

# Composition analysis and identification of ancient glass products

Jiaming Li<sup>#,\*</sup>, Yongsen Geng<sup>#</sup>, Fanyu Meng<sup>#</sup>

Institute of Geophysics & Geomatics, China University of Geosciences, Wuhan, hubei, 430000  
China

\*Corresponding author: 3348349877@qq.com

<sup>#</sup>These authors contributed equally.

**Abstract:** In this paper, we take the composition analysis and identification of lead-barium glass and high-potassium glass as the starting point, and analyze and solve the problems related to the weathering and chemical composition of the two glass types through machine learning and random forest. In this paper, the mean value and content interval of the given data are initially processed, and the classification basis can be roughly determined. Then the random forest classification model is used to classify them, and the importance of each feature is obtained. The K-mean clustering algorithm is adopted to solve the problem, and the reasonableness test is performed using the fuzzy clustering algorithm, and the results of both are similar, indicating good reasonableness. Finally, a random forest model was used to train the new sample data, and then a sobol sensitivity test was performed on them according to the classification results to obtain a higher sensitivity of alumina to the classification results of high-potassium glass and a higher sensitivity of lead oxide to the classification results of lead-barium glass.

**Keywords:** Machine Learning; Random Forest; Clustering Algorithm; Glass Component Identification.

## 1. Introduction

Glass is a valuable physical evidence of the early trade exchanges along the Silk Road in China. Glass was introduced into our country from abroad, our country to absorb technology, local materials, and foreign glass products are similar in appearance but different chemical composition. Glass in the refining, due to different fluxes, its main chemical composition is different, lead barium glass with a high content of lead oxide, barium oxide[1]. Potassium glass is fired with a high content of potassium substances such as grass wood ash as a flux. This paper analyzes the classification rules of high potassium glass and lead-barium glass; for each category, we select the appropriate chemical composition to classify them into subcategories, give the specific classification methods and results, and analyze the rationality and sensitivity of the classification results[2].

To investigate the classification pattern of high potassium glass and lead-barium glass, we need to investigate which chemical components play a key role in their classification, so we need to do statistics and analysis of their content distribution. Then the random forest classification model was used to test the classification and obtain the importance of each feature, so as to analyze which feature plays an important role in its classification[3]. Next, the subclasses of high potassium glass and lead-barium glass are divided, because the division is based on the sample set and the classification cannot be trained, so the K-mean clustering algorithm is adopted to solve the problem, and the fuzzy clustering algorithm is used to test the rationality, and if the results obtained by the two algorithms are similar, it means that the rationality is good; finally, the random forest model is used to train the new sample data, and then according to the classification results, the sobol sensitivity test is performed.

## 2. Model assumptions and notation

### 2.1 Assumptions [4-6]

1 It is assumed that the only factors that may affect the degree of weathering are the type, decoration, and color of the glass artifacts.

2 It is assumed that the small amount of data does not have a significant impact on the study of this problem.

3 It is assumed that the change in the chemical content of the glass artifacts is only affected by the degree of weathering

4 It is assumed that all chemical components in the glass artifacts have been detected.

### 2.2 Notations

Important notations used in this paper are listed in Table 1.

Table 1 Notations

Symbols	Definition
$A$	Real Value
$T$	Theoretical Value
$\chi^2$	Cardinality
$X_i$	The $i$ th value of the $X$ series
$Y_i$	$i$ -th value of $Y$ -series
$r$	Pearson coefficient
$\bar{X}$	$X$ Sequence average
$\bar{Y}_i$	$Y$ Sequence average
$P$	Significance

## 3. Model construction and solving

### 3.1 Analysis of the classification law of high potassium glass and lead-barium glass

In the process of classification, each feature is independent of the data between them, which just meets the conditions of the random forest classification model, and the algorithm is applied to the classification + regression model in general, which not only prevents overfitting and fast classification, but also can estimate the importance of features in the classification and has a strong resistance to noise interference. Subclassification of glass artifacts is unsupervised, and k-means clustering (k-means) algorithm can be used, which is a clustering algorithm based on sample set division and can accurately and effectively classify glass artifacts according to their common characteristics[7].

