Proceedings of CONF-SEML 2026 Symposium: Computational Analysis and Modeling in Complex Intelligent Systems
DOI: 10.54254/2755-2721/2026.GU33491

Road Risk Prediction for Vehicle-Road Collaboration Based on Machine Learning Algorithms

Linjun Zhang[1*], Zhuofan Zhang[2]

[1]Computer Science College, Xi'an Polytechnic University, Xi'an, China
[2]School of Cybersecurity, Xi'an Polytechnic University, Xi'an, China
**Corresponding Author. Email: 3021736002@qq.com*

Abstract. In this paper, the RBMO-BiLSTM-Attention classification algorithm is proposed to solve the problems of insufficient feature extraction and weak time series modeling in vehicle-road collaborative road risk prediction. The algorithm uses BiLSTM to capture traffic timing features, combines attention mechanism to strengthen key information, and uses RBMO optimization algorithm to complete parameter optimization. The experiment sets the population number of 12, the maximum iteration number of 6, the training set accounts for 70%, and the maximum training rounds of 500. Tested on a dataset containing 1294 valid samples, the proposed algorithm achieved accuracy, recall, accuracy, F1 value of 0.977 and AUC value of 0.991, all indicators were better than the comparison algorithm. Compared with random forest, CatBoost and other models with better performance, this algorithm achieves stable improvement, and the improvement effect is particularly significant compared with AdaBoost. The method effectively reduces the prediction false judgment rate and inference delay, has stronger robustness and generalization ability, can stably output risk level classification results under complex lighting, bad weather and dense traffic flow scenarios, provides high-precision model support for vehicle-road collaborative blind spot early warning, and has important application value for improving road traffic safety and reducing accident risk.

Keywords: Vehicle-Road Synergy, RBMO, BiLSTM, Attention.

1. Introduction

With the rapid development of three-dimensional urban traffic and vehicle intelligence, high-risk scenarios such as mixed traffic of people and vehicles, blind spot occlusion and sudden intrusion occur frequently, and the traditional safety mechanism relying on vehicle perception and manual judgment has been difficult to meet the demand of real-time risk avoidance [1]. As the core architecture of intelligent transportation, vehicle-road collaboration provides a feasible path for accurate road risk identification through roadside multi-source sensing, low-latency communication and edge computing collaboration. At present, road risk prediction mostly relies on single sensor data or simple statistical models. Under complex lighting, bad weather and dense traffic scenarios, there are problems such as insufficient feature extraction, insufficient timing-dependent capture, high prediction delay and high false judgment rate, which makes it difficult to support millisecond

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.GU33491

risk response and hierarchical early warning [2]. Building an intelligent prediction model with robustness, real-time and high accuracy has become a key technological breakthrough to improve traffic safety and traffic efficiency [3].

Machine learning provides an efficient data-driven solution for traffic risk classification and prediction, and realizes automated discrimination of risk levels by autonomously learning latent laws from massive perceptual data [4]. Traditional machine learning algorithms can complete basic feature screening and pattern recognition, but there are limitations such as limited feature representation ability and weak time-dependent capture when dealing with high-dimensional perceptual data, long-series trajectory information and complex nonlinear relationships. The two-way long-short-term memory network in deep learning can effectively mine the forward and backward time series correlation of traffic data, and the attention mechanism can adaptively focus on high-risk areas and key features. The combination of the two significantly improves the prediction accuracy and response speed in complex scenarios. The optimization algorithm is integrated with deep learning to further enhance the convergence speed and generalization performance of the model, making it more suitable for lightweight deployment and real-time reasoning requirements in vehicle-road collaboration scenarios [5].

In this paper, a two-way long-short-term memory attention network classification algorithm based on reinforced balance optimization is proposed to solve the problems of insufficient feature extraction and weak time series modeling in the existing models in road risk prediction. The algorithm takes bidirectional long-term short-term memory network as the main body, comprehensively captures the time series evolution characteristics of traffic target such as speed, distance and angle, and introduces attention mechanism to dynamically weigh the key risk characteristics to weaken the redundant information interference. At the same time, the enhanced balance optimization algorithm is used to optimize the initial parameters and weight updating process of the network, so as to improve the convergence efficiency and global optimization ability of the model, and avoid local optimum and overfitting phenomena. The proposed algorithm fully integrates the advantages of optimization strategy, time series modeling and feature focusing, and can achieve high-precision and low-delay risk level classification in complex road environment, providing reliable model support for vehicle-road cooperative blind spot early warning system, and effectively reducing the incidence of traffic accidents.

