IRF-HTID-BO-LSTM: Classification Model of Curve Shape Index for Mountainous Highways and Intelligent Traffic Incident Detection Method

Xun Gu[1*], Shuai Dai[2]
[1]Graduate School, People's Public Security University of China, Beijing 100038, China
[2]Policy Planning Research Office of road traffic safety research center of the Ministry of public security, Beijing 100038, China
E-mail: 201621350038@stu.ppsuc.edu.cn; blue80520@163.com
*Corresponding author

Keywords: highway structure, indicator grading; traffic incident detection; long short-term memory network; characteristic variables

Received: August 30, 2024

In response to the special geographical environment and traffic conditions of mountainous highways, reasonable highway structure design can significantly improve traffic safety and reduce traffic accidents. Therefore, a grading model and traffic incident detection method for mountainous highway curve line indicators are developed. By analyzing the traffic conditions and highway structure of mountainous highways, a classification algorithm based on highway curve structure indicators is proposed, and a mountainous highway curve structure grading model is constructed. Then, a long short-term memory network is introduced to design a highway traffic incident detection algorithm on the basis of Bayesian optimization. The results showed that the correlation fitting degree of curve index classification based on the classification model was 87.3%. With the increase of feature variables in the data set, the classification accuracy of the traffic incident detection method for different events showed a steady increase and reached a stable state of 92.8%. The accuracy of the most advanced method was only 90%, and the accuracy of the research method was higher than that of the most advanced method. The comprehensive performance showed that the area under the curve value of the proposed method was as high as 0.982, which was larger than other comparison algorithms. In addition, the area under the curve value of the most advanced method was 0.962. The above results demonstrate that the designed algorithm has good performance, which can effectively segment the curve shape indicators of highway structures, and accurately detect traffic incidents.

Povzetek: Opisan je model za klasifikacijo oblike krivulje gorskih avtocest in inteligentna metoda za zaznavanje prometnih nesreč (IRF-HTID-BO-LSTM). Pristop izboljšuje zaznavanje prometnih nesreč na gorskih cestah.

1 Introduction

The modernization of transportation is an important symbol of national modernization. For a long time, under the leadership of the Party, China's transportation has achieved remarkable achievements around the center and serving the overall situation [1]. China is accelerating the construction of transportation infrastructure. Focusing on the central task of the Party, China is striving to promote the high-quality development of transportation and construct the transportation power, guaranteeing Chinese path to modernization with modern transportation services with more Chinese characteristics, Chinese style, and Chinese style [2-4]. However, the traffic conditions of mountainous highways are constrained by the complex structure of mountainous highways. The highway structure of mountainous highways is closely related to traffic conditions, which poses challenges to highway design and precise management of traffic safety. A scientifically reasonable index structure of highway structure not only affects the safety, but also directly affects traffic smoothness and service efficiency [5-6]. Moreover, in the management of mountainous highways, the highways' structure is characterized by complex curved lines, frequent sharp turns, and continuous curved structures, which leads to more complex traffic conditions and traffic management work. In addition, according to statistics, there were 256,409 incident accidents in China in 2022, with an average of over 700 incidents per day and an average of 166 deaths per day. This has caused huge losses to people's property and life safety [7]. In addition, current traffic detection methods mainly focus on highways, rural highways, etc. Mountainous highways are difficult to achieve efficient and accurate detection and feedback due to complex highways conditions and poor signals. Therefore, the main research question is how to introduce intelligent traffic incident detection technology, which plays an important role in promoting intelligent management of transportation and ensuring the safety and stability of the transportation system. Meanwhile, how feature selection and Bayesian optimization LSTM can improve the accuracy of mountainous highways traffic incident detection. In response to the above issues, the study first designs a classification model for the curve shape indicators of mountainous highways, and then uses the Improved Random Forest Algorithm (IRF) for feature variable selection. An improved Long Short-Term Memory (LSTM) based on Bayesian Optimization (BO) algorithm is proposed. Based on the above content, the Intelligent Highway Traffic Incident Detection Algorithm (IRF-HTID-BO-LSTM) is designed. The research objective is to design a classification model for mountainous highway curve indicators and an intelligent traffic incident detection method. By exploring the structural constraints of mountainous highways, it is expected to improve real-time traffic incident detection capabilities, avoid secondary traffic incidents, ensure smooth and safe operation of highways, and enhance the intelligent management level of mountainous highways. The innovation of research mainly includes the following two aspects. Firstly, the model structure optimization method and highway curve level grading method for the design of flat curve length grading for mountainous highways are used to provide stronger support for traffic management decisions. In addition, the BO is introduced to optimize the hyper-parameters of the LSTM model, and the mixed sampling technology is used to reconstruct the imbalanced traffic data set. By constructing an initial variable set, more sensitive features to traffic incidents are determined to ensure the safety and stability of mountain highways, promoting the intelligence of traffic management.

2 Related works

With the development goal of building a comprehensive transportation power proposed, mountainous highway transportation has experienced rapid development. However, the complex driving environment of mountainous highways places higher demands on vehicle performance, driving skills, and attention than non mountainous highways, resulting in a high incidence of traffic accidents on mountainous highways and particularly prominent traffic safety accidents. Numerous scholars have conducted in-depth analysis and exploration on this matter. S. Cafiso et al. used warning signs to alert drivers to external changes in flat direction and speed to improve cornering safety. A unified curve standard on a two-lane road was built, allowing drivers to adjust their speed based on actual wind speed. The relative changes in collision rates of various risk categories were analyzed, and the factors affecting collisions were estimated [8]. G. Ashley et al. found that traffic incidents cause billions of dollars in losses to the United States every year. Therefore, the study utilized machine learning for collision analysis to identify driver, vehicle, and road related factors that affect driving risks in various location types. The research results showed that drivers who performed visual tasks at uncontrolled intersections were 2.7 times more likely to have a collision than drivers who did not perform the aforementioned tasks. The above findings further proved that establishing a safety awareness project for intersection safety was imperative [9]. M. R. Fatmi et al. developed a Logit model based on latent segmentation to analyze the severity of traffic collision injuries using collision data reported in Nova Scotia, Canada from 2007 to 2011. There was a segmentation of high-risk and low-risk damage severity. Moreover, high-risk road sections generated higher levels of injury severity, while low-risk road sections generated lower levels of injury severity [10]. D. E. Monyo et al. found that in areas with complex road features and frequent traffic conflicts, older drivers had an increased risk of making mistakes. Overpass is a highway location that presents more driving challenges than other basic road sections. Therefore, based on the traffic accident data from Florida from 2016 to 2018, this study used latent category clustering analysis and penalty logistic regression to explore the factors that affect older drivers' driving errors on interchanges. The results revealed that factors such as distracted driving, area type, and speed limit were all important in specific clusters [11].

To ensure the efficient and safe operation of mountainous highways, intelligent traffic incident detection methods are gradually being applied in the transportation field. However, current detection methods have shortcomings such as low efficiency, untimely feedback, and work intensity, which cannot adapt to complex mountainous highways. S. B. Li et al. developed an incident detection method on the basis of toll station data to ensure the smooth operation of highways, reduce traffic congestion, and avoid secondary accidents. A case study experiment was conducted on the highway network in Shandong Province. The method effectively detected highway incidents, dynamically evaluated the status of the transportation network, and provided suggestions for highway management departments [12]. X. Zhang et al. designed five methods for establishing and calculating traffic accident management measurements to manage highway accidents and reduce their impact. The research method could identify the advantages and disadvantages of accident management strategies and modify practices accordingly [13]. P. H. L. Rettore et al. designed a method for enriching highway data. Data from heterogeneous data sources was fused to enhance the service framework of intelligent transportation systems and improve the description of traffic conditions through location-based social media data. The traffic incident detection model achieved a score of over 90% [14]. M. Won et al. proposed an outlier analysis process to alleviate traffic congestion caused by traffic incidents and restore traffic system performance as safely and quickly as possible. This process was used to estimate the outliers of each detected event and utilized such outlier information to improve the prediction accuracy of incident duration. Through application examples, the research method improved the accuracy of estimating the duration of traffic incidents and detected potential system defects related to incident response, data recording, resource management, etc [15]. Following the above literature summary, Table 1 is compiled.

Table 1: Summary of literature results.

Author	Research method	Research results
S. Cafiso et al.	Constructing uniform curve standards on two-lane roads allows drivers to adjust their speeds to actual wind speeds	The system needs to be revised, both on actual risk classification and how it is managed
G. Ashley et al.	Using machine learning for crash analysis to identify driver, vehicle, and road-related factors that affect the risk of driving in different locations and then analyzing the most important factors derived from the machine learning analysis	Drivers who perform visual manual tasks at uncontrolled intersections were 2.7 times more likely to be involved in a crash than drivers who do not perform the above tasks
M. R. Fatmi et al.	A Logit model based on potential segmentation was developed to analyze the severity of traffic collision injuries	There are segments of high risk and low risk injury severity, and the linear road alignment produces higher injury severity at high risk sections and lower injury severity at low risk sections
D. E. Monyo et al.	The factors influencing the driving error of elderly drivers on the interchange were investigated by potential category cluster analysis and penalty logistic regression	Variables that are significant in a particular cluster, and in factors such as distracted driving, area type, speed limit, etc. are important in all collisions versus a few specific clusters
S. B. Li et al.	An event detection method based on toll station data is proposed	Taking the expressway network of Shandong Province as an example, a numerical example test is carried out. This method can effectively detect highway accidents, dynamically estimate traffic network status, and provide suggestions for highway management departments
X. Zhang et al.	Establishment and calculation method of five traffic accident management measures corresponding to the establishment and improvement of traffic incidents by Kentucky Transportation Cabinet	The method can identify the strengths and weaknesses of accident management strategies and modify practices accordingly
P. H. L. Rettore et al.	A road data enrichment approach is proposed to enhance the framework of intelligent transportation system services by fusing data from heterogeneous data sources, and to improve the description of traffic conditions through location-based social media data	The traffic incident detection model of the study method obtained a score of more than 90%
M. Won et al.	An outlier analysis procedure is proposed to estimate the outlier for each detected event and use such outlier information to improve the predictive accuracy of event duration estimates	The research method can improve the accuracy of traffic incident duration estimation and detect potential system deficiencies related to incident response, data recording, and resource management