### 3.2 Model solving

#### 3.2.1 Data analysis and random forest model validation

Preliminary analysis of the data shows that silica content has a great influence on its classification, silica content lower than 50% will only be lead-barium glass, higher than 80% will only be high potassium glass, and higher silica content is more likely to be high potassium glass. Secondly, potassium oxide and lead oxide also play a great role in its classification, and the following Table 2 to Table 5 shows that potassium oxide has a higher content in high potassium glass and lead oxide has a higher content in lead-barium glass.

Table 2 High potassium glass silica content

Ingredient content	Quantity	Ratio
50-60	2	16.67%
60-70	7	58.33%
70-80	2	16.67%
80-90	1	8.33%
Total	12	100.00%

Table 3 High potassium glass potassium oxide content

Ingredient content	Quantity	Ratio
0	1	8.33%
5-10	6	50.00%
10-15	5	41.67%
Total	12	100.00%

Table 4 Lead barium glass potassium oxide content

Ingredient content	Quantity	Ratio
30-40	2	15.38%
40-50	0	0.00%
50-60	2	15.38%
60-70	7	53.85%
70-80	2	15.38%
Total	13	100.00%

Table 5 Lead barium glass calcium oxide content

Ingredient content	Quantity	Ratio
0-5	3	27.27%
5-10	8	72.73%
Total	11	100.00%

In order to verify the analysis, the random forest model will be used next to analyze the data, take the sampling with put-back to construct sub-datasets, ensure the same order of magnitude between different sub-sets (elements can be repeated), use the sub-datasets to construct sub-decision trees, put the data to be predicted into each sub-decision tree, output one result for each sub-decision tree, count the voting results of the sub-decision trees, and the one with more votes is the random The output result of the forest[8]. In this problem, the variables are first assigned to 0 for high potassium and 1 for lead-barium, 0 for no weathering and 1 for weathering, and 0 to 13 for each feature component in Form II. The random forest analysis based on the data information of each feature quantity in the table yields[9]as shown in Figure 1.

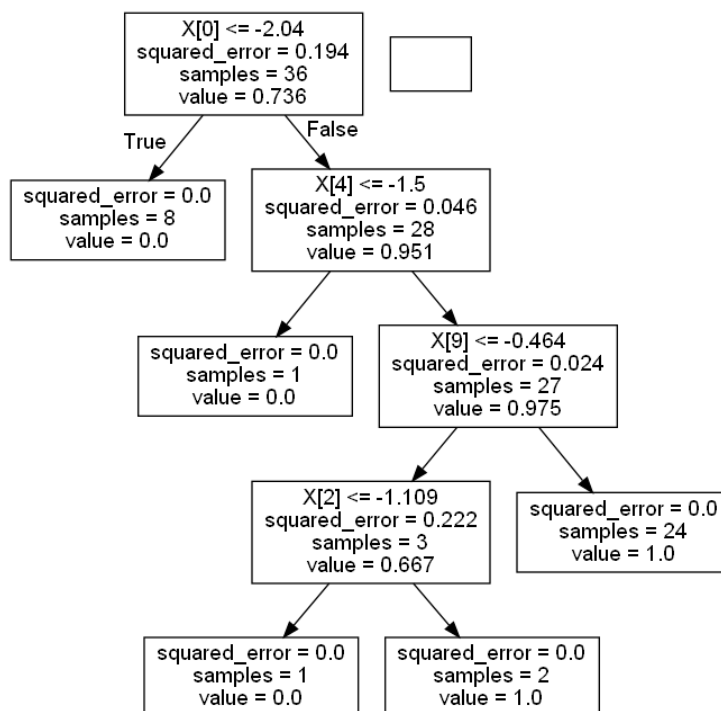


Figure 1 Decision Tree Results for Random Forest

The results show that silica plays a key role in the classification of high potassium glass and lead-barium glass, and magnesium oxide, sodium oxide, and copper oxide also have some influence on their classification. The importance of each feature is obtained. The importance ranking chart was generated by sorting them in descending order of feature importance as shown in Figure 2.

