

Research and Analysis of Glass Weathering Based on Ergodic K-Prototype Algorithm

Sifan Li^{*,†}, Rizheng Zhang[†], Yang Guo[†]

Xi'an University of Architecture and Technology, School of information and control engineering,
Xian Shaanxi

* Corresponding Author Email: 1574101037@qq.com

[†] These authors contributed equally.

Abstract. When making glass products, adding fluxes with different chemical compositions can form different glass types. Affected by the burial environment, the composition proportion of ancient glass products will change over time, thus affecting the correct judgment of their types. Based on decision tree, traversal K-prototype algorithm, grey correlation analysis and other methods, this paper studies and analyzes the classification and identification related to glass weathering and the statistics of chemical composition laws, the relationship between the surface weathering of glass relics and the glass type, decoration and color is analyzed. In this paper, after data pretreatment, the surface weathering and the other three indicators are preliminarily observed through contingency analysis, and then the correlation between the variables is further clarified by using the chi-square test. It is found that the surface weathering is significantly related to the overall glass type, and lead barium glass is more prone to surface weathering than high potassium glass; There is no obvious correlation between surface weathering and decoration and color. Then, according to the glass types, the glass relics are divided into four glass types: lead-barium weathering/non-weathering, high-potassium weathering/non-weathering, and the statistical analysis is established after the pretreatment of the chemical composition of these four types. Then, combined with the independence t-test, we can further obtain the specific element differences, and give a ranking table of the degree of influence. Finally, based on the above statistical exploration, the model of chemical composition law of glass relics before weathering is built according to the mechanism analysis of the four composition laws of high-potassium glass Si and its chemical composition that significantly affects the weathering, the chemical composition that significantly affects the weathering of lead-barium glass, and the chemical composition that has no effect on the weathering of glass. According to the test data of weathering point, the prediction table of chemical composition content before weathering is given. The interval test of the model results shows that the failure rate of high potassium prediction is high. The revised model of chemical composition law of glass relics before weathering is obtained by analyzing and considering the modification of the value rule of high potassium Si.

Keywords: Classification of glass weathering, statistical mechanism, decision tree.

1. Introduction

In the manufacturing process of glass products, because of its high melting point, it is necessary to add flux to reduce the melting temperature. The addition of different fluxes will form glass types with different chemical compositions. Due to the influence of the burial environment, the internal elements of ancient glassware will exchange with the environmental elements, resulting in the change of its composition proportion, thus affecting the correct judgment of its type. Therefore, for ancient glass products, this paper studies the type identification related to weathering and the laws of corresponding chemical composition.

2. Establishment and solution of the model

Statistical analysis of chemical properties. Based on the methods of independence test and mechanism analysis, this paper excavates the relationship and statistical laws, and establishes the

model of chemical composition law of glass relics to reduce the chemical composition content before weathering. Finally, the model is properly tested and possibly optimized.

2.1. Relationship between weathering and glass type, decoration and color

In the analysis of the relationship between the surface weathering of glass relics and its glass type, decoration and color, after data pretreatment, this paper first makes preliminary observation of the surface weathering and the other three indicators through contingency analysis, and then uses the chi-square test to further clarify the correlation between the variables, as shown in Table 1.

Table 1. Missing value data

number	Ornamentation	type	colour	Surface weathering
19	A	Lead barium		Weathered
40	C	Lead barium		Weathered
48	A	Lead barium		Weathered
58	C	Lead barium		Weathered

It can be seen from the analysis of the above missing data that the data with color gaps are all lead-barium glass with weathered surface, and the decoration type is A or C, showing obvious non-random loss characteristics. At the same time, the statistics of the proportion of missing data show that the proportion of missing data in lead-barium is $4/40=10.0\%$, and the proportion of all data is $4/58 \approx 6.90\%$, which accounts for a large proportion. The missing data cannot be simply deleted, this paper temporarily ignores the processing of missing data, and then carries out specific processing when missing values are used in subsequent analysis as shown in Fig 1.

