Data-Driven Intervention Strategies for Mitigating Illegal Wildlife Trade: a Case Study of the United States

Pan Xue, Tianchang Zhou, Hui Sun, Jihao Song, Xiaoliang Guo and Zhiwei Shao

July 15, 2024

Data-Driven Intervention Strategies for Mitigating Illegal Wildlife Trade: A Case Study of the United States ⋆

Pan Xue[1]0000-0001-6681-3248, Tianchang Zhou[2], Hui Sun[1], Jihao Song[1], Xiaoliang Guo[1], and Zhiwei Shao[1]

[1] School of General Education, Xi'an Eurasia University, Xi'an, Shaanxi 710065, China

[2] School of Communication and Information Engineering, Xi'an University of Posts Telecommunications, Xi'an, Shaanxi 710121, China

Abstract. Given the escalating threat posed by illegal wildlife trade to global biodiversity and ecological equilibrium, this study endeavors to forecast the potential effects of a five-year, data-driven initiative on curbing illicit wildlife trade within the United States over the forthcoming half-decade.This paper establishes a complete system of evaluation indicators to assess the progress of the program, and establishes a system dynamics model based on $Vensim$ modeling and simulation to predict a 14% increase in law enforcement after the implementation of the data-driven program. A weighted optimization model with $ARMA$ and multiple linear regression weights was established to predict the volume of illegal wildlife trade in the next five to ten years with or without intervention, the weighted prediction model underwent optimization using the particle swarm algorithm, resulting in enhanced convergence speed and accuracy of the model's solution.

Keywords: Illegal wildlife trade · Evaluation index system · Weighted optimization prediction · Particle swarm algorithm.

The illegal wildlife trade is a global problem with far-reaching negative impacts on biodiversity, the societyeconomy and the environment. According to reports by organizations such as the World Conservation Union ($IUCN$) and the World Wide Fund for Nature (WWF), the illegal wildlife trade includes the illegal hunting, transporting and selling of wildlife and their products, and encompasses trade in everything from ivory, rhinoceros horns and tiger skins to rare plants and timber.

In the case of the ivory trade, for example, there is a lack of effective law enforcement control over the entire trade chain, from Africa to transit countries

⋆ This work is supported in part by the Basic Research Program of Natural Science of Shaanxi Province (No.2022JM-055), the Science Research Program of Xi'an Eurasia University (No.2022GCPY01), and also by the school-level projects: Higher Mathematics Visualization teaching research under digital background (No. 2023YB015).

to final consumer markets. Combating the illegal trade in wildlife is a complex global issue, and by strengthening international cooperation, making full use of modern science and technology and increasing public participation, we can hope to combat this criminal activity more effectively and protect the Earth's precious wildlife resources. In the paper [1], the authors present specific measures for a data-driven project to combat illegal wildlife trade. In the paper [2], the author compiled an indicator evaluation system table, deeply discussed the relationship between each major indicator and illegal wildlife trade, predicted the impact of the implementation of data-driven projects, and explored the implementation of data-driven work to combat illegal wildlife. feasibility of trade. The authors aim to provide a comprehensive understanding of the complexities involved in solving this problem and shed light on the various stakeholders and stakeholders, and propose a new approach to solving this type of problem model. For more information on wildlife conservation, one can refer to references [3],[4]and[5].

1 Index Evaluation System

The data in this article come from the World Bank website, Wind database and National Bureau of Statistics, as well as journal search websites such as *Knowledge.com*, *PubMed*, *Google*, and *Scholar*.

Based on Python data crawler, the keyword illegal wildlife is used to search the government in the field of academic journals. The United States appears most frequently, so we believe that the United States is the most representative example.

Using illegal wildlife trade as the keyword, search on journal search websites such as *PubMed*, *Sci*, *Hub*, *Google*, *Scholar*, etc., and then conduct a secondary search using illegal wildlife trade as the keyword. American keywords. Finally, the 22 indicators that appear most frequently in American academic journals were obtained, which were divided into three categories: power, resources, and interest based on the *NLP* model, as shown in Table 1. Based on these indicators, an indicator evaluation system model was established to analyze the relationship between main indicators and illegal wildlife trade.