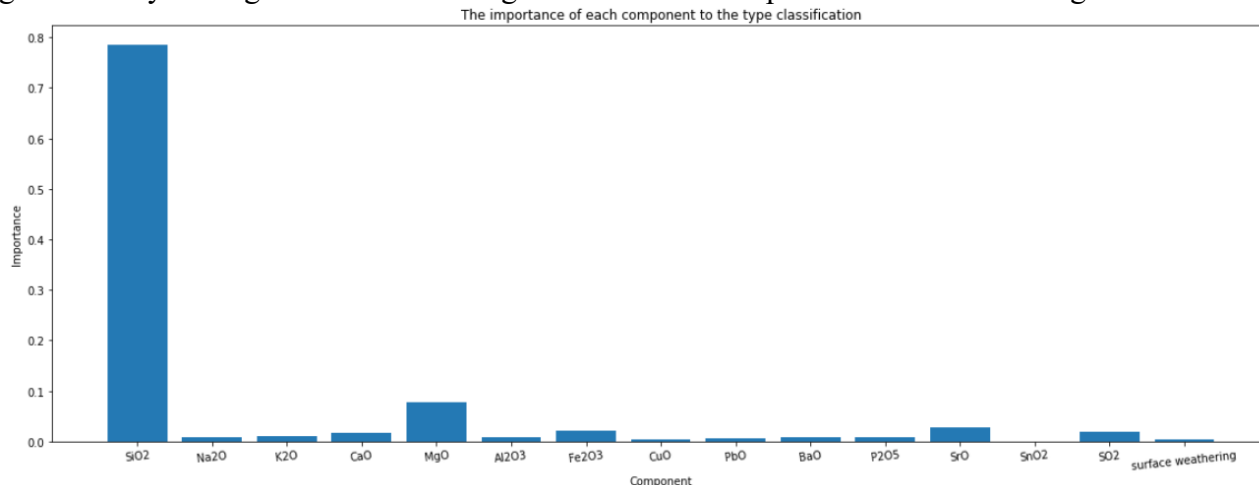


Figure 2 Importance of each chemical component

### 3.2.2 Subcategories of glass artifacts

High potassium glass.

Table 6 High potassium glass artifacts are divided into proportions according to composition

Chemical composition content (%)		Quantity	Proportion (%)
Aluminum oxide	3-8	9	75.00
	8-12	3	25.00
Total		12	100.00
Calcium Oxides	0-5	3	27.27
	5-10	8	72.73
Total		11	100.00

Lead barium glass.

Table 7 Lead barium glass artifacts according to the composition of the proportion

Chemical composition content (%)		Quantity	Proportion (%)
Lead oxide	0-20	6	46.15
	20-40	7	53.85
	Total	13	100.00
Alumina	1-2	3	23.08
	2-3	3	23.08
	3-4	3	23.08
	4-5	3	23.08
	5-6	1	7.69
	Total	13	100.00

According to the division of the content of each chemical composition, it is initially presumed that the high potassium glass can be divided into two subclasses according to the content of aluminum and calcium, and the lead-barium glass can be divided into four subclasses according to the content of lead and aluminum. Because the above chemical components have a certain amount in each content range and are more evenly distributed, it is reasonable to use them as the basis for classification.

For each type of subclassification, an unsupervised learning clustering algorithm is used to verify the speculation. k-means clustering (k-means) is a clustering algorithm based on sample set partitioning, in which the data are divided into K groups and K objects are randomly selected as the initial clustering centers, and then the distance between each object and the clustering center is calculated so that each object is assigned to the clustering center closest to it[9].The specific implementation process is as follows.

(1) Data source: get the excel sub-table of high potassium type in calcium oxide (CaO), aluminum oxide (Al<sub>2</sub>O<sub>3</sub>) in the excel sub-table.

(2) Cluster analysis: The high potassium is divided into two major classes, so the a priori method is used to define k=2, and the following Figure 3 is obtained.

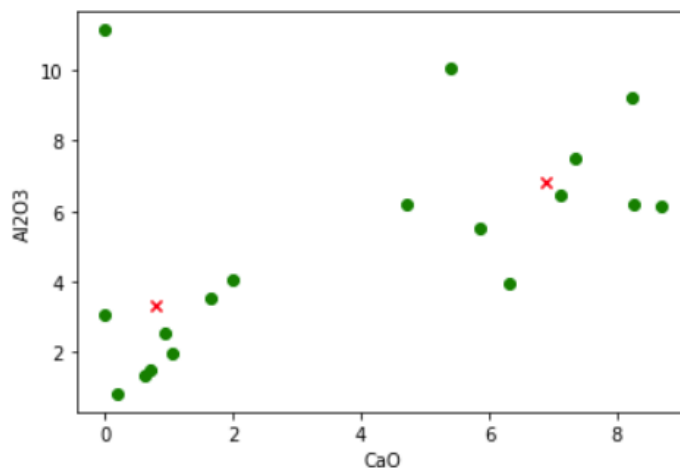


Figure 3 k-means clustering results

Where green is the sample point and red is the center of mass, whose coordinates are (0.9037499999999999, 2.3362499999999997), (6.1979999999999995, 7.233) respectively, so it is possible to divide the high potassium glass into two subclasses.

