WOA-XGBoost Based Railway Accident Type Prediction and Cause Analysis

Yuhan Xie, Gensuo Mi, Ding Tan, Chenning Liu

Abstract—**Aiming at the problems of unbalanced railway accident categories and incomplete reason analysis for different types of accidents, a railway accident type prediction method based on the Whale Optimization Algorithm (WOA) to improve the Extreme Gradient Boosting (XGBoost) is proposed. The SHAP method is adopted to analyze the reasons for diverse types of railroad accidents. Firstly, railway accidents are divided into six types such as derailment and collision. The SMOTE algorithm is used to resample unbalanced accident data. Secondly, WOA-XGBoost is constructed to classify and predict different types of railway accidents. Finally, SHAP (Shapley Additive Explanation) is introduced to explain the WOA-XGBoost model with the best performance, identify the important factors affecting the model output, and analyze the causes of different types of railroad accidents. The results demonstrate that the accuracy of the proposed model is 0.817, the precision is 0.809, and the F1 value is 0.761. The WOA-XGBoost model outperforms the decision tree, the gradient boosting decision tree, and the XGBoost model in terms of performance. According to the SHAP method, mechanical failure, human error, and speed are the important factors leading to different types of accidents.**

Index Terms— **railway accident type prediction; extreme gradient boosting; whale optimization algorithm; SHAP cause analysis**

I. Introduction

Railroad accidents contain risk factors, accident types, severity, and other information. Determining the key factors that need to be focused on controlling from plenty of recorded accident information is significant for accident prevention. The prediction and reasons analysis of railroad accidents is an important aspect of railroad safety. Therefore, based on the historical data of railway accidents, it holds substantial practical implications to forecast what types of railroad accidents will occur in a specific situation, identify the risk factors that need to be strengthened in different types of railroad accidents such as derailment and collision, and explore the reasons of railroad accidents, to improve the level of railway safety management.

Manuscript received October 08, 2024; revised January 26, 2025.

This work was supported in part by the National Natural Science Foundation of China (Nos. 61763025).

Yuhan Xie is a postgraduate student at School of Automation and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China. (Corresponding author, e-mail: yuhanxiexie@126.com).

Gensuo Mi is a Professor at School of Automation and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China. (e-mail: 3573901427@qq.com).

Ding Tan is an assistant engineer of China Railway SIYUAN Consulting and Engineering Group Co., Ltd., Nanning 530000, China. (e-mail: 1524121781@qq.com).

Chenning Liu is a postgraduate student at School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China. (e-mail: 1985164936@qq.com).

The existing railroad accident prediction research took human error, management organization, environment, equipment failure, and other influencing factors as independent variables and took accident severity [1], risk level, number of accidents, loss amount, or accident type as dependent variables. In the existing research, traditional algorithms such as Decision Tree (DT) [2], Support Vector Machine [3], and Logistic Regression [4] were mainly used to build accident prediction models. However, the prediction of railway accident types has the problem of category imbalance, which will make the model biased toward most class samples, thus affecting the model's performance [5]. The above algorithms are mainly applicable to simple and balanced data. However, for railway accident records with large sample sizes and unbalanced categories, the performance of the prediction model needs to be enhanced. The focus of current research is on further improving the model's performance.

In comparison to traditional algorithms, the ensemble learning algorithm improves the prediction performance by merging numerous models and performs better on accident data with unbalanced categories. Moreover, it can evaluate the contribution of independent variables to the model through feature importance. Therefore, relevant scholars have introduced ensemble learning algorithms into road traffic safety [6], [7], ship safety analysis [8], and railway safety research. Zhou et al. [9] adopted the Random Forest (RF) algorithm to predict accidents at highway-railroad crossings. The results demonstrated that the RF model can mitigate the adverse impacts caused by imbalanced data. R. Bridgelall et al. [10] adopted the Extreme Gradient Boosting (XGBoost) to predict derailment and non-derailment and identified the important factors of derailment accidents through feature importance and principal component analysis. Zhong et al. [11] proposed an algorithm based on the Gradient Boosting Decision Tree (GBDT) for predicting the types of railroad accidents and analyzing the causes. The conclusion showed that derailment and collision accidents are closely related to location, train speed, temperature, and personnel factors. H. Meng et al. [12] proposed an AdaBoost-Bagging algorithm to predict railway accidents and examined the reasons based on the importance of features. Although these studies have provided valuable insights and useful references for railway accident analysis, they have limitations in analyzing accident causes in depth. Specifically, they have mainly focused on using feature importance to identify important factors in models. However, the influence of variable values on accidents is different, and they fail to analyze the influence of variable values on accidents in detail. Therefore, a more comprehensive approach is essential to analyze the causes of diverse categories of railroad accidents.

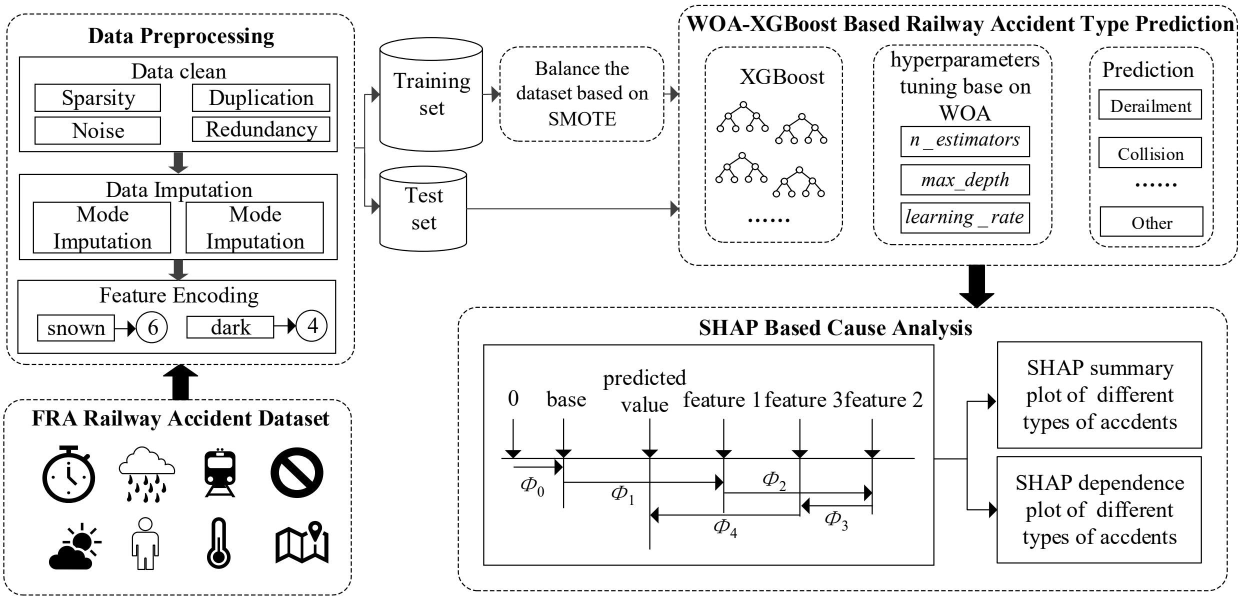


Fig. 1. Railway accident type prediction and cause analysis framework

In recent years, scholars have applied the Shapley Additive Explanations (SHAP) to road and railway safety research to analyze accident factors more comprehensively. The SHAP method can calculate the contributions of variables and analyze the degrees of positive and negative impacts of different variable values on the prediction results. H. L. Ding et al. [13] analyzed the causes of bicycle collisions based on SHAP, and the results showed that population density is an important factor in bicycle accidents. R. Bridgelall et al. [14] applied machine-learning algorithms to study human-caused railway accidents. Through the SHAP game-theoretic model analysis, they concluded that human-caused accidents are usually correlated with derailments. J. Liu et al. [15] utilized SHAP to explain the influence of different features on the accidents and concluded that speed, visibility, and total tonnage are the key variables in the model.