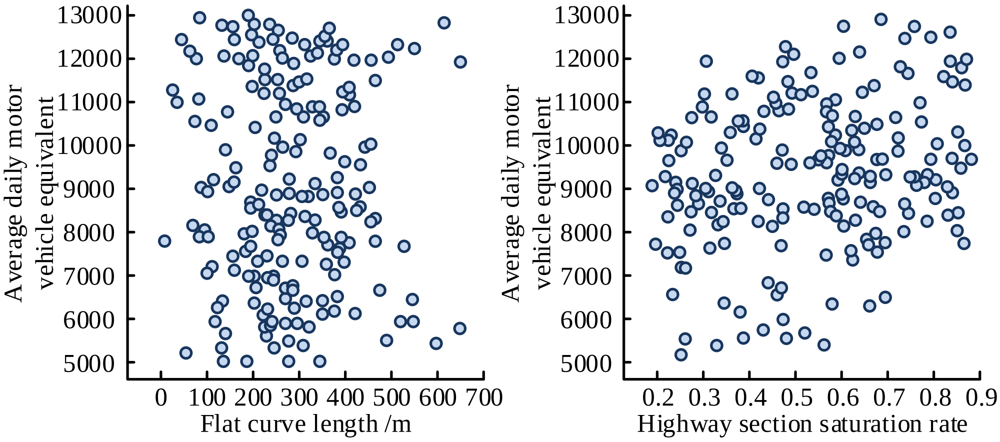
Based on the above content, the current research results mainly focus on the correlation between road traffic conditions and structure conditions and intelligent traffic incident detection methods. The SOTA method with the best performance is the intelligent traffic service framework integrating heterogeneous data sources, but it still has certain limitations, that is, it is difficult to accurately analyze the road structure and traffic safety conditions of mountain highways. Its subsequent management cannot get timely feedback, usually in traffic accidents, and it takes more time to deal with. Therefore, an index optimization method and a highway curve grading method for the flat curve structure grading of mountainous highways are developed, and an IRF-HTID-BO-LSTM method is designed to detect traffic incidents.

3 Construction of a grading model for curve shape index of mountainous highways and a traffic incident detection method

The study first proposes an index optimization method for grading the structure of highway flat curves and a method for grading highway curve levels. Then, a feature variable selection method based on IRF is designed, and an LSTM model based on BO is constructed. Finally, a mixed sampling method is combined with the above content to obtain the final IRF-HTID-BO-LSTM method.

3.1 Optimization method for index grading of highway flat curve structure and grading method for highway curve levels

The complex geometric conditions of highways often become one of the important factors that induce traffic incidents. Identifying geographical features and establishing practical connections between traffic conditions can help optimize highway safety risk assessment, and improve the accuracy and practicality of traffic management work [16-17]. The research first collects related highway alignment data, including G109, Beijing-Lhasa Highway (Ningxia Section), G110, Beijing-Qingtongxia Highway (Ningxia Section), G211, Yinchuan-Rongjiang Highway (Ningxia Section), G244, Wuhai-Jiangjin Highway (Ningxia Section), G307, Huanghua-Shandan Highway (Ningxia Section), G309, Qingdao-Lanzhou Highway (Ningxia Section), G312, Shanghai-Khorgos Highway (Ningxia Section), G327, Lianyungang-Guyuan Highway (Ningxia Section), G338, Haixing-Tianjun Highway (Ningxia Section), G341, Jiaonan-Haiyan Highway (Ningxia section), G344, Dongtai-Lingwu highway (Ningxia section), G566, Xiji-Tianshui Highway (Ningxia section). Furthermore, survey data on highway traffic conditions from 2022 to 2023 are collected from the aforementioned highways. Based on this, a geographic information system-based database for highway alignment and traffic condition management is constructed. Firstly, the correlation between traffic conditions and other factors is analyzed, as shown in Figure 1.



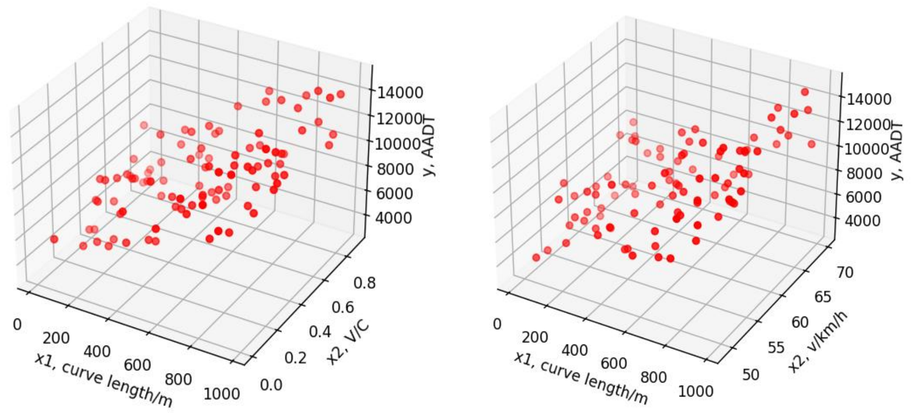
(a) Correlation of different curve line lengths

(b) Correlation at different saturation rates

Figure 1: The correlation between traffic conditions and other factors.

Figures 1 (a) and 1 (b) respectively show the correlation between different curve lengths and saturation rates, and both exhibit consistent distribution trends. The relationship between the curve shape and traffic conditions, as well as other factors, is shown in Figure 2.

In Figure 2, highway sections with longer flat curve lengths often exhibit higher saturation rates and larger average daily traffic volumes. Within a certain length range of a flat curve, the saturation rate and average daily traffic volume of the highway section fluctuate within a certain range. A longer flat curve allows vehicles to pass at higher and more stable speeds, exhibiting a higher saturation rate of traffic flow on the highway section. After introducing the average highway speed factor, the correlation between its distribution is less obvious. Based on the above analysis, combined with the modeling concept of highway traffic safety analysis in the interactive highway safety design model and the actual situation of traffic conditions on mountainous highways in China, as well as the correction coefficient of flat curve indicators, a suitable structural analysis model for flat curve indicators on mountainous highways is constructed. In addition, based on the specific situation of



(a) The relationship between the length of the flat curve and traffic conditions

(b) The relationship between curves and other factors

Figure 2: The relationship between curves and traffic conditions and other factors.

Table 2: Highway curve grading method.

Model level reference	**Design speed (km/h)**	**Minimum length of flat curve (m)**
C_L^0	40	≤200
C_L^1	60	(200, 300]
C_L^2	80	(300, 400]
C_L^3	100	(400, 500]
C_L^4	120	(500, 600]

mountainous highways, and the collected incident and geometric data of mountainous highways, an accident prediction and correlation model suitable for mountainous highways is established, as shown in equation (1).

$$Quality = AADT = v\beta^1 + C\_L\beta^2 + V/C\beta^3 + ... + Var\beta^n + \varepsilon \quad (1)$$

In equation (1), $Quality$ and $AADT$ represent the robustness of the road network and the average daily traffic volume, respectively. v, β, C_L, V/C, Var and ε correspond to the average vehicle speed, regression coefficient, flat curve length, road saturation rate, other reference factors, and errors of the road section, respectively. The regression method is applied to consider the relationship between geographical features and traffic flow under road network indicators, which can evaluate the applicability of indicators in traffic flow research. Deepening the grading optimization of the flat curve structure and indicators has a promoting effect on further improving its application value in practical decision-making. Therefore, the study summarizes curve type data as the classification basis, and further supplements and expands the interpretable flat curve length by introducing a grading expansion model for optimization, as expressed in equation (2).

$$Quality = AADT = C\_L^0\beta^0 + C\_L^1\beta^1 + ... + C\_L^n\beta^n + ... + \varepsilon \quad (2)$$

From this, grading processing can be carried out. The specific content of the highway curve classification method is shown in Table 2.

3.2 Feature variable selection for traffic incidents detection based on improved random forest

There is a close relationship between highway structure and traffic safety. To further achieve intelligent detection of traffic incidents on mountainous highways, it is necessary to first determine the effective feature variables for the HTID algorithm. Secondly, based on the Traffic Flow Fluctuation (TFF) theory, the traffic flow change characteristics under the highway traffic incidents should be analyzed to establish a comprehensive initial variable set for the HTID algorithm. Then, the key variables sensitive to traffic incident detection are screened. Finally, the feature variable set of the HTID algorithm is obtained.

