Composition analysis and identification of glass products based on Pearson correlation analysis

Zhijin Cai *, Zhonghan Zheng, Xin Jiang

College of Automation, Jiangsu University Of Science And Technology, Zhenjiang, Jiangsu , 212100, China

* Corresponding Author Email: 3051418298@qq.com

Abstract. Ancient glass objects record the cultural exchanges between East and West along the Silk Road, but they have been subjected to varying degrees of weathering over time. First, to investigate the relationship between surface weathering and glass type, decoration and color of glass artifacts, this paper investigated the correlation using Kappa consistency test and judged the difference using chi-square test, and found that there was significant correlation and difference between surface weathering and glass type. For the statistical pattern between the presence or absence of weathering on the surface and the chemical composition content, the content of each chemical composition was compared between weathered and unweathered on the surface by visualizing its chemical composition content. Secondly, to analyze the classification pattern of high potassium and lead-barium glass, it is subclassed. In this paper, the chemical constituents relevant to the classification were first selected by Pearson correlation analysis, followed by subclassification of the two types of glass artifacts by establishing a K-Means clustering subclassification model based on the elbow method to determine the number of cluster centroids. Thirdly, in order to identify the unknown types of glass artifacts, this paper establishes a random forest regression algorithm for glass type identification model and derives the results by prediction. Finally, different categories of glass artifacts are analyzed to distinguish glass from four specific categories in terms of decoration, type, color and surface weathering, and the differences in chemical composition correlation relationships between different categories are obtained by comparing the chemical composition correlation tables of different categories.

Keywords: Chi-square test, Pearson correlation analysis, K-Means clustering, random forest regression algorithm

1. Introduction

The Silk Road [1] is the embodiment of the value of cultural exchanges between China and the West, and the early ancient glass was made in China by taking indigenous building materials after absorbing Western technology, but it is highly susceptible to weathering from the environment of burial and preservation, leading to changes in composition.

Li Liuliu in his study concluded the results that atmospheric foul acidic gases cause different degrees of deterioration on the surface of carbonate artifacts [2], and Du Rong [3] studied that the natural environment causes extremely serious disease phenomena in artifact buildings, with nearly 90% of artifact buildings producing varying degrees of disease, and the disease phenomena are prominent and extremely widely distributed. In this paper, based on a batch of data related to ancient glass products in China, we studied and explored various aspects of their surface weathering products and morphology, and the specific problems are as follows.

(1) The relationship between the surface weathering of glass artifacts and their glass type, decoration and color; combined with the type of glass, the analysis of the surface of the artifact samples with and without the weathering of chemical composition content of the statistical law, and according to the weathering point detection data, to predict their chemical composition content before weathering.

(2) Analyze the classification laws of high potassium glass and lead-barium glass; for each category, choose the appropriate chemical composition to classify them into subcategories, give

specific classification methods and classification results, and analyze the reasonableness and sensitivity of the classification results.

(3) The chemical composition of unknown categories of glass artifacts to identify the type to which they belong, and the sensitivity of the classification results.

(4) For different categories of glass artifact samples, analyze the correlations between their chemical compositions and compare the differences of chemical composition correlations between different categories.

2. Data pre-processing

We filtered the data through 58 groups of glass artifacts related to type, decoration, color and surface weathering, using EXCEL software to eliminate component ratio outliers, the results are shown in Table 1.

Table.1. Glass artifact sample composition ratio outliers

Artifact Number	Chemical composition ratio/%				
Heritage Sampling Sites 15	Silicon Dioxide	Sodium oxide	Potassium oxide	Magnesium oxide	Aluminum oxide
	61.87	3.21	7.44	1.02	3.15
	Iron oxide	Copper oxide	Lead oxide	Barium oxide	Phosphorus pentoxide
	1.04	1.29	0.19	0	0.26
Heritage Sampling Sites 17	Silicon Dioxide	Sodium oxide	Potassium oxide	Magnesium oxide	Iron oxide
	60.71	2.12	5.71	0.85	1.04
	Copper oxide	Lead oxide	Barium oxide	Phosphorus pentoxide	Tin oxide
	1.09	0.19	0	0.18	0

3. Model building and analysis

3.1. Analysis of the relationship between weathering and glass type, decoration and color

Since the relationships between the data include correlations and differences, the relationship between surface weathering and its glass type, ornamentation and color is explored in two aspects. The above process is plotted as a flow chart as shown in Figure 1.

