworth-Heinemann: Oxford, UK, 2015.
20. Kadri, N.; Ellouze, A.; Ksantini, M.; Turki, S.H. New LSTM Deep Learning Algorithm for Driving Behavior Classification. *Cybern. Syst.* **2023**, *54*, 387–405. [[CrossRef](#)]
21. Turki, A.; Kahouli, O.; Albadran, S.; Ksantini, M.; Aloui, A.; Amara, M.B. A sophisticated Drowsiness Detection System via Deep Transfer Learning for real time scenarios. *AIMS Math.* **2024**, *9*, 3211–3234. [[CrossRef](#)]
22. Lin, X.; Huang, Y. Short-term high-speed traffic flow prediction based on ARIMA-GARCH-M model. *Wirel. Pers. Commun.* **2021**, *117*, 3421–3430. [[CrossRef](#)]
23. Gao, Y.; Zhou, C.; Rong, J.; Wang, Y.; Liu, S. Short-term traffic speed forecasting using a deep learning method based on multitemporal traffic flow volume. *IEEE Access* **2022**, *10*, 82384–82395. [[CrossRef](#)]
24. Xing, Z.; Liu, X.; Fang, R.; Zhang, H.; Liu, Z. Research on qualitative classification method of drivers' driving style. In Proceedings of the 2021 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA), Dalian, China, 27–28 August 2021; pp. 709–716.
25. Zylius, G. Investigation of route-independent aggressive and safe driving features obtained from accelerometer signals. *IEEE Intell. Transp. Syst. Mag.* **2017**, *9*, 103–113. [[CrossRef](#)]
26. Niyang, X.; Yunbing, Y.; Xuequan, T.; Lin, Z. Research on Driving Style Recognition Model Considering Different Traffic Density and Driving Behavior. In Proceedings of the 2022 6th CAA International Conference on Vehicular Control and Intelligence (CVCI), Nanjing, China, 28–30 October 2022; pp. 1–8.
27. Murphey, Y.L.; Milton, R.; Kiliaris, L. Driver's style classification using jerk analysis. In Proceedings of the 2009 IEEE Workshop on Computational Intelligence in Vehicles and Vehicular Systems, Nashville, TN, USA, 30 March–2 April 2009; pp. 23–28.
28. Greenshields, B.D.; Bibbins, J.; Channing, W.; Miller, H. A study of traffic capacity. In *Highway Research Board Proceedings*; National Research Council: Rockville, MD, USA, 1935; pp. 448–477.
29. Mazloumian, A.; Geroliminis, N.; Helbing, D. The spatial variability of vehicle densities as determinant of urban network capacity. *Philos. Trans. R. Soc. A* **2010**, *368*, 4627–4647. [[CrossRef](#)]
30. Hsueh, Y.-L.; Yang, Y.-R. A short-term traffic speed prediction model based on LSTM networks. *Int. J. Intell. Transp. Syst. Res.* **2021**, *19*, 510–524. [[CrossRef](#)]
31. Ben-Hur, A.; Horn, D.; Siegelmann, H.T.; Vapnik, V. A support vector clustering method. In Proceedings of the Proceedings 15th International Conference on Pattern Recognition. ICPR-2000, Barcelona, Spain, 3–7 September 2000; pp. 724–727.
32. Kennedy, J.; Eberhart, R. Particle swarm optimization. In Proceedings of the Proceedings of ICNN'95-International Conference on Neural Networks, Perth, WA, Australia, 27 November–1 December 1995; pp. 1942–1948.
33. Chandola, V.; Banerjee, A.; Kumar, V. Anomaly detection: A survey. *ACM Comput. Surv. CSUR* **2009**, *41*, 1–58. [[CrossRef](#)]
34. Wang, Q.; Yu, S.; Qi, X.; Hu, Y.; Zheng, W.; Shi, J.; Yao, H. Overview of logistic regression model analysis and application. *Zhonghua Yu Fang Yi Xue Za Zhi* **2019**, *53*, 955–960.
35. Murugesan, M.; Thilagamani, S. Efficient anomaly detection in surveillance videos based on multi layer perception recurrent neural network. *Microprocess. Microsyst.* **2020**, *79*, 103303. [[CrossRef](#)]
36. Charbuty, B.; Abdulazeez, A. Classification based on decision tree algorithm for machine learning. *J. Appl. Sci. Technol. Trends* **2021**, *2*, 20–28. [[CrossRef](#)]
37. Mohammadi, M.; Rashid, T.A.; Karim, S.H.T.; Aldalwie, A.H.M.; Tho, Q.T.; Bidaki, M.; Rahmani, A.M.; Hosseinzadeh, M. A comprehensive survey and taxonomy of the SVM-based intrusion detection systems. *J. Netw. Comput. Appl.* **2021**, *178*, 102983. [[CrossRef](#)]
38. Peng, B.; Li, X.; Gao, J.; Liu, J.; Chen, Y.-N.; Wong, K.-F. Adversarial advantage actor-critic model for task-completion dialogue policy learning. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 6149–6153.

39. Cui, J.; Chen, Z.; Tian, L.; Zhang, G. Overview of user and entity behavior analysis technology based on machine learning. *Comput. Eng.* **2022**, *48*, 10–24.
40. Krajewski, R.; Bock, J.; Kloeker, L.; Eckstein, L. The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 2118–2125.
41. Computer and Network Security. White Paper on Autonomous Driving Data Security. Available online: [https://m.sohu.com/a/453643278\\_653604](https://m.sohu.com/a/453643278_653604) (accessed on 12 November 2023).
42. Abbas, M.; Safar, M.; Salem, A. Anomaly detection system for altered signal values within the intra-vehicle network. In Proceedings of the 2020 15th Design & Technology of Integrated Systems in Nanoscale Era (DTIS), Marrakech, Morocco, 1–3 April 2020; pp. 1–6.
43. Ren, G.; Zhu, S. Parameter Calibration of Traffic Flow Speed-Density Model Based on K-means Clustering Algorithm and Least Square Method. In Proceedings of the 21st COTA International Conference of Transportation Professionals: Advanced Transportation, Enhanced Connection, CICTP 2021, Xi'an, China, 16–19 December 2021; pp. 66–76.

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