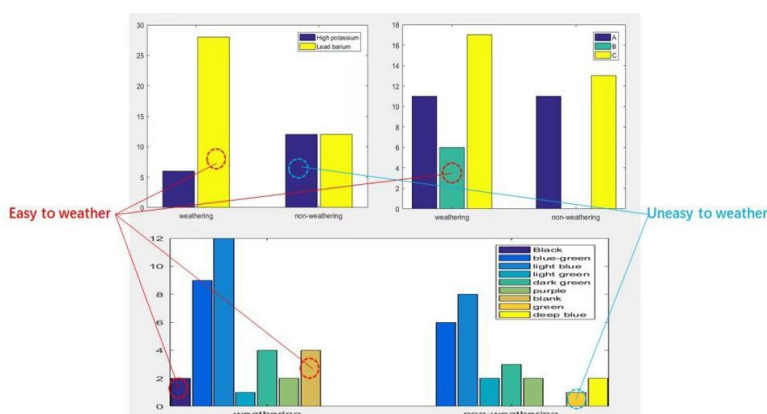


Figure 1. Preliminary exploration of surface weathering trend

Through preliminary observation in Fig 1, Considering the type of glass, lead-barium glass is more prone to surface weathering, and high-potassium glass is less prone to surface weathering, Surface B is completely weathered, while the proportion of weathered and unweathered surface A and C is slightly different, the surface of black, white and blank items is completely weathered, and the green surface is not weathered. The other color surfaces are slightly weathered.

2.2. Correlation analysis of surface weathering based on chi-square test

Because the classification information of glass relics analyzed this time is discrete data, this paper uses the chi-square test method to analyze the correlation between discrete variables, as shown in Table 2.

Table 2. Marginal frequency contingency table

	hyperkalemia	Lead barium	total	scale
Weathered	6	28	34	17.65%
No weathering	12	12	24	50.00%
total	18	40	58	31.03%

List the corresponding table of theoretical values:

Table 3. Corresponding Table of Theoretical Values

	hyperkalemia	Lead barium	total
Weathered	11	23	34
No weathering	7	17	24
total	18	40	58

Square the difference between the actual value and the theoretical value:

$$\chi^2 = \frac{(6-11)^2}{11} + \frac{(28-23)^2}{23} + \frac{(12-7)^2}{7} + \frac{(12-17)^2}{17} = 8.402 \quad (1)$$

In this study, the degree of freedom $d f=1$, $\alpha=0.05$, the confidence interval is 95%.

Because of $\chi^2 = 8.402 > \chi_{95\%}^2(1) = 3.841$, it falls into the reject domain and rejects the original assumption of H_0 . It is concluded that there is a correlation between surface weathering and glass type. Next, further analysis of the specific correlation and statistical analysis of the data show that lead barium glass is more prone to surface weathering, and high potassium glass is less prone to surface weathering.

3. Statistical law of weathering chemical composition content based on independent t test

Based on the types of glass, this paper divides the glass relics into four types: lead-barium weathering, lead-barium weathering, high-potassium weathering and high-potassium weathering, and carries out statistical analysis on the chemical composition of these four types. First, the data is preprocessed to facilitate the subsequent regular statistics.

3.1. Normalization

The purpose of normalization here is to unify the total chemical composition of each glass relic data to 100%. The data with chemical composition less than 100% will be expanded or reduced at the same time according to the original chemical content proportion of each component, so that its total proportion will approach 100% as far as possible.

According to the analysis of this paper, it is found that in the lead-barium weathering glass data, there are several groups of data that detect the chemical composition content of the weathering part, so the measurement data of the weathering part is put into the lead-barium weathering data group for statistical analysis.

3.2. High potassium without weathering

First of all, this paper makes descriptive statistics on the chemical components of 12 groups of high-potassium non-weathering cultural relics as shown in Table 4:

Table 4. Statistics of chemical composition of high-potassium non-weathering cultural relics

Variable name	Maximum	minimum value	average value	standard deviation	median	CV
SiO ₂	87.05	59.01	67.984	8.755	65.530	0.129
Na ₂ O	3.38	0	0.695	1.287	0	1.852
			⋮			
SO ₂	0.47	0	0.102	0.186	0	1.825

According to coefficient of variation:

$$C.V = (SD \div MN) \times 100\% \quad (2)$$

Where, SD is the standard deviation of the sample, and MN is the average value of the sample.

For the chemical composition of high-potassium non-weathering cultural relics, only the coefficient of variation (CV) value of Si is $0.129 < 0.15$. It can be judged that there is a small probability of

abnormal value in the current data, and the average value can be used for descriptive analysis, while the coefficient of variation of other chemical components is greater than 0.15, which requires further analysis.

The same method is used to carry out descriptive statistics on the high potassium weathering type, and the same conclusion is obtained. Only Si can be described and analyzed with the average value, and the other chemical components need further analysis as shown in Fig 2.

