A new method for the composition analysis and classification of ancient glass products before and after weathering

Jiayi Zhao [1], Ziyang Tang [1, *], Jianyue Deng [2]

[1] SDU-ANU Joint Science College, Shandong University, Shandong, WeiHai, 264209

[2] School of Astronautics, Harbin Institute of Technology, Harbin, China, 150001

* Corresponding author: 202000700064@mail.sdu.edu.cn

Abstract. This paper focuses on the composition analysis of ancient glass artifacts. With the help of chemical composition data from different locations of glass artifacts, the chemical compositions of weathered and unweathered points are compared using the chi-square test as well as the visualization of R. We find that the chemical compositions of SiO_2 and K_2O have significant changes before and after weathering. Then the composition of weathered points before unweathering was predicted by using the feature point ratio method and neural network prediction. Further, with the help of tools such as Python, modeling methods such as support vector machine and random forest classification were used to dichotomize high potassium and lead-barium glass and compared with the actual results. Then, by using principal component analysis, clustering heat map, and K-means clustering, the composition-based subclassification of glass was established, and the rationality of the subclassification and classification model was verified with the literature, and finally, the sensitivity and rationality of the classification method was analyzed with the defined sensitivity function. The work of this paper will hopefully contribute to the archaeological identification of the composition of glass products.

Keywords: classification problem, principal component analysis, cluster analysis, neural network prediction.

1. Introduction

The analysis and research on the composition of ancient glass in China began in the middle of the last century, because the materials of ancient glass in China originated from the local area and learned from western technology, therefore, although the appearance of glass products and foreign glass products are similar, there are differences in the types and contents of the chemical composition of the glass[1] [2]. There are mainly categories of potassium silicate glass (using substances with high potassium content such as oxalic ash as flux), soda calcium silicate, alkali-containing calcium silicate glass, and lead barium glass (adding lead ore as a flux in firing). The data relating to ancient glass products in China obtained in this paper have been classified into two types of high potassium glass and lead-barium glass by the testing means of archaeologists, and the chemical composition types and contents of the samples have been detected, and the data between 85% and 105% are valid data [3]. However, the problems of weathering on the surface of the glass of cultural relics and the weathering of sampling points in relation to the type, decoration, and color of the glass, the prediction of the chemical composition of the glass samples before weathering, and the classification rules of the two types of glass have not been effectively solved. Based on this, this paper uses correlation analysis, principal component analysis, and cluster analysis while conducting mathematical modeling to solve the problem of analyzing and identifying the composition of ancient glass products.

2. Glass chemical composition prediction model establishment and solution

2.1. Model building and solving

2.1.1. Correlation analysis of weathering with type, ornamentation, and color

In the article, we use SPSS software to perform a chi-square test and correlation analysis [4]. The following is a detailed presentation of the statistical and chi-square test results through SPSS.

Table 1. Surface weathering * Ornamentation Cross-tabulation

			Ornamentation			Total
			A	B	C	
Surface weathering	Weathering	Counting	11a	6a	17a	34
		Expectation count	12.9	3.5	17.6	34.0
		Percentage of tattoos	50.0%	100.0%	56.7%	58.6%
		Adjusted residuals	-1.0	2.2	-.3	
	No weathering	Counting	11a	0a	13a	24
		Expectation count	9.1	2.5	12.4	24.0
		Percentage of tattoos	50.0%	0.0%	43.3%	41.4%
		Adjusted residuals	1.0	-2.2	.3	
Total		Counting	22	6	30	58
		Expectation count	22.0	6.0	30.0	58.0
		Adjusted residuals	100.0%	100.0%	100.0%	100.0%
Each subscript letter indicates a subset of the ornament categories, and at the .05 level, the column proportions for these categories are not significantly different from each other.						

Through Table 1, we can further obtain the results of the chi-square test, that is, the value and p-value, the value is 4.957, the degree of freedom of this problem (3-1)*(2-1)=2, and we can check the corresponding p-value is 0.084. It is greater than our significance index, so we cannot reject the original hypothesis, that is, there is no relationship between surface weathering and ornamentation. Similar to the previous section, we can get the statistical counts, and expected counts of the two types of glass weathering, and finally conclude by calculation that there is also a significant relationship between surface weathering and glass type. We also get the conclusion that the lead-barium type of glass is more likely to be weathered, while the high potassium type of glass is more difficult to be weathered. Similarly, after obtaining the cross-tabulation table and chi-square test table for surface weathering and color, it was concluded that there was no significant relationship between color and weathering.

2.1.2. Data statistics of chemical composition with and without weathering

Combined with the type of glass, the statistical pattern of the content of chemical composition with and without weathering on the surface of the artifact samples is analyzed. First we have to combine the type of glass, that is, the data is divided into two groups of high potassium and lead barium are discussed separately, their data comparison in unweathered and weathered chemical composition, and finally, get the data comparison characteristics before and after weathering as shown in the following table.

a. Statistical law of high potassium glass

Table 2. Means and variances of unweathered and weathered compositional data for high potash glass