Table 1: Dimension Categorization of Indicators (Part 1)

Power	**Resource**
Legal rights	Imports of ores and metals
High tariff rates	Gross domestic savings
Average bound rates	Human capital index
Tax share	Education financing
Customs and other import duties	Net energy import rate
Other taxes	Inventory change

Table 2: Dimension Categorization of Indicators (Part 2)

Interest	**Trade**
Population growth	Percentage of Cargo in Wildlife Trade Seized
System coverage	Estimated Annual Number of Elephants Illegally Killed
Credit system coverage	Worldwide Seizures of Wildlife Products
Share of IT services	Number of Poaching Incidents in Africa
Gini coefficient	Species Diversity Involved in Wildlife Trafficking

The scores of these 22 indicators were quantified on a scale of 1-10 and the weights were multiplied by the frequency of occurrence to obtain a system of evaluation indicators.Calculate the frequency of occurrence weighted score and denote

$$W_s = S_i \times \ F_i \tag{1}$$

and normalize

$$N_i = \frac{W_S - W_{\min}}{W_{\max} - W_{\min}} \tag{2}$$

where i represents the index of the indicator, F_i represents the frequency of the indicator i appearing in relevant research, S_i represents the quantitative score, W_S is the weighted score, Based on this quantification, the evaluation index system is obtained.

2 Multiple Linear Regression Model

2.1 Data-driven Project

Based on reptile and world wildlife big data analysis According to the big data obtained by Reptile and the analysis of the World Wildlife Crime Report released by the United Nations, the United States not only plays an important role in combating the global illegal animal trade, but is also in the vortex of global wildlife crime[6]. The U.S. government can implement data-driven programs to combat illegal wildlife trade by:

1. Strengthen law enforcement: The U.S. government already has laws such as the Endangered Species Act and the Wildlife Conservation Act, which can further increase the cost of violating the law and strengthen law enforcement.
2. International cooperation: As a signatory to CITES, we promote international cooperation to jointly combat illegal wildlife trade.
3. Funding and resource investment: Governments should allocate financial resources to support research, surveillance, law enforcement, and public education.
4. Raise public awareness: Raise public awareness of illegal wildlife trade issues through education programs and public events.

5. Research and analysis support: Use data collection and analysis capabilities to monitor trade patterns and changes in wildlife populations to guide policy development and resource allocation.
6. Cost-benefit analysis: Analyze the benefits of investment in combating illegal wildlife trade and demonstrate long-term benefits to the public.

Through these measures, the U.S. government can effectively combat illegal wildlife trade, protect biodiversity, and maintain global ecological balance and human welfare.

2.2 Principal Component Analysis

First, the KS test was performed on the data to remove outliers, and Newton interpolation was used to supplement the data. Eliminate data that has no linear relationship and does not conform to the normal distribution, and apply indicator principal component analysis and data dimensionality reduction to extract the main features of the data[7]. Let

$$KMO = \frac{\sum_{i=1}^{p}\sum_{j=1,j\neq i}^{p} r_{ij}^2}{\sum_{i=1}^{p}\sum_{j=1,j\neq i}^{p} r_{ij}^2 + \sum_{i=1}^{p}\sum_{j=1,j\neq i}^{p} a_{ij}^2} \tag{3}$$

where p represents the number of variables, r_{ij} represents the correlation coefficient between variables i and j, and a_{ij} represents the average correlation coefficient between variables i and j. We denote

$$\chi^2 = \frac{(n - 1 - \frac{2p+5}{6}) \times \sum_{i=1}^{p}(\sum_{j=1,j\neq i}^{p} \ln(R_{ij}^2))}{1 + \frac{p(p-1)}{6(n-1)}} \tag{4}$$

where n is the sample size, p is the number of variables, R_{ij} is the correlation coefficient between the variables i and j, and χ^2 is the chi-square statistic of the Bartlett's test of sphericity. By analyzing the results, we chose to perform