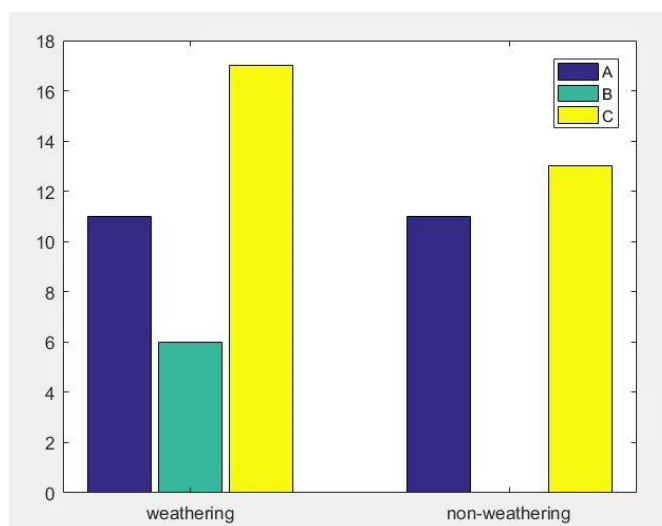


Figure 2. High potassium glass analysis visualization

For high-potassium glass, the proportion of non-weathering is 61.11%, and the proportion of weathering is 38.89%. For high-potassium weathered glass and high-potassium non-weathering glass, the average value can be used to describe the Si. After comparison, the proportion of Si in high-potassium glass after weathering will increase significantly, with an increase of 25.979%.

In order to further analyze the differences of various chemical components under the conditions of high potassium weathering and high potassium non-weathering, we set a categorical variable to determine whether it is weathered. The non-weathering number is 0, the weathering number is 1, and the independence t test is performed with the quantitative field (the proportion of all chemical components). Due to the large proportion of Si content and the small proportion of other components, observation together will lead to distortion of other components, which is difficult to analyze. Therefore, it is tested separately and the results are as follows Table 5:

Table 5. Single t test of high potassium glass Si

Variable name	Weather or not	sample size	average value	standard deviation	t	P	Average difference	Cohen's d value
SiO_2	1	6	93.963	1.734	7.656	0.000***	26.935	3.736
	0	12	67.984	8.415				
	total	18	75.109	14.478				

Note: * * * represents the significance level of 1%

The average value of high potassium weathering and non-weathering on Si is 93.963/67.984 respectively; The P value of F test result is 0.000 * *, so the statistical results are significant, indicating that there is a significant difference between weathering and non-weathering of high-potassium glass on silicon dioxide (Si). The range of difference Cohen's d value is 3.736>0.800, which is very large.

The mean value results of the independent sample t-test for the remaining components are shown in Fig 3. By comparing the mean values, can explore their differences.

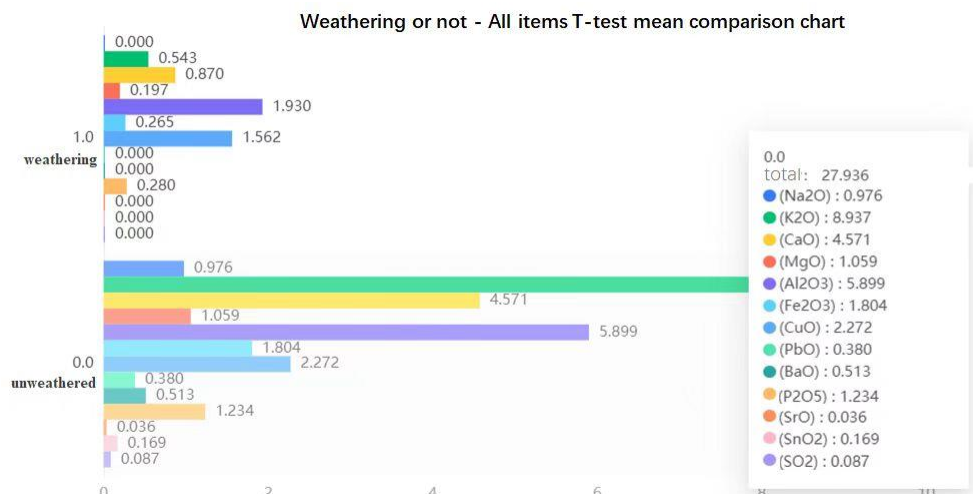


Figure 3. Comparison of t-test mean values of chemical composition for weathering of high-potassium glass

According to the mean value of the detailed table and the mean value of the F test results, this paper obtains that the chemical components that have differences between high-potassium weathering and high-potassium non-weathering are: potassium oxide (O), calcium oxide (CaO), magnesium oxide (MgO), aluminum oxide (), and iron oxide (); The chemical components that do not differ include sodium oxide (), copper oxide (CuO), lead oxide (PbO), barium oxide (BaO), phosphorus pentoxide (), strontium oxide (SrO), tin oxide (Sn), and sulfur dioxide (S). The influence degree of different chemical components is ranked as shown in Table 6.