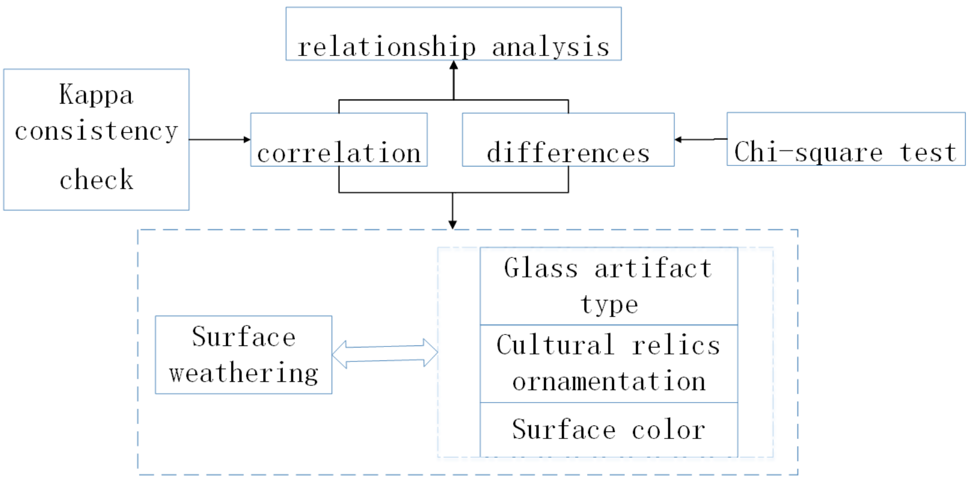

Figure 1. Flow chart of the relationship between surface weathering and glass type, decoration and color

3.1.1 Correlation - Kappa consistency test [4]

Suppose the number of true samples in each category is $a_1, a_2, \cdots, a_c$, and the number of predicted samples in each category is $b_1, b_2, \cdots, b_c$, and the total number of samples is n, then we have

$$p_e = \frac{a_1 \times b_1 + a_2 \times b_2 + \cdots + a_c \times b_c}{n \times n} \tag{1}$$

The results of the Kappa consistency test are shown in Table 2.

Table.2. Kappa consistency test table

	p-value	Standard Error	Z-value	Kappa value
Glass decoration and glass surface weathering	0.492	/	/	/
Glass type and glass surface weathering	0.024**	0.13	2.25	0.292
Glass color and glass surface weathering	0.674	/	/	/

Note: ***, **, and * represent 1%, 5%, and 10% significance levels, respectively.

According to the Kappa consistency test table, the significance p-value of glass decoration is 0.492>0.05, which does not present significance at the level and cannot reject the original hypothesis, indicating that there is no correlation between glass decoration and glass surface weathering; the significance p-value of glass type is 0.024**<0.05, which presents significance at the level and rejects the original hypothesis, indicating that there is a correlation between glass type and There is a correlation between glass surface weathering; the significance p-value of glass color is 0.674>0.05, which does not present significance at the level, and the original hypothesis cannot be rejected, indicating that there is no correlation between glass color and glass surface weathering.

3.1.2 Variance-Cardinal test [5]

The chi-square test is the degree of deviation between the actual observed value and the theoretical inferred value of the statistical sample. The degree of deviation between the actual observed value and the theoretical inferred value determines the size of the chi-square value.

The data on the decoration, type and color of each glass artifact were processed using SPSSPRO software to count the amount of surface weathering, and the results were obtained as shown in Tab.3.

Table.3. Cardinality test results table

Category	Name	Surface weathering/pc		Total/pc	X^2	Correction X^2	Significance P
		No weathering/pc	Weathering/pc				
Ornamentation	A	11	11	22	4.941	4.941	0.085*
	B	0	6	6			
	C	11	17	28			
Type	Lead Barium	12	28	40	5.061	3.790	0.024**
	High Potassium	10	6	16			
Color	NAN	4	0	4	9.962	9.962	0.268
	light green	1	2	3			
	Pale Blue	12	6	18			
	Dark Green	4	3	7			
	Deep Blue	0	2	2			
	Purple	2	2	4			
	Green	0	1	1			
	Blue-Green	9	6	15			
	Black	2	0	2			

Among them, it is clear from the question that some artifacts may be too weathered to identify their colors, so the table replaces their colors with NAN.

According to the chi-square test results table, the significance p-value of glass decoration is 0.085*> 0.05, which does not present significance at the level, so the original hypothesis is accepted and there is no significant difference between glass surface weathering and glass decoration; the significance p-value of glass type is 0.024**<0.05, which presents significance at the level, so the original hypothesis is rejected and there is no significant difference between glass surface weathering and glass The p-value of significance for glass type is 0.268> 0.05, which does not show significance at the level, so the original hypothesis is accepted, and there is no significant difference between glass surface weathering and glass color.

In summary, combining correlation and difference, it can be seen that the surface weathering of glass artifacts has correlation with glass type, but not with glass decoration and glass color; at the same time, the surface weathering of glass artifacts before and after weathering has difference with glass type, but not with glass decoration and glass color.

3.2. Analysis of chemical composition content with and without weathering

3.2.1 Statistical law of the content of chemical components with and without weathering

In order to clearly visualize the data changes of chemical composition content of high potassium and lead barium glass with and without weathering on the surface, we use line graphs to show the data of chemical composition content of high potassium and lead barium glass before and after weathering as shown in Figure 2~5.