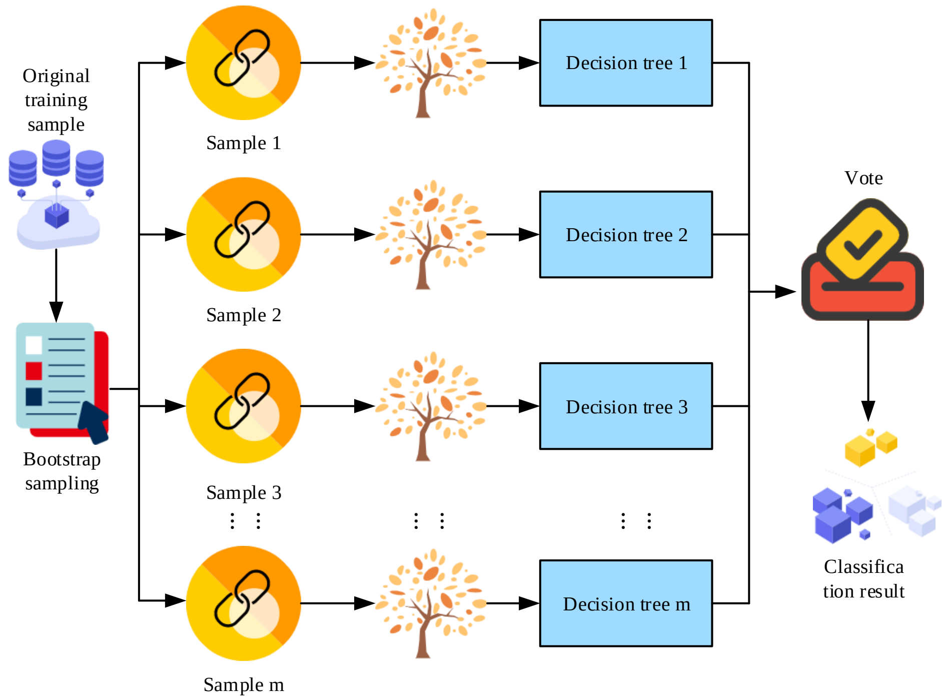


Figure 3: Flow chart of RF algorithm.

The TFF theory simulates the continuity equation of fluids through the basic principles of fluid mechanics, constructs the continuity equation of traffic flow, and seeks the theoretical relationship between traffic flow, density, and speed [18-19]. The specific application process of TFF theory is as follows. Firstly, the speed, traffic flow, and density of section A are v_A, f_A, and ρ_A. Then, based on the principles of TFF theory, the impact of traffic flow parameters on mountainous highways is analyzed in the event of a traffic accident [20-22]. When an accident occurs, the traffic capacity of section A will decrease to less than f_A, and the f_A reaching the downstream section will also decrease until it becomes the traffic capacity for a highway traffic incident. The sudden change in traffic status downstream of the road section can be marked as D for the traffic flow state here. Because the speed change corresponding to the transition state from A to D is relatively small, a forward wave will be generated. The waveform can be abbreviated as v_{AD}, as calculated in equation (3).

$$v_{AD} = \frac{f_A - f_D}{\rho_A - \rho_D} \tag{3}$$

In equation (3), f_D and ρ_D correspond to the traffic flow and density at state D, respectively. At upstream of the traffic incident point, the speed and flow will decrease accordingly, creating a high-density range represented by points A to E. At this time, the traffic flow status of the road section is denoted as E. A backward wave will be generated at point E, and the corresponding wave is abbreviated as v_{AE}, as shown in equation (4).

$$v_{AE} = \frac{f_A - f_E}{\rho_A - \rho_E} \tag{4}$$

In equation (4), f_E and ρ_E correspond to the traffic flow and density at state E, respectively. As time progresses, the area around the traffic incident will be divided into four sections. The traffic flow in the upstream and downstream parts of the entire section will still maintain its original turning direction. The upstream direction at point A will be greatly affected, while f_D and ρ_D at point D will decrease, resulting in congestion. At E, f_E will decrease, but ρ_E will maintain a high value, which constrains the traffic capacity. If A is not promptly handled, the impact of highway traffic time on traffic volume will continue to expand over time, ultimately leading to the shock waves and diffusion waves upstream and downstream, respectively. Through the analysis of TFF on the above-mentioned highway traffic incidents, it can be concluded that there are certain patterns in the changes that occur. Therefore, the basic parameters of traffic flow can be used as input parameters for subsequent intelligent detection algorithms. To more significantly represent changes in traffic flow, various relevant parameters can be combined, or different parameters of upper and lower detectors can be combined.

In addition, the study constructs an initial variable set, which includes the actual traffic flow parameter values obtained by detector detection, the product of the differences between different traffic flow parameters of upstream and downstream detectors, the ratio of measured traffic flow parameter values, and the difference and ratio between the measured traffic flow parameters of the same detector and the predicted values. The predicted values are obtained through the moving average method. To better select the feature variables of

HTID algorithm, the IRF is proposed for feature variable selection. The IRF algorithm mainly consists of two parts, the RF part and recursive feature elimination. The specific process is as follows. Firstly, in the RF section, M samples are randomly and selectively selected from K original samples using Bootstrap. The selected and unselected samples form the decision tree $g_m(m=1,2,...,q)$ and Out of Bag (OOB) M_m^{OOB}, respectively. Secondly, k_{ty} initial variables are randomly

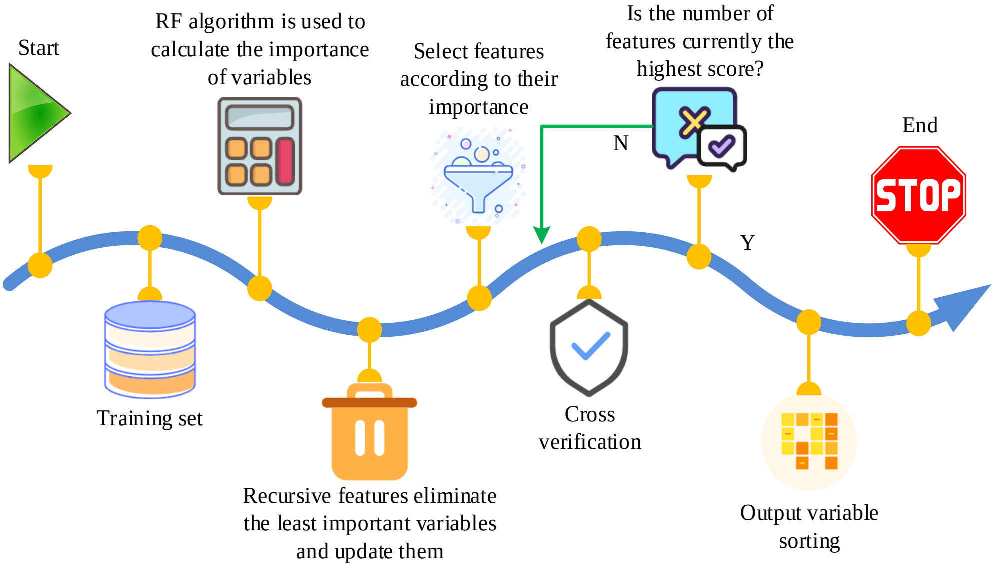


Figure 4: IRF process diagram.

selected from the initial variables, and the optimal node is selected from k_{ty} for each split of the tree [23-25]. In the growth process of the forest, each tree does not need pruning and will continue to split. By repeating the above operation q times, RF $f=(a_1,a_2,...,a_q)$ can be obtained. For each decision tree, the classification accuracy P_m is calculated using OOB data G_m^{OOB}. In the training set, each initial variable is denoted as α_b, and random noise is introduced into the α_b of G_m^{OOB} to obtain new data $\hat{G}_m^{OOB}$. The $\hat{\alpha}_b$ of each a_m for the hidden $\hat{G}_m^{OOB}$ is calculated. Finally, the importance ID of α_b is calculated, as shown in equation (5).

$$ID=\frac{1}{q}\sum_{j=1}^{q}(\alpha_b-\hat{\alpha}_b) \qquad (5)$$

On the basis of the above content, the RF algorithm is displayed in Figure 3.

After the RF part is completed, ID can be used as a basis to select each number of features one by one, and then cross validate the selected feature set. Finally, the number of features with the highest average score is obtained and determined [26-27]. Combining the RF part and recursive feature elimination, a complete schematic diagram of the IRF process can be obtained, as shown in Figure 4.

In Figure 4, the ID of α_b is first calculated and sorted, and then the recursive feature elimination method is used to extract features while updating the feature set. The above steps are repeated until all features are traversed and the feature set with the best accuracy is selected. Then, the cross-validation method is used to select the highest feature score set. Finally, the feature variable set and sorting are obtained.

3.3 Highway traffic incident detection algorithm based on LSTM optimized by bayesian optimization

After the feature variable selection is completed, the HDIT algorithm can be constructed based on the grading model. Usually, recurrent neural networks are applied in short-term information prediction, but they cannot meet the requirements of higher accuracy prediction. However, LSTM can solve the gradient vanishing in the above neural networks and perform long-term learning of relevant information tasks. The structure of LSTM only adds cell states on the basis of recurrent neural networks, which are used to store previously learned information and sequences, and achieve information exchange through special forms. This structure can increase memory implementation, so it can present good results in many problems [28-30]. The core part of LSTM is the neuron state. The information in the neuron state is controlled through a gating mechanism, which mainly includes four parts: forget gate, input gate, update gate, and output gate. In the forget gate, it is implemented through the Sigmoid layer, as expressed in equation (6).

$$F_t=\sigma\left[W_F\cdot(h_{t-1},x_t)+p_F\right] \qquad (6)$$

In equation (6), F_t is the output value of the forget gate. σ, W_F, h_{t-1}, and p_F correspond to the activation function, weight value, output of the previous neuron, and bias value, respectively. In the input gate, candidate vectors are generated through the tanh layer. The required finer values are determined by the sigmoid function. The output I_t and update content $\tilde{C}_t$ of the input gate are calculated using equation (7).

$$\begin{cases} I_t = \sigma\left[W_I \cdot \left(h_{t-1}, x_t\right) + p_I\right] \\ \tilde{C}_t = \tanh\left[W_C \cdot \left(h_{t-1}, x_t\right) + p_C\right] \end{cases} \quad (7)$$

