An Investigation into the Composition Analysis and Identification of Ancient Glass Artifacts

Lianjie Wang *

Department of engineering, Beijing Forestry University, Beijing, China

* Corresponding Author Email: cunima@outlook.com

Abstract. This study employs association analysis and chi-square tests to examine the relationship between weathering and glass type, color, and patterns in ancient glass artifacts. The findings reveal a stronger association between lead-barium glass and weathering, with weaker associations for color and patterns. Combining both methods demonstrates that glass type exhibits the strongest correlation with weathering, followed by color, while patterns remain unrelated. Utilizing kernel density plot analysis, the research investigates variations in chemical composition pre- and post-weathering, predicting component content before weathering. The classification of glass is based on intrinsic and extrinsic characteristics, highlighting notable differences in potassium oxide, lead oxide, and barium oxide between high-potassium and lead-barium glass. Scatter plots further illustrate these compositional disparities. Applying a decision tree, glass is classified according to patterns, color, and surface weathering, determining the influence of various factors. Hierarchical clustering is then used to categorize chemical compositions, identifying five categories for lead-barium glass and nine for high-potassium glass.

Keywords: Ancient Glass, Association Analysis, Decision Tree Approach.

1. Introduction

Chinese silk, as a principal commodity, played a pivotal role in trade between China and West and Central Asia, fostering the exchange of diverse goods, among which glassware featured prominently. The trade of silk facilitated intercultural and economic interactions between these regions, underscoring its significance in historical global commerce [1].

Early Chinese unearthed glass exhibits complex compositions, including sodium-calcium silicate, lead-barium silicate, and potassium-calcium silicate glasses, reflecting the diversity of ancient glassmaking techniques [2]. In antiquity, the addition of fluxing and stabilizing agents facilitated the production of glass with varying chemical compositions, such as lead-barium and potassium glasses, demonstrating diverse ancient glassmaking methodologies. The abundance of unearthed glass artifacts within China yields extensive and intricate data, necessitating meticulous documentation of diverse information associated with these relics [3].

Predominantly lead-barium silicate glassware represents a unique and indigenous Chinese glassmaking innovation. The glass industry in China has been influenced by the West, and this is closely related to the Central Asian ancestry of He Chou [4]. Analysis of incomplete data from Chinese ancient glass reveals two primary types: high-potassium and lead-barium glass. Detection constraints can cause variations in compositional ratios, so this research deems data falling between 85% and 105% as acceptable for examination.

2. Researching the relationship between weathering and glass.

2.1. Preprocessing based on existing data

Upon organizing and analyzing the existing data, it has been discovered that the color information is missing for the glass artifacts with identification numbers 19, 40, 48, and 58. An examination of the relationship between color and surface weathering in the data reveals a positive correlation between the degree of weathering and color depth. All artifacts with missing color data belong to the weathered category; hence, we have assigned the color value of black for these artifacts.

2.2. Analysis model based on association analysis method

The grey relational analysis method can provide quantitative basis for the dynamic changes of data. Therefore, this method is suitable for analyzing the dynamic process of data [5]. The evaluation objects can be compared and ranked based on the magnitude of the correlation between each alternative and the optimal solution. Patterns, categories, and surface differentiation can be considered as three sets: patterns = {A, B, C}, categories = {high potassium, lead-barium}, and colors = {blue-green, purple, light blue, dark green, light green, green, black, dark blue}.

Support refers to the ratio of support count to the total number of transactions, which is:

$$support(x \to y) = p(x, y) = \frac{A}{B} \tag{1}$$

A is the number of itemsets that simultaneously contain both x and y; B is the total number of itemsets.

For the rule {x} → {y}, the confidence is calculated by dividing the support count of {x, y} by the support count of {x}, which is:

$$confidence(x \to y) = p(y|x) = \frac{\{x,y\}}{\{x\}} \tag{2}$$

Lift is used to compare the difference between the probability of Y occurring when it is derived from the association rule {x} →{y} and the occurrence of Y in the total itemsets. The grey relational analysis method determines the weight coefficients by considering the differences among evaluation indicators, which effectively avoids interference from subjective factors. As a result, it can provide a more objective reflection of the roles of various evaluation indicators in the comprehensive evaluation index system [6]. Which is:

$$lift(x \to y) = \frac{p(y|x)}{p(y)} = \frac{p(x,y)}{p(x)p(y)} = \frac{\{x,y\}}{\{x\} \times \{y\}} \tag{3}$$