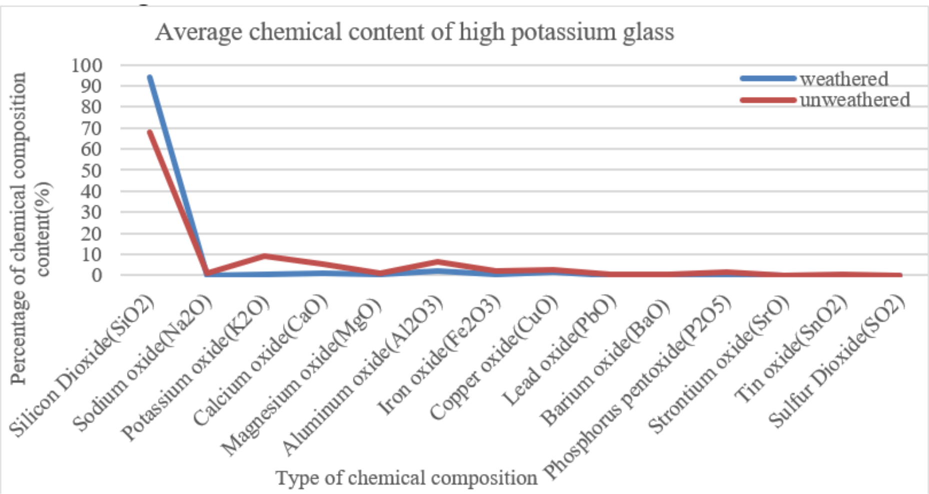

Figure 2. Average chemical content of high potassium glass

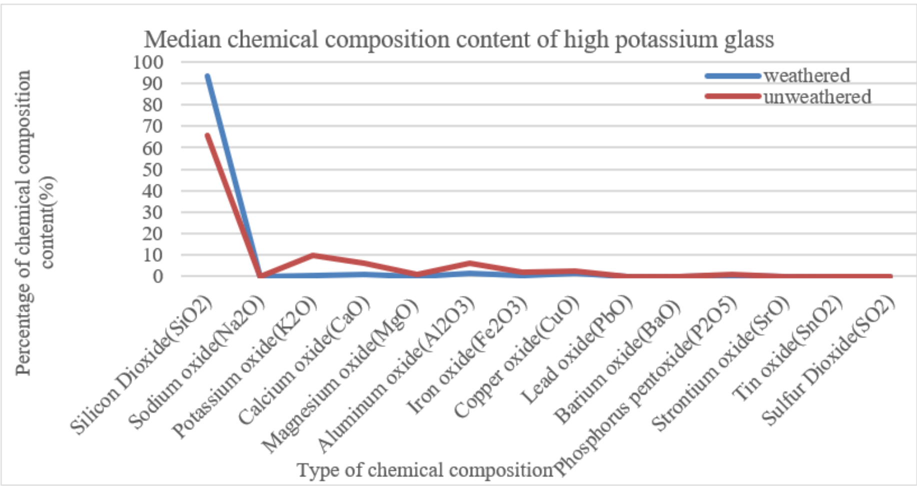

Figure 3. Median chemical composition content of high potassium glass

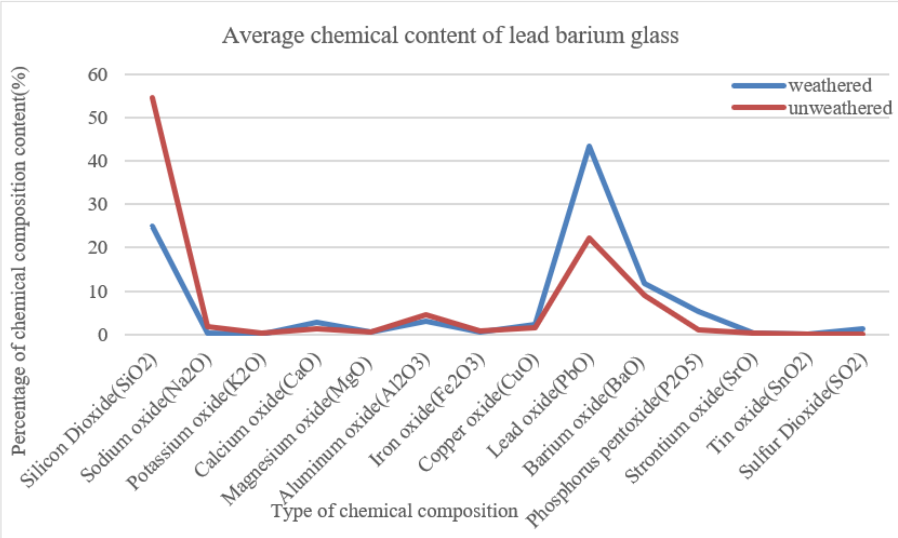

Figure 4. Average chemical content of lead barium glass

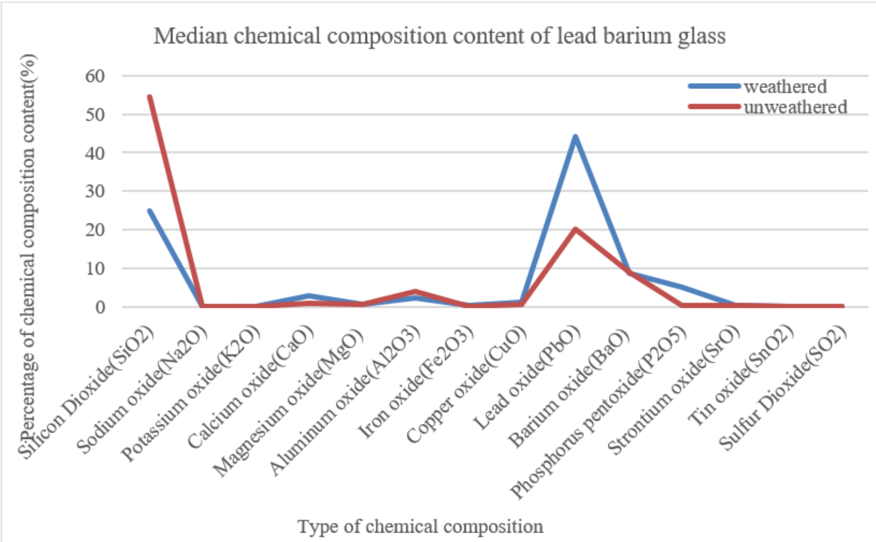

Figure 5. Median chemical composition content of lead barium glass

From different glass types, the table describing the content of chemical components on the surface of artifact samples with and without weathering and the mean and median line graphs of component content in Figures 2~5 can be roughly seen that the chemical components of high potassium glass are basically smaller in the case of weathering on the surface than in the case of unweathered, and the chemical components of lead-barium glass are basically larger in the case of weathering on the surface than in the case of unweathered.

3.2.2 Prediction of chemical composition content of high potassium and lead-barium glass before weathering

Based on the mean and standard deviation of the content of each chemical component of high potassium glass and lead-barium glass in the study of statistical laws, we develop a prediction model for the content of chemical components before weathering.

$$\frac{u_{\text{wij}} - \overline{u_{\text{wj}}}}{\sigma_{\text{wj}}} = \frac{u_{\text{uij}} - \overline{u_{\text{uj}}}}{\sigma_{\text{uj}}} (i - \text{sample serial number}, j - \text{chemical composition}) \tag{2}$$

Where u_{wij} is the proportion of a chemical component in the total component content of a sample after weathering, $\overline{u_{\text{wj}}}$ is the mean of the proportion of a chemical component in the total component content after weathering, and σ_{wj} is the standard deviation of the proportion of a chemical component in the total component content after weathering.

Deformation of equation (3) yields:

$$u_{\mathrm{uij}} = \frac{u_{\mathrm{wij}} - \overline{u_{\mathrm{wj}}}}{\sigma_{\mathrm{wj}}} \cdot \sigma_{\mathrm{uj}} + \overline{u_{\mathrm{uj}}} \tag{3}$$

The data on the content of each chemical component of high potassium glass and lead-barium glass after weathering were substituted into equation (4), respectively, and the results were obtained as shown in Table 4.