Ingredients	SiO_2	Na_2O	K_2O	CaO	MgO	Al_2O_3	Fe_2O_3	CuO	PbO	BaO	P_2O_5	SrO	SnO_2	SO_2
Pre-mean	67.98	0.70	9.33	5.33	1.08	6.62	1.93	2.45	0.41	0.60	1.40	0.04	0.20	0.10
Forward Difference	8.76	1.29	3.92	3.09	0.68	2.49	1.67	1.66	0.59	0.98	1.43	0.05	0.68	0.19
Post-mean	93.96	0.00	0.54	0.87	0.20	1.93	0.27	1.56	0.00	0.00	0.28	0.00	0.00	0.00
Posterior variance	1.73	0.00	0.45	0.49	0.31	0.96	0.07	0.93	0.00	0.00	0.21	0.00	0.00	0.00
Changes	25.98	-0.70	-8.79	-4.46	-0.88	-4.69	-1.67	-0.89	-0.41	-0.60	-1.12	-0.04	-0.20	-0.10

From Table 2, we can see that for the high potassium glass, the content of SiO_2, K_2O, Al_2O_3, CaO, etc. changes significantly, and the last column shows that only the content of SiO_2 increases significantly after weathering, and the other components basically become a decreasing trend.

We consider that SiO_2 is almost the main component of the glass before and after weathering, and the change of its content will greatly affect the content of other elements, so the following definition of the percentage content of the remaining components excluding SiO_2 can be obtained by calculation.

Comparison of the percentage content of the remaining components of high potassium excluding SiO_2 before and after weathering. Finally, the following conclusions were obtained: the content of Na_2O and K_2O decreased before and after weathering (with a serious decrease in K_2O), the content of CaO, Al_2O_3, CuO, and P_2O_5 increased, and the rest of the chemical composition remained almost unchanged. It can be seen that the content of K_2O is the criterion to judge whether the high potassium glass is weathered or not.

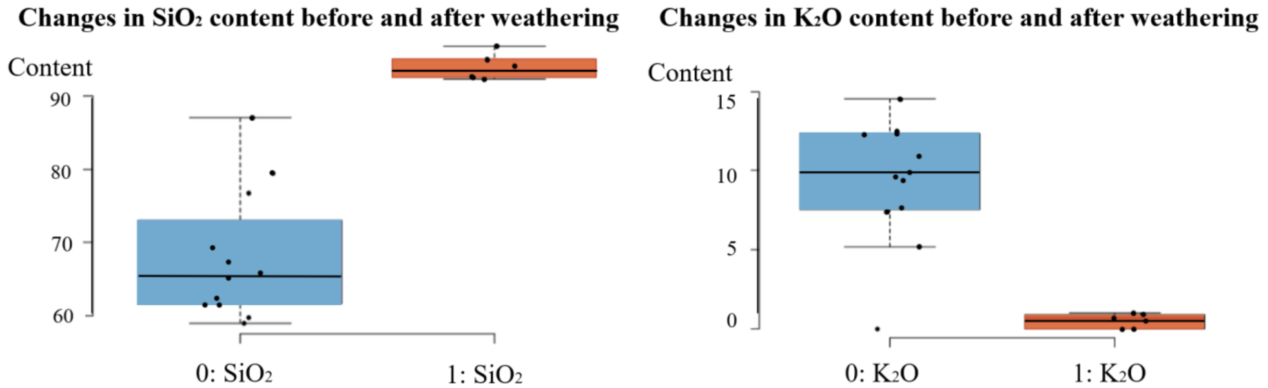

Figure 1. Changes in SiO_2 content before and after weathering of high potassium glass (left panel) (right panel) K_2O content

In addition, it can be observed from Figure 1 the box and three-point plots that the SiO_2 content of the high potassium glass increases substantially before and after the weathering whereas the K_2O content decreases.

In spite of what was plotted before, we performed a t-test with the null hypothesis that there is no significant difference before weathering and then calculated p-values to compare with the significance index to further illustrate the significant difference between the two sets of data. We found that the p-values of both t-tests are less than 0.05, which proves that we can reject the original hypothesis, that is, the content of SiO_2 and K_2O is clearly observable before and after weathering.

b. Statistical law of lead-barium glass

For lead-barium glass, the same method was applied to obtain relevant conclusions, and it was finally found that the content of SiO_2, Na_2O, and Al_2O_3 decreased before and after the occurrence of weathering in lead-barium glass (where the content of SiO_2 was severely reduced, and the p-value was less than 0.05 to be able to observe obvious differences), and the content of PbO, BaO, CaO, P_2O_5, and SO_2 increased (especially the change of PbO was obvious, and the p-value was less than 0.05 can observe a significant difference, many of the data are not detected SO_2, so take the average value will make the amount of change in SO_2 content distortion, analyze the data containing SO_2 in the table one by one can be seen that the more serious the weathering of lead barium glass, the higher the SO_2 content, serious weathering parts SO_2 content even up to 15%), the rest of the chemical composition almost no change. It can be seen that the content of SiO_2 and SO_2 is the criterion to determine whether the weathering of high potassium glass occurs.

To sum up, we can get all kinds of glass before and after the weathering of the chemical composition according to the laws of statistical data are more obvious changes in the law.

(1) high potassium glass K_2O content is significantly reduced, SiO_2 significantly increased, PbO and BaO content does not change much.

(2) lead-barium glass with significantly reduced SiO_2 content, greater changes in PbO, and no significant difference in SO_2 content.

Both CaO, CuO, and P_2O_5 content have increased, which can be regarded as the common weathering of the glass after the composition change pattern.

2.1.3. Predicting the content of weathering points before weathering

a. Neural network method prediction

The feature point method is a simple prediction method, but it is easy to apply and explain. Here we further give a more complex neural network method and use it for comparison. This is obtained by analyzing the given data. A total of 57 sets of valid data are given, including 19 sets of data for high potassium glass and 48 sets for lead-barium type glass, which are analyzed in the neural network analysis [5].