Confidence is defined as a measure used to explore the association between X and Y, which determines the likelihood of Y occurring given X, through the conditional probability of the association rule {x} → {y}.

$$conviction(x \to y) = \frac{P(\bar{y})}{p(\bar{y}|x)} = \frac{1-\{y\}}{1-\{x\}\to\{y\}} \tag{4}$$

Association rules reflect a strong link between two items and must satisfy either a minimum confidence or support threshold. The Apriori algorithm can optimize the process of finding such rules, which is the ultimate goal of association analysis. The algorithm steps are as follows:

(1) Set the minimum support threshold, calculate the support for all 1-itemsets, and filter out the frequent itemsets L1.

(2) Combine the 1-itemsets in L1 to obtain 2-itemsets, calculate their support, and filter out the frequent itemsets L2.

(3) Combine the 2-itemsets in L2 to obtain 3-itemsets, calculate their support, and filter out the frequent itemsets L3.

(4) Mine association rules from frequent itemsets, identifying the probability of one itemset appearing when another itemset appears.

(5) Set the minimum confidence threshold, calculate the confidence of association rules, and filter out high-confidence rules.

In this study, the algorithm is implemented in Python, utilizing the Apriori model from the efficient library, and setting minimum support and confidence thresholds. Taking a minimum support of 0.2 as an example, frequent itemsets L1, L2, and L3 are obtained. By employing this algorithm, potential association rules in the data can be uncovered, providing valuable information for data analysis, as shown in Table.1.

Table 1. All frequent itemsets and their occurrence counts in the itemset.

Frequent itemsetL1	Occurrence count.
{A}	22
{C}	30
{ High potassium }	18
{ Lead barium }	40
{Bule green}	15
{ Light blue }	20
{ Weathering }	34
{ Unweathered }	24

In this research, all association rules are derived from the frequent itemsets L2 and L3, and a minimum confidence threshold of 0.7 is set to filter out association rules with confidence greater than or equal to the minimum confidence. The analysis reveals three types of association rules:

(1) Lead-barium glass category is more prone to surface weathering;

(2) Glass with type C patterns and belonging to the lead-barium category is more susceptible to surface weathering;

(3) Light blue glass in the lead-barium category is more prone to surface weathering.

Glass type significantly affects surface weathering, particularly for lead-barium glass. Surface weathering is also influenced by patterns and colors, with light blue glass and type C patterns having a higher probability of weathering under lead-barium conditions. Increasing the minimum support threshold strengthens the association of association rules, resulting in fewer but stronger rules, with only {lead-barium} → {weathering} remaining. This suggests that the association between glass type and weathering is stronger than that between patterns/colors and weathering.

2.3. Analysis model based on chi-square test

The present study also attempted to use chi-square test to quantitatively determine the relationship between surface weathering and glass type, decoration, and color. A larger chi-square value indicates a stronger correlation between the observed and expected values, and thus a smaller independence [7].

Non-conforming data is removed, null values set to black and 0, and data merged to obtain a preprocessed input file before model establishment. Part of the data is shown in Table.2.

Table 2. An example image of merged partial results

Sampling points of cultural relics	Decoration.	Type.	color	Degree of surface weathering.	SiO_2	K_2O	CaO	MgO	AI_2O_3
01	C	High potassium	Blue green	Unweathered	69.33	9.99	6.32	.87	3.93
02	A	Lead barium	Light blue	Weathering	36.28	1.05	2.34	1.18	5.73
03 part I	A	High potassium	Blue green	Unweathered	87.05	5.19	2.01	.00	4.06
03 part II	A	High potassium	Blue green	Unweathered	61.71	12.37	5.87	1.11	5.50

04	A	High potassium	Blue green	Unweathered	65.88	9.67	7.12	1.56	6.44
05	A	High potassium	Blue green	Unweathered	61.58	10.95	7.35	1.77	7.50
06 part I	A	High potassium	Blue green	Unweathered	67.65	7.37	.00	1.98	11.15
06 part II	A	High potassium	Blue green	Unweathered	59.81	7.68	5.41	1.73	10.05
07	B	High potassium	Blue green	Weathering	92.63	.00	1.07	.00	1.98
08	C	Lead barium	Purple	Weathering	20.14	.00	1.48	.00	1.34
08 Severely weathered points	C	Lead barium	Purple	Weathering	4.61	.00	3.19	.00	1.11

The data was imported into IBM SPSS Statistics software, and chi-square distribution statistics were performed. Partial results are shown in the Fig 1.