This paper takes advantage of the fast convergence speed and strong optimization capability of the Whale Optimization Algorithm (WOA) to effectively improve the performance of XGBoost model. Meanwhile, the SHAP approach provides a more adequate explanation for the predictive model, enabling a more comprehensive analysis regarding the reasons for different categories of accidents. Therefore, to tackle the problems of unbalanced accident types, poor performance of prediction models, and insufficient cause analysis, a railroad accident type prediction and reason analysis method based on ensemble learning is proposed. Firstly, the synthetic minority oversampling technique (SMOTE) is adopted to process the accident dataset. Then, an improved XGBoost algorithm grounded in the WOA algorithm is put forward for predicting railway accident types. Finally, SHAP is introduced to explore the causes of railroad accidents. The overall framework is shown in Fig. 1.

II. SMOTE ALGORITHM

Oversampling techniques can address the problem of category imbalance for railway accident data. The traditional random over-sampling (ROS) technique balances the class distribution by randomly replicating minority samples, but this method tends to cause over-fitting in the model [16]. Compared with the ROS technique, SMOTE considers the similarity between samples of the same category and synthesizes new samples within the original minority sample space through specific strategies, which can effectively avoid the overfitting problem caused by simple replication [17]. The SMOTE algorithm is applied to mitigate the impact of data imbalance on the test results. The process is illustrated in Fig. 2.

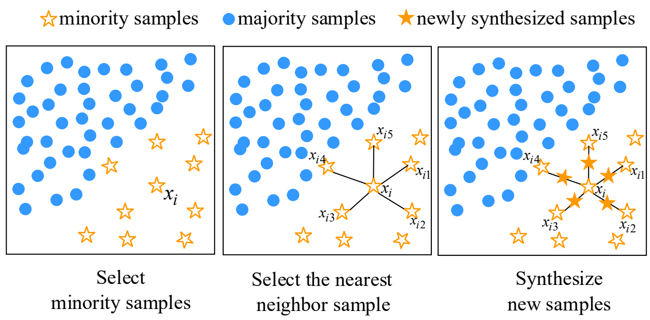


Fig. 2. SMOTE algorithm diagram

The SMOTE algorithm flow is as follows.

(1) Select minority sample: Select sample x_i from minority samples as the benchmark for synthesizing new samples, calculate the distance between the sample and other samples in the minority class based on the Euclidean distance, and find the k-nearest neighbor.

(2) Select the nearest neighbor sample: The sampling rate is set according to the unbalanced proportion of the sample, and a neighbor x_{ij} is randomly selected from the k-nearest neighbors for x_i as the auxiliary sample. The operation is repeated N times.

(3) Synthesize new samples: Each randomly selected nearest neighbor sample x_{ij} is calculated according to (1). Interpolation is carried out between the reference sample and the nearest neighbor sample to construct a new sample. Finally, N synthetic samples are produced.

$$X_{synthetic} = x_i + rand(0,1) \times (x_{ij} - x_i) \tag{1}$$

Where x_i is the i-th sample within the minority class, x_{ij} is the j-th nearest neighbor sample of x_i. $X_{synthetic}$ is a sample synthesized by interpolation between x_i and x_{ij}, $rand(0,1)$ represents a randomly generated number between [0,1].

III. Accident Type Prediction Based on WOA-XGBOOST

A. XGBoost

XGBoost is an improved ensemble learning algorithm based on GBDT [18]. XGBoost improves the prediction performance of the overall model by serially integrating multiple decision trees, which can provide a better prediction effect for railway accident data with numerous samples and imbalanced categories [19]. Based on the traditional loss function, the objective function adds a regular term to control the model's complexity and improve the model's generalization capacity. In the training process, the objective function is expanded by second-order Taylor expansion to improve the accuracy of the model [20].

A railway accident-type prediction model is constructed based on the XGBoost algorithm. The dependent variable is the accident type, and the independent variables include speed, visibility, mechanical equipment failure, and other factors. There is a dataset $D=\{(x_i,y_i)\}$, which has n railway accident samples and b independent variables. The objective function of XGBoost is shown in (2) and the complexity is shown in (3).

$$Obj=\sum_{i=1}^{n}l\left(y_i,\hat{y}_i\right)+\sum_{k=1}^{K}\Omega\left(f_k\right) \tag{2}$$

$$\Omega\left(f_t\right)=\gamma T+\frac{1}{2}\lambda\sum_{j=1}^{T}\omega_j^2 \tag{3}$$

Where $l\left(y_i,\hat{y}_i\right)$ is the loss function, which is the deviation of the predicted value from the actual value of the railway accident type. n is the number of samples imported into the k-th tree. The complexity $\Omega\left(f_t\right)$ adjusts the penalty strength by γ and λ to regulate the model's complexity, with the aim of averting overfitting of the model. T denotes the number of leaf nodes. ω_j denotes the weight of the j-th leaf node.

The second-order Taylor expansion of (2) is carried out. In the t-th iteration, the objective function is transformed into (4).

$$\begin{aligned}Obj^{(t)}=&\sum_{i=1}^{n}\left[l\left(y_i^t,\hat{y}_i^{(t-1)}\right)+g_if_t\left(x_i\right)+\frac{1}{2}h_if_t^2(x_i)\right]\\&+\sum_{k=1}^{t-1}\Omega\left(f_k\right)+\Omega(f_t)\end{aligned} \tag{4}$$

Where g_i and h_i denote the first-order and second-order partial derivatives of the loss function, t denotes the index of the tree, and f_t is the model corresponding to the t-th tree.

In the t-th iteration, the value of $\hat{y}_i^{(t-1)}$ is known. Then $l(y_i,\hat{y}_i^{(t-1)})$ and $\sum_{k=1}^{t-1}\Omega(f_k)$ are also known constants. These constants are independent of optimizing the objective function, so (4) can be converted into (5).

$$Obj^{(t)}=\sum_{i=1}^{n}\left[g_if_t\left(x_i\right)+\frac{1}{2}h_if_t^2\left(x_i\right)\right]+\Omega\left(f_t\right) \tag{5}$$

In (4), the objective function is traversed on the accident samples. To discover the optimal tree during the t th iteration, the objective function is converted into the form of traversal on the leaf. The size of the objective function is directly related to the total count of leaf nodes, as shown in (6).

$$Obj^{(t)}=\sum_{j=1}^{T}\left[\omega_jG_j+\frac{1}{2}\omega_j^2\left(H_j+\lambda\right)\right]+\gamma T \tag{6}$$

Where $G_j=\sum_{i\in I_j}g_i$, $H_j=\sum_{i\in I_j}h_i$.

The partial derivative of ω_j is made to be equal to 0, and then $\omega_j=-\frac{G_j}{H_j+\lambda}$ is obtained and substituted back to (6). The result is shown in (7).

$$Obj^{(t)}=-\frac{1}{2}\sum_{j=1}^{T}\frac{G_j^2}{H_j+\lambda}+\gamma T \tag{7}$$

As the value of the objective function decreases, the accuracy of the model improves, and the reliability of the classification increases. When the objective function reaches its minimum value, the optimal model can be obtained.

B. WOA-XGBoost Model

The XGBoost algorithm encompasses hyperparameters like the learning rate, the number of base learners, and the maximum depth of the tree. An improper setting of these hyperparameters will result in a decline in the accuracy of accident type prediction. Specifically, the settings of $n\_estimators$ and max_depth affect the complexity of the accident-type prediction model. The model is simpler when the values assigned to these two hyperparameters are relatively small. However, the model may not be fully trained, leading to under-fitting. Conversely, the model becomes more complex when the values set for these two hyperparameters are relatively large. Nevertheless, this increased complexity may reduce the model's generalization ability, giving rise to overfitting. The learning rate can control the model's iteration rate and prevent overfitting.

WOA has strong global search ability and fast convergence speed [21], which can effectively search for the optimal parameter combination of the XGBoost algorithm and improve the performance of railway-accident prediction model. Thw WOA is used to iteratively optimize the above three important hyperparameters of XGBoost.