Table 3: Results of KMO test and Bartlett's test of sphericity

Index Name	**illicit trade**	**the powers**	**resources**	**interest**
KMO value	0.633	0.332	0.511	0.528
chi-square pproximation	87.933	7.192	91.593	103.863
df	6	6	10	15
P	0.000*	0.532	0.000***	0.000***

principal component analysis on the two variables of resources and illegal wildlife trade for dimensionality reduction. This selection was based on their performance in terms of correlations and significance levels in factor analysis. Taking the illegal wildlife trade with the highest KMO value as an example, we calculated the contribution rate as the principal component as the weight to obtain the indicators with the largest influencing factors[8]. Denote

$$q_{ij} = \frac{(1 + y_i - y_j^2)^{-1}}{\sum_{k \neq l} (1 + y_k - y_l^2)^{-1}} \tag{5}$$

where y_i denotes the data point i in the low-dimensional space, and q_{ij} is the conditional probability between the data point i and j. In order to select principal components, we calculate the contribution of the principal components η_i and the cumulative contribution γ_i:

$$\eta_i = \frac{\lambda_i}{\sum_{j=1}^{p} \lambda_j} \tag{6}$$

$$\gamma_i = \sum_{j=1}^{i} \eta_j \tag{7}$$

and calculate the principal component Y_{ik} score for each sample:

$$Y_{ik} = \sum_{j=1}^{p} Z_{ij} v_{jk} \eta_k. \tag{8}$$

Based on the scree plot analysis, two principal components were selected and the following factor loading coefficient table was obtained. Among them, P,

Table 4: Factor Loading Coefficients

	Principal Component 1	**Principal Component 2**	**Communalities**
P	0.723	0.684	0.991
E	-0.971	0.181	0.975
W	0.925	-0.07	0.861
N	0.891	-0.285	0.875

E, W and N respectively represent Percentage of Cargo in Wildlife Trade Seized, Estimated Annual Number of Elephants Illegally Killed, Worldwide Seizures of Wildlife Products, Number of Poaching Incidents in Africa. According to the factor loading coefficient table, multiple indicators of illegal wildlife trade show significant correlations with principal component 1 and principal component 2, revealing the complex relationship between illegal trade and different potential factors.

For the indicator of rights, we choose to introduce person correlation for analysis.We denote the sample covariance as

$$Cov(x, y) = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{n - 1} \tag{9}$$

and the sample standard deviation as

$$S_{\mathrm{x}} = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n-1}} \tag{10}$$

where n is the number of samples, x_i and y_i are the values of the corresponding indicators, and $\bar{x}$ and $\bar{y}$ are the mean values of the corresponding indicators.

$$r_{xy} = \frac{\mathrm{Cov}(X,Y)}{S_X S_Y} \tag{11}$$

Through the dimensionality reduction results, we found that there is a clear correlation between illegal wildlife trade indicators, motivation indicators, resource efficiency indicators, and time indicators. According to the description of the three first-level indicators in the U.S. Statistical Yearbook, and on the basis of the description of the three main indicators in ACS, we select other parts of the domestic economy and energy use as new indicators to establish a broader evaluation indicator. system. For newly introduced indicators, a new indicator evaluation system will be established.

Table 5: New Indicator Evaluation System

Year	**Trade**	**Right 1**	**Right 2**	**New Energy Indicator**	**Interest**
2000	-1.361099688	193.2577842	332.0528762	0.973167251	1.115770614
2001	-1.256141762	-86.5564173	513.274681	0.830152497	0.510649735
...	...	...	...	...	...
2021	1.20330976	679.1377848	190.888172	0.430271389	-0.048572938
2022	2.20330976	379.1377848	190.888172	0.430271389	-0.048572938

2.3 Model Building

The following multiple linear regression model is established here, with illegal animal trade as the dependent variable$Y$, rights 1, rights 2, interest and energy as independent variablesX_1, X_2, X_3, X_4 as follows,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon \tag{12}$$

where β_0 is the intercept, $\beta_1, \beta_2, \beta_3, \beta_4$ is the coefficient of the independent variable, and ε is the error term.

The prediction results are shown in Figure 22. The fitting results of the model show that the mean square error (MSE) of the multiple linear regression model is 0.073, and the value of R^2 (coefficient of determination) is 0.781, indicating that the model explains the variation between illegal animal trade and other indicators better.