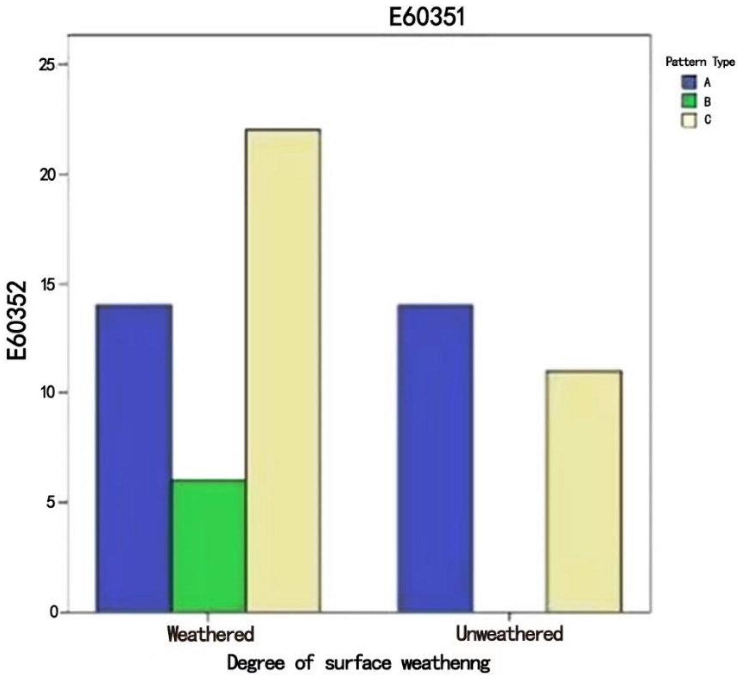

Fig 1. Degree of surface weathering, type of decoration, type, color.

Using SPSS for chi-square test analysis, the relationships between patterns, types, colors, and surface weathering were examined. The results show that the difference between patterns and weathering is not statistically significant (X2=5.722, P=0.057>0.05); however, the difference between types and weathering is statistically significant (X2=9.066, P=0.003<0.05), indicating significant differences in weathering outcomes for different categories. The difference between colors and weathering is also statistically significant (X2=14.358, P=0.045<0.05), suggesting significant differences in weathering outcomes for different colors.

From these findings, we can conclude that the surface weathering of glass artifacts is related to the glass type and color, but not to the patterns.

2.4. Statistical model for the content of weathering chemical components on sample surfaces

The initial step excludes data not meeting compound content requirements. Glass types are differentiated into high-potassium and lead-barium and further categorized based on weathering conditions. Certain chemical components with over 1/3 values equal to 0 are excluded to ensure statistical validity, reduce data processing, and improve efficiency.

Frequency distribution histograms are established for the four categories of data. Taking non-weathered lead-barium as an example, the frequency distribution histogram for silicon dioxide is shown in Fig 2.

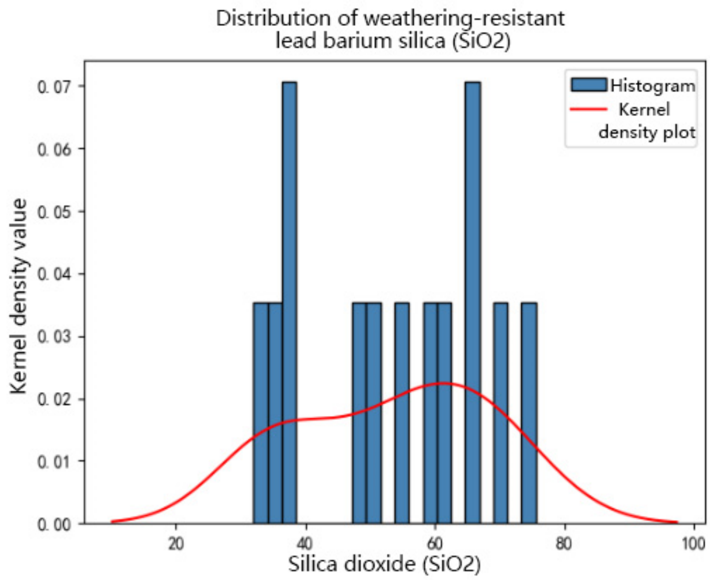

Fig 2. Unweathered lead barium silica (SiO_2)

2.5. Predictive model for chemical composition content prior to weathering

2.5.1 Model establishment

In this study, the content of a certain chemical component before and after weathering within the same glass type is represented graphically. Using Python for processing, we obtain Fig 3.-Fig .4 (taking non-weathered high-potassium category data as an example).