Table 6. Ranking of influence degree of different chemical composition of high potassium glass

order	chemical composition	Impact degree (difference range)
1	silica	3.736
2	K_2O	2.621
3	MgO	1.552
4	Al_2O_3	1.528
5	CaO	1.261
6	Fe_2O_3	1.155

3.3. Weathering Analysis of Lead Barium Glass

Descriptive statistics are made for the weathered and non-weathered parts of lead and barium. According to the coefficient of variation, it is found that all chemical components have abnormal values, and the average value cannot be used for descriptive analysis, so it is necessary to use independent t-test for further analysis.

The independence t-test of lead-barium glass was carried out in the same way as that of high-potassium glass, and the results are shown in Table 7:

Table 7. Single t test of lead barium glass Si

Variable name	Weather or not	sample size	average value	standard deviation	t	P	Average difference	Cohen's d value
SiO_2	1	24	26.642	9.038	-9.149	0.000***	28.018	2.670
	0	23	54.660	11.829				
	total	47	40.353	17.555				

Note: * * * represents the significance level of 1%

The average value of lead and barium weathering and non-weathering on silicon dioxide (Si) is 26.642/54.660 respectively; The P value of F test result is 0.000 * *, so the statistical results are significant, indicating that there is a significant difference between weathering and non-weathering of

lead barium glass on silicon dioxide (Si). The range of difference Cohen's d value is $2.67 > 0.8$, and the range of difference is very large.

The mean value of the independent sample t-test of the other components is as follows. By comparing the mean value, the difference relationship can be mined.

Based on the mean value in the detailed table and the mean value of the F-test results, we obtain that, the chemical components that differ between lead and barium weathering and non weathering are: sodium oxide, calcium oxide (CaO), lead oxide (PbO), phosphorus pentoxide, and silicon dioxide (Si); The chemical components that do not differ include potassium oxide (K₂O), magnesium oxide (MgO), aluminum oxide, iron oxide, copper oxide (CuO), barium oxide (BaO), strontium oxide (SrO), tin oxide (Sn), and sulfur dioxide (S).

The ranking of the degree of influence of different chemical components is shown in Table 8:

Table 8. Ranking of influence degrees of different chemical compositions in lead barium glass

order	chemical composition	Impact degree
1	silica	2.670
2	Lead oxide	2.131
3	Phosphorus pentoxide	1.219
4	calcium oxide	0.879
5	Sodium oxide	0.848

4. Model solving

Due to the interval form of the chemical composition rule, this article uses the average value of the upper and lower limits of the interval as the representative of the interval, and the (unweathered representative - weathered representative)/weathered representative is the adjusted ratio, as shown in Table 9.

Table 9. Floating adjustment of indicators affecting weathering (%)

	(K ₂ O)	(MgO)	(Al ₂ O ₃)	(CaO)	(Fe ₂ O ₃)
High potassium glass	+9.298	+5.333	+2.109	+1.586	+19.609
	(SiO ₂)	(PbO)	(P ₂ O ₅)	(CaO)	(Na ₂ O)
Lead barium glass	+0.223	+3.790	-0.364	-0.327	-0.667

Note: +indicates that the chemical composition needs to be increased from weathering to non weathering, and - indicates that it needs to be decreased

"Then, the indicators that affect weathering may vary greatly or slightly. In order to ensure that the cumulative sum of component proportions is between 85% and 105%, it is necessary to adjust the non impact indicators to a certain extent. It is advisable to redistribute the remaining proportions according to the four types of proportion rules."