Table 6: Multiple Linear Regression Coefficients

Coefficient	Description	Value
β_0	Intercept	-0.056
β_1	Right 1	3.357×10^{-4}
β_2	Right 2	1.860×10^{-4}
β_3	Interest	-0.466
β_4	Energy	0.584

2.4 Modeling System Dynamics to Assess Project Impacts

We build the following dynamic system,

$$\frac{dS}{dt} = rS\left(1 - \frac{S}{K}\right) - \alpha SL \tag{13}$$

$$\frac{dD}{dt} = \beta D\left(1 - \frac{P}{P_0}\right) \tag{14}$$

$$\frac{dL}{dt} = \gamma\left(\frac{D}{D_0} - 1\right) \tag{15}$$

$$\frac{dP}{dt} = \epsilon\left(\frac{S}{K} - \frac{D}{D_0}\right) \tag{16}$$

In the formula, the regeneration rate is (r), the carrying capacity is (K), the market demand elasticity is β, the execution impact coefficient is γ, and the price adjustment rate is ϵ. Based on Vensim system dynamics modeling and simulation, it is expected that the US government's law enforcement intensity will increase by 14% after the implementation of the project.

The project proposal is based on data-driven decision-making, cost-benefit analysis and other methods to provide strong support for the US government to reduce illegal wildlife trade. Overall, the project is consistent with the goals and responsibilities of the U.S. government, and has urgency and potential value for implementation.

3 Weighted Optimization Prediction Model

3.1 ARMA Model

First, after making the first difference of the data, a new stationary sequence can be obtained. Use Eviews to perform the unit root test. According to the results, it can be seen that the sequence is a stationary time series.

Determine the number of differentials,

$$Y_t^* = (1 - B)^d Y_t \tag{17}$$

Table 7: Unit Root Test Result

Variable	Order	t	P	AIC	1%	5%	10%
−1.260099588	0	−0.867	0.725	1.391	−3.839	−2.954	−2.627
	1	−2.919	0.024**	2.541	−3.589	−2.954	−2.627
	2	−3.429	0.005***	−1.727	−3.784	−3.405	−2.282

where Y_t^* is the differenced sequence and B is the lag operator. Plot the ACF and $PACF$ of the differenced sequence and determine the order of the AR and MA. Let

$$\rho_k = \frac{\gamma_k}{\gamma_0} \tag{18}$$

where γ_k is the ACF and ρ_k is the autocorrelation coefficient. Selection of the number of seasonal differences

$$Y_t^* = (1 - B^s)(1 - B)^d Y_t \tag{19}$$

After comparison, we found that the $arima(0, 1, 1)$ model has the best prediction effect, and the results are as follows.

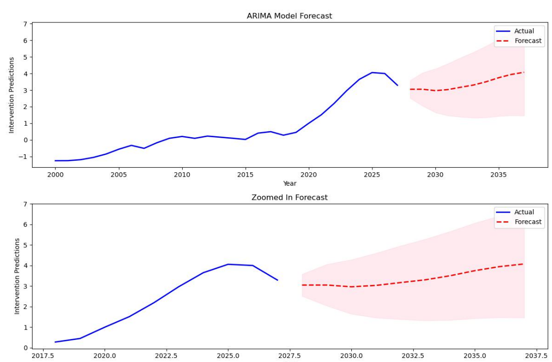

Fig. 1: ARIMA (0,1,1) forecast model

3.2 Solving the Weighted Prediction Model Based on Particle Swarm Optimization

To improve the accuracy of forecasts, we use the $ARIMA$ (0, 1, 1) model and linear regression for data-weighted forecasts. Based on a weighted optimization algorithm, the accuracy of the overall prediction model is improved by adjusting the weight of each model prediction[9].

We define the weighted predicted value,

$$J = x_1\hat{y}_1 + x_2\hat{y}_2 + \cdots + x_m\hat{y}_m \tag{20}$$

among them,

$$\sum_{i=1}^{m} x_i = 1, \quad x_i \geq 0, \quad (i = 1, 2, \cdots, m)$$

$$e_{ij} = \hat{y}_{ij} - y_j$$

represents the error between the i prediction method and the j simulated value of historical data, and denoted

$$E_j = \begin{pmatrix} e_{1j}^2 & e_{1j}e_{2j} & \cdots & e_{1j}e_{mj} \\ e_{1j}e_{2j} & e_{2j}^2 & \cdots & e_{2j}e_{mj} \\ \vdots & \vdots & \ddots & \vdots \\ e_{1j}e_{mj} & e_{2j}e_{mj} & \cdots & e_{mj}^2 \end{pmatrix} \tag{21}$$