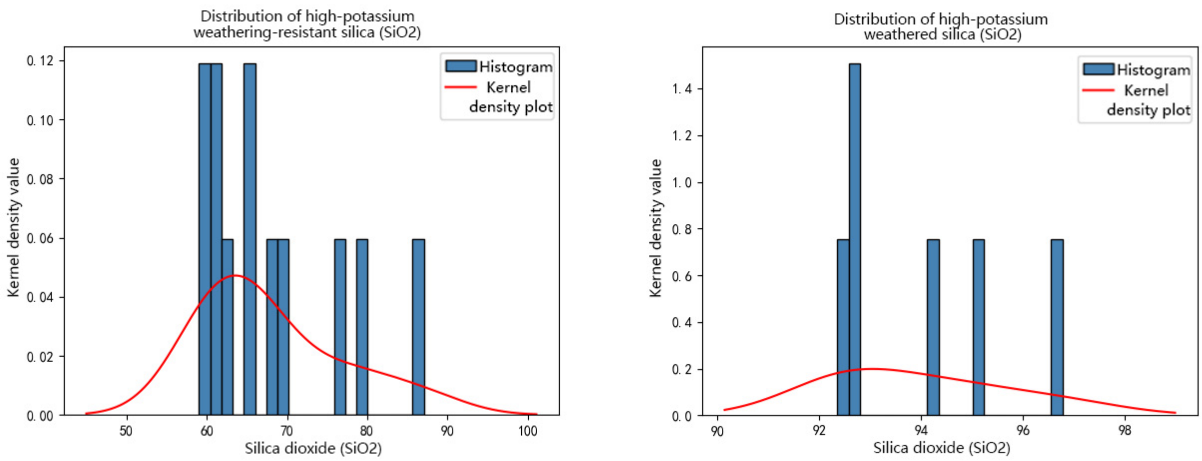

Fig 3. Distribution of unweathered and weathered SiO_2 with high potassium

From the above kernel density plot, it can be observed that the probability of silicon dioxide content in non-weathered high-potassium glass is highest in the range of 60%-70%, while the probability of silicon dioxide content in weathered high-potassium glass is highest at 92% or higher. Based on this comparison, it can be inferred that the silicon dioxide content in high-potassium glass is considerably lower before weathering than after weathering.