(1) Neural network structure diagram

In the neural network, we divided the data into 11 groups of training data, 3 groups of validation data, and 3 groups of test data. Using Matlab's neural network toolbox, we used the quantized conjugate gradient method for training and built a neural network with 12 inputs, 12 outputs, and 20 hidden neurons. Since the amount of data is relatively small, several training sessions are required and the optimal network model is obtained after debugging. The network structure(in Figure 2), output error, and goodness-of-fit are shown in Figure 4.

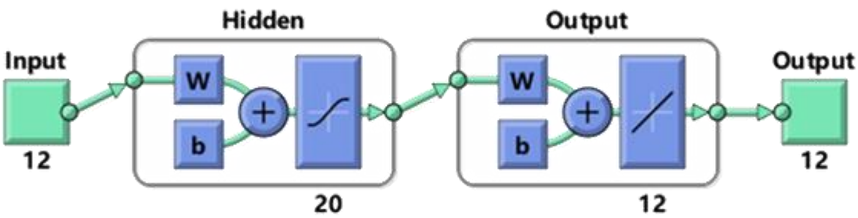

Figure 2. Structure of the neural network

(2) Analysis of the error

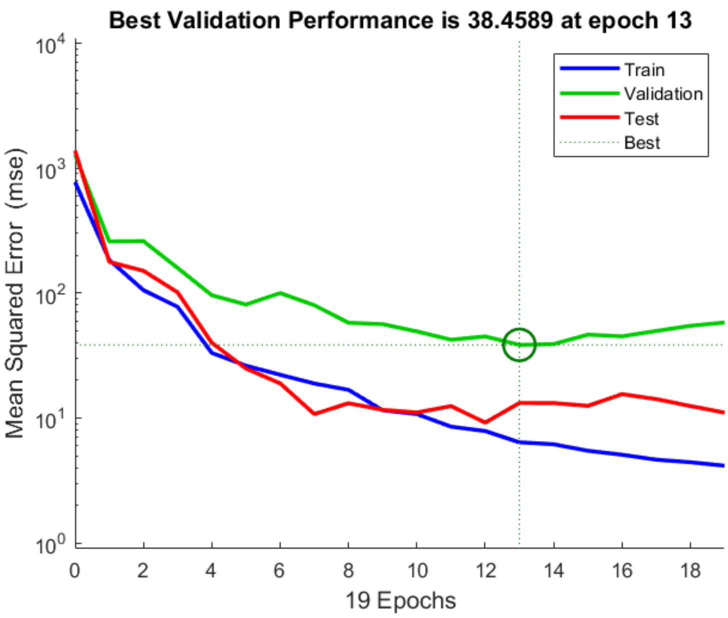

Figure 3. Neural network error curve

From Figure 3, we can see that the neural network was trained for 19 cycles, and due to the phenomenon of overfitting, it reached the optimum in the thirteenth cycle, at which time the MSE for the goodness-of-fit was about 38.46, which is already high considering that the output data of the neural network is 12 groups. Therefore, we choose the data from training 13 cycles as our prediction parameters.

(3) Analysis of the goodness-of-fit

As observed in the figure, the goodness of fit of the neural network is close to or greater than 95% for all three types of data given, which has a good prediction effect.

