Transactions on Materials, Biotechnology and Life Sciences
ISSN: 2960-1932, eISSN: 2960-2319 | Volume 4
NSCBM 2024

Study on Ancient Glass Classification Based on Fisher Linear Discriminant Analysis and Systematic Cluster Analysis

Yifan Fu [1, *, #], Yudi Lin [2, #], Jian Chen [3], Jiaxin Chen [4], Zhenni Chen [3]

[1] Department of Economics and Finance, Sanya University, Sanya, China, 572022

[2] College of Finance and Economics, Sanya University, Sanya, China, 572022

[3] Saxo FinTech Business School, Sanya University, Sanya, China, 572022

[4] College of Humanities and Arts, Xi'an International University, Xi'an, China, 710077

* Corresponding Author Email: 15501938775@163.com

[#] These authors contributed equally.

Abstract. This paper discusses the classification of high potassium glass and lead-barium glass, and puts forward the model establishment and solution method using Fisher linear discriminant analysis and system cluster analysis. This PAPER points out THE obvious changes of chemical composition in glass cultural relics, and explores the classification rules of glass according to the different weathered states. The establishment of the model includes the steps of index selection, training group and test group division, Fisher linear discriminant function construction and cluster analysis. Through independent sample t-test and data analysis, the chemical components affecting the classification of glass were selected, and the Fisher linear discriminant function was successfully constructed, which reached 100% classification accuracy on the training group data. In the systematic cluster analysis, the sub-categories of high potassium glass and lead-barium glass are realized by distance calculation and grouping merging. The results of the model have been verified and sensitivity analysis show that its high reliability and robustness have the potential to accurately classify glass in practical applications. These methods provide effective analytical tools and strategies to solve the problem of ancient glass classification.

Keywords: Linear discriminant analysis, systematic cluster analysis, ancient glass classification.

1. Introduction

Influenced by the Silk Road, China had frequent trade and close exchanges with foreign countries in ancient times. Therefore, the foreign popular glass jewelry was also introduced into our country early. In ancient times, our country absorbed its technology and made its own glass with local materials [1]. Therefore, although the ancient glass of our country and foreign glass products looks similar, but the chemical composition is very different.

The main chemical component of glass is silicon dioxide (SiO_2), which has a high melting point. In order to facilitate preparation, we need to add flux in refining to reduce the melting temperature [2]. There are many kinds of flux commonly used in our country in ancient times. Due to the flux, the content of lead oxide (PbO) and barium oxide (BaO) is high; High potassium glass is fired with substances with high potassium content such as grass and wood ash as flux, which has high potassium content [3].

Ancient glass is highly susceptible to weathering due to the influence of burial environment. Weathering, that is, the process of large exchanges between internal elements and environmental elements. This process will lead to changes in the proportion of glass components, which will affect the judgment of its category. If there is no weathering on the surface of some cultural relics, the color and decoration can still be clearly seen, but the possibility of shallow weathering in some parts cannot be ruled out. However, the surface of some cultural relics is seriously weathered, and large areas are grayish-yellow and purple, which are called obvious weathered areas (weathering layer) and general weathered surfaces respectively, but there may also be unweathered areas [4]. In this study, we deeply explored the application of Fisher linear discriminant analysis in the classification of ancient glass, combined with the systematic cluster analysis method, aiming to improve the accurate classification of ancient glass samples [5]. As one of the important research objects in archaeology and cultural heritage, the classification and identification of ancient glass has always been one of the focuses of museums and cultural relics institutions [6]. However, ancient glass samples are often diverse and complex, and the traditional classification methods have been unable to meet the needs of rapid and accurate classification of large-scale samples. In this context, we will introduce the idea of Fisher linear discriminant analysis and combine it with systematic cluster analysis to propose a comprehensive method, in order to improve the accuracy and efficiency of ancient glass sample classification [7]. Through the problem analysis ideas of this study, we will explore how to effectively apply these methods, so as to bring new enlightenment and methods to the field of ancient glass classification.