The whale optimization algorithm, which takes whale hunting behavior as a reference, is a mathematical model that simulates contraction encirclement, spiral encirclement, and global random search of prey by whales [22]. When the whale shrinks and encircles the prey, supposing that the optimal solution represents the hunting target. The whale position closest to the target is the optimal individual spatial position. The rest of the whales update the position based on this as shown in (8).

$$X(t+1)=X^*(t)-A\,|\,C\cdot X^*(t)-X(t)\,| \tag{8}$$

Where $X^*(t)$ is the optimal individual space position vector, $X(t)$ is the current individual position vector. A is calculated by (9) and C is calculated by (9).

$$A = 2a \cdot r_1 - a \tag{9}$$

$$C = 2 \cdot r_2 \tag{10}$$

Where a is a linearly decreasing convergence factor in the range of [0,2]. r_1 and r_2 represent random vectors whose components fall within the interval [0,1].

The spiral-up position update is mathematically modeled based on the separation between the present individual and the prey [23], with the expression shown in (11).

$$X(t+1) = D' \cdot e^{bl} \cdot \cos(2\pi l) + X^*(t) \tag{11}$$

Where D' represents the distance between the current individual and the prey, b represents a defined constant, and l represents a randomly-generated number between [0,1].

According to probability p, WOA selects how to update the individual position and chooses to adopt the contraction-surrounded method or the spiral rising position update method. The specific model is shown in (12).

$$X(t+1) = \begin{cases} X^*(t) - A\,|C \cdot X^*(t) - X(t)| & p < 0.5 \\ D \cdot e^{bl} \cdot \cos(2\pi l) + X^*(t) & p \geq 0.5 \end{cases} \tag{12}$$

Where p denotes a random number in the range of $(0,1)$.

It achieves the process of shrinking by reducing the value of a. When a in (9) gradually decreases, A also decreases. When $|A|<1$, the whale shrinks and surrounds the prey. When $|A|\geqslant 1$, the whale randomly searches for prey in the global range, and its mathematical model is shown in (13).

$$X(t+1) = X_r - A \cdot D \tag{13}$$

Where $D = |C \cdot X_r - X(t)|$, X_r represents the random individual's position in the whale population.

The flowchart of WOA-XGBoost is illustrated in Fig. 3.

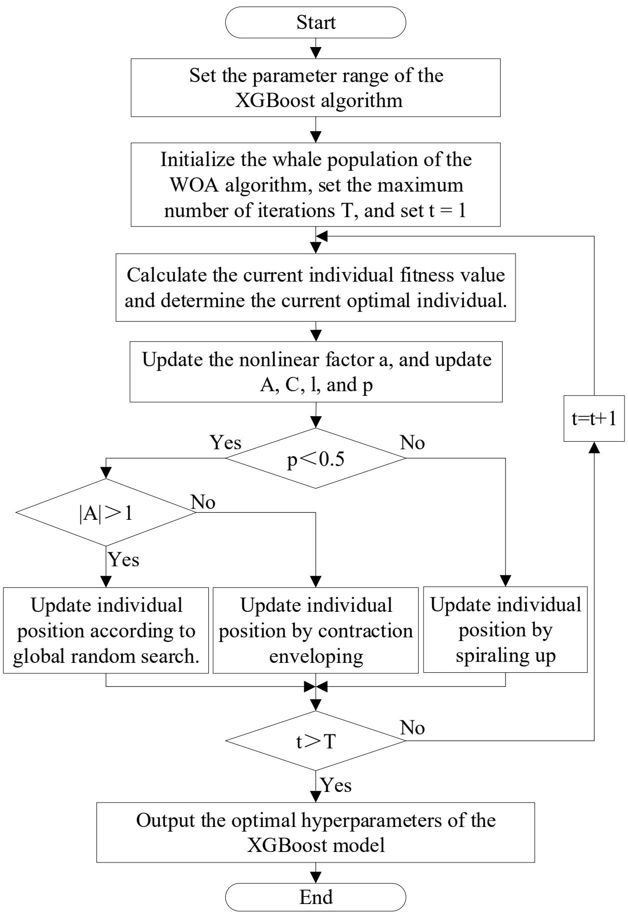


Fig. 3. WOA-XGBoost flowchart

IV. Railway Accident Cause Analysis Method Based on SHAP

To improve the interpretability of the accident type prediction model, the SHAP method is introduced to visualize the decision-making mechanism of the prediction model. SHAP can more precisely capture the interaction between features. It can find out the key features affecting the model output and analyze the reasons for different types of railway accidents.

SHAP explains the output of the black box model by calculating the contribution of the variables [24]. The contribution value $\phi_{i,j}$ is calculated as (14).

$$\phi_{i,j} = \sum_{S \subseteq B \setminus \{x_{i,j}\}} \frac{|S|!(b-|S|-1)!}{b!} \left[f_{x_i}\left(S \cup \{x_{i,j}\}\right) - f_{x_i}(S) \right] \tag{14}$$

Where M is a set of all variables in the railway accident data set, with a total of b independent variables. S is a subset selected from the set B. $|S|$ denotes the size of M. When the variable set S is selected, the model's prediction value of the i-th accident sample is $f_{x_i}(S)$. When the $x_{i,j}$ is added to the set S, the model's prediction value of sample x_i is $f_{x_i}\left(S \cup \{x_{i,j}\}\right)$.

SHAP provides an additive interpretation model based on Shapley value in game theory [25]. The prediction result of each sample is the linear sum of the contribution values of all independent variables in the sample, modeled as shown in (15).

$$f(x_i) = F(z') = \phi_0 + \sum_{j=1}^{m} \phi_{i,j} z'_{i,j} \tag{15}$$

Where $f(x_i)$ is the prediction result of the model for sample x_i. b is the number of features. If $\phi_{i,j} > 0$, it means that the j-th independent variable of the i-th accident sample acts positively on the outcome of the model. On the contrary, if $\phi_{i,j} < 0$, the j-th independent variable acts negatively on the prediction result.

V. Simulation and Results Analysis

A. Railway Accident Data Processing

The data are derived from the railway equipment accident database published by the Federal Railroad Administration (FRA) [26]. The accident records from 2012 to 2023 are selected as experimental data. The database records railway accidents since 1975 in tabular form, including 144 attributes such as accident type, location, speed, load, number of casualties, and loss amount.

According to the actual situation of railway accidents, referring to the FRA accident type classification standard [27] and historical literature research [28], [29], the types of railway accidents are divided into six types, such as derailment and collision. The coding and statistics of railway accident types are shown in Table I. The number of type 0 (Derailment) is the largest, and the number of type 4 (Fire/Violent rupture) and type 5 (Other) is very small, indicating that the categories of the data set are extremely unbalanced.