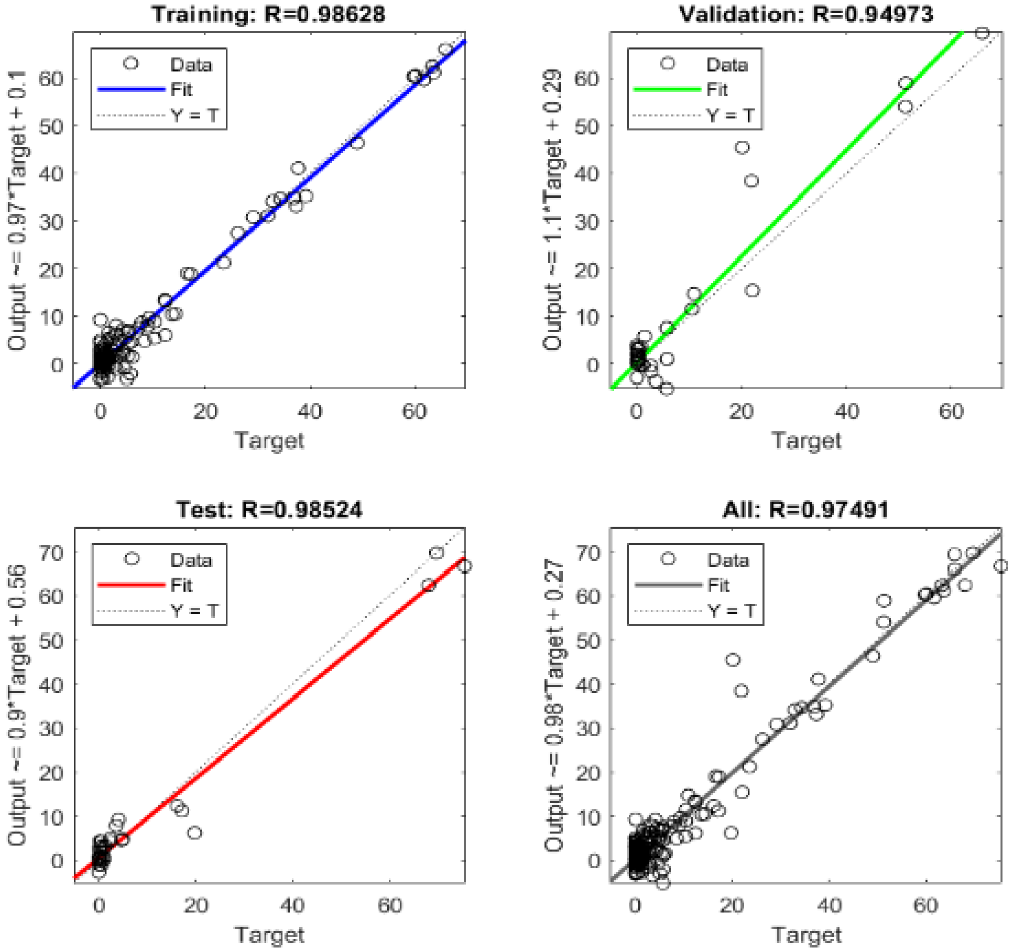

Figure 4. Neural network goodness-of-fit graph

3. The classification model of high potassium and lead-barium glass is established and solved

3.1. Classification law of high potassium and lead-barium glass

First, we build the classification models for both high potassium and lead-barium categories. Since the corresponding category labels of the data are known, we prefer to use machine learning with supervised learning for the classification. Here we mainly used the support vector machine, decision tree, and random forest algorithms for classification.

Support vector machine is a binary classification model, which is well suited for this problem to classify two types of glass, high potassium and lead barium. It is mainly divided into linear and nonlinear classifiers, and in this paper, the linear approach is mainly used [6].

In this study, each sample x represents a heritage sample, and its data characteristics are the content of various chemical components, while y corresponds to two categories of classification, i.e., high potassium and lead-barium, for example, so that high potassium is -1 and lead-barium is 1. It is hoped that this method can find a hyperplane that can well separate the two types of data, i.e., satisfy the above condition. However, because the sample has multiple features, it is a high-dimensional problem, and it is not convenient for us to show the effect of SVM classification visually.

In order to show it intuitively, we use the method of principal component analysis (PCA), which is implemented in python, to reduce the high-dimensional data to two dimensions, which can be shown intuitively in the plane, where the specific principle of the principal component analysis method is developed in detail in the next section, and here we only use the results after the dimensionality reduction into two dimensions and draw the scatter plot of the data as shown in Figure 5 on the left.

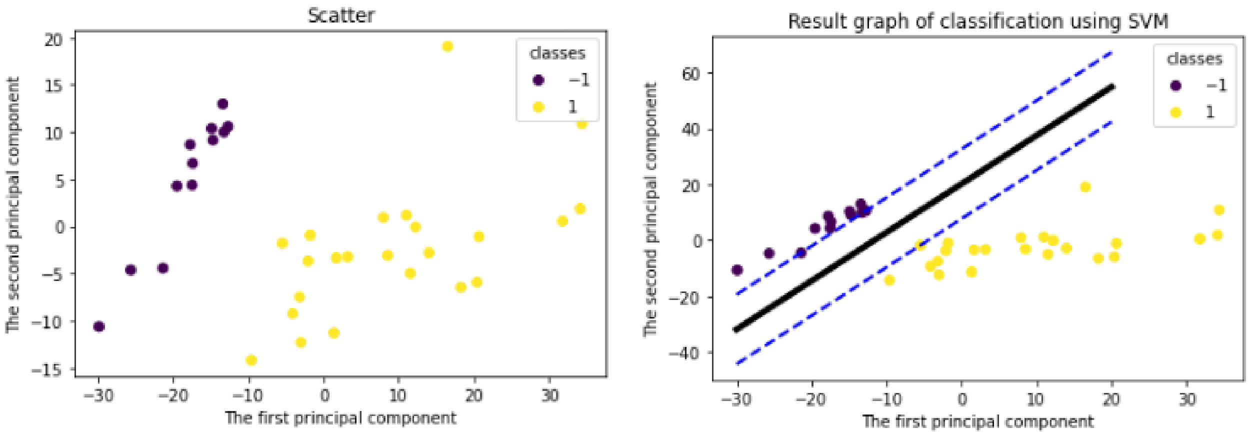

Figure 5. Schematic diagram of SVM principle

From the left Figure 5, we can see that the two types of artifacts have good boundaries to distinguish, so the method of SVM is appropriate, while the right Figure 5 is a schematic diagram of the results after SVM for binary classification, where the black line is the hyperplane needed for the solution, i.e., the first equation of the room in the formula, and the blue line represents the decision boundary, i.e., the second equation that satisfies the formula.

In the support vector machine classification method, we take 80% of the data points as the training set and 20% of the data points as the test set. The test set is evaluated against the trained SVM model. Finally, we obtained that the binary classification model obtained by the SVM model worked better and its classification was successfully predicted on both test sets, so that both high potassium and lead-barium can be classified by the SVM model.