Since the composition content of lead oxide in lead barium is uniformly distributed in five intervals, the k value cannot be defined directly by the a priori method, so the elbow method is used to determine the k value[10], and the graph of the elbow method is obtained by running the Figure 4 of lead barium as follows.

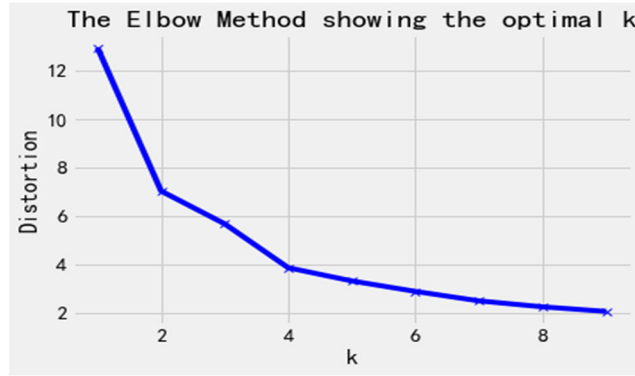


Figure 4 Elbow method to determine the center of mass

Therefore,  $k=4$  is defined, and similarly the K-mean clustering algorithm is used to obtain 4 prime points, which are  $[2.2, 61.001999999999995]$ ,  $[2.4458333333333333, 29.614166666666666]$ ,  $[3.02375, 43.890625]$ ,  $[5.68625, 16.9675]$ , as follows Figure 5.

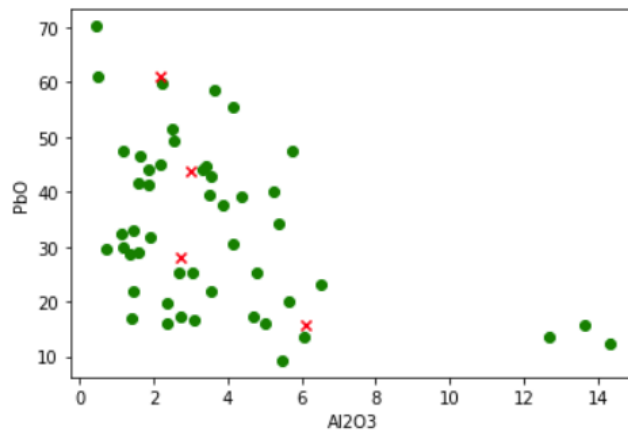


Figure 5 k-means clustering results

Initially, it was judged that it was reasonable to divide lead barium into four subclasses, and in order to verify the rationality of the model, the fuzzy mean clustering algorithm was used to perform cluster analysis, and by running the sample data of lead barium, the following Figure 6 was obtained

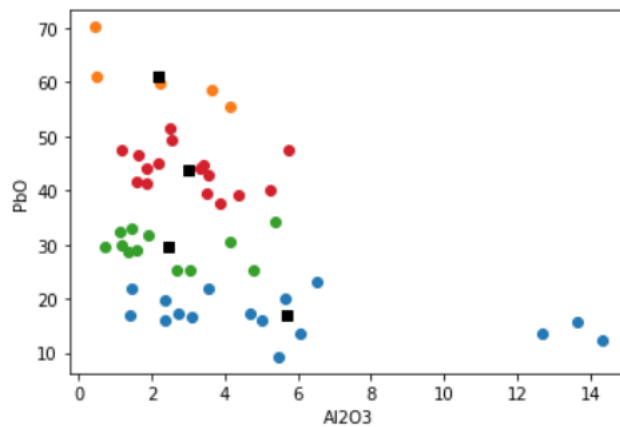


Figure 6 Fuzzy mean clustering results

It can be seen that the results of the two types of algorithms are basically the same, which proves the reasonableness of the model. In the high potassium glass the low chalcogenide glass is marked as 1 and the high chalcogenide glass is marked as 2, and in the lead barium glass it is divided into pb1, pb2, pb3 and pb4, which are marked as 1, 2, 3 and 4 respectively, and the subclasses of high potassium glass and lead barium glass according to the above model can finally obtain the following results as shown in Table 8 and Table 9.