TABLE I
CODE AND PROPORTION OF TYPE

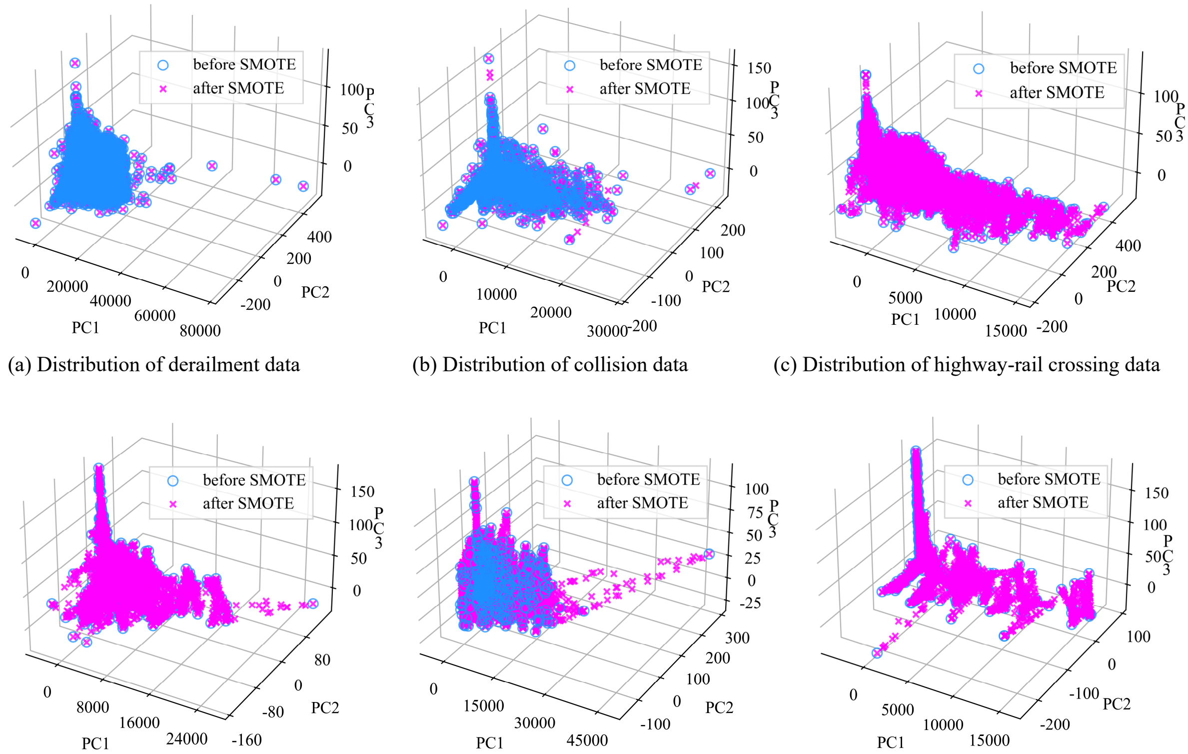
Code	Accident type	Proportion
0	Derailment	57.87%
1	Collision	24.59%
2	Highway-rail crossing	8.53%
3	Obstruction	3.73%
4	Fire/Violent rupture	1.86%
5	Other	3.43%

Data preprocessing primarily involves three main aspects: data cleaning, feature encoding, and missing value imputation. Firstly, data cleaning is carried out to delete samples with outliers and eliminate sparse, redundant, and irrelevant features. Secondly, the unstructured data is encoded into structured data by Label-Encoding. Finally, the missing discrete data is filled by mode, and the missing continuous data is filled by mean. The preprocessed dataset contains 27118 accident records and retains 20 independent variables, including time, speed, weather, and the total number of cars. Some examples of categorical features and their coding descriptions are shown in Table II.

To reduce the influence of category imbalance on the model's prediction performance, the pre-processed FRA railway accident records are used as the data set. 80% of the data is chosen as the training dataset, while the remaining 20% serves as the test dataset. The SMOTE algorithm is then employed to balance the accident type within the training dataset. Moreover, the principal component analysis (PCA) is applied to decrease the dimensionality of both the original training set and the balanced training set. The first three principal components are chosen to display the data distribution visually. Because different accidents have different data distribution characteristics and change trends, six data distributions are obtained for different accident types, as illustrated in Fig. 4. Fig. 4 demonstrates that the data distribution characteristics before and after sampling are similar, and the trend of change is consistent, indicating that the minority samples synthesized by SMOTE are close to the actual accident samples.

TABLE II
CATEGORICAL FEATURES

Feature	Code
season	1 - spring, 2- summer, 3 - autumn, 4 - winter
weather	1-clear, 2-cloudy, 3-rain, 4-fog, 5-sleet, 6-snown
type_of_consist	1 - freight train, 2 - passenger train, 3 - locomotive, 4 - cars, 5 - work train, 6 - yard
type_of_track	1 - main, 2 - yard, 3 - siding, 4 - industry
trackcwr	0 - other, 1 - continuous welded rail track
risk	1 - brake, 2 - trailer or container on flatcar, 3 - body, 4 - draft system, 5 - truck components, 6 - axles and journal bearings, 7 - wheels, 8 - locomotives, 9 - doors, 10 - general mechanical electrical failures, 11 - use of brakes, 12 - employee physical condition, 13 - flagging, fixed, hand and radio signals, 14 - general switching rules, 15- main track authority, 16 - train handling / train make-up, 17 - speed, 18 - use of switches, 19 - miscellaneous, 20 - roadbed, 21 - track geometry, 22 - rail, joint bar and rail anchoring, 23 - frogs, switches and track appliances, 24 - other structure, 25 - signal and communication, 26 - environmental conditions, 27 - extreme weather conditions, 28 - loading procedures, 29 - violation of highway-rail crossing rules by highway users, 30 - obstacles on the track, 31 - others



(a) Distribution of derailment data
(b) Distribution of collision data
(c) Distribution of highway-rail crossing data
(d) Distribution of obstruction data
(e) Distribution of fire/violent rupture data
(f) Distribution of other data

Fig. 4. The data distribution corresponding to different types of accidents before and after SMOTE

B. Accident Type Prediction Results

The WOA-XGBoost model is trained based on the balanced training set incident data. The whale population size of WOA is set to 20, the maximum iteration count is set to 100, and the accuracy of the model on the test set is taken as the fitness value. The convergence curve of the hyperparameters optimization process is shown in Fig. 5. The results of hyperparameters optimization are as follows: max depth is 8, number of estimators is 567, and learning rate is 0.09789.

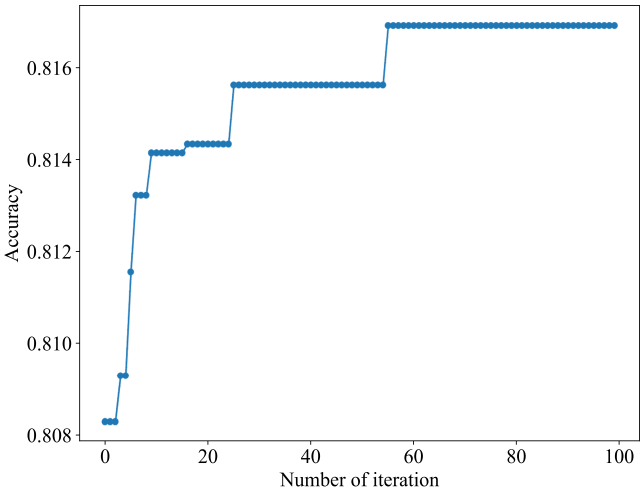


Fig. 5. Hyperparameters optimization process of WOA-XGBoost

The WOA-XGBoost model is compared with the Decision Tree, GBDT, and XGBoost models. The prediction results of different models are visualized through the confusion matrix, as illustrated in Fig. 6. The horizontal coordinate of the confusion matrix is the predicted accident type, and the vertical coordinate is the actual accident type.

In addition, Accuracy, Precision, Recall and F1-score are selected as the evaluation indicators to measure the performance of the model. The four indicators are computed according to the confusion matrix, as presented in (16) ~ (19).