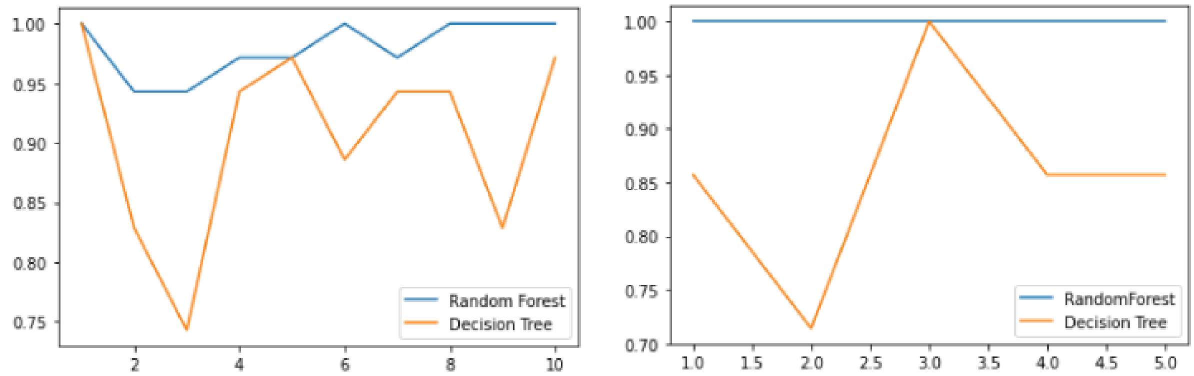

Figure 6. Accuracy of multi-fold cross-validation

By the cross-validation method, we get the results shown in Figure 6. The left Figure 6 shows a five-fold cross-fold validation with no fold accuracy, blue corresponds to the random forest method and orange corresponds to the decision tree method. From Figure 6, it is obvious that the random forest method is much more accurate than the decision tree method and has good accuracy (minimum 95% accuracy).

3.2. Classification of high potassium and lead-barium glass subclasses

For the classification of high potassium and lead-barium glass subclasses, we do not have a direct data label or even a unified standard, so about subclass classification is an unsupervised model, and we need to start from the data itself to explore the intrinsic characteristics of the data. Usually, we use clustering for unsupervised model classification, and here we take the commonly used K-means clustering, hierarchical clustering dendrogram, and clustering heat map [6].

The first thing we introduce is the hierarchical clustering dendrogram as well as the clustering heat map. The clustering heat map is most commonly used in biology to plot the expression values of differential genes into a clustering heat map, where the higher the expression value the lighter the color in the heat map. Heatmaps are used to see how differentially expressed genes are across samples or to compare differences between different clustered groups.

Here we introduce the clustering heat map because we want to get a subclass division of the glass, and here the color of the heat map represents how much of each chemical component is present. The goal of clustering is achieved by comparing the differences between the data and grouping similar data into one category. In Python, the scipy. cluster function package can help us achieve this goal very well. And in the function, we can also choose different clustering methods, and in this paper, we choose Ward's minimum variance clustering.

Before drawing the clustering heat map, it is worth mentioning that the data need to be normalized, because the content of different chemical components varies greatly, and what we need is the analysis of the differences between data and data, we need to normalize each column of data, that is, the data of the same chemical content of different samples, and then clustering, which better reflects the characteristics of the data. Similarly, the sklearn. preprocessing function in Python provides the normalization process, and it should be noted that we are normalizing each column of the data here.

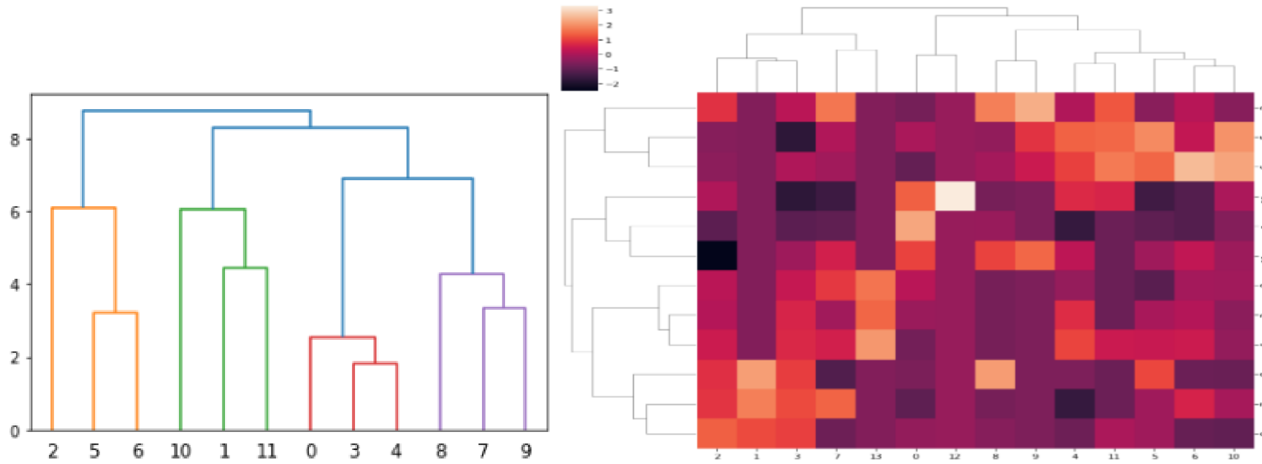

Figure 7. Hierarchical clustering diagram and heat map of high potassium glass

As shown in Figure 7, we have obtained the clustering heat map of high potassium glass. In the left clustering map, we can clearly see the sample data with similar characteristics, while the heat map on the right gives a more intuitive comparison. For example, three data samples 7, 8, and 9 are grouped together because they are similar in SiO_2, Na_2O, and K_2O, and the colors that can be visualized in the heat map are similar. Based on this, we can divide them into subcategories.