Table.4. Predicted content of each chemical component of high potassium glass and lead-barium glass before weathering (partial)

Type	Heritage Sampling Sites	Silicon Dioxide	Sodium oxide	Potassium oxide	Calcium oxide	Magnesiu m oxide	... ...
High Potassium	07	61.25058	0	4.546199	6.600506	0.645049	... ...
High Potassium	09	73.32054	0	9.741783	3.747493	0.645049	... ...
High Potassium	10	82.15838	0	12.64779	1.148082	0.645049	... ...
Lead Barium	08	49.33645	0.76167	0.046248	0.379705	0.137357	... ...
Lead Barium	08 Severe weathering point	32.0154	0.76167	0.046248	1.703276	0.137357	... ...
Lead Barium	11	64.33761	0.76167	0.317592	1.950962	0.686873	... ...
...	......	......	......	......	......	......	... ...

Since the content ratio of each component has a practical meaning and cannot be less than 0, the predicted value is directly assigned to 0 for the case where it is less than 0.

The total component content of the heritage sampling sites for 08 severely weathered sites and 26 severely differentiated sites is less than 85%, which is below the effective value range, but based on the severe weathering of the glass surface, which may lead to increased detection difficulty or other difficulties, resulting in a severe under-representation of the total percentage, the predicted results are considered to be within the acceptable range.

The percentage of total composition content of the artifact sampling points for 02, 36, 38 and 48 is greater than 105%, which is higher than the range of valid values. By exploring the original data, it is found that the silica content of these cases are significantly larger than other records, and it is considered that the percentage of weathered chemical composition content of the original records may have a record or detection error. Also the method itself may have some error due to the large predicted volume.

3.2.3 Reasonableness Analysis - Predicted Content Based on XGBoost Regression Test [6]

To determine the reasonableness of the predicted values of high potassium and lead-barium glass on the chemical composition before weathering, the average relative error was calculated using XGBoost regression model to verify.