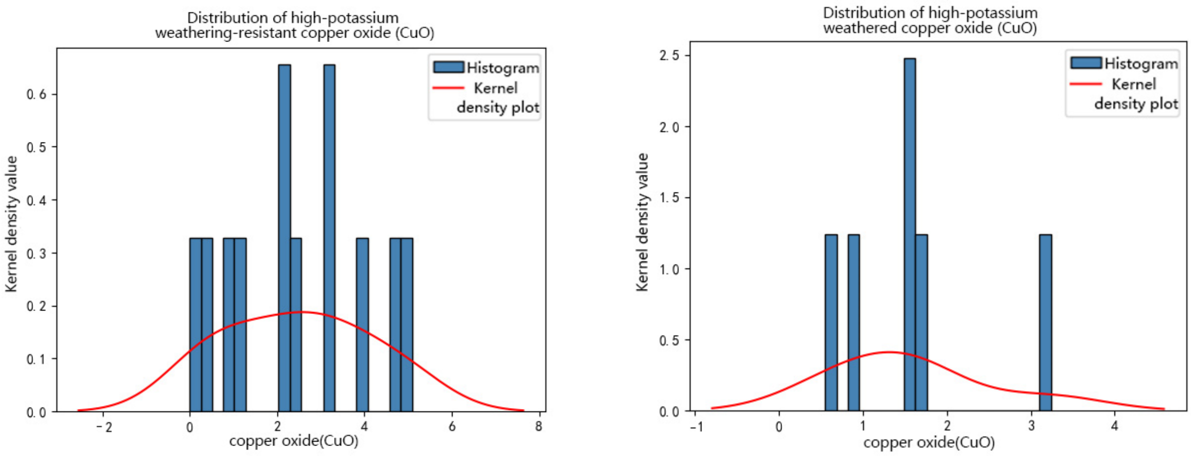

Fig 4. Distribution of unweathered and weathered CuO with high potassium

Kernel density plots show a slight increase in copper oxide content in non-weathered high-potassium glass compared to weathered glass. Histograms and kernel density plots for various chemical components in high-potassium and lead-barium glass reveal predicted contents for non-weathered glass. Components with over 1/3 values of 0 are excluded.

The predicted contents for non-weathered high-potassium glass are: SiO_2 63%, P_2O_5 1%, CaO 7.2%, K_2O 10.6%, Al_2O_3 6%, MgO 1.5%, FeO 2%, CuO 2.5%, with remaining contents at 0. Non-weathered lead-barium glass contents are: SiO2 63%, P_2O_5 0.2%, BaO 8%, CaO 0.8%, Al_2O_3 2.5%, PbO 19%, CuO 0.75%, with remaining contents at 0. The highest point of the curve in kernel density plots indicates the most probable content value.

2.5.2 Analyzing the classification rules of high-potassium glass and lead-barium glass

Presenting the chemical composition content of each sampling point of cultural relics in the form of graphs and charts, which are shown in Fig 5 and Fig 6.

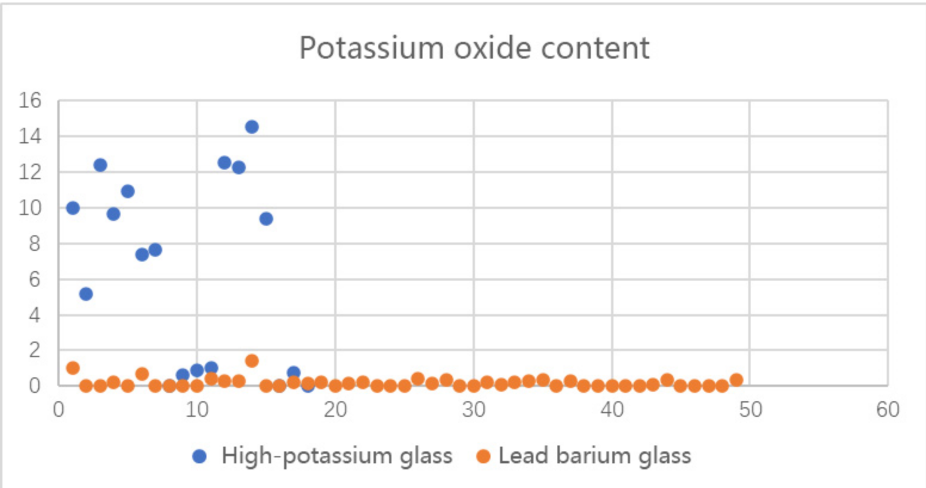

Fig 5. Graph of the potassium oxide content in high potassium glass and lead barium glass.

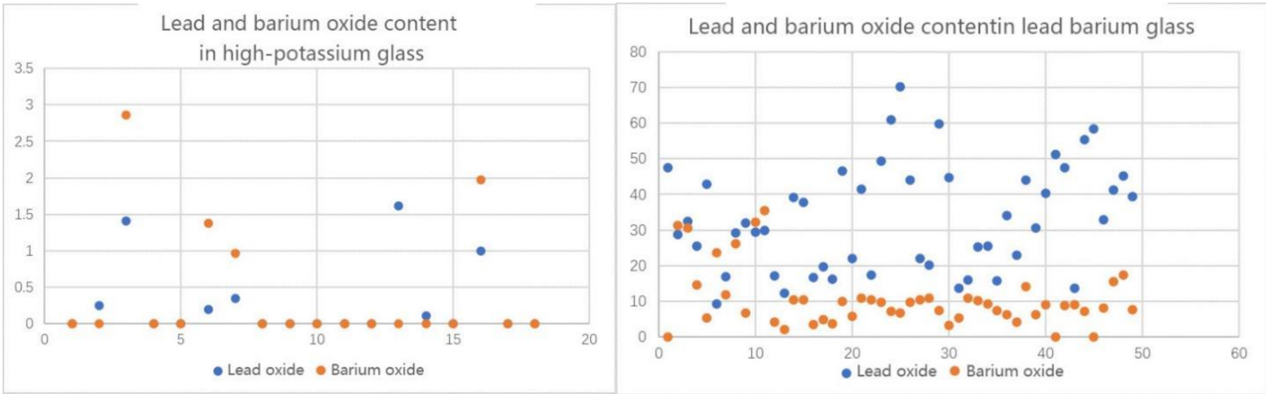

Fig 6. Graphs of the lead oxide and barium oxide content in high potassium glass and lead barium glass.