Table 8 Classification of high potassium glass

Type	Calcium oxide (CaO)	Aluminum oxide (Al <sub>2</sub> O <sub>3</sub> )	Subclass
High Potassium	6.32	3.93	1
High Potassium	2.01	4.06	2
High Potassium	5.87	5.5	2
High Potassium	7.12	6.44	2
High Potassium	7.35	7.5	2
High Potassium	0	11.15	1
High Potassium	5.41	10.05	2
High Potassium	1.07	1.98	1
High Potassium	0.62	1.32	1
High Potassium	0.21	0.81	1
High Potassium	0.72	1.46	1
High Potassium	8.7	6.16	2
High Potassium	8.23	9.23	2
High Potassium	8.27	6.18	2
High Potassium	0	3.05	1
High Potassium	4.71	6.19	2
High Potassium	1.66	3.5	1
High Potassium	0.94	2.51	1

Table 9 Classification of lead barium glass

Type	Aluminum oxide (Al <sub>2</sub> O <sub>3</sub> )	Lead oxide (PbO)	Subclass
Lead barium	5.73	47.43	3
Lead barium	1.34	28.68	2
Lead barium	1.11	32.45	2
Lead barium	2.69	25.39	2
Lead Barium	3.57	42.82	3
Lead Barium	5.45	9.3	1
Lead Barium	1.42	16.98	1
Lead Barium	1.59	29.14	2
Lead Barium	1.9	31.9	2
Lead Barium	0.7	29.53	2
Lead Barium	1.18	29.92	2
Lead Barium	4.7	17.14	1
Lead Barium	14.34	12.31	1
Lead Barium	4.35	39.22	3
Lead Barium	3.86	37.74	3
Lead Barium	3.11	16.55	1
Lead Barium	2.36	19.76	1
Lead Barium	2.35	16.16	1
Lead Barium	1.62	46.55	3
Lead Barium	1.44	22.05	2
Lead Barium	1.6	41.61	3
Lead Barium	2.72	17.24	1
Lead Barium	2.57	49.31	3
Lead Barium	0.5	61.03	4
Lead Barium	0.45	70.21	4
Lead Barium	3.33	44.12	3
Lead Barium	3.53	21.88	1
Lead Barium	5.66	20.12	1
Lead Barium	2.25	59.85	4
Lead Barium	3.41	44.75	3
Lead Barium	12.69	13.61	1
Lead Barium	5	15.99	1
Lead Barium	4.79	25.25	2
Lead Barium	3.06	25.4	2
Lead Barium	13.65	15.71	1
Lead Barium	5.38	34.18	2
Lead Barium	6.5	23.02	1
Lead Barium	1.87	44	3

Lead Barium	4.16	30.61	2
Lead Barium	5.25	40.24	3
Lead Barium	2.51	51.34	3
Lead Barium	1.16	47.42	3
Lead Barium	6.06	13.66	1
Lead Barium	4.15	55.46	4
Lead Barium	3.65	58.46	4
Lead Barium	1.45	32.92	2
Lead Barium	1.85	41.25	3
Lead Barium	2.18	45.1	3
Lead Barium	3.52	39.35	3

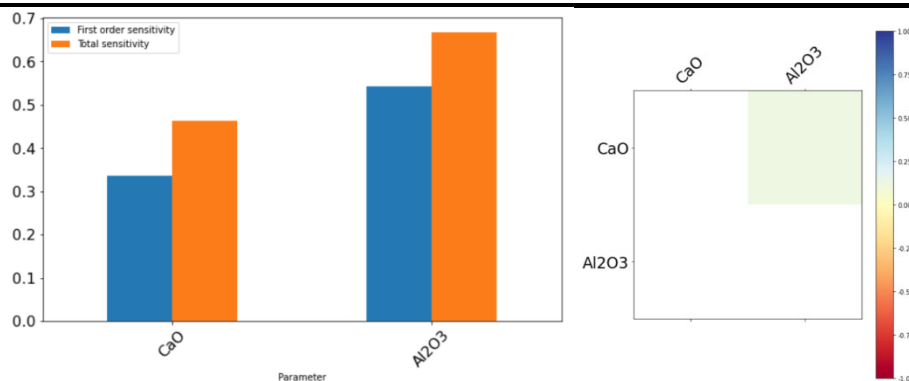


Figure 7 Sensitivity of subclassification of high potassium glass

Then the new sample data is trained with the random forest model, and then a sobol sensitivity analysis is performed on it according to the classification results, and the following Figure 7.