2. Establishment and solution of the model

2.1. Selection of classification indicators

Because there are many types of chemical components contained in glass cultural relics, and the differences of some components in different types of glass are not significant, it is not suitable for the classification standard of glass types. Therefore, in this paper, independent sample t-test was first conducted on glass types and chemical components to find out the chemical components that showed significant differences in different types of glass, and to explore the classification rules of high-potassium glass and lead-barium glass [8].

The weaved and unweaved data were imported into SPSS (27) for independent sample t-test, and the final results are shown in Table 1:

Table 1: Independent sample t-test results

Name of variable	P value for weathered glass	P value for unweathered glass
(SiO_2)	0.000***	0.026**
(Na_2O)	0.300	0.998
(K_2O)	0.010**	0.000***
(CaO)	0.010***	0.000***
(MgO)	0.160	0.021**
(Al_2O_3)	0.371	0.001***
(Fe_2O_3)	0.232	0.077*
(CuO)	0.705	0.711
(PbO)	0.000***	0.000***
(BaO)	0.002***	0.000***
(P_2O_5)	0.009***	0.954
(SrO)	0.000***	0.017**
(SnO_2)	0.603	0.284
(SO_2)	0.456	0.767
Note: ***, **	* represent the significance level of 1%, 5%	10%, respectively

According to the above results, under the condition of weathering, there are significant differences in the following seven chemical components of different types of glass: SiO_2, K_2O, CaO, PbO, BaO, P_2O_5 and StO. Therefore, the above seven chemical components can be used to classify the weathered glass. However, in the case of unweathered glass, different types of glass have significant differences in the following eight chemical components SiO_2, K_2O, CaO, MgO, Al_2O_3, PbO, BaO, SrO, so the above eight chemical components can be used to classify the unweathered glass.

2.2. Set up the training and test groups.

In order to test the accuracy of the model and prevent the occurrence of overfitting, this paper divides the data into training group and test group according to the ratio of 8:2, uses the data of training group to estimate the model, and uses the data of test group to test. Taking the weathering data as an example, the selection method of training group and test group is described: in the weathering glass data, there are 6 data of high potassium type, among which one data is randomly selected as the test group, and the rest as the training group; There are a total of 21 data with the type of lead and barium, from which 4 data are randomly selected as the test group and the rest as the training group. The same is true for unweathered data.

2.3. Fisher linear discriminant analysis.

In the case that the glass has been weathered, the classification rules of high-potassium glass and lead-barium glass were found, and the data of the weathered glass were imported into SPSS (27) for Fisher linear discriminant analysis. The results are as follows (1 represents high-potassium glass, 0 represents lead-barium glass); In the case of weathering, the Fisher linear discriminant function of high potassium glass is shown in Equation 1:

$$\begin{aligned} F1 &= 82.659*SiO_2 + 18.681*K_2O + 284.708*CaO - 63.007*PbO \\ &+ 90.094*BaO + 54.207*P2O5 - 118.291*SrO - 4017.357 \end{aligned} \tag{1}$$

The Fisher linear discriminant function of lead-barium glass is shown in Formula 2:

$$\begin{aligned} F2 &= 71.107*SiO_2 + 18.289*K_2O + 250.934*CaO - 55.093*PbO \\ &+ 78.573*BaO + 46.337*P2O5 - 130.913*SrO - 3000.056 \end{aligned} \tag{2}$$

Then the constructed Fisher linear discriminant function was used to classify the data in the training set, and the results are shown in Table 2:

Table 2: Classification of the training set of weathering data by Fisher linear discriminant functions F1 and F2

		Type = high potassium	0	1	Total amount
The original	Count up	0	18	0	18
		1	0	5	5
		Unclassified cases	3	1	4
	%	0	100	0	100
		1	0	100	100
		Unclassified cases	75	25	100