Lead-barium glass samples outnumber high-potassium glass samples, with higher lead oxide and lower potassium oxide content in lead-barium glass. Lead-barium glass can be identified by high lead oxide and approximately 10.49% barium oxide content.

A decision tree model was used for classification, with decoration types A, B, and C defined as 1, 2, and 3, respectively; high-potassium and lead-barium types as 1 and 2, and colors as ordinal numbers from 1 to 8. Black and purple colors, decoration type A, and weathering are significantly associated with lead-barium glass, while blue-green color, decoration types B and C are more related to high-potassium glass. The Jupyter NoteBook was used to run the decision tree code, indicating a significant relationship between certain variables and glass type.

3. Establishment and solving of a sub-classification model based on cluster analysis method

3.1. Establishment of a classification model using cluster analysis method

The 18 high potassium data were divided into 5 categories based on 14 chemical conditions, while the remaining 49 data were divided into 4 categories, with the number of categories determined by the dendrogram.

Hierarchical clustering is a mature and commonly used clustering method [8]. Through the hierarchical clustering analysis dendrogram, we can see which variables or samples have a stronger correlation [9]. Calculate the distance *dij* between each pair of the 18 sample data groups. The distance is calculated using the Euclidean distance, that is:

$$d(\vec{x_i}, \vec{x_j}) = \sqrt{\sum_{k=1}^{p} (x_{ik} - x_{jk})^2} \tag{5}$$

Where *i* and *j* represent two distinct samples from the 18 samples, and *k* represents a chemical composition, $p = 14$. The distance matrix $D_0 = (d_{ij})$ n x n, where $n = 18$, is constructed using the Euclidean distance. Only the first 5 rows and 5 columns are shown below:

$$\begin{array}{cc} \begin{bmatrix} 0 & & & & \\ 19.22 & 0 & & & \\ 8.87 & 27.29 & 0 & & \\ 4.75 & 22.56 & 6.79 & 0 & \\ 8.76 & 27.18 & 4.76 & 4.78 & 0 \end{bmatrix} & \begin{matrix} G1 \\ G2 \\ G3 \\ G4 \\ G5 \end{matrix} \end{array} \tag{6}$$

There are a total of 18 samples, so we construct 18 classes represented as *Gn*, where *n* = 1, 2, 3, ..., 18. Each class contains only one sample point, and the distance between different classes is the distance between samples. The distance between all different classes is calculated using the class average method, which is expressed as:

$$D(G_1, G_2) = \frac{1}{n_1 n_2} \sum_{x_i \in G_1} \sum_{y_i \in G_2} d(x_i, y_i) \quad (7)$$

In this study, the average distances between 18 categories were calculated, where G1 and G2 represent the two categories, and n_1 and n_2 denote the number of sample points in each category, respectively. The results show that the minimum inter-class distance is D(G1, G4) = 4.75, so they are merged into a new class G19. By continuously performing clustering, calculating the distances between classes, and constructing a new distance matrix D1, a comprehensive category is eventually obtained.

By analogy, clustering is continuously conducted, and the distances between various categories are calculated, followed by the construction of a new distance matrix. This process is repeated until a final new category is ultimately acquired.

3.2. Solving of a classification model using cluster analysis method

Based on a literature survey [10], The modern glass-making process involves SiO_2, BaO, Al_2O_3, ZnO, CaO, MgO, Fe_2O_3, CuO, Na_2O, and K_2O, with Fe_2O_3 and CuO as the main colorants. Lead-barium glass samples outnumber high-potassium glass samples, with higher lead oxide content in the former. High-potassium glass is characterized by high potassium oxide content, while lead-barium glass has low or zero potassium oxide content. A decision tree is used for classification, with pattern types A, B, and C defined as 1, 2, and 3, high-potassium and lead-barium classes as 1 and 2, and colors black, blue-green, green, light blue, light green, dark blue, dark green, and purple as 1 to 8, and unweathered and weathered as 0 and 1, respectively.