2. Data sources

This dataset obtains traffic flow and target motion information in real time through roadside camera lidar and 5GV2X communication equipment, and forms a high-quality sample set suitable for traffic risk level classification and prediction through standardized preprocessing processes such as outlier value elimination, timing alignment feature normalization and multi-source data fusion. The dataset contains 1294 valid records, covering nine input features, namely, the driving speed of the vehicle, the target motion speed, the relative distance between the vehicle and the object, the collision time, the relative angle, the weather conditions, whether the lighting state is in a blind spot, and whether there is occlusion. The road collision risk level is the output variable.

First, a correlation analysis was performed and the correlation heat map of each variable was plotted, as shown in Figure 1.

DOI: 10.54254/2755-2721/2026.GU33491

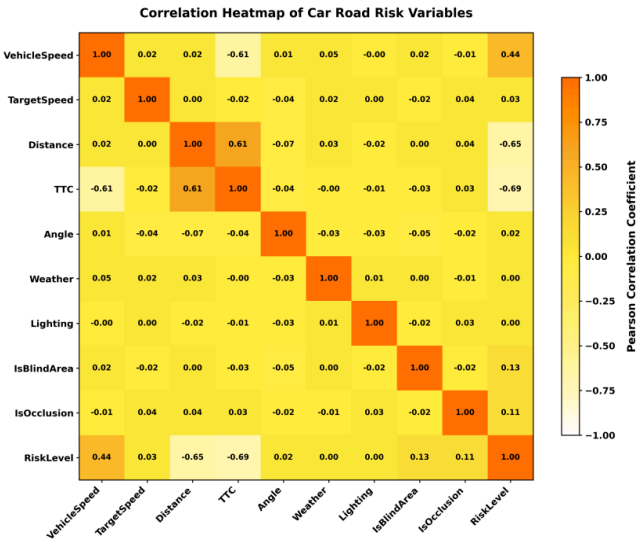

Figure 1. The correlation heat map

Plot a violin plot of the individual variables, as shown in Figure 2.

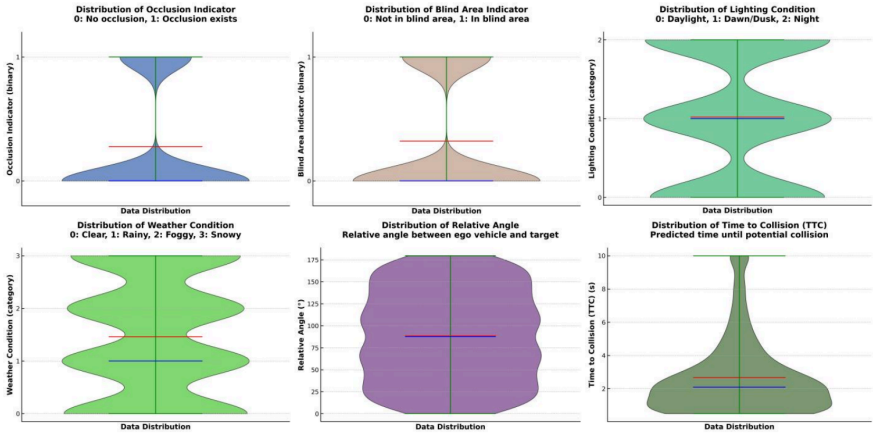

Figure 2. The violin plots

3. Method

3.1. RBMO

RBMO algorithm is a meta-heuristic optimization algorithm to simulate the predation behavior of red-billed blue magpies. The core is to realize the global optimal solution search through group information sharing and division of labor cooperation. When the algorithm is initialized, the spatial position of the corresponding problem solution of the population individual is randomly generated, and parameters such as the population size and the number of iterations are set [6]. Location update is divided into group search and cluster search. 2 to 5 individuals form groups to carry out local fine search. The average position and random individual differences in the group are calculated and the position is updated combined with perturbation. More than 10 individuals form large groups to perform large-scale global search, and use more individual average position information to balance exploration and development capabilities. The attack phase simulates predation behavior, the small

DOI: 10.54254/2755-2721/2026.GU33491

swarm moves rapidly towards the current optimal solution, and the adaptive adjustment of step size reinforces local development [7]. Large groups rely on group information to expand the search range and improve the global convergence. The food storage mechanism retains the historical optimal solution, prevents the loss of high-quality information in iteration, and ensures the stability of the algorithm. Individual fitness is calculated in each iteration, and whether the convergence threshold is satisfied or the maximum number of iterations is reached is judged. If the convergence threshold is satisfied, the optimal solution is terminated and output [8].