Then, the objective function can be defined as

$$\begin{aligned} L &= \sum_{j=1}^{n}(\hat{J}_j - y_j)^2 = \sum_{j=1}^{n}(x_1\hat{y}_{1j} + x_2\hat{y}_{2j} + \cdots + x_m\hat{y}_{mj} - y_j)^2 \\ &= \sum_{j=1}^{n}(x_1e_{1j} + x_2e_{2j} + \cdots + x_m e_{mj})^2 = \sum_{j=1}^{n} x^T E_j x. \end{aligned} \tag{22}$$

Thus, we obtain a multi-objective optimization model based on error minimization. Among them, x_i represents the weight of the i-th prediction method, $\hat{y}_{ij}$ represents the prediction value of the i-th method for the j-th historical data, y_j represents The actual value of the j-th historical data, e_{ij} represents the prediction error, E_j represents the error matrix, and L represents the objective function.

Since directly solving the weighted model will face huge computational workload and consume a lot of time, we use the particle swarm optimization algorithm to make up for some defects in the model establishment process. The algorithm has the characteristics of fast convergence speed and simple parameter setting, which can effectively reduce the amount of computation. Through the planning model, each particle can determine its individual optimal solution, thereby obtaining the global optimal solution. We use the minimum error as the objective function and the weight sum as 1 as the constraint, and use the multi-objective particle swarm optimization algorithm to solve the weighted model.Here, we introduce particle position and distance. For convenience, we provide the following definitions and formulas:

$$z_{ij} = z_{ij} + d_{1ij} \times \mathrm{rand}(0) \times (\mathrm{pbest}_{ij} - x_{ij}) \quad + d_{2ij} \times \mathrm{rand}(0) \times (\mathrm{pbest}_{ij} - x_{ij}) \tag{23}$$

$$z_{\mathrm{ij}}=\omega \times z_{\mathrm{ij}}+d_{1 \mathrm{ij}} \times rand(0) \times\left(\text { pbest }_{i j}-x_{i j}\right)+d_{2 \mathrm{ij}} \times \text { rand }(0) \times\left(\text { pbest }_{i j}-x_{i j}\right) \tag{24}$$

Pareto optimal solution set:

$$\text{Pareto Rank} \quad R(i)=\min \{R(j) \mid j \neq i, \operatorname{fitness}(i) \prec \operatorname{fitness}(j)\}$$

$$\text{Crowding Distance} \quad D(i)=\max \{d_j(i) \mid j=1,2, \ldots, n\}$$

$$d_j(i)=\min \{|f_{j,min}(i)-f_{j,min}(k)|\ | k \neq i\}$$

Linear decreasing weight strategy,

$$\omega^{(t)}=\left(\omega_{ini}-\omega_{end}\right)\left(G_k-\mathrm{g}\right) / G_k+\omega_{end} \tag{25}$$

where G_k is the maximum number of iterations, ω_{mi} is the initial inertia weight,ω_{end} and is the inertia weight when it iterates to the maximum evolution number. The results after MOPSO optimization are shown in Figure 2.

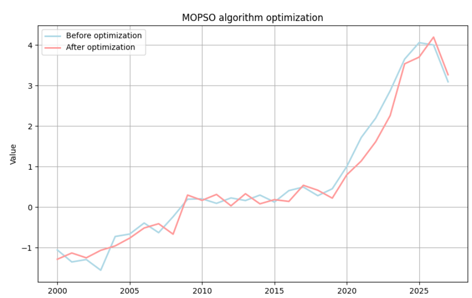

Fig. 2: Comparison of the accuracy of three models

This paper uses root mean square error ($RMSE$), mean absolute percentage error ($MAPE$) and coefficient of determination (R^2) as evaluation indicators to test the model prediction results. Assume n is the sample size, y_i is the actual observed value, $\hat{y}_i$ is the corresponding predicted value,

$$RMSE=\sqrt{\frac{1}{n} \sum_{i=1}^n\left(y_i-\hat{y}_i\right)^2} \tag{26}$$

The above formula is the root mean square error ($RMSE$), which is expressed as the average arithmetic square root of the sum of square errors between the predicted value and the true value.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100 \tag{27}$$

The above formula is the mean absolute percentage error (MAPE), which represents the average absolute value of the ratio of the error between the predicted value and the true value and the true value. Let

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}. \tag{28}$$

The above formula is the coefficient of determination (R^2), which represents the ratio of the sum of squares of errors to the total sum of squares. The value range is $0 \leq R^2 \leq 1$.