Solution of reducing the proportion of non weathered components from the proportion of weathered components

Float adjustment of indicators affecting weathering, the average value of the upper and lower limits of the interval is used as the representative of the interval, and the adjusted ratio is (unweathered representative - weathered representative)/weathered representative, Calculate the weathering value of important chemical components: (100%+rising or falling ratio) weathering value, Adjust the proportion of non impact indicators, difference between the total proportion of previously weathered components and the total proportion of important components at this time, distribute the difference values proportionally to obtain the results. So far, this article has obtained the coincidence interval of each component element under four classifications.

Table 10. Conformance intervals for each constituent element under four classifications

	High potassium weathering	High potassium fresh	Lead barium weathering	Lead and barium unweathered
SiO_2	93.96	67.03	(11.43, 41.85)	(36.00, 73.20)
K_2O	(0, 1.29)	(2.70, 15.96)	(1.41, 0)	(1.41, 0)
Na_2O	(0, 3.38)	(0, 3.38)	(0, 1.19)	(0, 5.70)
CaO	(0.05, 1.70)	(0.11, 10.56)	(0, 5.50)	(0, 3.50)
MgO	(0, 0.71)	(0, 2.22)	(0, 2.73)	(0, 2.73)
Al_2O_3	(0.30, 3.56)	(2.40, 10.83)	(0, 14.43)	(0, 14.43)
Fe_2O_3	(0.15, 0.38)	(0, 4.74)	(0, 4.59)	(0, 4.59)
CuO	(0, 5.09)	(0, 5.09)	(10.57, 0)	(10.57, 0)
PbO	(0, 1.62)	(0, 1.62)	(23.39, 64.60)	(8.12, 35.87)
BaO	(0, 2.68)	(0, 2.68)	(0, 26.23)	(0, 26.23)
P_2O_5	(0, 4.5)	(4.5, 0)	(0, 12.49)	(0, 4.16)
SrO	(0.12, 0)	(0.12, 0)	(0, 1.12)	(0, 1.12)
SnO_2	(0, 2.36)	(0, 2.36)	(0, 1.31)	(0, 1.31)
SO_2	(0, 0.47)	(0, 0.47)	(0, 3.66)	(0, 3.66)

5. Prediction Results

Due to the large size of the results table, this article only temporarily presents the prediction results of sampling point 22 (high potassium glass) and sampling point 38 (lead barium glass), as shown in Table 11.

Table 11. Prediction of chemical composition before weathering (partial)

Cultural relics sampling point	22 (High potassium)	38 (Lead barium)
SiO_2	71.362	33.003
Na_2O	3.599	1.375
K_2O	0.809	0.257
CaO	1.686	0.675
MgO	0.674	0.498
Al_2O_3	3.574	2.631
Fe_2O_3	0.419	0.837
CuO	5.419	1.927
PbO	1.725	51.179
BaO	2.853	4.782
P_2O_5	4.791	0.478
SrO	0.128	0.022
SnO_2	2.513	0.239
SO_2	0.500	0.667

6. Prediction result test

Using the weathering free composition rule of high potassium glass and lead barium glass, judge the predicted value, obtain the chemical composition range of weathering free glass from the regular model (Formula 4), and program to determine whether the predicted result is within the regular range. The judgment results are as follows Fig 4:

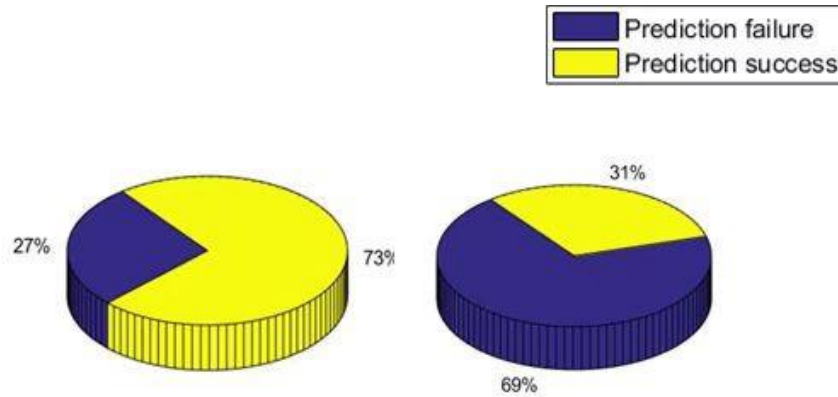


Figure 4. Predicted Percentage, Lead Barium (Left), High Potassium (Right)

It can be seen from the results that the prediction of lead and barium in this article is relatively successful and reliable, but the prediction of high potassium is unreliable. Next, this article analyzes the reasons for the failure of the prediction of high potassium.