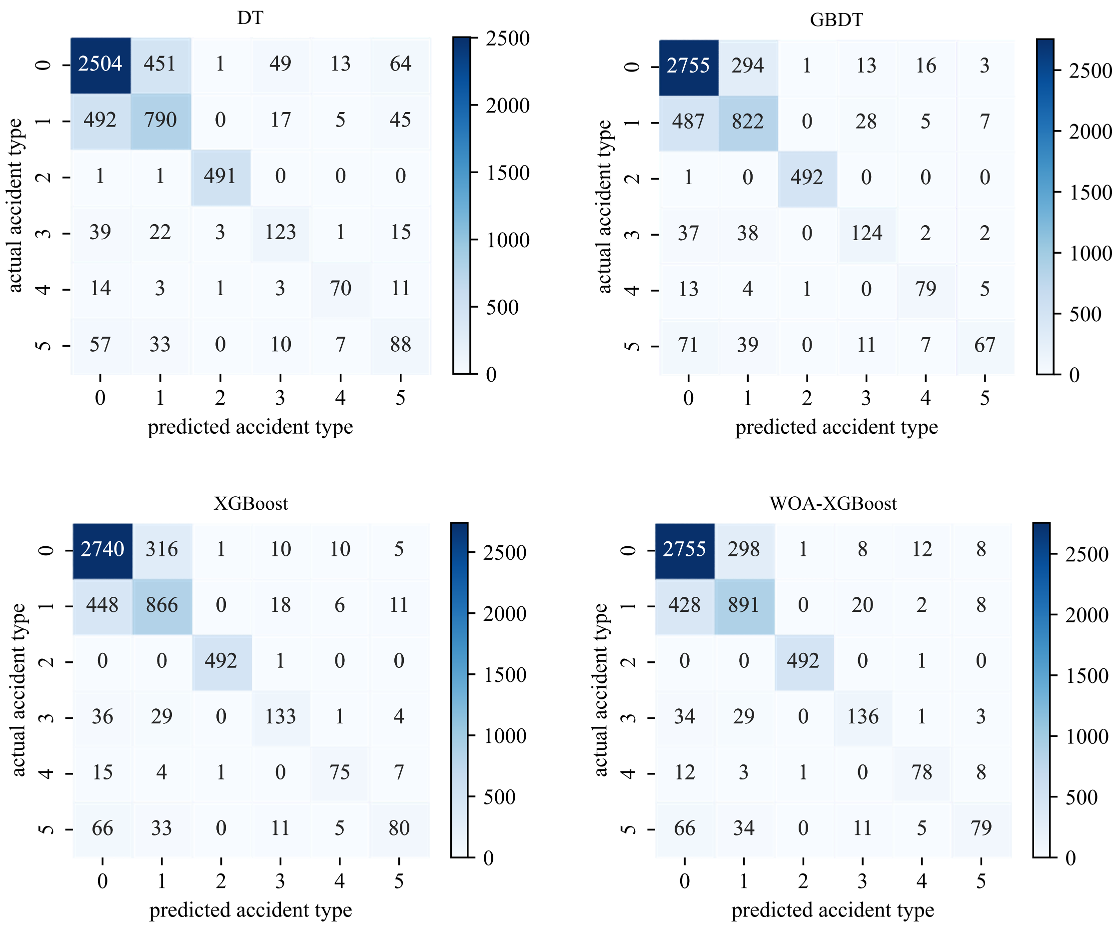
$$Accuracy = \frac{1}{C}\sum_{c=1}^{C}\frac{TP_c + TN_c}{TP_c + TN_c + FP_c + FN_c} \tag{16}$$

$$Precision = \frac{1}{C}\sum_{c=1}^{C}\frac{TP_c}{TP_c + FP_c} \tag{17}$$

$$Recall = \frac{1}{C}\sum_{c=1}^{C}\frac{TP_c}{TP_c + FN_c} \tag{18}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{19}$$

Where C is the total number of railway accident types, in this paper $C = 6$. TP_c denotes the number of those rightly predicted to be the c-th type, TN_c denotes the number of those rightly predicted not to be the c-th type, FP_c denotes the number of those incorrectly predicted to be the c-th accident type, FN_c denotes the number of those wrongly predicted not to be the c-th accident type.



DT	0	1	2	3	4	5
0	2504	451	1	49	13	64
1	492	790	0	17	5	45
2	1	1	491	0	0	0
3	39	22	3	123	1	15
4	14	3	1	3	70	11
5	57	33	0	10	7	88

GBDT	0	1	2	3	4	5
0	2755	294	1	13	16	3
1	487	822	0	28	5	7
2	1	0	492	0	0	0
3	37	38	0	124	2	2
4	13	4	1	0	79	5
5	71	39	0	11	7	67

XGBoost	0	1	2	3	4	5
0	2740	316	1	10	10	5
1	448	866	0	18	6	11
2	0	0	492	1	0	0
3	36	29	0	133	1	4
4	15	4	1	0	75	7
5	66	33	0	11	5	80

WOA-XGBoost	0	1	2	3	4	5
0	2755	298	1	8	12	8
1	428	891	0	20	2	8
2	0	0	492	0	1	0
3	34	29	0	136	1	3
4	12	3	1	0	78	8
5	66	34	0	11	5	79

Fig. 6. Confusion matrix

The performance comparison of four prediction models is shown in Table III and Fig. 7. The results show that the precision of the WOA-XGBoost model is 0.8087, which is 11.93% higher than the precision of decision tree algorithm, and other evaluation indexes are 4%~7% higher than that of the single decision tree, which proves that the performance of the ensemble learning model outperforms that of the single model. The WOA-XGBoost model achieves an accuracy of 0.8169, a recall of 0.732, and an F1 value of 0.761. These values are better than those of the DT, GBDT, and XGBoost models.

TABLE III
COMPARISON OF MODEL PERFORMANCE

Model	Accuracy	Recall	Precision	F1
DT	0.7496	0.6896	0.6894	0.6890
GBDT	0.8000	0.7050	0.7881	0.7302
XGBoost	0.8086	0.7216	0.8014	0.7521
WOA-XGBoot	0.8169	0.7320	0.8087	0.7610

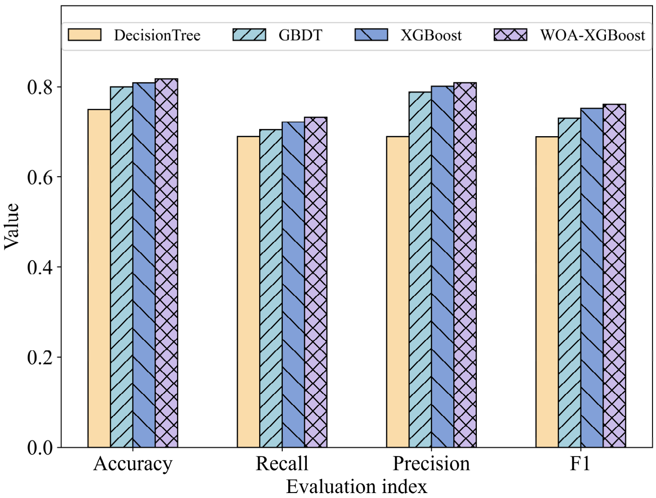


Fig. 7. Performance comparison of four models

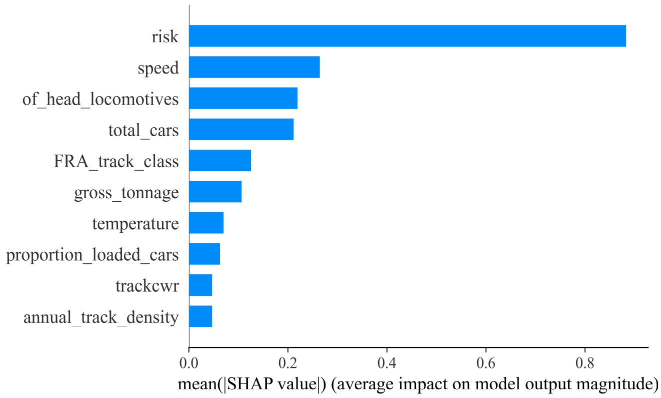
C. Accidents Cause Analysis

Compared with the traditional feature importance analysis of ensemble learning algorithms, SHAP can not only calculate the contribution of variables and rank the importance of features on the global model, but also qualitatively analyze the positive and negative influence of different values of individual variables on the prediction results, so that it can more comprehensively analyze the accident causative factors.