Similarly, we can obtain the aggregation heat diagram of lead barium glass. Then, before implementing K-means clustering, the first thing we need to do is to reduce dimensions by using principal component analysis.

Table 3. Cumulative contribution of principal components for data on two types of glass with high potassium lead and barium

Principal Components	1	2	3	4	5	6	7	8
RatioCumulative (K)	0.766	0.889	0.941	0.973	0.985	0.991	0.996	0.999
RatioCumulative (PbBa)	0.700	0.891	0.949	0.973	0.987	0.992	0.995	0.997

Here we show the gravel plot of the principal component analysis on the two types of glass with high potassium-lead-barium without weathering in Figure 8, from which it is clear that the cumulative contribution of 2 principal components has been greater than 85% and the cumulative contribution of 4 principal components has been greater than 95%. It is also obvious that the effect of dimensionality reduction is achieved through principal components. In addition, it shows that we are able to illustrate the nature of the data better by 2-4 features, which also provides a general range for the subsequent subcategory classification, that is, under each major category of high potassium and lead-barium, we can further refine the subcategory into 2-4 subcategories, which can basically include most of the data characteristics.

To further demonstrate the effect of clustering, we first take only the first two principal features for graphical display.

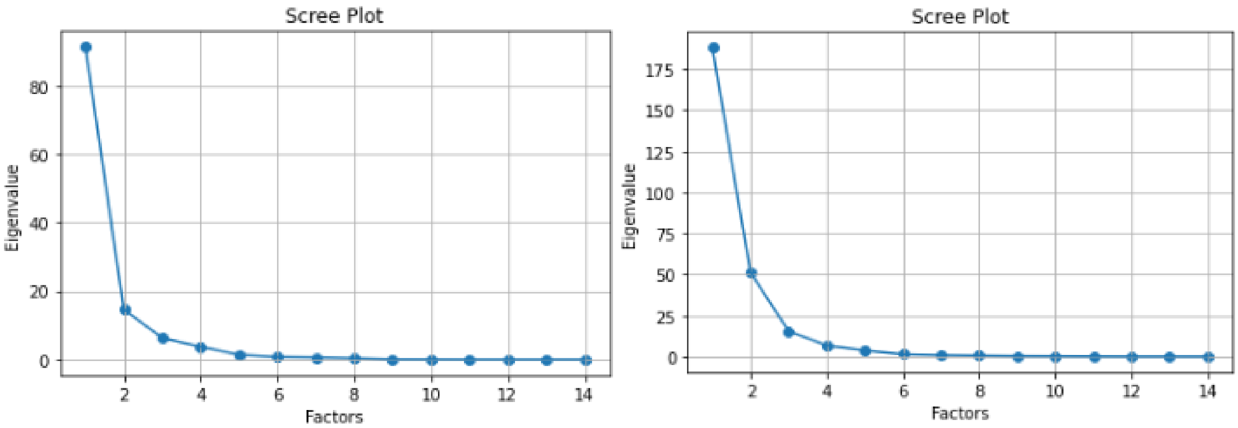

Figure 8. Gravel plot for principal component analysis

Through our previous analysis, the subclass division can be performed by using the method of principal component analysis, but considering that if all 14 chemical components are involved in the principal component algorithm, the obtained principal component results are too complicated to meet the needs of the actual situation, therefore, the importance of each chemical component is determined according to its content, and six chemical components are selected (SiO_2, CuO, PbO, BaO, Al_2O_3, P_2O_5) were selected to participate in the principal component analysis, which simplified the analysis and made it more convenient for us to get the subclasses. Their contents are, respectively, and the following results are obtained in Figure 9.

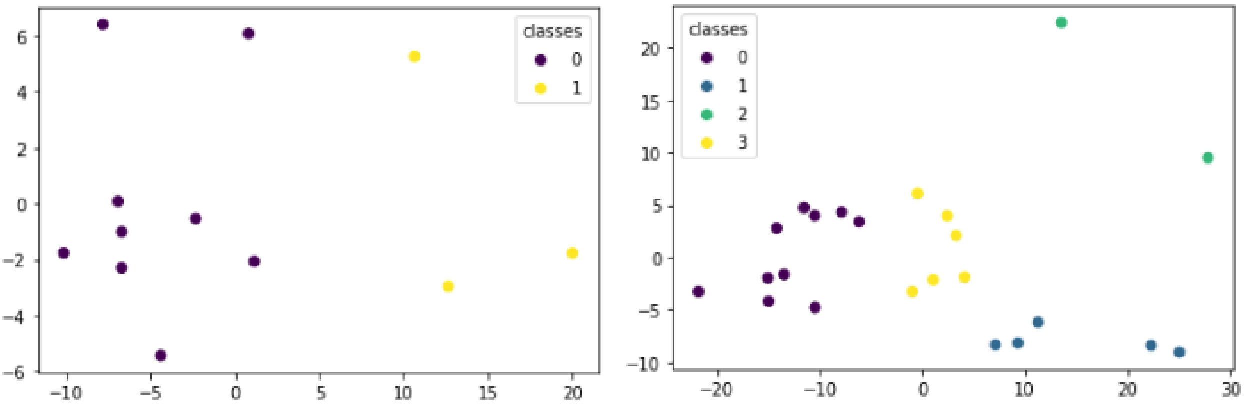

Figure 9. (a) High potassium glass divided into two clusters (b) Lead barium glass divided into four clusters