Under the condition that the glass is not weathered, the classification rules of high potassium glass and lead barium glass are found, and the data of unweathered glass are imported into SPSS (27) for Fisher linear discriminant analysis. The results are as follows (where 1 represents high potassium glass and 0 represents lead barium glass):

In the case of no weathering, the Fisher linear discriminant function of high potassium glass is shown in Equation 3:

$$\begin{aligned} F3 = 15.065 * SiO_2 + 19.230 * K_2O + 20.128 * CaO - 8.548 * MgO + 24.429 \\ * Al_2O_3 + 14.307 * PbO + 25.837 * BaO - 111.053 * SrO - 721.571 \end{aligned} \tag{3}$$

The Fisher linear discriminant function of lead-barium glass is shown in Formula 4:

$$\begin{aligned} F4 = 15.693 * SiO_2 + 19.458 * K_2O + 18.388 * CaO - 15.240 * MgO + 27.843 \\ * Al_2O_3 + 15.383 * PbO + 26.665 * BaO - 112.365 * SrO - 783.552 \end{aligned} \tag{4}$$

Then the constructed Fisher linear discriminant function was used to classify the data in the training set, and the results are shown in Table 3:

Table 3: Fisher linear discriminant function F3 and F4 classification on the training set of unweathered data

		Type = high potassium	0	1	Total amount
The origina	Count up	0	10	0	10
		1	0	8	8
		Unclassified cases	2	2	4
	%	0	100	0	100
		1	0	100	100
		Unclassified cases	50	50	100

3. Analysis and test of the results

In the case of weathering and unweathering, the constructed Fisher linear discriminant function can achieve 100% classification accuracy for the training group data, and the model performs well. In determining the type of glass that has weathered on its surface, Select it SiO_2, K_2O, CaO, PbO, BaO, P_2O_5, SrO the data of are respectively substituted into F1 and F2 We get a constant in f1 and f2. If f1 > f2. then the glass is divided into high potassium glass [9]. If f1 < f2, the glass is divided into lead-barium glass. When judging the type of surface unweathered glass, select it SiO_2, K_2O, CaO, PbO, BaO, P_2O_5, SrO. The data of are respectively substituted into F3 and F4 get a constant in f3 and f3. If f3 > f4. the glass is divided into high potassium glass. If f3 < f4, the glass is divided into lead barium glass.

The established model is used to classify the test group data left above, and the data of different surface weathering conditions are respectively put into the corresponding fisher linear discriminant function, and the function values of different glass types are calculated and compared. The glass types corresponding to the test group data are shown in Table 4 and Table 5 respectively:

Table 4: Validation of the surface weathering test group

(SiO_2)	(K2O)	(CaO)	(PbO)	(BaO)	(P_2O5)	(SrO)	Surface weathering	Type of judgment
92.72	0	0.94	0	0	0.36	0	weathering	High potassium
29.15	0	1.21	41.25	15.45	2.54	0	weathering	Barium lead
25.42	0	1.31	45.1	17.3	0	0	weathering	Barium lead
30.39	0.34	3.49	39.35	7.66	8.99	0.24	weathering	Barium lead

Table 5: Validation of the surface unweaved test group

(SiO_2)	(K_2O)	(CaO)	(MgO)	(Al_2O_3)	(PbO)	(BaO)	(SrO)	Surface weathering	Actual type	Type of judgment
79.46	9.42	0	1.53	3.05	0	0	0.07	unweathered	High potassium	High potassium
76.68	0	4.71	1.22	6.19	1	1.97	0	unweathered	High potassium	High potassium
51.54	0.29	0.87	0.61	3.06	25.4	9.23	0.85	unweathered	Barium lead	Barium lead
49.01	0	1.13	0	1.45	32.92	7.95	0	unweathered	Barium lead	Barium lead

It can be seen that the classification results are completely consistent with the actual situation, and the classification accuracy of the test group data reaches 100%, indicating that the established Fisher linear discriminant function has strong reliability, and can accurately reflect the classification law of high potassium glass and lead barium glass.