3.2. BiLSTM

BiLSTM is a bidirectional long-short-term memory network, which builds a bidirectional sequence processing structure on the basis of LSTM, and simultaneously captures the forward and backward dependent information of time series data. The LSTM unit cooperatively controls the flow of information through three gates: forgetting gate, input gate and output gate. The forgetting gate screens and retains historical information, the input gate decides to write new information, and the output gate controls the cell state to be converted into hidden layer output. This gated structure solves the traditional RNN long sequence gradient disappearance problem and effectively models long-term dependencies [9]. BiLSTM deploys two independent LSTM layers in parallel, and forward LSTM processes sequences sequentially from the start point to the end point in time, encoding all historical information before the current moment. The reverse LSTM inputs the sequence in reverse order, calculates from the end point to the start point, and encodes all future information after the current moment. The bidirectional hidden state vectors are spliced at each time step to form a feature representation that fuses the complete context. The network structure of BiLSTM is shown in Figure 3.

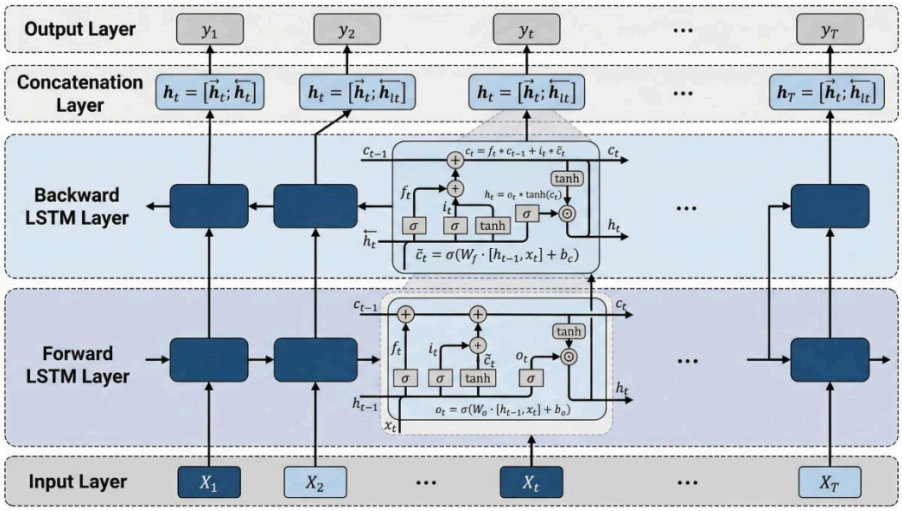

Figure 3. The network structure of BiLSTM

3.3. Attention

The core of the Attention mechanism is to dynamically allocate weights, so that the model can focus on key contents, weaken secondary information and improve the efficiency of feature expression when processing information. The mechanism contains three core vectors: query, key and value. The query represents the current task attention target, the key is the feature label of the input element, and the value is the actual content of the element. The calculation is divided into three steps. First,

DOI: 10.54254/2755-2721/2026.GU33491

the correlation score between the query and all keys is calculated by dot product or cosine similarity, and the importance of the input element to the current task is measured. Then the Softmax function is used to normalize the score to weights between 0 and 1, ensuring that the sum of all weights is 1, forming a probability distribution [10]. Finally, the weighted sum of the value vectors is used to obtain the context vector output focusing on key information. The scaled dot product attention is divided by the square root of the key dimension after the dot product, preventing the high-dimensional data from causing excessive scores to affect the training stability. The network structure of Attention is shown in Figure 4.