Table 4. MATLAB principal component analysis results

Accumulative Contribution Rate	0.4384	0. 6673	0.8468	0.9734	0.9957	1.0000
Eigenvector matrix (Percentage of components)	-0.5327	-0.0982	0.3396	0.3649	-0.1681	-0.6556
	0.4904	-0.3694	0.3181	0.0486	-0.7200	0.0334
	0.2735	0.7142	-0.2201	-0.2039	-0.3132	-0.4764
	0.5550	-0.2727	0.1756	-0.0808	0.5665	-0.5093
	-0.2647	-0.4591	-0.3807	-0.6927	-0.1715	-0.2549
	0.1513	-0.2423	-0.7478	0.5801	-0.0700	-0.1331

Through the results of principal component analysis in Table 4, we can analyze that: principal component one consists of silica, barium oxide, and copper oxide with the largest proportion of composition, and it is known from the main composition of glass and the firing process of lead-barium glass that principal component two mainly describes the most important compound composition of lead-barium. Silica carries a large negative load, while barium oxide and copper oxide carry a relatively small positive load, so principal component one can be used to reflect the proportion of flux added in firing; principal component two is mainly composed of lead oxide and aluminum trioxide, and principal component three is mainly composed of phosphorus trioxide and aluminum trioxide, which together reflect the weathering characteristics of lead-barium glass[7].

After the dimensionality reduction of the data, K-means clustering was used to classify the subclasses. Further, in order to better determine how many subclasses we need to classify, we need to use elbow diagrams to help us determine exactly how many classes we need to classify to get a reasonable classification effect. The calculated values of the three principal components of each artifact are imported into SPSS for cluster analysis, and the elbow diagram for the classification of lead barium glass according to the spectral map is shown in Figure 10.

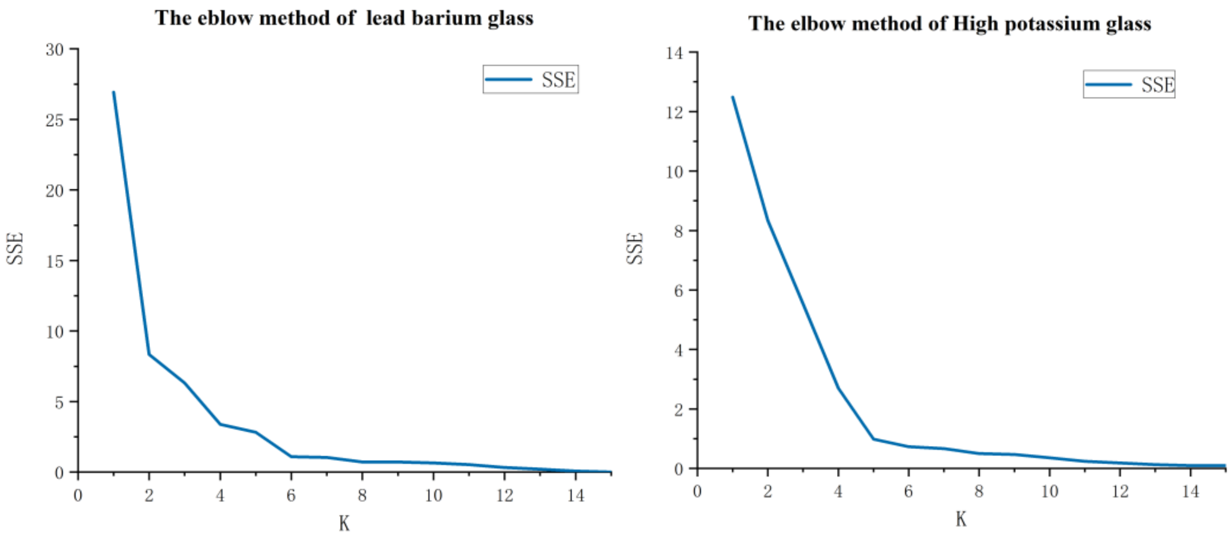

Figure 10. Elbow diagram of cluster analysis for Pb-Ba and HK

After extracting the coefficients of the system clustering, the elbow diagram of the coefficients is drawn in descending order, and according to the analysis for the elbow diagram, a relatively small slope is chosen as a reasonable number of classifications, which is 4 classes.

The classifications are introduced as follows.

(1) SiO_2-CuO-PbO-BaO glass, which contains more CuO in addition to more silica and flux, but there is only one class among all undifferentiated glasses.

(2) SiO_2-PbO-BaO-Al_2O_3 glass, in addition to containing more silica and combustion aids, but also contains more Al_2O_3, and similarly only one class.

(3) high melting point SiO_2-PbO-BaO glass, which contains less PbO and BaO, as PbO and BaO are converted from fluxes, this kind of glass may have a high melting point.

(4) Low melting point SiO_2-PbO glass, in which the PbO and BaO content is higher, and since PbO and BaO are converted from fluxes, this glass may have a lower melting point. The latter two occupy a larger variety.

Similar to lead-barium glass, for high potassium glass, according to the problem a content analysis of compounds in high potassium glass, after removing compounds with low content and relatively stable composition in high potassium glass, five compounds of silica, sodium oxide, potassium oxide, calcium oxide, and aluminum oxide were retained, and they were subjected to principal component analysis, and two principal components were chosen to be retained. The data obtained from the principal component analysis were then imported into spss for systematic cluster analysis, and the spectrograms are shown in the appendix. According to the spectral diagram, combined with the elbow diagram and the complexity of the classification, the high-potassium glasses were divided into two subclasses.