This model helps determine the type of glass based on various characteristics. The preprocessed data file is used to run the decision tree code in Jupyter NoteBook, obtaining the decision tree as shown in Fig 7.

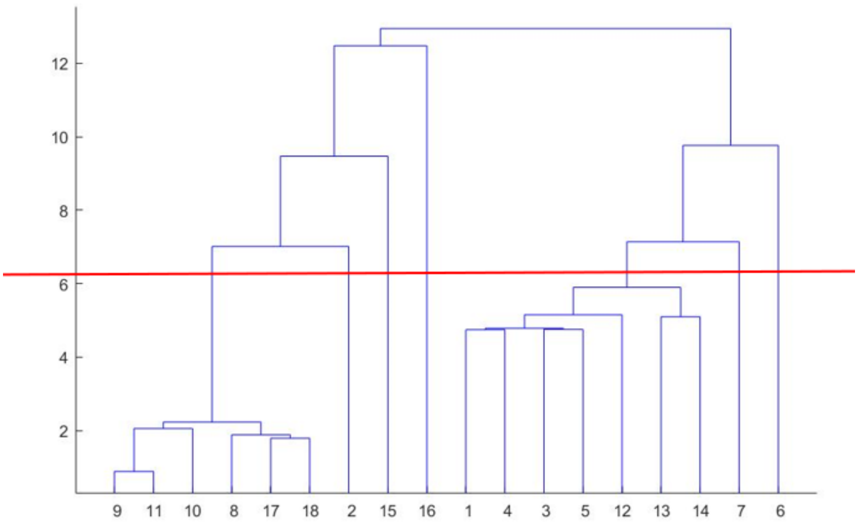

Fig 7. Dendrogram of high potassium glass data.

Here, K value was set to 8, resulting in the classification of all high potassium glass samples into 5 groups. Combining tables 1 and 2, the resulting table was generated as shown in Table.3.

Table 3. Table of high potassium glass grouping.

Table of high potassium glass grouping.	Samples of cultural relics.
First group	3part I, 7, 9, 10, 12, 22, 27
Second group	18
Third group	21
Fourth group	1, 3part II, 4, 5, 6part II, 13, 14, 16
Fifth group	6part I

From the table, it can be seen that the first and fourth groups have a larger number of samples, but in the second, third, and fifth groups, there is only one sample in each group. This indicates that the chemical composition of these three sampling points is significantly different from that of other cultural relic sampling points, and there are also significant differences between them.

The same method was used to process the sample data of lead barium glass, and the clustering diagram was drawn using MATLAB in Fig 8.

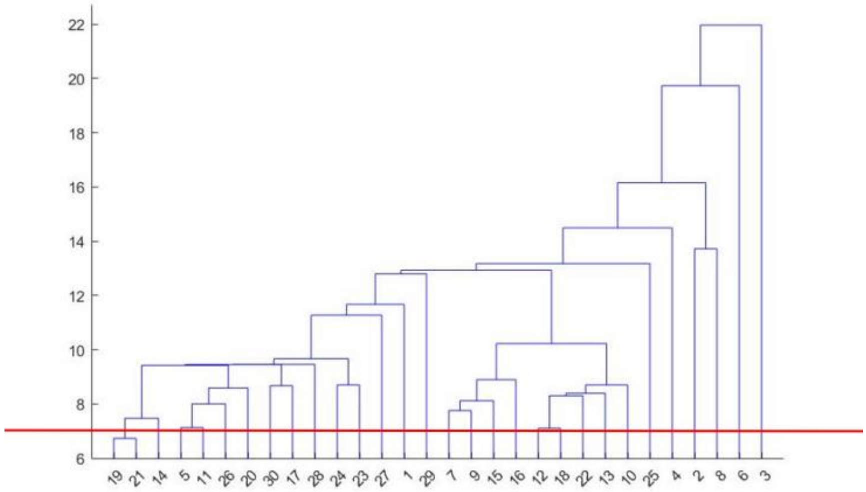

Fig 8. Dendrogram of lead barium glass sample data

Here, all lead barium glass samples were divided into 9 groups, combining tables 1 and 2, the data is shown in Table.4.