For a dataset containing n m-dimensional entries, the XGBoost model can be expressed as

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i),\ f_k \in F(i = 1, 2, \cdots, n) \tag{4}$$

where $F=\left\{f(x)=w_q(x)\right\}(q:R^m \to \left\{1,2,\cdots,U\right\}, w \in R^U$ is the set of CART decision tree structures, q is the tree structure of the sample mapping to the leaf nodes, U is the number of leaf nodes, and w is the real number fraction of leaf nodes. The objective function of the XGBoost model can be divided into an error function term L and a model complexity function term Ω. The objective function can be written as

$$Obj = L+\Omega, L=\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2, \Omega=\gamma T+\frac{1}{2}\lambda\sum_{j=1}^{T}w_j^2 \tag{5}$$

Where γT is the L_1 regular term and $\frac{1}{2}\lambda\sum_{j=1}^{T}w_j^2$ is the L_2 regular term.

The XGBoost classification algorithm was used to predict the classification of samples with a gbtree learner, its learner number was 100, the L1 regular term was 0, the L2 regular term was 1, the maximum depth of the tree was 10, the sampling rate of the sample levy was 1, the sampling rate of the tree features was 1, the sampling rate of the node features was 1, and the minimum weight of the samples in the leaf nodes was 0. Taking the percentage of the chemical composition of silica For example, using SPSSPRO software, 70% of the chemical composition content was used as the training set, and the remaining 30% was used as the test set to test, and the predicted values of the chemical composition content percentage of high potassium and lead-barium glass were obtained, as shown in Table 5,6.

Table.5. Comparison of power load forecasting of 403 line

Heritage Sampling Sites	7	9	10	12	22	27
Silicon Dioxide	69.202194	86.56498	61.70094	65.843506	84.28146	69.202194

Table.6. Predicted silica content of lead-barium glass as a percentage (partial)

Heritage Sampling Sites	2	8	08 Severe weathering point	11	19	26	...
Silicon Dioxide	51.798	53.779	31.938305	50.607	67.918	63.233	...

Based on the percentage of silica content of high potassium and lead-barium glass calculated using the XGBoost model in Table 6 and Table 7, and the percentage of silica content of high potassium and lead-barium glass calculated using the prediction model in 5.2.3, the average relative error of silica content was calculated:

$$\bar{\delta}=(\sum_{i=1}^{i_w}\frac{u_{wij}-XG_{ij}}{u_{wij}})/i_w \tag{6}$$

Where, XG_{ij} denotes the predicted values tested by XGBoost regression and denotes the number of samples weathered.

The average relative error of silica content of high potassium glass is -8.95% and that of lead-barium glass is -3.85%, both of which are less than 10% in absolute value. In summary, we believe that the model predictions of the percentage of silica content before weathering for high potassium and lead-barium glass are more reasonable. The remaining chemical components were examined in the same way as silica, and will not be described in this paper. Therefore, the model predicting the percentage of content of each chemical component before weathering is more reasonable.

3.3. Sub-category classification of different glasses

For each category to be divided into subcategories, we can classify them by clustering and analyze the classification results from the perspective of rationality and sensitivity. In dealing with the rationality, we consider two aspects: the profile coefficient and the comparison of the characteristics of each type of chemical composition content, and the sensitivity considers the stability of the judgment of the impact on the results after changing the chemical composition content, and the robustness of the impact on the results after changing the algorithm parameters. The specific flow chart is shown in Figure 6.

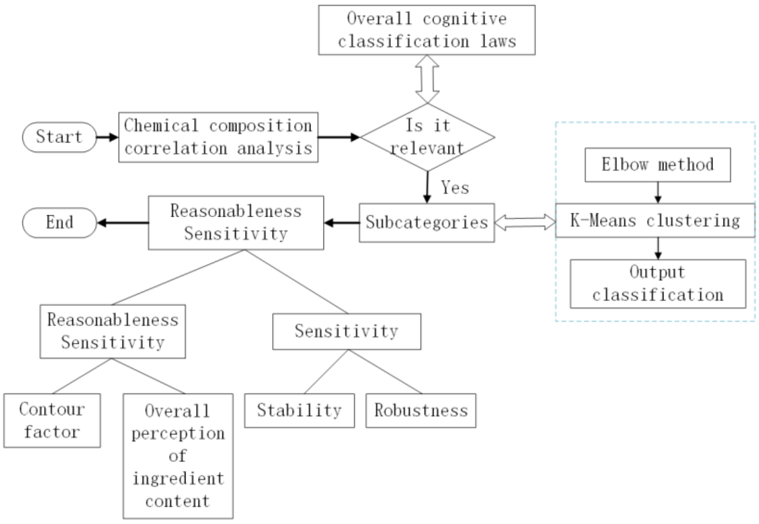

Figure 6. Neural network structure

3.3.1 Exploring the classification law of high potassium and lead-barium glass

In order to investigate the classification pattern of high potassium glass and lead-barium glass, we quantified the types and analyzed the classification pattern of high potassium glass and lead-barium glass by analyzing the correlation between their types and chemical composition and visualizing the content of each chemical composition to process the overall perception in two aspects.