(1) ordinary high potassium glasses

(2) High potassium glasses with the presence of CaO-K_2O-Al_2O_3. In addition to the compounds found in ordinary high potassium glasses, these glasses also contain much higher amounts of CaO, K_2O, and Al_2O_3, which we presume are elements added during refining to enhance the color and properties of the glass.

We finally obtained the overall classification results schematically shown in Figure 11.

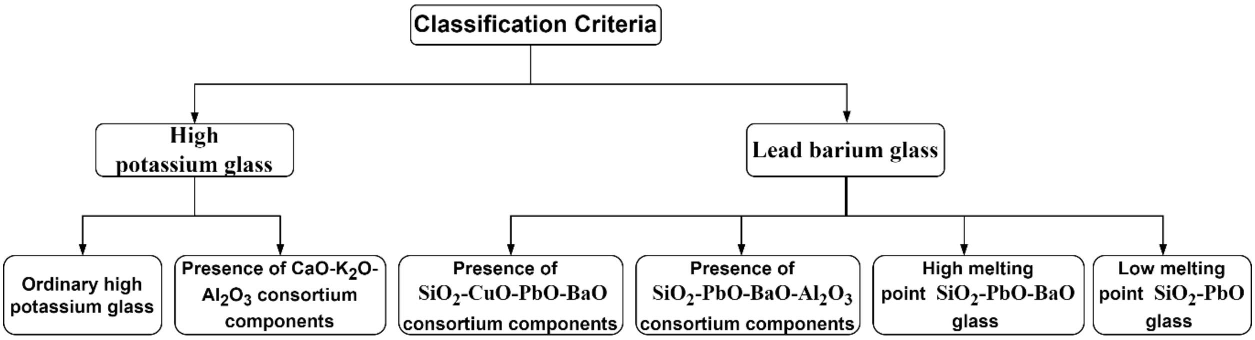

Figure 11. Glass sub-category classification criteria diagram

3.3. Rationalization of classification results and sensitivity analysis

Combining the above classifications and reviewing the data shows that the composition of glass before weathering is closely related to the age of glassmaking. Therefore, a more detailed subclassification of glass is beneficial to locate the dynasties of glass production. In the above analysis, we use archaeological data to analyze the content of compounds in glass to achieve a general position on the age of glassmaking. In general, we have reduced the dimensionality of the data by correlation analysis and principal component analysis, and then divided the subtypes by clustering methods, and the results obtained are consistent with the division of the literature to some extent. We can say that our classification is very reasonable[7-9].

4. Conclusion

In this paper, we used a combination of data analysis methods, chi-square test, correlation test, and R visualization to analyze the data characteristics and differences of glass before and after weathering in detail; we analyze and predict the data by establishing a mathematical model. Here we use a combination of the method of characteristic point ratio and neural network algorithm to achieve the prediction of the content of each chemical component of each glass type weathering point before weathering； In this paper, we also further established a classification model.

In the future, this study has the potential to further deepen the research, firstly, more data are used for analysis, secondly, in this study, the main elements are selected, and the analysis of trace elements still needs to be improved in the future, thirdly, the weathering point data are not considered in the subclass classification in this paper, and the future research on the weathering point data needs to be more in-depth, for example, to support the prediction results, and the weathering point can also be subclassed. Future research on weathering point data needs to be more in-depth, such as supporting the prediction results, and subclassifying weathering points.

References

[1] Gan Fu Xi, Cheng Huan Sheng, Li Qing Hui. The origin of ancient glass in China--a study of the earliest ancient glass in China [J]. Chinese Science (E Series: Technical Science), 2007 (03): 382 - 391.

[2] Zhao Hualing. Research on the historical development of ancient glass in Zibo area [D]. Suzhou University, 2008.

[3] Huang Qisan. An overview of the chemical composition of ancient glasses of the world[DB]. http://www.gxmuseum.cn/a/science/31/2011/1073.html

[4] Li QH, Huang KJ, Li F, Gan FX. Chemical composition of a group of ancient Warring States glass samples excavated in China [J]. DOI:10.16334/j.cnki.cn31-1652/k.2006.02.002.

[5] Yang Lu. Research on Neural Network Forecasting Methods [D]. Beijing Jiaotong University, 1998.

[6] Qiyuan Jiang, Jinxing Xie, Jun Ye, Mathematical Models (3rd ed.) Beijing, Higher Education Press, 2006.9.12

[7] Li QH, Huang KJ, Li F, Gan FX. Chemical composition of a group of ancient Warring States glass samples excavated in China[J]. DOI:10.16334/j.cnki.cn31-1652/k.2006.02.002.

[8] Gan Fu Xi, Zhao Hong Xia, Li Qing Hui, Li Ling, Cheng Huan Sheng. Scientific and technological analysis and research of Warring States glassware excavated in Hubei Province [J]. Jianghan Archaeology, 2010 (02): 108 - 116+151+0.

[9] Wen Rui, Zhao Zhiqiang, Ma Jian, Wang Jianxin. Composition analysis of glass beads excavated from the Shirenzigou site group in Balikun, Xinjiang[J]. Spectroscopy and Spectral Analysis, 2016, 36 (09): 2961 - 2965.