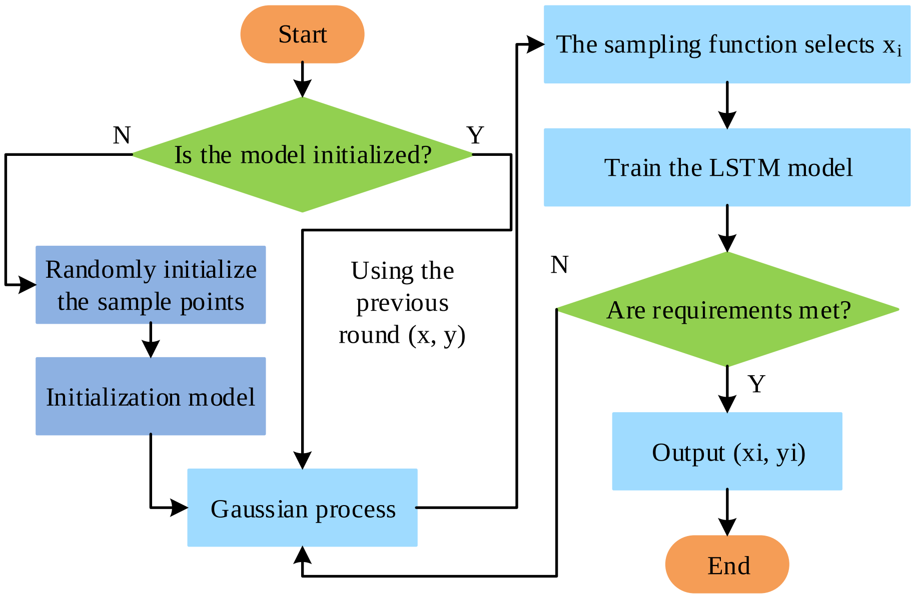


Figure 5: The process of optimizing LSTM hyper-parameters based on BO.

In equation (7), tanh is the tanh function. W_I and W_C correspond to the weights of input gates and cell states, respectively. p_I and p_C are the bias values of the input gate and cell state. In the update gate, it is necessary to update the cell state, as calculated in equation (8).

$$C_t = F_t \cdot C_{t-1} + I_t \cdot \tilde{C}_t \quad (8)$$

In the output gate, the expression for the output result O_t is shown in equation (9).

$$O_t = \sigma\left[W_O \cdot \left(h_{t-1}, x_t\right) + p_O\right] \quad (9)$$

In equation (9), W_O and p_O are the weight and bias values of the output gate, respectively. The final output result h_t is calculated using equation (10).

$$h_t = O_t \cdot \tanh\left(C_t\right) \quad (10)$$

LSTM has significant advantages in processing long time series tasks. It is more in line with actual situations, which facilitates subsequent classification tasks. All hidden units in the last layer are output, and then linked to the fully connected layer to complete binary classification. Due to the influence of hyper-parameters on the model performance, it is crucial to determine the appropriate combination of hyper-parameters. The BO algorithm is a hyper-parameter optimization method that can intelligently select the next evaluation point based on historical observation results, achieving parameter configuration close to the optimal solution in fewer iterations, and overcoming the time-consuming and unstable random results of other hyper-parameter setting methods. Therefore, it is used to optimize LSTM to achieve better performance in highway traffic incident detection. The specific process of optimizing LSTM model with BO algorithm is as follows. Firstly, the range of hyper-parameters is set and initialized to obtain the corresponding hyper-parameter data set. A set of data is randomly selected for Gaussian process regression to establish a probability distribution function and fit the objective function. The prior distribution of the Gaussian process is updated by the loss value, and the surrogate model is modified. Then, the sampling function is applied to select the next optimal sample point, which is the point x_i to be evaluated. The above points are input into LSTM for training, which can obtain the new output value y_i of the objective function, update it to the sample set $Q = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, and update the lining model. Finally, the condition judgment is completed through the loss value. If the value satisfies the requirements, the loss value and the current optimal hyper-parameter combination can be output. Otherwise, Q is updated, and the iterative correction process can be continued until it meets the requirements. From this, the process of optimizing LSTM hyper-parameters based on BO can be obtained, as shown in Figure 5.

In Figure 5, the first step is to determine whether the model has completed initialization. If it has, the sampling function selection step can be entered. Otherwise, the initialization step can be entered. Next, the initial sample points are randomly selected and used to initialize the LSTM model. Then, a Gaussian process is applied to

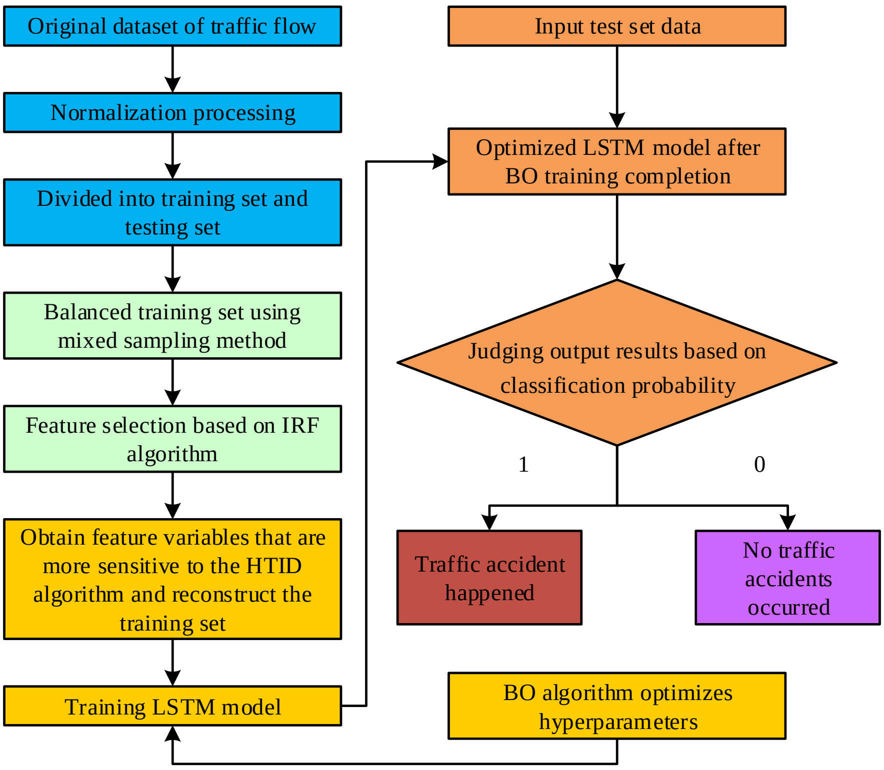


Figure 6: Schematic diagram of IRF-HTID-BO-LSTM method flow.

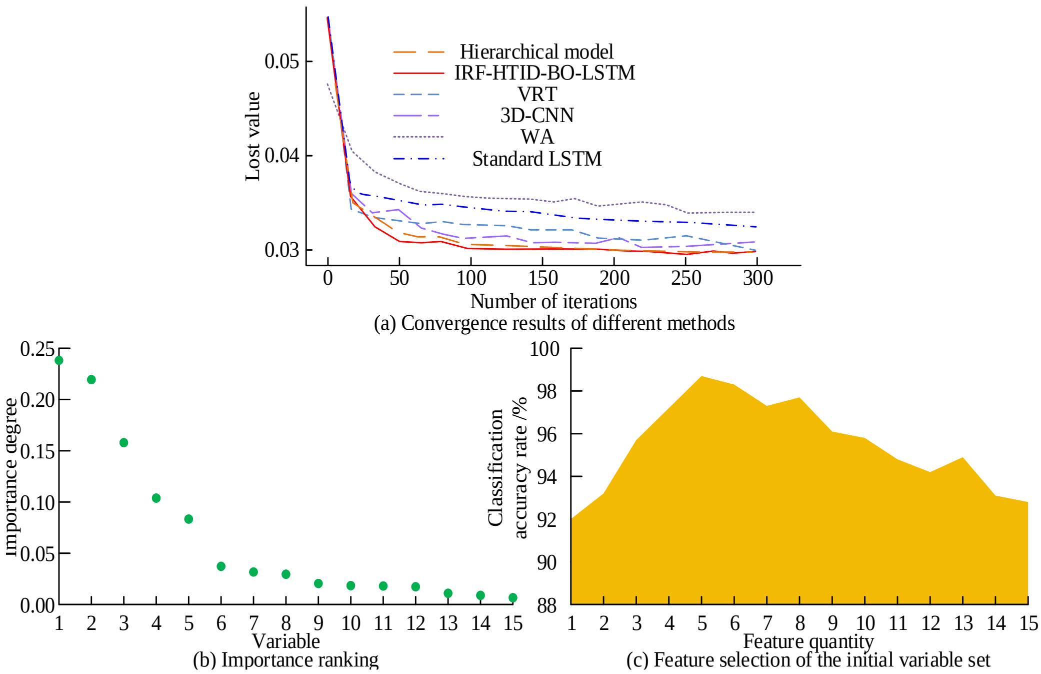


Figure 7: The convergence performance of different methods and the results of initial characteristic variable selection.

establish a surrogate model for predicting LSTM hyper-parameters. Finally, a new combination of hyper-parameters is selected through the sampling function and used to train the LSTM model. Finally, the performance is checked to check whether it meets the requirements. If it does not meet the requirements, the Gaussian process is returned. If it meets the requirements, the process can be ended. During the optimization process, the designed hyper-parameters include learning rate, batch size, iteration count, number of hidden layer nodes, and time step size. The values of learning rate are [0.01, 0.001, and 0.0001], the range of batch size values is [32, 64, 128, 256, 512], and the settings of other hyper-parameters are based on past experience. During training, to balance the samples, a mixed sampling processing method is combined with IRF-based feature variable selection and BO-optimized LSTM to obtain the final IRF-HTID-BO-LSTM algorithm. The specific implementation process is as follows. Firstly, the determined feature variables are applied to construct the training set and obtain the output matrix, as shown in equation (11).