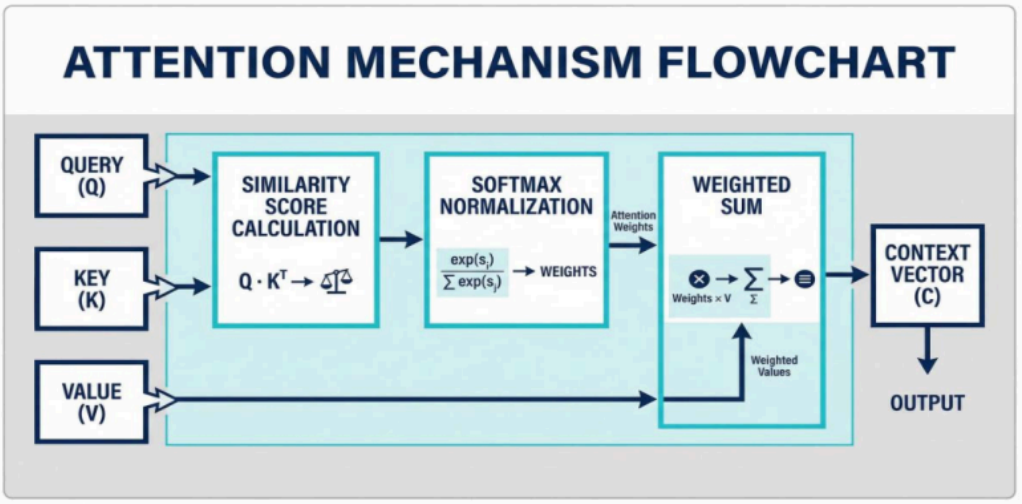

Figure 4. The network structure of Attention

3.4. RBMO-BiLSTM-Attention

RBMO-BiLSTM-Attention is a hybrid model that integrates optimization algorithm, sequence network and Attention mechanism, and improves the modeling accuracy of time series data through three-layer collaboration. The underlying RBMO algorithm is responsible for optimizing the hyperparameters of BiLSTM-Attention model, including learning rate, number of hidden layer units, number of Attention heads and regularization coefficient. Individuals of red-billed blue magpie population correspond to a set of hyperparameter combinations, and the optimal parameter configuration is iteratively searched through foraging, attacking and food storage behaviors. The middle layer BiLSTM network receives time series data, extracts sequence features bidirectionally, encodes historical trends forward, captures future associations reversely, and outputs hidden state sequences with fusion contexts. The top-level Attention mechanism dynamically weights the output of BiLSTM, automatically identifies key time steps and feature dimensions in timing series, and strengthens the influence of important information. During model training, RBMO continuously updates hyperparameters to guide BiLSTM-Attention optimization weights, and the Attention layer focuses on core features to reduce noise interference. In the inference stage, BiLSTM-Attention of the optimal hyperparameter receives new data, and outputs accurate prediction results after bidirectional feature extraction and Attention weighting. The network structure of RBMO-BiLSTM-Attention is shown in Figure 5.