1. Chemical composition and type quantitative correlation analysis

(1) Principle of Pearson Correlation [7] Analysis Model

The chemical composition contents of the glass artifacts all satisfy a normal distribution, so the correlation analysis was performed by Pearson correlation analysis. The chemical composition content data of high potassium glass $X:\{X_1, X_2, \cdots, X_n\}$, and the chemical composition content data of lead-barium glass $Y:\{Y_1, Y_2, \cdots Y_n\}$ are the overall data, then the overall mean value

$$E(X)=\frac{\sum_{i=1}^{n}X_i}{n}, E(Y)=\frac{\sum_{i=1}^{n}Y_i}{n}, \tag{7}$$

Overall covariance:

$$Cov(X,Y)=\frac{\sum_{i=1}^{n}(X_i-E(X))(Y_i-E(Y))}{n} \tag{8}$$

Overall Pearson coefficient:

$$\rho_{XY}=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}=\frac{\sum_{i=1}^{n}\frac{(X_i-E(X))}{\sigma_X}\frac{(Y_i-E(Y))}{\sigma_Y}}{n} \tag{9}$$

Where σ_X is the standard deviation of X and σ_Y is the standard deviation of Y.

$$\sigma_X=\sqrt{\frac{\sum_{i=1}^{n}(X_i-E(X))^2}{n}}, \sigma_Y=\sqrt{\frac{\sum_{i=1}^{n}(Y_i-E(Y))^2}{n}} \tag{10}$$

(3) Pearson correlation analysis model solution

According to the equations (7) ~ (10) combined with the chemical composition content data of high potassium and lead-barium glass, the correlation coefficients of high potassium and lead-barium glass categories and each chemical composition type were solved as shown in Table 7.

Table.7. Correlation coefficients of high potassium and lead-barium glass categories and each chemical composition type

Chemical composition	Correlation coefficient	Significance p-value
Silicon Dioxide	-0.694	0.000***
Oxidation Na	/	0.337
Potassium oxide	-0.717	0.000***
Calcium oxide	-0.345	0.004**
Magnesium oxide	/	0.441
Aluminum oxide	/	0.101
Iron oxide	-0.273	0.025**
Copper oxide	/	0.659
Lead oxide	0.757	0.000***
Barium oxide	0.535	0.000***
Phosphorus pentoxide	0.285	0.019**
Strontium oxide	0.535	0.000***
Tin oxide	/	0.437
Sulfur Dioxide	/	0.329

According to the correlation coefficients and significant p-values in the table, it was found that high potassium and lead-barium glass categories were correlated with silica, potassium oxide, calcium oxide, iron oxide, lead oxide, barium oxide, phosphorus pentoxide, and strontium oxide, where silica, potassium oxide, and calcium oxide were negatively correlated with glass categories, and lead oxide, barium oxide, phosphorus pentoxide, and strontium oxide were positively correlated with glass categories.

2.Visualization of chemical composition content

For the visualization of the content of silica, potassium oxide, calcium oxide, iron oxide, lead oxide, barium oxide, phosphorus pentoxide, and strontium oxide, the mean and median were calculated using EXCEL software processing as shown in Table 8.

Table.8. Mean and median chemical composition of high potassium and lead-barium glass

Data Type	Glass Type	Silicon Dioxide	Potassium oxide	Calcium oxide	Iron oxide	Lead oxide	Barium oxide	Phosphorus pentoxide	Strontium oxide
Average/%	High Potassium Glass	76.64	6.4017	3.845	1.376	0.274	0.3989	1.028	0.0279
	Lead barium glass	38.88	0.173	2.05	0.6559	33.349	10.490	3.292	0.348
Median/%	High Potassium Glass	73.01	7.525	3.36	0.46	0	0	0.68	0
	Lead barium glass	35.78	0	1.48	0.23	31.9	8.94	1.41	0.31

It can be visualized that the chemical composition content of silica of high potassium glass is much higher than that of lead-barium glass, and also much higher than that of other chemical compositions of lead-barium glass; while the chemical composition content of lead oxide and barium oxide of lead-barium glass accounts for much higher than that of high potassium glass.

3.3.2 High potassium, lead barium glass sub-classification division method and results

In order to subclassify each category, we selected eight chemical components with high relevance to the classification of high potassium and lead-barium types-silica, potassium oxide, calcium oxide, iron oxide, lead oxide, barium oxide, phosphorus pentoxide, and strontium oxide-and used the K-Means clustering [8] method to subclassify high potassium and lead-barium glass artifacts, respectively Classification.

1.Determining the optimal number of clusters using the elbow method [9]

The core metric of the elbow method is the SSE (sum of squared errors), which is

$$SSE = \sum_{i=1}^{k} \sum_{p \in C_i} |p - m_i|^2 \tag{11}$$

Where, C_i is the first i cluster, p is the sample points in C_i, m_i is the center of mass of C_i (the mean of all samples in C_i), and SSE is the clustering error of all samples, which represents the clustering effect.

Combining equation (11) with VSCode software and EXCEL to jointly describe the error sum of squares for high potassium and lead-barium glass when k is taken to different values, the elbow classification diagram is obtained as in Figure 7.