The figure shows that alumina is more sensitive to the classification results. Similarly, the same model training and sensitivity analysis is performed for lead-barium glass, and the following Figure 8 is obtained

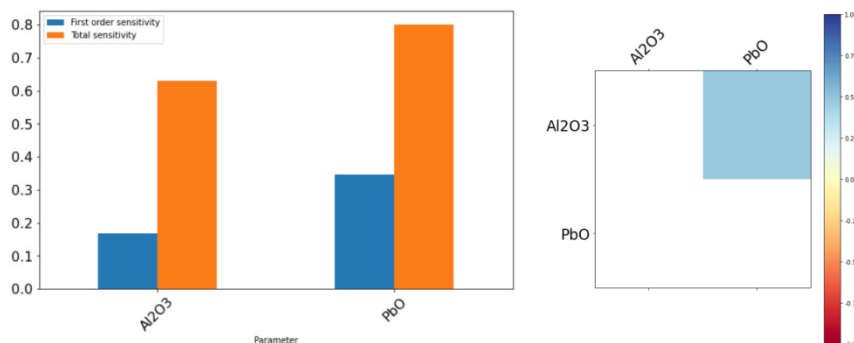


Figure 8 Sensitivity of lead barium glass subclass division

The graph shows that lead oxide is more sensitive to the classification results

## 4. Conclusion

In this paper, the given data are initially processed by mean value and content interval division to roughly determine the basis of their classification; then the classification is tested by random forest classification model to obtain the importance of each feature, so as to analyze which feature plays an important role in their classification. Next, the subclasses of high potassium glass and lead-barium glass are divided, because the division is based on the sample set and the classification cannot be trained, so the K-mean clustering algorithm is adopted to solve the problem, and the reasonableness test is conducted using the fuzzy clustering algorithm, and if the results obtained by the two algorithms are similar, it indicates good reasonableness; finally, the random forest model is used to train the new sample data, and then the classification results are based on the sobol sensitivity test is performed.

The random forest model based on this problem has high accuracy and can classify and identify glass artifacts in great detail. The random forest model based on the problem of this paper is fast to train and can obtain the importance ranking of variables. The amount of data is small and may produce error uncontrollable factors may affect the analysis of the model on the composition of glass artifacts

## References

- [1] Hui-Hui Liu. Random forest optimization and application based on cost-sensitive and swarm intelligence algorithms [D]. Hefei: Anhui University of Architecture, 2022(60).
- [2] Xing Yangyun. Research on text classification technology based on machine learning[D]. Qingdao: Qingdao University of Science and Technology, 2019(62).
- [3] Chen Guanlin. Analysis of GDP forecasting in three northeastern provinces based on gray model[J]. Economic and Management Science, 2022(3):10-13.
- [4] Luo Jingjing, Tang Weizhen, Ding Jiting. A study on data visualization under Pearson coefficient based control simulation training data independent and factor analysis[J]. Computer Science, 2021(48):623-628.
- [5] Xian Yiheng, Li Xintong, Zhou Xueqi, et al. Compositional analysis and origin discrimination of turquoise glass artifacts excavated from two sites in Xinjiang[J]. Spectroscopy and Spectral Analysis, 2020(40):967-970.
- [6] Li Meng, Bao Lei, Hu Yi, et al. Implementation of random number online detection method based on chi-square test[J]. Microelectronics, 2022(52):388-392.
- [7] Wang Jingxiang. Principle study and practical application of decision tree algorithm [J]. Computer Programming Skills and Maintenance, 2022(8) 54-56.
- [8] An Qiuzhou. Research on the classification of glass relics of the Grand Canal--Tianjin as an example[J]. Glass Relics Identification and Appreciation, 2021 (20) 104-106
- [9] Deng Hui, Wu Xiantan, Zhang Wenjiang, et al. Slope-Unit Scale Landslide Susceptibility Mapping Based on the Random Forest Model in Deep Valley Areas [J]. Remote Sensing,14(17):4245-4245.
- [10] Li Hui, Wu Zixuan, Yuan Xing, et al.The research on modeling and application of dynamic grey forecasting model based on energy price-energy consumption-economic growth[J]. Energy, 2022(257).