4. Establishment and solution of system clustering model

Considering that the type of cultural relic will not change before and after weathering, for example, lead-barium glass will not change into other types of glass after weathering, so it must still be lead-barium glass [10]. Therefore, only the non-weathered data are classified in this question. Firstly, we first classified the original data in "Form 2" according to whether there was weathering or not, screened out all the non-weathering data, and then classified the data to separate the data of non-weathering lead-barium glass and non-weathering high-potassium glass.

4.1. Treat each sample as a class and calculate the distance between pairs

The data of each group of samples are recorded as shown in Formula 5:

$$x_i = (x_{i1}, x_{i2}, ..., x_{il}), (l = 14) \tag{5}$$

Of which. $i = 1,2,...n$. If cluster analysis is carried out on high potassium glass. the $n = 10$. That is, there are 10 groups of samples to be clustered. If cluster analysis of lead-barium glass is carried out, the $n = 10$. That is, there are 12 groups of samples to be clustered. Consider each set of data as a class and write as $G_i (i = 1,2,...n)$. At this point, the distance between the two classes is the distance between the two samples. In this paper, the square Euclidean distance is selected as the distance between two samples, so the distance formula between two samples is shown in Formula 6:

$$d_{ij} = d(x_i, x_j) = \sum_{k=1}^{l} (x_{ik} - x_{jk})^2 \tag{6}$$

4.2. The two classes with the smallest distance are merged into a new class, and the distance between the new class and all classes is re-calculated

The hypothesis x_{io} and x_{j0} the distance between d_{i0j0} minimum distance, then will x_{io} and x_{j0} to a new class, let's call it G_{n+1}. and then calculate G_{n+1} 与 Distance between other classes. At this point, we choose to use the shortest distance method (or nearest element method) to calculate the distance between classes. The distance calculation formula is shown in Formula 7:

$$d(G_p, G_q) = \min d_{ij} = \min d(x_i, x_j), (x_i \in G_p, x_j \in G_q) \tag{7}$$

4.3. Repeat Step2 until all classes are finally merged into one

The above process is realized in SPSS (27) software, and the system clustering pedigree diagram can be finally obtained as shown in Figure 1 and Figure 2:

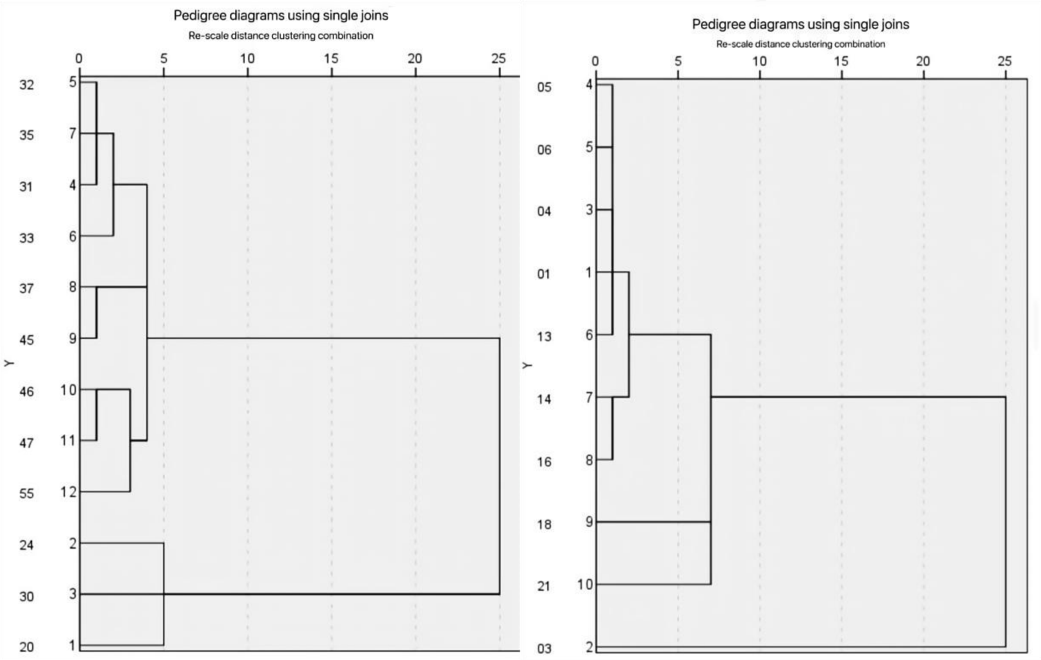

Figure 1: Cluster lineage diagram of lead-barium glass system and high potassium glass system

In order to determine the final number of clusters, this paper uses the elbow rule to roughly estimate the optimal number of clusters through graphics. Suppose n samples are divided into K classes（$k \le n-1$）, us $d(C_k(k=1,2,...K)$ represents the KTH class, and the center of gravity of this class is denoted as uk. Then the distortion degree of the KTH class is $\sum_{i \in C_k} |x_i - u_k|^2$. The total distortion degree of all classes is defined as shown in Equation 8:

$$J = \sum_{K=1}^{K} \sum_{i \in C_k} |x_i - u_k|^2 \tag{8}$$

In the above equation, J is also called the aggregation coefficient. Visualize the calculated aggregation coefficient, and get the line chart of aggregation coefficient as shown in Figure 2:

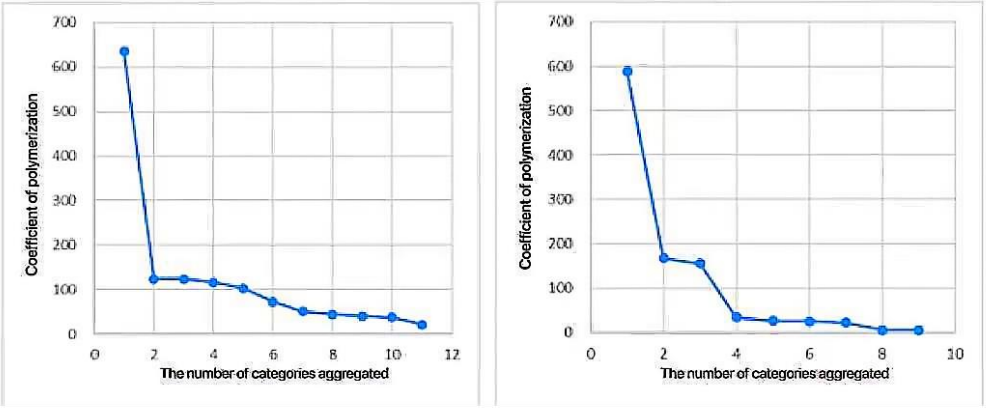

Figure 2: Plot of polymerization coefficient of lead-barium glass and high potassium glass

It can be seen from the above figure that when the number of categories of lead-barium glass is 2, the downward trend of the broken line is significantly slowed down, so the number of categories can be set as 2, that is, K=2; From the line chart of polymerization coefficient of high potassium glass, it can be seen that when the value of K is from 1 to 2, the distortion degree changes the most, and when the value exceeds 2, the distortion degree changes significantly, so the elbow is K=2, so the number of categories can be set as 2. After the number of clustered categories K is determined, the pedigree diagram above can be performed as shown in Figure 3:

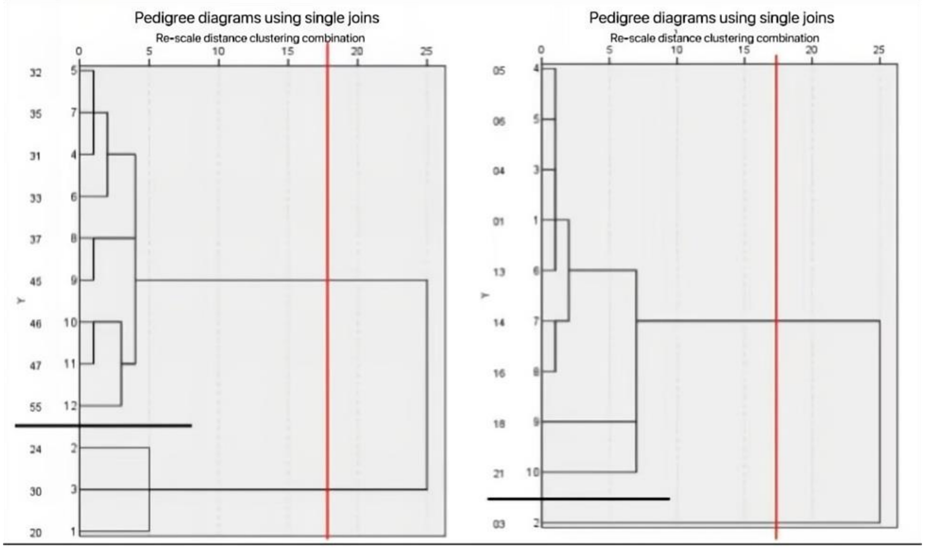

Figure 3: Division of the spectra of lead-barium glass and: high potassium glass

It can be clearly seen from the above figure that lead-barium glass can be divided into two sub-categories according to the content of SiO_2, CuO and BaO chemical components: the first category is low SiO_2 to high CuO and BaO, which includes three cultural relics numbered 20,24 and 30. The second category is high SiO_2 ~ low CuO, Ba0 classes, which include Numbers for 32, 35, 31, 33, 37, 45, 46, 47, 55 nine cultural relics. High potassium glass can be divided into two sub-categories according to the content of SiO_2, PbO and P_2O_5 chemical components: the first category is low SiO_2 to high PbO and P_2O_5, which only includes one cultural relic numbered 03. The second category is high SiO_2 ~ low PbO, P_2O_5 classes, including such number 05, 06, 04, 01, 13, 14, 16, 18, 21 nine cultural relics.

5. Rationality analysis of system cluster model

According to the above conclusions, according to the results of cluster analysis, we choose SiO_2, CuO, BaO three chemical components as the basis for the sub-classification of lead-barium glass; SiO_2, PbO and P_2O_5 were selected as the basis for the sub-classification of high potassium glass. In order to verify the rationality of the above classification method, we used the graph constructor in SPSS software to draw the clustering effect diagram, as shown in Figure 4:

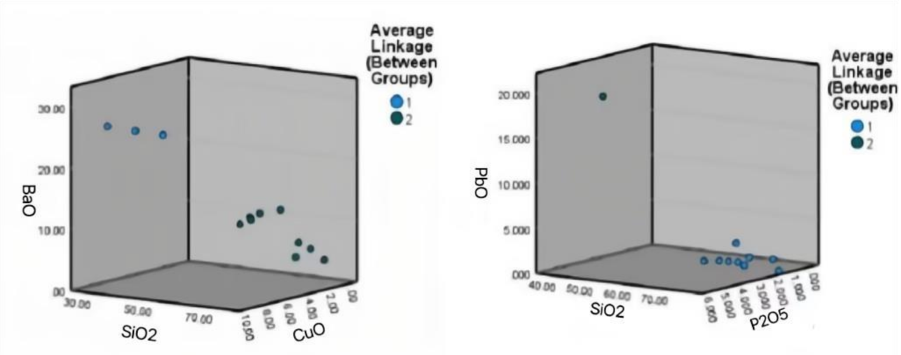

Figure 4: Clustering effect of lead-barium glass and high potassium glass

As can be seen from the figure above, for the three chemical components of lead-barium glass and high-potassium glass, the distribution of the two sub-categories in space is obviously a cluster of their own, independent distribution, no overlap occurs, and the space is far away, which indicates that the results of the clustering model established in this paper are relatively reasonable.