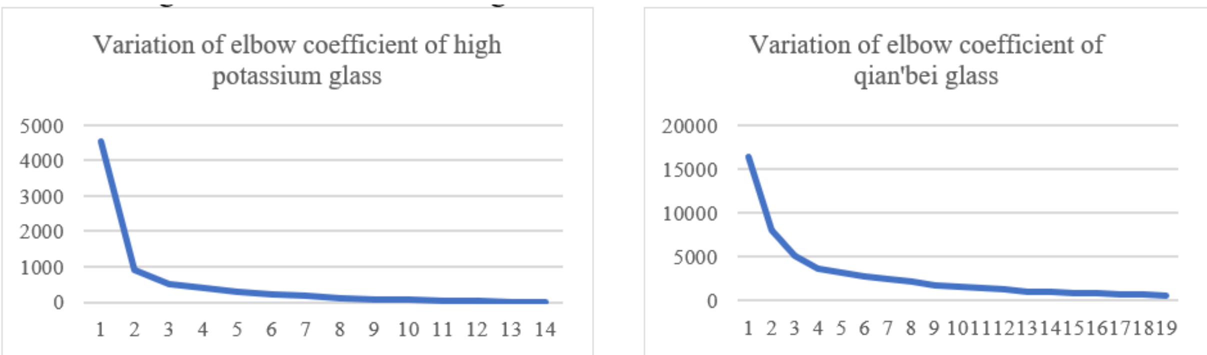

Figure 7. The elbow method classification chart of high potassium (k=1~14) and lead-barium (k=1~19) glass

According to Figure 9, it can be seen that the curve initially declines at an extremely fast rate, and then the decline rate gradually tends to level off after the elbow, and according to the horizontal coordinates corresponding to the elbow turning point, we can get the sub-category of high potassium glass into two better, and the sub-category of lead-barium glass into four better.

2.Select the initialized k samples as the initial clustering centers $a_i = a_1, a_2, \cdots, a_k (1 \le i \le k)$

3.Calculate the clustering center

4.Keep iterating until the clustering center no longer changes or stops when the maximum number of iterations is reached, and the iteration results are shown in Tables 9,10

Table.9. Subcategory clustering centers for high potassium glass

Category	Silicon Dioxide	Potassium oxide	Calcium oxide	Iron oxide	Lead oxide	Barium oxide	Phosphorus pentoxide	Strontium oxide
Type I	63.62	10.82	6.36	2.31	0.41	0.58	1.52	0.05
Type II	89.66	1.99	1.33	0.44	0.14	0.22	0.53	0.01

Table.10. Lead barium glass subcategory clustering center

Category	Silicon Dioxide	Potassium oxide	Calcium oxide	Iron oxide	Lead oxide	Barium oxide	Phosphorus pentoxide	Strontium oxide
Type A	1935	0.05	2.76	0.23	59.39	4.71	5.01	0.66
Type B	58.11	0.20	1.04	0.66	19.35	8.06	0.87	0.22
Type C	12.07	0.10	2.28	0.00	30.15	32.39	5.08	0.49
Type D	30.43	0.20	2.84	0.93	40.54	10.26	4.93	0.35

5.The final subclassification results were obtained, as shown in Tables 11,12

Table.11. High potassium glass sub-classification classification results

Category	Heritage Sampling Sites
Type I	01, 03 parts 2, 04, 05, 06 parts 1, 06 parts 2, 13, 14, 16
Type II	03 Part 1, 07, 09, 10, 12, 18, 21, 22, 27

Table.12. Lead barium glass sub-classification division results

Category	Heritage Sampling Sites
Type A	39, 40, 43 part 1, 51 part 2, 54, 54 severely weathered spots, the
Type B	20, 23 unweathered, 25 unweathered point, 28 unweathered point, 29 unweathered point, 31, 32, 33, 35, 37, 42 unweathered point 1, 42 unweathered point 2, 44 unweathered point, 45, 46, 47, 48, 49 unweathered point, 53 unweathered point, 55
Type C	08、08 severely weathered spots, 26、26 severely weathered spots
Type D	02, 11, 19, 24, 30 parts 1, 30 parts 2, 34, 36, 38, 41, 43 parts 2, 49, 50, 50 unweathered points, 51 parts 1, 52, 56, 57, 58

3.4. Prediction of unknown types of glass artifacts

The random forest regression algorithm is used, the initial number of parameter decision trees is set to 100, the minimum number of samples in the leaf nodes is 1, the minimum weight of samples in the leaf nodes is 0, the maximum depth of the tree is 10, the maximum number of leaf nodes is 50, 70% of the data is used as the training set, the remaining 30% of the data is used as the test set, and the MATLAB software is used to train the model, and the training evaluation results are obtained as shown in Table 13.