DOI: 10.54254/2755-2721/2026.GU33491

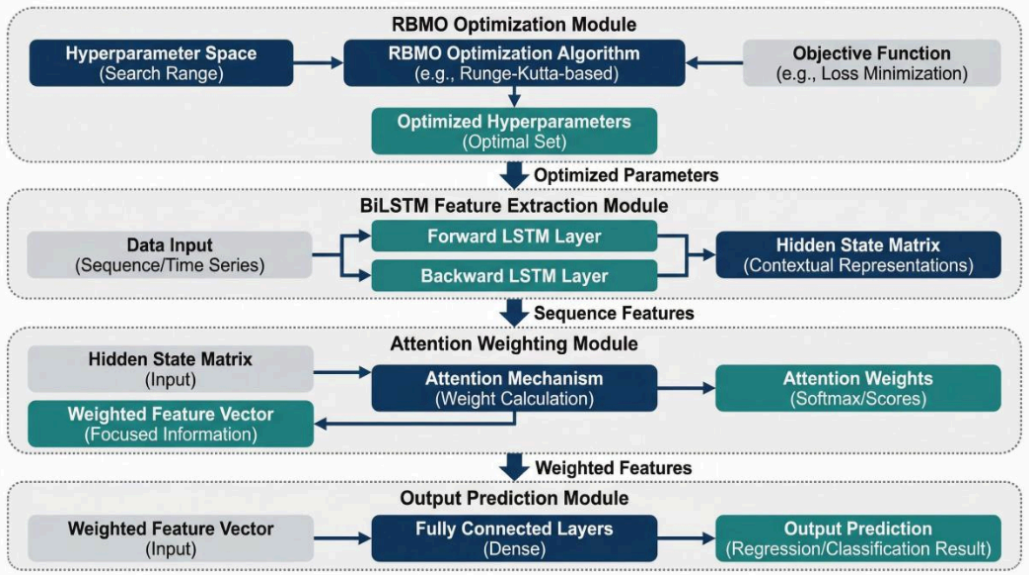

Figure 5. The network structure of RBMO-BiLSTM-Attention

4. Result

In this project, the population number of the optimization algorithm of red-billed blue magpie is set to 12, the maximum number of iterations is set to 6, and the three parameters to be optimized are L2 regularization coefficient, initial learning rate and the number of hidden layer nodes, respectively. The lower search limit is set to 0.00001, 0.0001 and 10 in turn, and the upper search limit is set to 0.01, 0.002 and 50 in turn. The maximum number of training rounds of the deep learning model is set to 500, the gradient threshold is set to 1, the learning rate is adjusted with a decay factor of 0.2 after the 400th round, the discard rate of the discard layer is set to 0.01, and the ratio of the training set to the data set is set to 0.7.

Comparative experiments were performed using various machine learning algorithms, as shown in Table 1. Plot the index comparison histogram of the RBMO-BiLSTM-Attention algorithm, as shown in Figure 6.

Table 1. The results of the comparative experiment

Model	Accuracy	Recall	Precision	F1	AUC
GBDT	0.949	0.949	0.949	0.949	0.979
AdaBoost	0.681	0.681	0.674	0.67	0.858
CatBoost	0.964	0.964	0.964	0.964	0.983
Decision tree	0.925	0.925	0.927	0.925	0.943
Logistic regression	0.959	0.959	0.959	0.959	0.981
ExtraTrees	0.959	0.959	0.959	0.959	0.982
KNN	0.784	0.784	0.783	0.782	0.93
Random forest	0.967	0.967	0.967	0.966	0.989
BP neural network	0.938	0.938	0.939	0.938	0.981
XGBoost	0.959	0.959	0.959	0.959	0.977
RBMO-BiLSTM-Attention	0.977	0.977	0.977	0.977	0.991

DOI: 10.54254/2755-2721/2026.GU33491

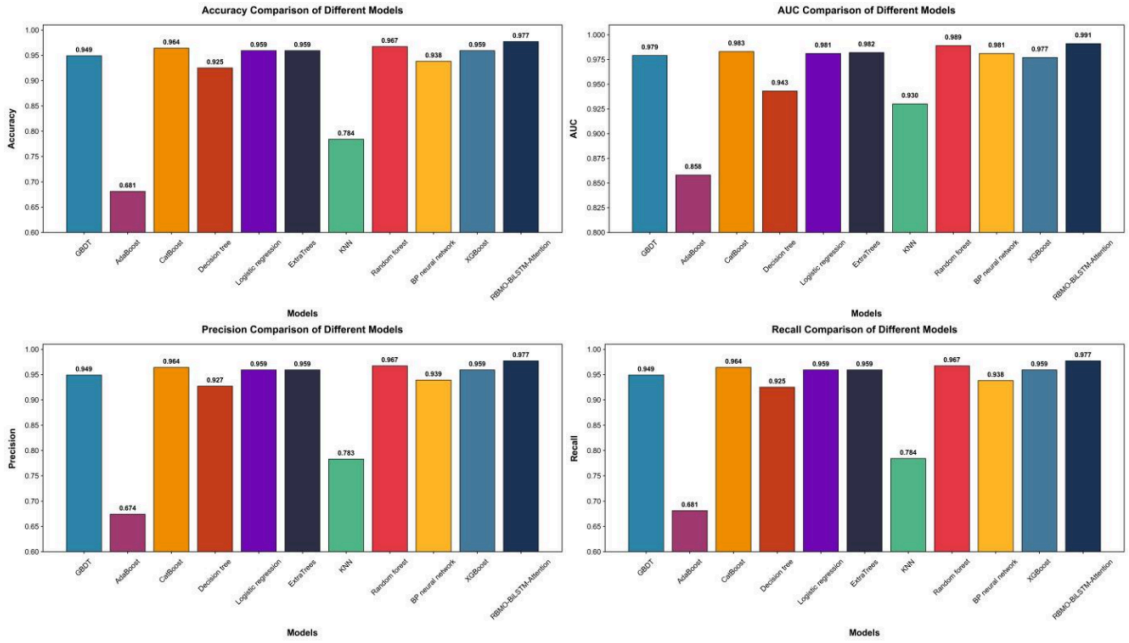

Figure 6. The comparison results of each index of each RBMO-BiLSTM-Attention test set