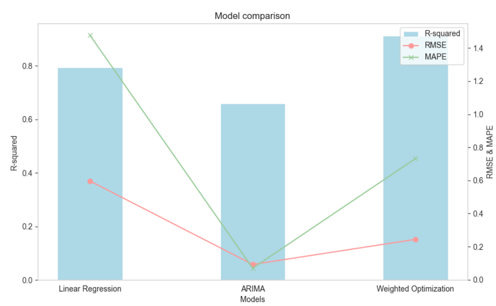

Fig. 3: Comparison of the accuracy of three models

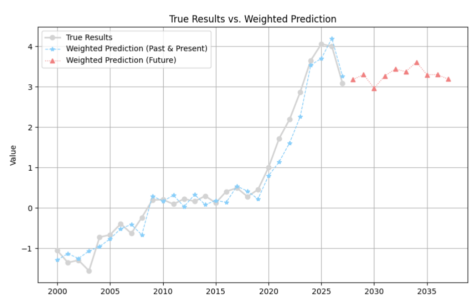

Fig. 4: Optimized model prediction

In summary, the prediction model based on weighted optimization is better than a single model. Since ARMIA's prediction of long-term data is not effective, we use the Newton interpolation method to train the model with the predicted data as a new data set. It is predicted that after the implementation of this project, the global illegal wildlife trade will decrease by 29% in 2033, as shown in Figure 4. In the long run, the implementation of this project will effectively reduce illegal wildlife trade activities and help achieve the goal of protecting global species diversity.

References

1. Jin L, Yesheng Z, Peng L, et al. Optimal design of recycled concrete mix proportion based on LSTM-MOPSO algorithm [J]. Journal of Taiyuan University of Technology, 2024,(5):1-12.
2. Kumar S G, Prakash C S, Bhim S, et al., Detection of fraudulent in contraband Shahtoosh trade: Unfair deal in illicit wildlife market[J]. Forensic Science International: Genetics,2022,59102713-102713.
3. Hao J T, William S S, Roman L C, Economic value of illegal wildlife trade entering the USA[J]. PLoS one, 2021, 16(10). DOI:10.1371/journal.pone.0258523
4. Simon D, Security and Conservation: The Politics of the Illegal Wildlife Trade Security and Conservation: The Politics of the Illegal Wildlife Trade By Rosaleen Duffy New Haven[M]. CT : Yale University Press, 2022.
5. Bernardo B, Gaetano C, Vincenzo D, Security and Privacy Violation Rules in Trigger-Action IoT Platforms with NLP Models.[J] IEEE Internet of Things Journal, 2022,6(10): 5607-5622. D10.1109/jiot.2022.3222615
6. Aref M-S, Waed T. K, Mohamadal B, Tarek L, Development of multiple linear regression, artificial neural networks and fuzzy logic models to predict the efficiency factor and durability indicator of nano natural pozzolana as cement additive.[J] Journal of Building Engineering, 2022, 52(15): 104475. DOI:https://doi.org/10.1016/j.jobe.2022.104475
7. Li B, Habbal F, Ortiz M, Optimal transportation meshfree approximation schemes for fluid and plastic flows.[J] International Journal for Numerical Methods in Engineering, 2010, 83(12): 1541-1579. DOI:10.1002/nme.2869
8. Zhang L, Zhicheng J, Dayong L, et al., A correlation flow data processing method based on K-means clustering analysis and multiple linear regression[J]. Petroleum Pipe and Instrument, 2024, 10 (01): 52-56+62. DOI:10.19459/j.cnki.61-1500/te.2024.01.010
9. Yong K, Research on Feature Selection and Model Optimization Algorithms for Random Forests[D]. Harbin Institute of Technology, 2008.