Taking high potassium glass as an example, it can be seen from the model establishment part of this question that the data of each sample is expressed as: $Xi-(x_{i1},x_{i2},...,x_{i,14})(i=1,2,...10)$. Without loss of generality, we choose for each sample x_{i1} Element adds the disturbance term ε, that is, $x_{i1}+\varepsilon$ in place of x_{i1}, Of which $\varepsilon \sim N(0,1)$. Keep up x_i Other elements remain unchanged, and cluster analysis is conducted again according to the perturbed data. Similarly, the treatment of lead-barium glass is the same as above. The above steps are repeated three times in SPSS, and the classification results before and after the disturbance are compared and analyzed. It can be found that the sub-categories of both high-potassium glass and lead-barium glass after the disturbance are completely consistent with those before the disturbance, which indicates that the model established in this paper has very low sensitivity and is very robust.

6. Conclusion

This study focuses on the application of Fisher linear discriminant analysis in the classification of ancient glass, combined with the systematic cluster analysis method, to improve the accurate classification of ancient glass samples. Through in-depth analysis of the problem, we show the limitations of traditional classification methods for complex and diverse ancient glass samples, and propose a comprehensive method combining Fisher linear discriminant analysis and systematic cluster analysis, aiming to improve the classification accuracy and efficiency. This study not only brings new ideas and methods to the field of ancient glass classification, but also provides enlightenment for similar complex sample classification problems. Through the practice and analysis of this research, we show innovative thinking and methods for traditional problems, and contribute new insights and possibilities to the development of ancient glass research and cultural relic protection.

References

[1] Tan Ke 'er, Tan Jie. Composition analysis and identification of ancient Glass products based on machine learning algorithm and optimization model [J]. Sino-arab Science and Technology Forum (Chinese and English), 2024, (05): 113-117.

[2] Li G Q, Lin X X, Peng M, et al. Composition analysis and identification of ancient glass products based on Random Forest and Support vector Machine [J]. Science and Technology Innovation and Application,2023,13(34):109-113.

[3] Liu, N., Cui, J. & Jiang, X. Xinjiang ancient glass research findings and technology [J]. Journal of western research, 2023, (4): 113-121 + 169.

[4] WANG Zidong. Mathematical model for composition analysis and identification of ancient glass products [J]. Modern information technology, 2023, 7 (14): 88-93 + 98.

[5] Cao J Y, Xu T Y, Liu Y, et al. Composition prediction and classification of ancient glass products based on RBF and SVM [J]. Science & Technology Innovation and Application,2023,13(18):41-43+48.

[6] Du, S., Le, J., Zhou, H. et al. Unearthed from the glass composition analysis and differential study [J]. Journal of northwest university for nationalities (natural science edition), 2023, 44 (02) : 8-16.

[7] Huang, H., Li, C., Liu, S. et al. Composition analysis and identification of ancient glass products based on composition data [J]. Mathematical Modeling and Applications,2023,12(02):52-62+124.

[8] Wan, H. & Deng, M. Statistical modeling of Composition analysis and identification of ancient glass products [J]. Mathematical Modeling and Applications,2023,12(02):27-40.

[9] Ma, P., Han, Y., Li, D. et al. Analysis and classification of ancient Glass products based on Composition data [J]. Mathematical Modeling and Applications, 2023, 12(02): 63-73. (In Chinese) DOI: 10.19943/J.2095-3070.jMMIA.2023.02.06.

[10] Chen, Z. & Ji, J. Based on hierarchy clustering model of ancient glass composition analysis and evaluation [J]. Journal of modern information technology, 2023, 7 (8): 122-125.