$$OP = [y_i] = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{5400} \end{bmatrix} \qquad (11)$$

In equation (11), $y_i \in \{0,1\}$ represents the label of the i -th input sample. 1 and 0 correspond to the event label and non-incident label, respectively. Based on the partitioning of the training set, the output corresponding to the first 2700 input samples is 1, and the remaining samples are 0. In the training and optimization part of LSTM, the input data set $\{x_s, x_s \in X_{t\_set}, y_s \in Y_{t\_set}\}$ is first determined. Then, the forget gate, input gate, update gate, and output gate are sequentially passed. After training each segment x_s through the highway traffic incident detection method, the final feature vector Y_s^{LSTM} output by LSTM can be obtained, as shown in equation (12).

$$Y_s^{LSTM} = L_{LSTM}\left(x_s; W_F, W_C, W_I, W_O\right) \qquad (12)$$

In equation (12), $L(\cdot)$ represents the mapping function of LSTM. Because highway traffic incident detection belongs to binary classification, the general loss function uses cross entropy, while the output category probability uses softmax function. The obtained probability $P_{mc}(x_s)$ and cross entropy expression are shown in equation (13).

$$\begin{cases} P_{mc}(x_s) = soft\max\left(Y_s^{LSTM}\right) \\ \zeta(W,p) = -\left\{\sum_{m=1}^{n}\left\{y_m \cdot \log\left[P_{mc}(x_s)\right] + (1-y_m) \cdot \log\left\{1-P_{mc}(x_s)\right\}\right\}\right\} \end{cases} \qquad (13)$$

In equation (13), mc represents the classification label. $\zeta(W,p)$ and y_m are the objective function and the true label, respectively. The dataset is divided into training and testing sets at a ratio of 5:5. The LSTM model is set as $Z(x_s; W, p)$. The input dataset is $x_s \in X_{test}$. It is compared with the obtained classification prediction probability values to complete the classification based on the predicted label results. The flowchart of the IRF-HTID-BO-LSTM method can be obtained, as shown in Figure 6.

In Figure 6, the original traffic flow data set is normalized and partitioned. Then, the training set is balanced using a mixed sampling algorithm of Borderline SMOTE over-sampling and Tomek Links under-sampling. The IRF algorithm is used for feature selection to obtain feature variables that are more sensitive to the HTID algorithm. The training set is reconstructed and input into the LSTM model for training. The BO algorithm optimizes the hyper-parameters of the LSTM, and the output results can be determined by the classification probability.

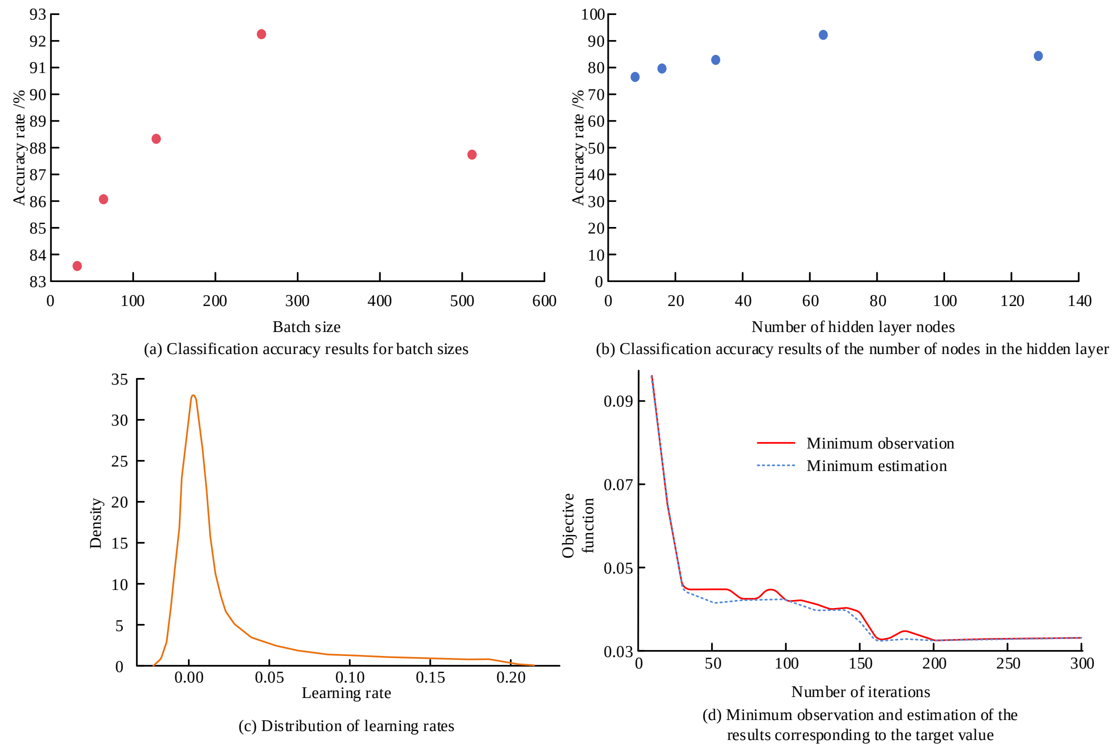


Figure 8: Training results of different hyper-parameters.

4 Results

To test the performance and application effectiveness of the proposed IRF-HTID-BO-LSTM method, the study first examines the effectiveness of the method, and explores the correlation fit of the grading model index optimization. Then, the comparative experiments are conducted to scientifically analyze the performance. Finally, its effectiveness in practical applications is evaluated.

4.1 Performance analysis of grading model construction and traffic incident detection methods

To evaluate the performance of the research method, experiments are conducted using software Matlab. The data set used for processing is as follows. Firstly, a total of 100 sets of mountain road traffic incidents were obtained, with a total of 10000 data points. Among them, the traffic incident data set contains 2224 data points, while the rest are non traffic incident data points. Then, the traffic incident data set is divided into training and testing sets in a 1:1 ratio. Then, a mixed sampling algorithm of Borderline SMOTE oversampling and Tomek Links under-sampling is used to balance the two types of samples in the training set, resulting in 5600 samples. Finally, the samples are divided into training and testing sets in a 1:1 ratio. Afterwards, the IRF is used to screen the initial variable set constructed, obtain feature variables that are more sensitive to the HTID algorithm, reconstruct the training set, and use it to train the LSTM network. The BO algorithm optimizer hyper-parameters are used. Finally, the output results can be used to determine whether a traffic incident has occurred on mountainous highways. In addition, the performance evaluation of the grading model is based on data collected from 487 curvature profiles of national highways, which are used for correlation degree calculation. To verify the performance of research methods more scientifically, the current mainstream methods are introduced for comparison, namely the standard LSTM method, the traffic incident detection method based on Video Recognition Technology (VRT), the detection method based on Three-Dimensional Convolutional Neural Networks (3D-CNN), and the detection method based on Wavelet Analysis (WA). In addition, commonly used indicators are used for evaluation, namely Detection Rate (DR), Mean Detection Time (MDT), False Alarm Rate (FAR), Comprehensive Performance Index (CPI), Receiver Operating Characteristic (ROC), and Area Under the Curve (AUC). MDT is the average time required to obtain test results, which is calculated by the average time required for all test times. FAR refers to the proportion of normal cases incorrectly reported as abnormal cases within a certain period of time. The lower the value, the better the specificity of the method. It is usually calculated by actually calculating the proportion of samples that are misclassified as positive. CPI is an indicator that comprehensively reflects the performance of the detection method. It takes into account DR, MDT and FAR three indicators, and is calculated by the product of MDT, FAR/100 and (1-DR/100). The study first tests the convergence performance of different methods and analyzes the results of the initial feature variables. The random feature variables and decision tree are set to 4 and 1000, respectively, and the results are shown in Figure 7.

Table 3: Statistical results of grading model structure optimization.

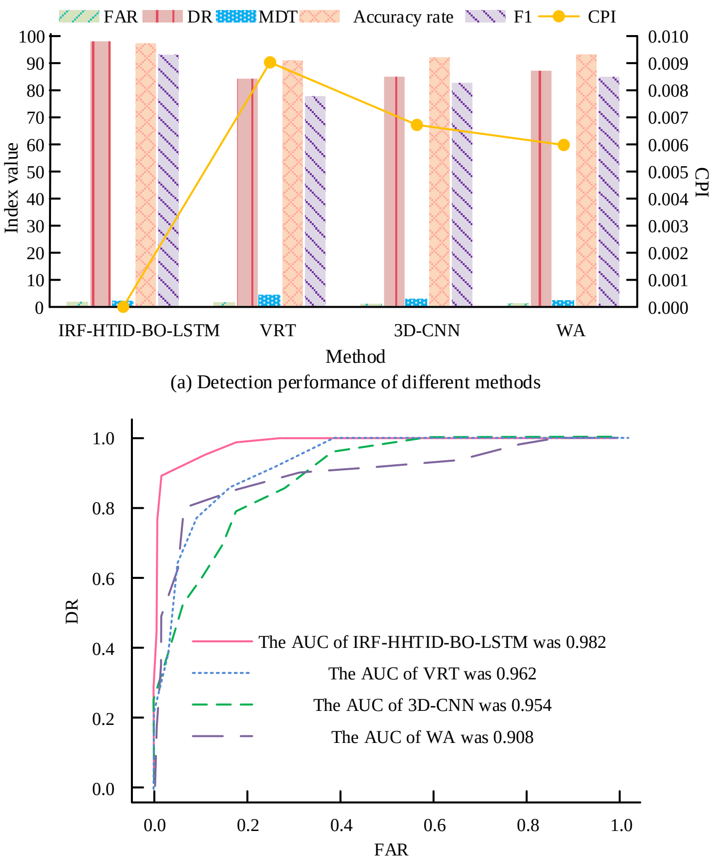
Model	Regression statistics	/
After improvement	Multiple R	0.873491573
	R Square	0.623749182
	Adjusted R Square	0.471134821
	Standard error	62.12849208
	Observed value	487
Before improvement	Multiple R	0.821998
	R Square	0.675681
	Adjusted R Square	0.554061
	Standard error	2526.453
	Observed value	487