Table.13. Assessment results for the eight glass artifacts

	Accuracy	Recall Rate	Accuracy rate	F1
Training set	1	1	1	1
Test set	1	1	1	1

Note: Accuracy rate: the proportion of correct predicted samples to the total samples, the larger the accuracy rate, the better; Recall rate: the proportion of predicted positive samples to the actual positive samples, the larger the recall rate, the better; Precision rate: the proportion of predicted positive samples to the actual positive samples, the larger the precision rate, the better; F1: the sum of precision rate and recall rate, precision rate and recall rate affect each other, although both Although a high precision rate and recall rate affect each other, although both are desirable, in practice it is often the case that a high precision rate results in a low recall rate, or a low recall rate results in a high precision rate. If a balance between the two is needed, then the F1 metric can be used.

As can be seen from the table, the test set has 1 for each evaluation index, and the model is trained very well.

Performing data prediction, the glass types of these eight groups of artifacts can be obtained as shown in Table 14.

Table.14. Assessment results for the eight glass artifacts

Artifact Number	A1	A2	A3	A4	A5	A6	A7	A8
Predicted results	High Potassium Glass	Lead barium glass	Lead barium glass	Lead barium glass	Lead barium glass	High Potassium Glass	High Potassium Glass	Lead barium glass

4. Conclusions

This paper introduces the classification standard of ancient glass cultural relics, and studies and analyzes its chemical composition. Firstly, the correlation and difference between surface weathering of glass cultural relics and glass type, decoration and color were studied by Kappa consistency test and chi-square test. Then, using the median and average value of each chemical composition, the statistical law of glass before and after weathering and each chemical composition was studied, and a relationship model before and after weathering was established to predict the proportion of each chemical composition before glass weathering. Secondly, the classification rule of high-potassium lead-barium glass was studied by Pearson correlation analysis, and the optimal number of clusters for high-potassium glass and lead-barium glass was determined by elbow method, and the sub-classification was carried out accordingly. The rationality and sensitivity of the model were studied by using the silhouette coefficient method and perturbing the proportion of chemical components in the model. Then, the prediction model of glass type is established by using random forest regression algorithm, and the sensitivity of the model is studied by analyzing the stability and robustness of the model. Finally, Pearson's correlation analysis was used to study the correlation between the chemical components of the given categories from four aspects: decoration, type, color and surface weathering, and obtained the difference of the correlation between the categories. It provides help for the subsequent identification and analysis of over-weathered glass cultural relics, and plays a positive role.

References

[1] Chen Shuyu and Hou Zhili. An analysis of the ancient Silk Road and ancient Chinese glass [EB/OL]. 2019-05-15[]. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2019&filename=MZBL201905082&uniplatform=NZKPT&v=F-tkx6_Ipxo1UlR97MTaRZrsv6IWguhm-OQGzkhaLAGHiZju7t4jkTgTcdBDGhtb.

[2] Li, Liu-Liu Li. Experimental study on the simulation of accelerated weathering of carbonate artifacts by atmospheric acidic gases [EB/OL]. 2020[]. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CDFD&dbname=CDFDLAST2020&filename=1019266049.nh&uniplatform=NZKPT&v=AjWL4Z-griROYQIJydQtymrOLcbAz6w4g_QfNNLc83fuME7GPNQEMH2zjSdK-xLI.

[3] Du Rong. Study on the disease of heritage buildings in the southeast coastal area [EB/OL]. 2020[]. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CMFD&dbname=CMFD202001&filename=1019689197.nh&uniplatform=NZKPT&v= VyvbRc4Jl7TMPOMcWEacA6jtmZAqL9DxXfeusuwP33kr6zJN_oLZ6EwCavlerwn6.

[4] CSDN. kappa consistency test tutorial_Kappa consistency test: Do the results of two diagnostic methods agree? [EB/OL]. 2021-01-09 [2022-09-26]. https://blog.csdn.net/weixin_36403128/article/details/112501404.

[5] Y.S. Shi, X.K. Wang, Y.T. Zhou, G.T. Jiang, T.Y. Xu. A multi-radar anti-spoofing interference method based on chi-square test and SVM[J]. Air and Space Defense,2022,5(01):108-114.

[6] Cai M., Sun J., Li P.D., and Bao Q. Comparative analysis of three machine learning algorithms in regression applications [EB/OL]. 2022-07-27[]. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2022&filename=DLXZ202208032&uniplatform=NZKPT&v=IFUK4IZLzpilCHvr7vFvfEIbJJhlG7xgrfRTrWxBQUPnG5f19of60dwvMB11s_rG.

[7] Understanding Pearson Correlation Coefficient (PCC) [EB/OL]. 2018-03-13[]. https://blog.csdn.net/ichuzhen/article/details/79535226.

[8] Huang Qunhui,Li Jin,Yang Chen. An insider threat detection method based on elbow method with clustering algorithm[J]. Journal of Hubei College of Nationalities (Natural Science Edition),2019,37(03): 331-335.DOI: 10.13501/j.cnki.42-1569/n.2019.09.021.

[9] Principles of K-Means clustering algorithm [EB/OL]. 2021-11-23[]. https://blog.csdn.net/weixin_38239039/article/details/121400342.

[10] Xie CH,Xu W. Prediction method of boiler contamination factor based on random forest regression algorithm[J]. Shanghai Electric Technology,2022,15(01):29-32.