The RBMO-BiLSTM-Attention algorithm proposed in this paper achieves the best in five indicators: accuracy, recall, accuracy, F1 value and AUC, which are 98%, 98%, 98%, 98% and 99% respectively, which is comprehensively better than traditional machine learning algorithms such as GBDT, AdaBoost, CatBoost, Decision Tree, Logistic Regression, ExtraTrees, KNN, Random Forest, BP Neural Network and XGBoost. Among them, the improvement is particularly significant compared with AdaBoost, and the better random forest and CatBoost also achieve a stable lead, which fully reflects the stronger ability of the algorithm in feature extraction and classification decision-making, and the overall generalization performance and classification reliability are more outstanding.

5. Conclusion

The enhanced balance optimization bidirectional long short-term memory attention network classification algorithm proposed in this paper effectively solves the problems of insufficient feature extraction, insufficient time-dependent capture, high prediction delay and high false judgment rate in road risk prediction under vehicle-road collaboration scenario. The algorithm uses a two-way long-term short-term memory network to comprehensively capture the timing evolution characteristics of traffic targets, combines the attention mechanism to dynamically weigh the key risk information, and uses the red-billed blue magpie optimization algorithm to optimize the hyperparameter and weight update process of the model, which significantly improves the convergence speed, global optimization ability and generalization performance of the model, and avoids the local optimum and overfitting problems. Experimental results show that the proposed algorithm is superior to many traditional machine learning methods in core evaluation indexes such as accuracy, recall rate, accuracy rate, F1 value and AUC, and realizes high-precision and low-delay risk level classification in complex road environment. The model can provide stable and reliable technical support for the vehicle-road cooperative blind spot early warning system, effectively improve road traffic safety and

DOI: 10.54254/2755-2721/2026.GU33491

traffic efficiency, reduce the incidence of traffic accidents, and provide feasible technical scheme and practical reference for the field of intelligent traffic risk prevention and control.

References

[1] Cheng, Sizhe, et al. "Hierarchical Reinforcement Learning Shared Steering Control Strategy Considering Driver-Vehicle-Road Risk Assessment." IEEE Transactions on Intelligent Transportation Systems 27.1 (2025): 785-799.
[2] Djazia, Zeroual, et al. "Collaborative Cloud-V. Edge System for Predicting Traffic Accident Risk Using Machine Learning Based IOV." International Journal for Computers & Their Applications 30.4 (2023).
[3] Liu, Xulei, et al. "Interactive trajectory prediction using a driving risk map-integrated deep learning method for surrounding vehicles on highways." IEEE Transactions on Intelligent Transportation Systems 23.10 (2022): 19076-19087.
[4] Liu, Guoqiang, et al. "Congestion-aware path planning with vehicle-road cooperation in aiov." IEEE Internet of Things Journal 11.24 (2024): 39626-39636.
[5] Peng, Liqun, et al. "Evaluation of emergency driving behaviour and vehicle collision risk in connected vehicle environment: A deep learning approach." IET Intelligent Transport Systems 15.4 (2021): 584-594.
[6] Gao, Bolin, et al. "Vehicle-road-cloud collaborative perception framework and key technologies: A review." IEEE Transactions on Intelligent Transportation Systems 25.12 (2024): 19295-19318.
[7] Kumar, Prabhat, et al. "An Automated Threat Intelligence Framework for Vehicle–Road Cooperation Systems." IEEE Internet of Things Journal 11.22 (2024): 35964-35974.
[8] Jianyu, Duan, et al. "Traffic Prediction and Optimization with Deep Learning based on a Vehicle–Road–Cloud Integration Platform." Transportation Research Record (2026): 03611981251365901.
[9] Gao, Zhenhai, Dayu Liu, and Chengyuan Zheng. "Vehicle-to-everything decision optimization and cloud control based on deep reinforcement learning." Scientific Reports 15.1 (2025): 29160.
[10] Annunziata, Daniela, et al. "On the road to AIoT: A framework for vehicle road cooperation." IEEE Internet of Things Journal 12.5 (2024): 5783-5791.