Figures 7 (a) -7 (c) respectively correspond to the convergence results of different methods, the importance ranking of the initial variable set, and their selection results. From Figure 7, the proposed grading model and traffic incident detection method did not exceed 150 iterations on the training dataset to achieve a stable convergence state. However, the VRT method, 3D-CNN method, and WA method required 265, 273, and 284 iterations, respectively, to achieve a stable convergence state and corresponding higher loss values. In addition, the standard LSTM model requires 280 iterations to reach a stable state, because the baseline model has not been specifically optimized for traffic incident detection tasks in complex mountainous highway environments, and the input features contain a large amount of redundant or irrelevant information. LSTM requires more iterations to identify and utilize effective information. After running recursive feature elimination through cross validation, the number of features with better classification accuracy can be obtained. Then, feature variables with lower rankings can be deleted, and finally the set with the highest feature score can be selected. Similarly, the detector corresponds to the difference between the measured and predicted occupancy values, the ratio of the predicted and measured speed values, and the difference between the measured and predicted speed values. Figure 7 (b) showed that when the number of features was 5, the corresponding classification accuracy was the highest at 98.7%. Based on the importance ranking of the feature variables, the final set of feature variables can be determined as the ratio of the measured occupancy values corresponding to the upstream and downstream detectors, the product of the difference in occupancy values and the speed difference. Similarly, the difference between the measured occupancy values and the predicted values corresponding to the detectors, the ratio of the predicted speed values and the measured speed values, and the difference between the measured speed values and the predicted values can also be determined. This showed the above features contain more information about changes in traffic flow status, and can help the model better distinguish traffic incidents and run more stably. In addition, the higher importance is due to its stronger interaction with other features. When the ratio of the measured occupancy values corresponding to the upstream and downstream detectors is the highest, it can reflect the changes in traffic flow between different detectors and is a sensitive indicator of changes in traffic flow status. The product of occupancy rate difference and speed difference contains information from both dimensions of speed and occupancy rate, which can more comprehensively describe the dynamic changes of traffic flow. To assess the effectiveness of the grading model structure optimization, the regression statistical method is used for evaluation, as displayed in Table 3.

According to Table 3, the fitting degree of the grading model before optimization was only 82.2%, indicating a strong positive correlation. The R-Square

Table 4: The statistical results of the correlation degree of the optimized grading indicator of mountainous highway flat curve.

/	/	Coefficients	Standard error	t Stat	P	Lower 95%	Upper 95%
/	Intercept	0.055430518	0.098029	0.565449	0.585587	-0.16633	0.277188
Flat curve index grading	C_L^4	0.001274648	0.001808	0.705061	0.038612	-0.00282	0.005364
	C_L^3	3.873047452	1.19E-05	3.264968	0.009761	1.19E-05	6.56E-05
	C_L^2	15.04555171	2273.262	6.618486	0.000166	9803.4	20287.7
	C_L^1	27.10997887	80.05872	0.338626	0.043608	-157.506	211.7257
	C_L^0	11.06338754	18.46679	0.599097	0.025683	-31.5211	53.64787



(a) Detection performance of different methods

(b) The combined performance of the different methods

Figure 9: Comparison of detection performance and comprehensive performance of different methods.

value showed that the model explained 67.6% of the variability of the dependent variable, indicating that the model has good explanatory power for the data. The adjusted R-Square value was 47.1%, indicating that the model has a reasonable explanatory power. The fitting degree of the optimized grading model has been significantly improved, and the fitting effect of the highway traffic condition correlation model for flat curve structure grading was improved to 87.3%, indicating a strong linear relationship between the independent and dependent variables. To further evaluate the correlation between the optimized grading indicators for mountainous highway flat curves, the specific results are displayed in Table 4.

From Table 4, C_L^2 and C_L^3 had a significant impact on traffic conditions when the length of the flat curve was between 300m-500m. When the length of the flat curve exceeded 500m or was less than 200m, the remaining grades exhibited marginal significance, and the intercept term was not significant, making a small contribution to the model. The above results verify that different classifications have a significant impact on the routes and traffic conditions in mountainous areas, indicating that the grading model has a good application effect on indicator grading. The indicator grading of the flat curve length of mountainous highways is reasonable. To obtain the optimal hyper-parameters for the IRF-HTID-BO-LSTM method, the study focuses on optimizing the time step and the number of hidden layer nodes using the BO algorithm, and updates the training parameters using the adaptive moment estimation algorithm (Adam). Due to the large data set size, the fixed learning rate is 0.001, the maximum number of iterations is 300, the batch sizes are 32, 64, 128, 256, and 512, the time step range is 1-30, and the hidden layer node range is 8, 16, 32, 64, and 128. In addition, the study uses 5 cross validations for training. The obtained training results for different hyper-parameters are shown in Figure 8.

Figures 8 (a) and 8 (b) show the classification accuracy results corresponding to batch size and the number of hidden layer nodes, respectively. Figure 8 (c) and 8 (d) correspond to the distribution of learning rates and the comparison of the minimum observation and estimation corresponding to target values. Figures 8 (a) and 8 (b) showed that the accuracy results exhibited a bell-shaped curve with the change of hyper-parameters. As the number of hidden layer nodes and batch size continue to increase, the classification accuracy shows a trend of first increasing and then decreasing. Moreover, the number of hidden layer nodes has a relatively large impact on the classification accuracy. When the number of hidden layer nodes and batch size were 64 and 256, respectively, the performance of the research method was optimal, indicating that the research method has stability and robustness. Figure 8 (c) shows the distribution of learning rates. The corresponding density reached its highest when the learning rate was 0.001. In Figure 8 (d), with the gradual increase of iterations, the BO algorithm converged when the search reached 18 times, indicating

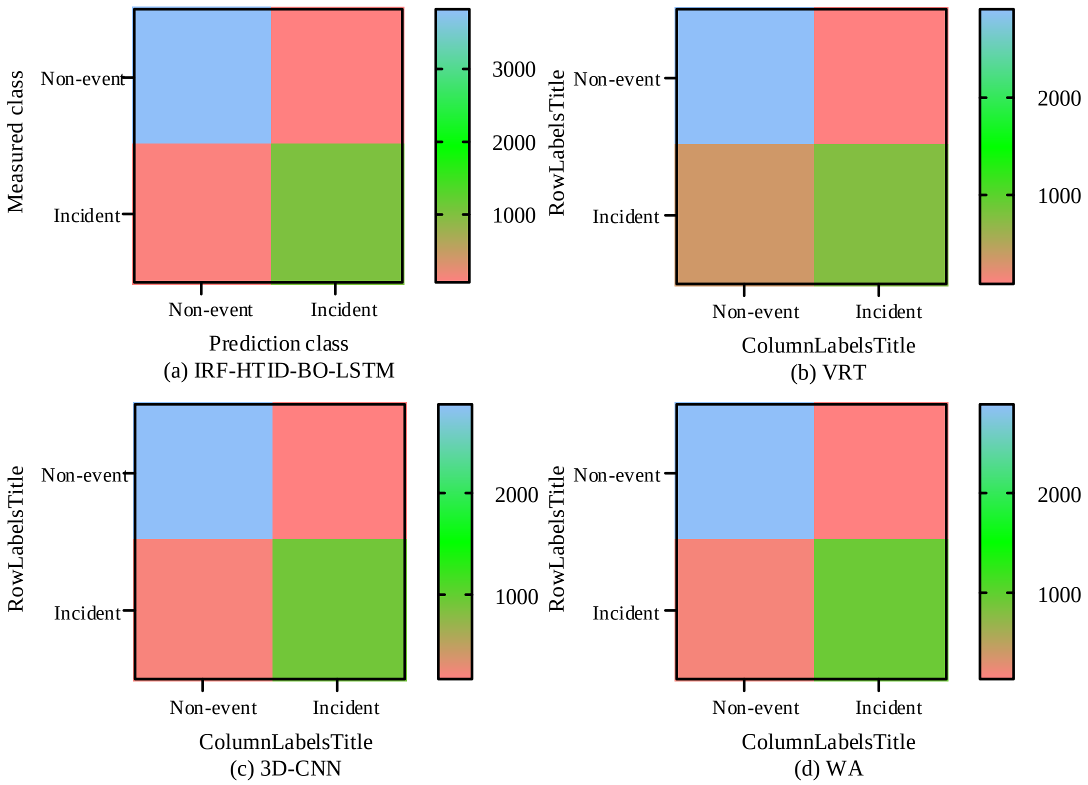


Figure 10: Comparison of classification performance results of different methods.

that the global optimal solution could be obtained. To further analyze the performance of the IRF-HTID-BO-LSTM method, experiments are conducted on different methods using DR, MDT, FAR, CPI, accuracy, F1 value, and AUC. The results are shown in Figure 9.

Figure 9 (a) and Figure 9 (b) respectively compare the detection performance and comprehensive performance results of different methods. From Figure 9, in the comparison results of detection performance, the research method had the best performance in all detection indicators except for FAR, which was 1.82%. DR, MDT, CPI, accuracy, and F1 value were 98.45%, 2.21%, 0.00159, 97.74%, and 93.52%, respectively. Compared with VRT, 3D-CNN, and WA, the detection performance of the research method has been significantly improved. Although mainstream methods sacrifice DR values to obtain smaller FAR values, there are still many traffic incidents that have not been detected. The research method can improve other detection performance while ensuring a lower MDT. Among them, when processing data of the same scale, the WA method had the highest computational efficiency, corresponding to an MDT value of 2.6 seconds, while the research method had relatively high computational efficiency. This means that the research method can significantly improve the response capability to emergencies, thereby improving road safety and traffic flow efficiency. In the comprehensive performance results, the AUC of the research method was the highest, at 0.982, while the AUC values of the VRT method, 3D-CNN method, and WA method corresponded to 0.962, 0.954, and 0.908, respectively. The proposed feature variable selection method can effectively improve the performance of the final detection method, and the overall performance of the research method is also the best. The comparison of classification performance results of different methods is shown in Figure 10.