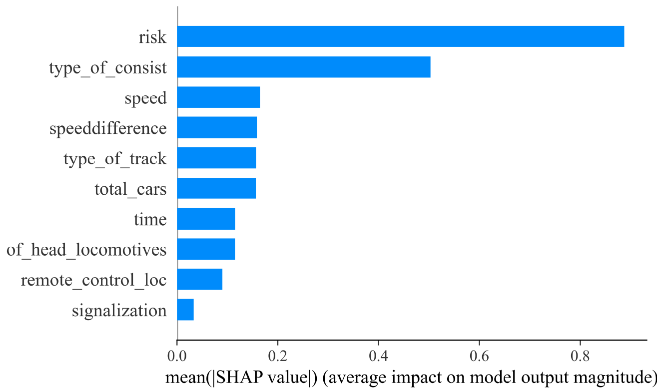
The SHAP method is introduced to explain the WOA-XGBoost model in this paper, and the important factors affecting the model's output are analyzed. Taking derailment and collision accidents as examples, we analyze the causes of railway accidents. Each row on the left side of the SHAP value summary plot represents a variable, arranged in descending order based on the magnitude of the variable's impact on the type of accident. The legend on the right side indicates the magnitude of the variable's value, and the horizontal coordinate indicates the SHAP value.

Fig. 8(a) summarizes the SHAP values of derailment accidents. Risk (including mechanical and electrical faults, track and subgrade factors, etc.) is the most influential feature of derailment accidents, followed by speed, the number of head locomotives, the total number of cars, and track class.

Fig. 8(b) summarizes the SHAP values of collision accidents. Risk (including communication and signal faults) is the most influential feature of collision accidents, followed by type of consist, speed, and speed difference (difference between speed and track speed limit).



(a) SHAP summary plot of derailment



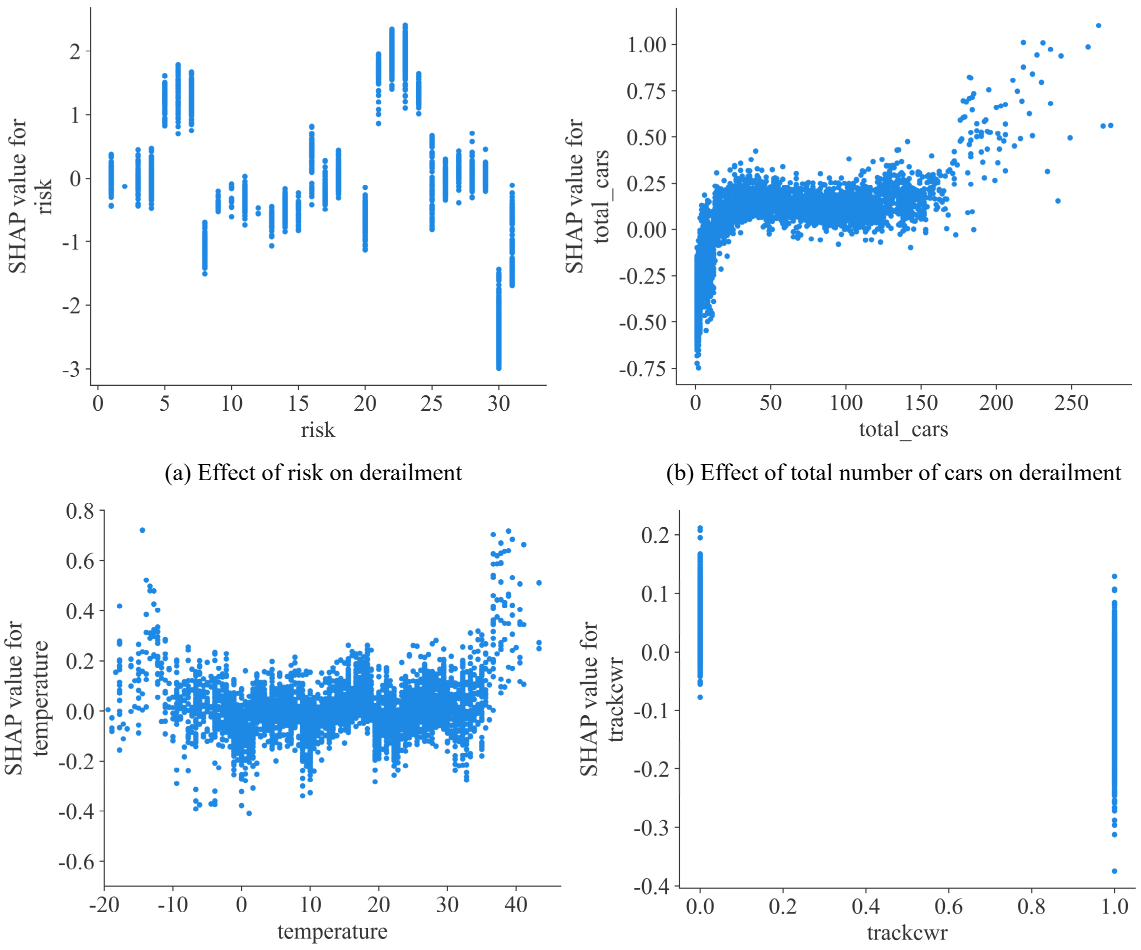
(b) SHAP summary plot of collision

Fig. 8. SHAP summary plot of derailment and collision

To better understand how variables affect the model's output, the SHAP summary plot is combined with the SHAP dependence plot, which can explain in more detail the effects of different values of variables on different types of accidents. In the SHAP dependence plot, each point represents an accident sample. The abscissa in the plot is the value of the feature, and the ordinate is the corresponding SHAP value.

Fig. 9 demonstrates the partial feature dependence plots of the derailment accident. Fig. 9(a) shows that the SHAP values for risk factors 5, 6, and 7 are positive. The result indicates that the failure of truck components, axles, and wheel components in the mechanical failure factor leads to derailment accidents. Risk factors 21, 22, 23, and 24 all belong to track and roadbed factors, and their SHAP values are all greater than 0, indicating that these factors lead to derailment.

Fig. 9(b) shows that the SHAP values of the number of cars in the range of 150-250 are significantly higher than those in the range of 1-150, indicating that the effect on derailment accidents is more significant when the number of cars is greater than 150. In addition, in the range [150,250], the SHAP value gradually increases with the number of cars, indicating that more cars increase the risk of collision.



(a) Effect of risk on derailment
(b) Effect of total number of cars on derailment
(c) Effect of temperature on derailment
(d) Effect of track welding situation on derailment

Fig. 9. SHAP feature dependence plots of derailment

Fig. 9(c) shows that the temperature corresponds to a positive SHAP value when it lies in the range of [-20, -10] and [35, 40]. This result indicates that high and low temperatures increase the likelihood of derailment accidents. Because the track material is affected by temperature, there is a risk of contraction of the track when the temperature is below -10°C. Moreover, when the external temperature exceeds 35°C, geometric deformation of the track may occur, further increasing the likelihood of derailment accidents.