Table 4. Groups of lead barium glass

lead barium glass	Samples of cultural relics.
First group	2,19,26 Severely weathered point, 30part I, 32, 34, 35, 36, 38, 39, 41, 42Non-weathered point.1, 42Non-weathered point.2, 43part II
Second group	43part I
Third group	23Non-weathered point, 25Non-weathered point, 26, 28Non-weathered point, 29Non-weathered point, 30part II, 31, 33, 37
Forth group	40
Fifth group	11
Sixth group	8
Seventh group	24
Eighth group	20
Ninth group	8 Severely weathered point

Nine groups of lead barium glass were identified, with the first and third groups having more samples and greater chemical similarity. Remaining groups had distinct chemical compositions and were classified separately. Systematic clustering can result in unsatisfactory classification outcomes. To analyze the rationality and sensitivity of the classification, the value of a certain chemical composition can be increased by 5% and then the systematic clustering can be performed again to observe whether there is a significant change in the classification results [11].

4. Conclusions

Glass analysis and identification can be categorized into external feature analysis and identification and internal feature analysis and identification for a more comprehensive and accurate process. The kernel density graph of the chemical composition of non-weathered glass can provide predicted values of the chemical composition, simplifying the prediction process and improving efficiency. Systematic clustering analysis of unknown glass with high potassium glass data may impact the division of high potassium glass itself and subsequently affect the identification of unknown glass. The selection of *K* in systematic clustering analysis is subjective and directly affects classification results; however, the elbow rule can be utilized to select the optimal number of clusters. Differences in colored glasses' chemical composition are mainly reflected in Fe_2O_3 and CuO. In the same color series, as the color becomes lighter, the correlation between Fe_2O_3 decreases, and CuO increases. SiO_2, PbO, BaO, P_2O_5, Al_2O_3, and CaO have a strong correlation with any color of glass, while the content of Fe_2O_3 and CuO varies significantly, as evidenced by the relationship matrix's first six columns. Comprehensive analysis of calculation data revealed that the relationship between chemical compositions has obvious regularity in different colored glasses.

References

[1] An Jiayao. The Silk Road and Glassware[J]. Cultural relics world, 2021, No.366(12): 87-92.

[2] Huang Xiaojiao. Comparative study of ancient glassware from China National Museum collection in China and the West[J]. Cultural relics world, 2023, No.381(03):86-91.

[3] Feng Bailing. Construction and Application of the Database of Ancient Glass Beads Excavated in China[D]. Northwest University, 2021. DOI: 10.27405/d.cnki.gxbdu.2021.001676.

[4] Shang Gang, Three Thousand Years of Chinese Glass [J]. Forbidden City, 2021, No. 317 (06): 14-31.

[5] Gai Yingli. Research on the Waterproof Reliability Evaluation of Subway Station Structures Based on Grey Relational Analysis [J]. Modern Urban Rail Transit, 2023 (04): 58-62.

[6] Bu Yinghao, Shi Hong. Research on the Evaluation of Innovation-Driven Ability in Guizhou's Advanced Equipment Manufacturing Industry: Based on Entropy Method and Grey Relational Analysis [J]. Productivity Research, 2023, No. 368 (03): 59-63+70. DOI: 10.19374/j.cnki.14-1145/f.2023.03.004.

[7] Li Meng, Bao Lei, Hu Yi. Implementation of online detection method for random numbers based on chi-square test[J]. Microelectronics, 2022,52(03): 388-392. DOI: 10.13911/j.cnki.1004-3365.210329.

[8] Jiang Qiyuan. Mathematical model[M]. Beijing: Higher Education Press,2007.

[9] Du Junpeng, Lv Jun,Wu Jisheng. Research on Water Quality Evaluation of Rivers Based on Principal Component Analysis and Systematic Clustering Analysis[J]. Water conservancy technology supervision,2022, No.182(12):216-220.

[10] Kan Yinghao. Technological Analysis of Excavated Glass from Glass Workshops in Boshan, Shandong, at the End of the Yuan Dynasty and the Beginning of the Ming Dynasty [D]. Shandong University, 2022. DOI: 10.27272/d.cnki.gshdu.2022.002831.

[11] Fu Deyin. Statistical testing issues in Q-type system clustering analysis[J].Statisyics and Information Forum,2007(03):10-14.