Table 5: Comparison of classification performance of different methods based on paired t-test.

Evaluating indicator	**Method**	**df**	**T-score**	**Sig**
Susceptibility	IRF-HTID-BO-LSTM	40	4.08	**
	VRT	40	5.38	**
	3D-CNN	40	7.46	**
	WA	40	8.18	**
	Standard LSTM	40	10.08	**
Specificity	IRF-HTID-BO-LSTM	40	16.30	**
	VRT	40	24.95	**
	3D-CNN	40	28.58	**
	WA	40	24.00	**
	Standard LSTM	40	28.02	**

Note: "* *" indicates $P<0.01$.

Table 6: Comparison of model accuracy results based on K-fold cross validation.

Method	**Sample size**	**Correlation coefficient**	**Root mean square error**	**Coefficient of variation**
IRF-HTID-BO-LSTM	2000	0.65	268	0.38
VRT	2000	0.30	375	0.51
3D-CNN	2000	0.29	305	0.49
WA	2000	0.31	229	0.27
Standard LSTM	2000	0.34	243	0.32

Figures 10 (a) -10 (d) show the confusion matrix results of the IRF-HTID-BO-LSTM method, VRT method, 3D-CNN method, and WA method, respectively. In Figure 10, the research method had the largest number of positive samples, which meant that the method identified the largest number of traffic incident samples. The WA method had the smallest positive class positive samples, indicating that its classification performance was the worst among mainstream methods. In summary, many indicators of the research method are superior to other mainstream algorithms, and the comprehensive performance and classification performance are the best, indicating that the feature selection and grading model can effectively improve the performance of the method. To more accurately quantify the performance effects of different methods, statistical tests are introduced to explore the differences between each method. Paired t-tests are used to analyze the classification effects of different methods, and the results are shown in Table 5.

From Table 5, the research method had extremely significant statistical significance compared to other benchmark models on sensitivity and specificity. The research method has a certain degree of reliability and lays a good foundation for subsequent practical applications. To test the robustness of the data set used in the study, k-fold cross validation is used to analyze different methods. The specific process is as follows. The data set is randomly divided into K equally sized subsets to ensure that each subset is as similar as possible in data distribution. Then, one subset is selected as the test set, and the remaining subsets are merged as the training set. Finally, the model is trained on the training set, and its performance is evaluated on the test set. The accuracy of the study is analyzed, and the results are shown in Table 6.

Table 6 shows that the accuracy performance from high to low is IRF-HTID-BO-LSTM, Standard LSTM, WA, VRT, and 3D-CNN. The accuracy of the research method is higher than other methods, which also indicates that the data set used in the study has stability and reliability.

4.2 Application effect analysis

To explore the effectiveness of the research method in practical applications, this study takes the traffic conditions of ordinary provincial highways as an example. The annual report data provided by Ningxia Highway Management Center in 2022 is imported into the research method. The graph analysis function in geographic information system is used to identify and mark key or abnormal road sections that exceed the indicator grading and cause abnormal traffic volume. Taking the Haitian Highway (western section of Ningxia) as an example, the key or abnormal point road section annotations obtained are shown in Figure 11.

From Figure 11, in the practical application of the research method, there were nine abnormal points detected in the Haitian line section, and the affected range of the abnormal point road section was relatively long. In addition, when the flat curve was too short, traffic conditions were more affected by the structure of the flat curve. In practical applications, the research method can be combined with the relevant functions of geographic information systems to visually display and label abnormal sections of mountainous highways, and timely feedback to relevant departments for management. The specific data results of the abnormal points in the Haitian section are displayed in Table 7.

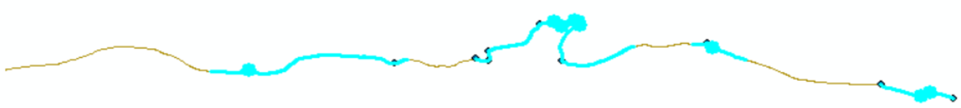


Figure 11: Mark and display map of key or abnormal points of Haitian highway (West Ningxia Section).

Table 7: The specific data results of the anomaly section of the Haitian section.

Road section name	**Section number**	**Section length of flat curve /m**	**Standardized_residuals**
Shapotou Unity Bridge	G338640521	10.59538775	2.952860532
Gutang village, Zhongning	G338640521	76.06553674	-2.620215777
Zhongning East Hedong	G338640521	31.59364867	-2.554101242
Shapotou tourism new town	G338640502	6.762924746	-2.488856025
Shapotou tourist area	G338640502	79.81748426	2.441351606
Shapotou Yangtan east	G338640502	81.97640697	2.340711859
Kantang	G338640502	283.6514657	2.306630542
Shapo head meets the water	G338640502	173.6766195	2.305995462
Intersection of Yingshui Railway Station	G338640502	9.164988211	-2.245316332
Mengjia Bay East	G338640502	90.77954661	2.234417005

According to Table 7, in practical applications, specific road segment numbers and length information of flat curved road segments where traffic incidents occur can be obtained through the research method. The standard error range was within ±3. The above results suggest due to the fact that the VRT method and 3D-CNN method may miss the traffic incident in some cases, while the WA method has significant problems in feature extraction or classification decision-making, resulting in the worst performance in traffic incident recognition. In summary, the research method still demonstrates good performance in practical applications, providing intelligent detection ideas for traffic incidents on mountainous highways and ensuring accurate detection of abnormal road sections even under complex road conditions.

5 Discussion

Mountainous highway terrain is complex, climate change, easy to appear traffic accidents. Intelligent traffic detection method can be based on real-time road conditions, and timely detection and early warning of potential safety hazards, improving driving safety. Therefore, an index optimization method and a highway curve classification method for mountainous highway flat curve structure are proposed in this paper. The IRF algorithm is used to select characteristic variables. BO algorithm and LSTM are combined to obtain IRF-HHTID-BO-LSTM algorithm.

The research results show that when the number of features is 5, the classification accuracy of the research method is the highest 98.7%. According to the importance ranking of the feature variables, the final feature variable set can be determined as the product of the ratio of the measured value of the share corresponding to the upstream and downstream detectors, the difference of the share and the velocity difference. The same detector corresponds to the difference between the measured value of occupancy and the predicted value, the ratio of the predicted speed and the measured speed and the difference between the measured speed and the predicted value. Among them, the ratio of the measured occupancy corresponding to the upstream and downstream detectors is the highest because it can reflect the change of traffic flow between different detectors. It is a sensitive indicator of the change of traffic flow state. The product of occupancy rate difference and speed difference contains information from both dimensions of speed and occupancy rate, which can more comprehensively describe the dynamic changes of traffic flow. The above results are generated because the research method combines BO algorithm and LSTM algorithm to adjust hyper-parameters and process time series data. Therefore, the research method can learn the patterns in the data more effectively, while other mainstream algorithms lack the corresponding complexity when processing complex mountainous highway traffic incident data.

The hyper-parameter sensitivity analysis results show that the accuracy results present a bell curve with the change of hyper-parameters. With the continuous increase of the node in the hidden layer and the batch size, the classification accuracy shows an initial increase followed by a decrease. Moreover, the number of nodes in the hidden layer has a relatively large impact on the classification accuracy. The performance of the research method is the best, which indicates that the research method has stability and robustness. The optimal hyper-parameter combination is obtained as follows. The time step, batch size and number of hidden layer nodes are 5, 64 and 256, respectively. The comparison of the detection performance and comprehensive performance results of different methods shows that the performance of other detection indicators is the best except for the FAR index of 1.82%. and the DR, MDT, CPI, accuracy and F1 are 98.45%, 2.21%, 0.00159, 97.74% and 93.52%, respectively. In addition, its detection performance is obviously better than other benchmark models, but the computational efficiency of the research method is relatively high, and the corresponding MDT value is 3.1s. The research can obtain a smaller FAR by sacrificing certain DR and computational efficiency. In addition, the best performing method in related work is the intelligent transportation service framework integrating heterogeneous data sources, whose accuracy rate is only 90%, because it is difficult to accurately analyze the road structure and traffic safety status of mountainous highways. Its subsequent management cannot get timely feedback. The comprehensive performance results show that the AUC values of the research method, VRT method, 3D-CNN method and WA method correspond to 0.982, 0.962, 0.954 and 0.908, respectively. The higher AUC value of the research method indicates because it can effectively remove redundant information through the improved feature variable selection method, improving the overall effect of the detection method.

In summary, the research method can still maintain a high comprehensive performance when facing complex mountainous highways. The effect is also excellent in practical application, but it sacrifices a certain degree of computational efficiency and has wrong classification. This may be due to the complexity and variability of mountain road traffic data. Therefore, in future, a more suitable model architecture for mountain road design can be selected according to the task characteristics, and the generalization ability of the model can be improved through data enhancement technology.