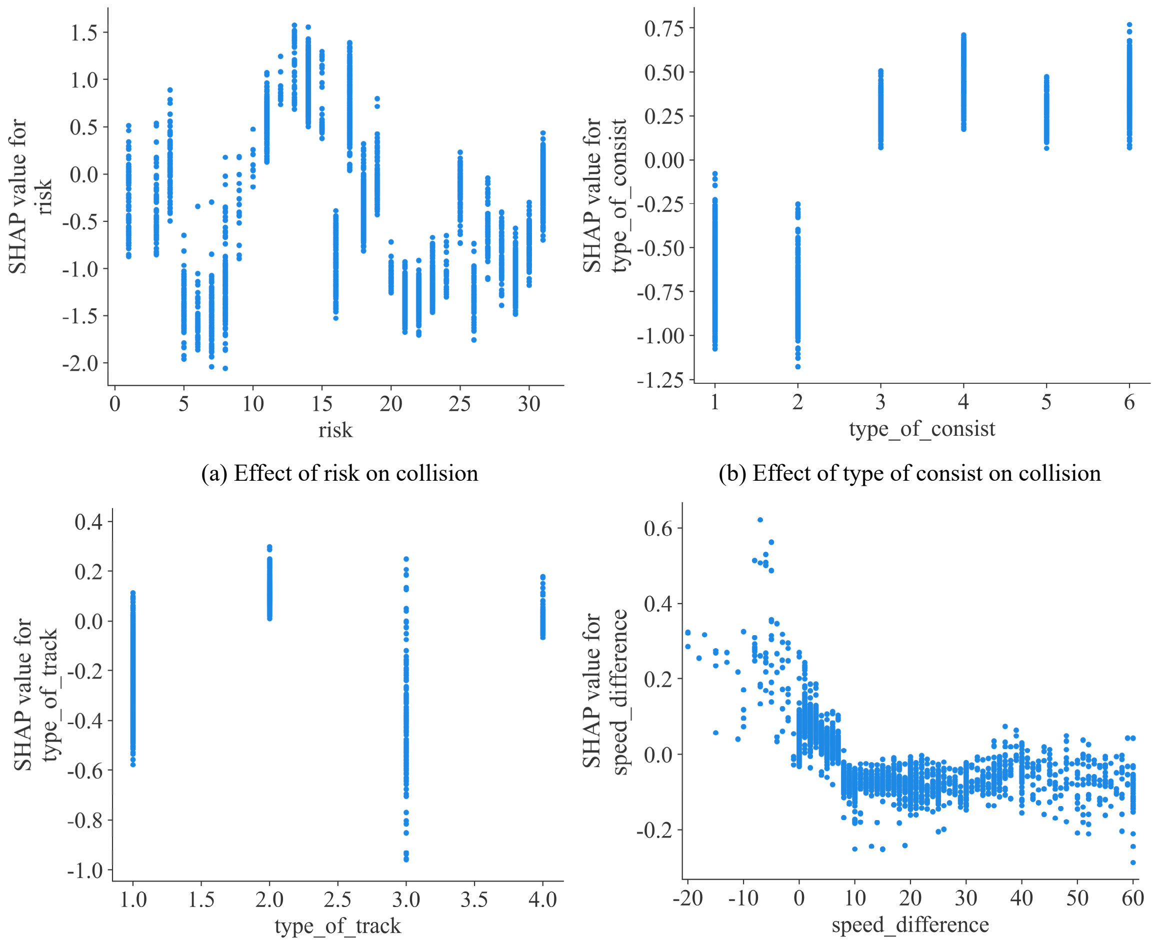
Fig. 9(d) shows that most SHAP values are positive when trackcwr=0, indicating that derailment accidents are more likely to occur on non-continuously welded tracks. Most of the corresponding SHAP values are negative when trackcwr=1, meaning that the continuously welded track plays a specific negative role in the occurrence of derailment accidents.

Fig. 10 shows the partial feature dependence plots of the collision accident. Fig. 10(a) demonstrates that the SHAP values exceed 0 when the risk factors are 11, 12, 13, 14, 15, and 17. These factors, namely improper use of braking devices, insufficient judgment ability and operational proficiency, improper use or disposal of signals, failure to comply with general switching rules, failure to comply with main track authority, and excessive speed, all contribute to collision accidents. Additionally, most of the corresponding SHAP values are negative when the risk factors are 1 and 4, indicating that the malfunction of the braking system and the coupler and draft gear system contribute to collision accidents.

Fig. 10(b) demonstrates that the SHAP values exceed 0 when the types of consist are 3, 4, 5, and 6. The result suggests that the likelihood of collisions increases when the consist type is locomotives, cars, work trains, or yards. Because the Federal Railroad Administration may have relatively less stringent safety requirements for shunting and work train operations.

Fig. 10(c) shows that the SHAP values corresponding to the track type of 2 are greater than 0, indicating that the track type of yard contributes to the occurrence of collision accidents. The characteristics of the yard track, such as its specific layout and traffic patterns, might increase the likelihood of collisions under certain circumstances. Moreover, most of the SHAP values corresponding to track type 4 are greater than 0, indicating that the track type of industry has a certain influence on causing collision accidents. For instance, the industrial track may involve complex operation procedures and frequent interactions, which could lead to a higher risk of collision accidents.

In this paper, a negative speed difference implies the train is overspeeding. Fig. 10(d) demonstrates that the SHAP values exceed 0 for a negative speed difference, indicating that overspeeding contributes to the collision. The majority of the SHAP values of the speed difference in the range of [0,7] are greater than 0, indicating that the train traveling speed close to the speed limit plays a role in the collision.



(a) Effect of risk on collision

(b) Effect of type of consist on collision

(c) Effect of type of track on collision

(d) Effect of speed difference on collision

Fig. 10. SHAP feature dependence plots of collision

D. Railway Safety Management Recommendations

The findings of this paper can provide some insights and references for railway safety agencies. The relevant authorities can reduce the occurrence of different types of railway accidents in the following ways.

(1) Infrastructure improvements. Railway management departments should improve track layouts and upgrade related systems and facilities to reduce derailment accidents. In addition, relevant authorities should improve warning devices, gates, lighting, and other equipment for highway railway crossings where accidents are frequent.

(2) Formulate a perfect equipment maintenance plan. Timely inspecting axles, sliding bearings, wheels, and other mechanical devices for failures can reduce derailment accidents. Since the failure of the braking system, coupler, and draft system can contribute to collisions, regular maintenance and overhaul are necessary. Besides maintaining mechanical devices, railway employees should strengthen the maintenance of track, roadbed, and other infrastructure, especially paying attention to the non-continuous welded portion of the track.

(3) Take timely measures in high-risk conditions. For example, in low-temperature conditions, clean up the snow in time and install snow-melting devices at turnout. Moreover, in high-temperature conditions, railway employees should check the geometric deformation of the track. Timely maintenance can effectively reduce the risk of derailment accidents. Employees should regularly use lighting devices in night conditions to reduce highway railway crossing accidents.

(4) Enhance employee training. Managers should develop comprehensive training programs to improve railway employees' ability to maintain faulty equipment. Relevant authorities need to emphasize the correct standard of operation regularly.

(5) Check the condition of employees before they go on duty. The railway department should strictly check the physical and mental conditions of the employees. Moreover, to ensure the employees' ability during work time, it should be strictly prohibited for employees to work under the influence of alcohol or excessive fatigue.

VI. CONCLUSIONS

A railway accident type prediction model based on WOA-XGBoost is proposed in this paper, and the SHAP method is introduced to analyze the causes of railway accidents. The following conclusions are obtained.

(1) SMOTE is used to synthesize new minority samples to address the problem of unbalanced data types of railway accidents in the paper. The distribution characteristics and change trends of the data synthesized by SMOTE are consistent with those of the original data. In this paper, the newly synthesized data is appended to the original data set to equalize the distribution of the accident dataset. Training the model with balanced data effectively reduces the impact

caused by the data imbalance problem and improves the model's predictive ability for minority-class samples.

(2) We improve the XGBoost algorithm by using the whale optimization algorithm. Then, the WOA-XGBoost model is used to predict six types of railroad accidents. The results suggest that WOA-XGBoost model's accuracy is 0.817, the recall is 0.732, the precision is 0.809, and the F1 value is 0.761. The accuracy of the WOA-XGBoost model is 6.7% higher than that of the DT model and 1.7% higher than that of the GBDT model. Regarding the F1 - score, the WOA-XGBoost model outperforms the DT model by 7.2% and the GBDT model by 3.1%. Overall, the WOA-XGBoost model outperforms the DT, GBDT, and XGBoost models in terms of performance.

(3) The SHAP method can not only mine the critical factors of diverse types of railroad accidents but also explore the impact and degree of each factor. This paper uses the SHAP method to analyze reasons for different types of accidents. The findings suggest that risk, speed, the number of cars, load, and temperature are important factors in derailment accidents. Specifically, failures of mechanical components such as bogies, axles, and wheels can lead to derailments. High temperatures (⩾ 35°C) and low temperatures (⩽-10°C) increase the risk of derailments. The impact on derailment is more significant when the total number of cars is greater than 150. The results indicate that risk, track type, and difference between speed and speed limit are significant factors in collision accidents. Improper use of braking devices in manual operation and failure to comply with shunting operation rules contribute to collision accidents. The tracks in the shunting yard area have a higher propensity for collisions. Train speeds approaching or exceeding track speed limits increase the likelihood of collisions. The SHAP method is suitable for analyzing different types of railway accidents, which is crucial for improving safety levels.