7. Model revision

In order to explore the reasons for the failure of high potassium prediction, this article compares the successfully predicted lead barium model with the high potassium model, and finds that the only difference between the two models in value selection is the value rule for Si. According to the results of the independence test, this article selects the average value of Si as the value taking rule for Si in the high potassium model, but selects the confidence interval as the value taking rule in the lead barium model to obtain correct prediction results. Based on this, this article makes the following revisions to the model:

The value of Si in the high potassium prediction model is modified to be determined by the confidence interval, and the revised model is:

$$C_i = \begin{cases} (Ra_i, Rb_i), v_i = 0 \\ (Ra_i, Rb_i), v_i = 1 \end{cases} \quad (3)$$

$$a_i \in \begin{cases} [\mu_1 - 1.69\sigma_1, \mu_1 + 1.69\sigma_1], v_i = 0 \\ [\mu_2 - 1.69\sigma_2, \mu_2 + 1.69\sigma_2], v_i = 1 \end{cases} \quad (4)$$

$$Rb_i \in [\min_i, \max_i] \quad (5)$$

8. Conclusion

In this paper, a subcategory division model is applied, which can be reasonably analyzed based on various attribute data of the target object, and divided into different types to facilitate the study of statistical or evolutionary laws. It can be applied to the study of various unknown unearthed cultural relics; Similarly, in question 1, this article mainly uses a prediction model for pre weathering data of weathered cultural relics. This idea of restoring their pre change attributes based on known laws can be used to study various restoration and speculation issues, such as inferring the original composition of damaged items.

References

- [1] Xu Zhenqi, Xu Chaojie, Zhang Yuezheng, Zhou Hao, Pang Kang, Zheng Yuejiu. Abnormal battery classification method based on decision tree [J/OL]. Journal of Automotive Engineering: 1-8 [2022-08-04 09: 20].

- [2] Wang Nengyang. Comprehensive Performance Evaluation of CNC Machine Tools Based on Improved Grey Correlation Analysis [J]. *Agricultural Equipment and Vehicle Engineering*, 2022, 60 (08): 142-145+154.
- [3] Tang Xuemei, Yu Lu, Li He, Ding Yinzong. Classification and component analysis of glazed building materials unearthed from the Dabaoen Temple in Nanjing [J]. *Proceedings of the Palace Museum*, 2022, (08): 101-110+142. DOI: 10.16319/j.cnki.0452-7402.2022.08.004.
- [4] Zhu Pengpeng, Cai Yuhe, Lou Guoliang, Mi Liang, Song Yuanyuan. Research on Highway Cargo Suitability Based on K-prototypes Algorithm [J]. *Highway*, 2021, 66 (09): 260-265.
- [5] Chen Wei, Wang Lei, Jiang Ziyun. Hybrid attribute data clustering algorithm based on K-prototypes [J]. *Computer Application*, 2010, 30 (08): 2003-2005+2110.
- [6] Wen Yun, Gao Anming, Meng Qinying, et al. Application of contingency tables and chi-square tests in cigarette consumption market research [J]. *China Business Journal*, 2017, No. 711 (08): 164-167. DOI: 10.19699/j.cnki.issn2096-0298/2017.08.073.
- [7] Hu Zhiling. Application and controversy of contingency analysis in memory system research [J]. *Psychological Dynamics*, 1993 (02): 17-22.
- [8] Deng Jinfu, Liu Cui, Di Yongjun, Feng Yanfang, Xiao Qinghui, Liu Yong, Ding Xiaozhong, Meng Guixiang, Huang Fan, Zhao Guochun, Wu Zongxu. Tectonic assemblage of tonalite trondhjemite granodiorite (TTG) rocks and their subclass classification [J]. *Geoscience Frontiers*, 2018, 25 (06): 42-50. DOI: 10.13745/j.esf.sf.2018.10.3.
- [9] Li Chunsheng, Jiao Haitao, Liu Peng, Liu Xiaogang. Improvement and application of decision tree classification algorithm based on C4.5 [J]. *Computer Technology and Development*, 2020, 30 (05): 185-189.
- [10] Shi Yulin, Liu Jiayi, Hu Xiaojuan, et al. A decision tree algorithm based on tongue pulse data for classification of non small cell lung cancer syndromes [J]. *World Science and Technology - Modernization of Traditional Chinese Medicine*, 2022, 24 (07): 2766-2775.