6 Conclusion

To reduce traffic accidents on mountainous highways and further enhance the ability of highways to maintain smooth traffic, an index optimization method for grading the structure of highway flat curves and a grading method for highway curve levels were designed. The IRF-HTID-BO-LSTM method was proposed. The proposed grading model and traffic incident detection method achieved stable convergence with lower loss values on the training data set without exceeding 150 iterations. The optimal hyper-parameter combination optimized by the BO algorithm was as follows. The time step, batch size, and number of hidden layer nodes were 5, 64, and 256, respectively. The research method only had a high FAR value of 1.82%, while the performance of other detection indicators was the best. DR, MDT, CPI, accuracy, and F1 value were 98.45%, 2.21%, 0.00159, 97.74%, and 93.52%, respectively. The highest AUC value was 0.982. In practical applications, the research method accurately obtained the specific section numbers and length information of flat curved road segments where traffic incidents occurred, and the standard error range was within ±3. In summary, the research method can effectively classify and segment the curve line indicators of highway structures, establish the curve line grading indicator of mountainous highway structures, and ensure the rapid and accurate detection of traffic incidents on mountainous highways, providing a certain reference for the detection of other traffic incidents. However, there are still shortcomings in the research. For the analysis of highway structure, certain parameters may not be suitable for this study. For example, the results indicate that the speed variable has not reached the level of statistical significance. Therefore, in future research, the model can be improved by introducing additional variables such as traffic intensity and lane occupancy rate.

7 Funding statement

This project was funded by National Key Research and Development Program of China (2023YFB4302701).

References

[1] A. Haydari, and Y. Yılmaz. "Deep reinforcement learning for intelligent transportation systems: A survey," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 1, pp. 11-32, 2022, https://doi.org/10.1109/TITS.2020.3008612.

[2] H. Zhao and A. Sharma. "Logistics distribution route optimization based on improved particle swarm optimization," Informatica, vol. 47, no. 2, pp. 243-251, 2023, https://doi.org/10.31449/inf.v47i2.4011.

[3] S. Murrar, F. M. Alhaj, and M. Qutqut. "Machine learning algorithms for transportation mode prediction: A comparative analysis," Informatica, vol. 48, no. 6, pp. 117-130, 2024, https://doi.org/10.31449/inf.v48i6.5234.

[4] C. Li and Z. Mu. "Analysis platform of rail transit vehicle signal system based on data mining," Informatica, vol. 47, no. 3, pp. 441-450, 2023, https://doi.org/10.31449/inf.v47i3.3942.

[5] J. Hawkins, and K. N. Habib. "A multi-source data fusion framework for joint population, expenditure, and time use synthesis," Transportation, vol. 50, no. 4, pp. 1323-1346, 2023, https://doi.org/10.1007/s11116-022-10279-8.

[6] H. Wang, L. Liao, W. Yi, and L. Zhen. "Transportation scheduling for modules used in modular integrated construction," International Journal of Production Research, vol. 62, no. 11, pp. 3918-3931, 2024, https://doi.org/10.1080/00207543.2023.2251602.

[7] C. Zhao, X. Chang, T. Xie, H. Fujita, and J. Wu. "Unsupervised anomaly detection-based method of risk evaluation for road traffic accident," Applied Intelligence, vol. 53, no. 1, pp. 369-384, 2023, https://doi.org/10.1007/s10489-022-03501-8.

[8] S. Cafiso, C. D'Agostino, and M. Kiec. "Investigating safety performance of the SAFESTAR system for route-based curve treatment," Reliability Engineering and System Safety, vol. 188, pp. 125-132, 2019, https://doi.org/10.1016/j.ress.2019.03.028.

[9] G. Ashley, O. A. Osman, S. Ishak, and J. Codjoe. "Investigating effect of driver-, vehicle-, and road-related factors on location-specific crashes with naturalistic driving data," Transportation Research Record, vol. 2673, no. 6, pp. 46-56, 2019, https://doi.org/10.1177/0361198119844461.

[10] M. R. Fatmi, and M. A. Habib. "Modeling vehicle collision injury severity involving distracted driving: Assessing the effects of land use and built environment," Transportation Research Record, vol. 2673, no. 7, pp. 181-191, 2019, https://doi.org/10.1177/0361198119849060.

[11] D. E. Monyo, H. J. Haule, A. E. Kitali, and T. Sando. "Are older drivers safe on interchanges? Analyzing driving errors causing crashes," Transportation Research Record, vol. 2675, no. 12, pp. 635-649, 2021, https://doi.org/10.1177/03611981211031232.

[12] S. B. Li, T. Sun, D. N. Cao, and. L. Zhang. "Incident detection method of expressway based on traffic flow simulation model," Communications in Theoretical Physics, vol. 71, no. 4, pp. 468-474, 2019, https://doi.org/10.1088/0253-6102/71/4/468.

[13] X. Zhang, R. R. Souleyrette, E. Green, T. Wang, M. Chen, and P. Ross. "Collection, analysis, and reporting of kentucky traffic incident management performance," Transportation Research Record, vol. 2675, no. 9, pp. 167-181, 2021, https://doi.org/10.1177/03611981211001077.

[14] P. H. L. Rettore, B. P. Santos, R. R. F. Lopes, G. Maia, L. A. Villas, and A. A. F. Loureiro. "Road data enrichment framework based on heterogeneous data fusion for ITS," IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 4, pp. 1751-1766, 2020, https://doi.org/10.1109/TITS.2020.2971111.

[15] M. Won. "Outlier analysis to improve the performance of an incident duration estimation and incident management system," Transportation Research Record, vol. 2674, no. 5, pp. 486-497, 2020, https://doi.org/10.1177/0361198120916472.

[16] M. Azari, A. Paydar, FB. Eizizadeh, and V. G. Hasanlou. "A GIS-based approach for accident hotspots mapping in mountain roads using seasonal and geometric indicators," Applied Geomatics, vol. 15, no. 1, pp. 127-139, 2023, https://doi.org/10.1007/s12518-023-00490-2.

[17] C. Y. Lin, Y. C. Lai, S. W. Wu, F. C. Mo, and C. Y. Lin. "Assessment of potential sediment disasters and resilience management of mountain roads using environmental indicators," Natural Hazards, vol. 111, no. 2, pp. 1951-1975, 2022, https://doi.org/0.1007/s11069-021-05126-5.

[18] H. Duan, and Y. Song. "Grey prediction model based on Euler equations and its application in highway short-term traffic flow," Nonlinear Dynamics, vol. 112, no. 12, pp. 10191-10214, 2024, https://doi.org/10.1007/s11071-024-09611-x.

[19] S. T. Zheng, R. Jiang, B. Jia, J. Tian, and Z. Gao. "Impact of stochasticity on traffic flow dynamics in macroscopic continuum models," Transportation Research Record, vol. 2674, no. 10, pp. 690-704, 2020, https://doi.org/10.1177/0361198120937704.

[20] Y. Liu, C. Lyu, X. Liu, and Z. Liu, "Automatic feature engineering for bus passenger flow prediction based on modular convolutional neural network," IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 4, pp. 2349-2358, 2021, https://doi.org/10.1109/TITS.2020.3004254.

[21] A. A. Chaudhari, K. K. Srinivasan, B. R. Chilukuri, M. Treiber, and O. Okhrin. "Calibrating Wiedemann-99 model parameters to trajectory data of mixed vehicular traffic," Transportation Research Record, vol. 2676, no. 1, pp. 718-735, 2022, https://doi.org/10.1177/03611981211037543.

[22] T. V. Geetha, A. J. Deepa, and M. M. Linda. "Deep learning method for efficient cloud IDS utilizing combined behavior and flow-based features," Applied Intelligence, vol. 54, no. 8, pp. 6738-6759, 2024, https://doi.org/10.1007/s10489-024-05505-y.

[23] M. Hasanvand, M. Nooshyar, E. Moharamkhani, and A. Selyari. "Machine learning methodology for identifying vehicles using image processing," Artificial Intelligence and Applications, vol. 1, no. 3, pp. 170-178, 2023, https://doi.org/10.47852/bonviewAIA3202833.

[24] H. Mokayed, T. Z. Quan, L. Alkhaled, and V. Sivakumar "Real-time human detection and counting system using deep learning computer vision techniques," Artificial Intelligence and Applications, vol. 1, no. 4, pp. 221-229, 2023, https://doi.org/10.47852/bonviewAIA2202391.

[25] J. Purohit, and R. Dave. "Leveraging deep learning techniques to obtain efficacious segmentation results," Archives of Advanced Engineering Science, vol. 1, no. 1, pp. 11-26, 2023, https://doi.org/10.47852/bonviewAAES32021220.

[26] I. Salman and J. Vomlel. "Learning the structure of bayesian networks from incomplete data using a mixture model," Informatica, vol. 47, no. 1, pp. 81-94, 2023, https://doi.org/10.31449/inf.v47i1.4497.

[27] G. S. Ohannesian and E. J. Harfash. "Epileptic seizures detection from EEG recordings based on a hybrid system of gaussian mixture model and random forest classifier," Informatica, vol. 46, no. 6, pp. 105-116, 2022, https://doi.org/10.31449/inf.v46i6.4203.

[28] C. Sivanandam, V. M. Perumal, and J. Mohan. "A novel light GBM-optimized long short-term memory for enhancing quality and security in web service recommendation system," Journal of Supercomputing, vol. 80, no. 2, pp. 2428-2460, 2024, https://doi.org/10.1007/s11227-023-05552-1.

[29] J. Wu, J. Tang, M. Zhang, J. Di, L. Hu, X. Wu, G. Liu, and J. Zhao. "PredictionNet: A long short-term memory-based attention network for atmospheric turbulence prediction in adaptive optics," Applied Optics, vol. 61, no. 13, pp. 3687-3694, 2022, https://doi.org/10.1364/AO.453929.

[30] W. Yang, W. Chang, Z. Song, F. Niu, X. Wang, and Y. Zhang. "Denoising odontocete echolocation clicks using a hybrid model with convolutional neural network and long short-term memory network," Journal of the Acoustical Society of America, vol. 154, no. 2, pp. 938-947, 2023, https://doi.org/10.1121/10.0020560.