(4) This paper puts forward suggestions on railway safety management in five aspects. These suggestions can provide certain references for the relevant departments. Railways should enhance infrastructure facilities, maintenance programs, and staff training. Railway employees should implement preventive actions in high-risk situations such as low-temperature, high-temperature, and low-light conditions. The results can help railway departments formulate safety management strategies and prevent railway accidents.

(5) Future work will use the method presented in this paper to study more types of railway accidents, enabling a more detailed analysis of the causes of a wider range of accidents. In addition, although the method in this paper achieves good performance in accident prediction and causal analysis, it is highly dependent on the data quality. More advanced data mining methods will be combined with profound railway expertise to enhance the quality of railway accident data. Therefore, analyzing railway accidents by combining advanced data mining techniques is also a direction for further research.

References

[1] S. M. Mostafa, S. A. Salem, S. M. Habashy, "Predictive Model for Accident Severity," IAENG International Journal of Computer Science, vol. 49, no.1, pp110-124, 2022.

[2] Z. J. Zheng, P. Lu, D. Tolliver, "Decision Tree Approach to Accident Prediction for Highway–Rail Grade Crossings: Empirical Analysis," Journal Article, vol. 2545, no. 1, pp. 115-122, 2016.

[3] S. Wang, K. Li, F. Shao and H. Huang, "An improved principal component analysis method for railway accident prediction using uncertain data," 2016 6th International Conference on Electronics Information and Emergency Communication, 17-19 June 2016, Beijing, pp. 249-252.

[4] S. B. Jin, J. W. Lee, "Study on accident prediction models in urban railway casualty accidents using logistic regression analysis model," Journal of the Korean Society for Railway, vol. 20, no. 4, pp. 482-490, 2017.

[5] M. Z. H. Jesmeen, J. Hossen, S. Sayeed, C.K. Ho, K. Tawsif, A. Rahman, et al. "A survey on cleaning dirty data using machine learning paradigm for big data analytics," Indonesian J. Electr. Eng. Comp. Sci, vol. 10, no. 3, pp. 1234-1243, 2018.

[6] A. B. Parsa, A. Movahedi, H. Taghipour, S. Derrible, A. K. Mohammadian, "Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis," Accident Analysis & Prevention, vol. 136, pp. 105405, 2020.

[7] M. Guo, Z. Yuan, B. Janson, Y. Peng, Y. Yang, W. Wang, "Older pedestrian traffic crashes severity analysis based on an emerging machine learning XGBoost," Sustainability, vol. 13, no. 2, pp. 926, 2021.

[8] Q. B. Wang, W. L. Wang, R. Zhao, "Prediction of ship accident risk and type based on improved XGBoost," Journal of Safety and Environment, vol. 22, no. 6, pp. 3227-3234, 2022.

[9] X. Y. Zhou, P. Lu, Z. J. Zheng, D. Tolliver, A. Keramati, "Accident Prediction Accuracy Assessment for Highway-Rail Grade Crossings Using Random Forest Algorithm Compared with Decision Tree," Reliability Engineering & System Safety, vol. 200, pp. 106931, 2020.

[10] R. Bridgelall, D. D. Tolliver, "Railroad accident analysis using extreme gradient boosting," Accident Analysis & Prevention, vol. 156, pp. 106126, 2021.

[11] M. H. Zhong, W. L. Zhang, Y. R. Li, Z. F. Zhu, Y. Zhao, "GBDT based railway accident type prediction and cause analysis," Acta Automatica Sinica, vol. 48, no. 2, pp. 470-479, 2022.

[12] H. Meng, X. Tong, Y. Zheng, G. Xie, W. Ji, X. Hei, "Railway accident prediction strategy based on ensemble learning," Accident Analysis & Prevention, vol. 176, pp. 106817, 2022.

[13] H. L. Ding, R. Q. Wang, T. T. Chen, "A hybrid approach for modeling bicycle crash frequencies: Integrating random forest based SHAP model with random parameter negative binomial regression model," Accident Analysis & Prevention, vol. 208, pp. 107778, 2024.

[14] R. Bridgelall, D. D. Tolliver, "Railroad accident analysis by machine learning and natural language processing," Journal of Rail Transport Planning & Management, vol. 29, pp. 100429, 2024.

[15] J. Liu, Y. S. Wang, C. B. Deng, C. B. Deng, "A new procedure for assessing and predicting the severity of accidents: A case study on freight-train derailments," Journal of Loss Prevention in the Process Industries, vol. 94, pp. 105511, 2024.

[16] N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," vol. 16, no. 1, pp. 321-357, 2002.

[17] S. Arwatchananukul, R. Saengrayap, S. Chaiwong, "Fast and Efficient Cavendish Banana Grade Classification using Random Forest Classifier with Synthetic Minority Oversampling Technique," IAENG International Journal of Computer Science, vol. 49, no.1, pp46-54, 2022.

[18] F. Femi-Oyewole, V. Osamor, D. Okunbor, "Ensemble Machine Learning Approach For Identifying Real-Time Threats In Security Operations Center," IAENG International Journal of Computer Science, vol. 51, no. 12, pp2094-2122, 2024.

[19] Y. Xie, G. Mi, "Extreme Gradient Boosting Based Railway Accident Type Prediction and Influencing Factors Analysis," 2024 6th International Conference on Communications, Information System and Computer Engineering (CISCE), 10-12 May 2024, Guangzhou, China, pp. 978-982.

[20] T. Chen, C. Guestrin, "Xgboost: A scalable tree boosting system," Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 13-17 August 2016, San Francisco, pp. 785-794.

[21] L. Wang, H. Li, C. Du, W. Hong, "Optimizing hydrogen yield in sorption-enhanced steam methane reforming: A novel framework integrating chemical reaction model, ensemble learning method, and whale optimization algorithm," Journal of the Energy Institute, vol. 114, pp. 101649, 2024.

[22] S. Mirjalili, A. Lewis, "The whale optimization algorithm," Advances in engineering software, vol. 95, pp. 51-67, 2016.

[23] L. H. Kang, X. Q. Li, L. L. Wu, "Predicting Stock Closing Price with Stock Network Public Opinion Based on AdaBoost-IWOA-Elman Model and CEEMDAN Algorithm," IAENG International Journal of Computer Science, vol. 49, no.4, pp975-984, 2022.

[24] K. Aas, M. Jullum, A. Løland, "Explaining individual predictions when features are dependent: More accurate approximations to Shapley values," Artificial Intelligence, vol. 298, pp. 103502, 2021.

[25] S. Mangalathu, S. H. Hwang, J. S. Jeon, "Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach," Engineering Structures, vol. 219, pp. 110927, 2020.

[26] FRA (Federal Railroad Administration), "Rail Equipment Accident/Incident Data," (Online). Available: Accident/Incident Dashboards & Data Downloads | FRA (dot.gov)

[27] FRA (Federal Railroad Administration), "FRA Guide for Preparing Accident/Incident Reports 2003," (Online). Available: FRA Guide for Preparing Accident/Incident Reports 2003 (Effective May 1,2003) | FRA (dot.gov)

[28] C. Y. Lin, M. Rapik Saat, C. P. Barkan, "Quantitative causal analysis of mainline passenger train accidents in the United States," Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, vol. 234, no. 8, pp. 869-884, 2020.

[29] X. Chen, X. Ma, L. Jia, Z. Zhang, F. Chen, R. Wang, "Causative analysis of freight railway accident in specific scenes using a data-driven Bayesian network," Reliability Engineering & System Safety, vol. 243, pp. 109781, 2024.