Research on Composition Analysis and Identification of Ancient Glass Products

Jiewen Zhong*, Jieli Chen, Zenghao Chen

School of Economics, Guangdong University of Technology, Guangzhou, China, 510000

* Corresponding Author Email: 452887806@qq.com

Abstract. Glassware was the witness of the ancient Silk Road trade. The paper focuses on the component analysis and identification methods of ancient glass products. The decision tree model was established and analyzed to obtain the classification rules of high potassium glass and lead barium glass. Based on the cluster analysis method, appropriate chemical elements were selected for each type to conduct cluster analysis. The result showed that high potassium glass and unweathered lead barium glass can be classified by their content of silicon dioxide. However, weathered lead barium glass needed to be classified by their content of lead oxide. The random forest regression model was used to classify glass cultural relics by their chemical elements. The chemical element data of glass cultural relics with known types were used as samples, and the chemical element data of glass cultural relics with unknown types were inputted into the trained model. Finally, the types of glass cultural relics were predicted and sensitivity analysis was conducted.

Keywords: Ancient glass products, Decision tree, Cluster analysis, Random Forest.

1. Introduction

1.1. Background

The Silk Road was an important bridge between the Eastern and Western cultures in ancient times. Chinese ancestors absorbed glass technology and produced glass from local materials in China. Therefore, compared with foreign glassware, local glassware had similarities in appearance but different chemical composition. The most effective way to judge the age and source of ancient glass was to test and study the chemical composition and physical characteristics of ancient glass products. At the same time, it also provided a strong scientific basis for solving and judging the source and age of ancient glass.

1.2. Literature Review

With the development of modern archaeological excavation technology, many ancient cultural relics have been excavated along the Silk Road, attracting many scholars' attention [1-3]. Among the excavated cultural relics, the appearance and internal elements of ancient glass have changed greatly due to the influence of the environment [4-6]. As a result, scholars have made in-depth research on the internal chemical composition changes of ancient glass samples. In this respect, most scholars used qualitative correlation factor analysis for classification and related tool analysis [7-10]. However, there are few quantitative and substantive research methods.

Liu Song et al. [11] used PXRF method to quantitatively analyze the surface elements of ancient glass samples, and pointed out the flux concentration changes in different parts of the glass samples. Balvanovic [12] and others described the papers and other documents of the late ancient Serbian glass, concluded that Foy 3.2 glass evolved into Foy 2.1 glass, and explained the relationship between the chemical composition of the two glasses. Zou Ying et al. [13] established a multivariate time series model to predict the content of the chemical composition of ancient glass before weathering, and converted the relevant data into time series analysis, but ignored the time interval between the initial production and the excavation of the glass. Currently, it can be seen that chemical composition prediction of ancient glass, the exploration of quantitative analysis methods is insufficient. This paper focused on the prediction of chemical composition of glass samples before and after weathering through data algorithms and modeling analysis. Decision tree, cluster analysis and random forest regression model would be used to predict the chemical composition of glass before and after weathering.

There was a batch of ancient Chinese glass products, and it was required to explore the classification law of high potassium glass and lead barium glass by detecting the chemical composition of these cultural relics samples. We adopted appropriate classification methods, selected suitable chemical components for each category, and divided them into subcategories. We presented the classification results and analyzed the rationality and sensitivity of the classification results. Finally, we analyzed the chemical composition of glass relics of unknown category in the data, so as to identify the specific type to which it belongs, and the sensitivity analysis of classification results was carried out.

1.3. Main Work

We used the decision tree model to analyze the classification of high-potassium glass and lead-barium glass. Based on the clustering analysis method, we selected the relevant chemical components of each type. We found the two most similar classes among each class and classified them into one class. We calculated the similarity and analyzed the rationality and sensitivity of the classification results by changing the number of classifications.

We established the random forest regression model and trained it with samples of known composition to determine the types of glass artifacts with unknown categories based on their chemical composition. Next, we used the model to predict the types of the unknown samples and analyzed how they correlate with their chemical composition. We adjusted the proportion of the components that had a high correlation (as long as the total percentage is valid) and fed them back into the model to obtain new predictions. Finally, we conducted sensitivity analysis by observing how these changes in components affect the prediction outcomes.

1.4. Arrangement

This paper was organized as follows. Section I introduced the research background, literature review, research problem and analysis of this paper. Section II presented the principles and assumptions of the three models used in this paper: decision tree, cluster analysis and random forest regression and put forward the basic assumptions of the paper. Section III applied the decision tree model to classify high-potassium glass and lead-barium glass based on their chemical composition. Section IV performed a subclass analysis using cluster analysis with different chemical components as criteria. The classification results and their rationality and sensitivity were discussed. Section V used the random forest regression model to predict the type of unknown glass artifacts and conducted a sensitivity analysis of the prediction results. Section VI concluded this paper by summarizing the research questions, methods, results and implications of this paper.

2. Preliminary

2.1. Decision Tree

This paper mainly used decision classification tree. Classification tree can process discrete data which is the data with limited data types[14]. It outputs the category regression tree of the sample, which can predict the continuous value. Assuming t is a node, the calculation formula of GINI coefficient of this node is:

$$GINI(t) = 1 - \sum_{k} [p(C_k|t)]^2 \tag{1}$$

2.2. Cluster analysis model

Cluster analysis is the use of mathematical methods to follow a certain index of similarity (or dissimilarity) according to the attributes of the samples themselves, then quantitatively determine the lipophilicity between samples and cluster the samples by this degree of lipophilicity. How well clustering works relies on methods and clustering algorithms that measure distance [15].

2.3. Random Forest model

Random forest algorithm adopts Bootstrap self-sampling method to obtain different sample sets for model construction, so as to increase the difference degree between models and the ability of extrapolation prediction [16]. The principle of random forest algorithm was shown in Figure 1.

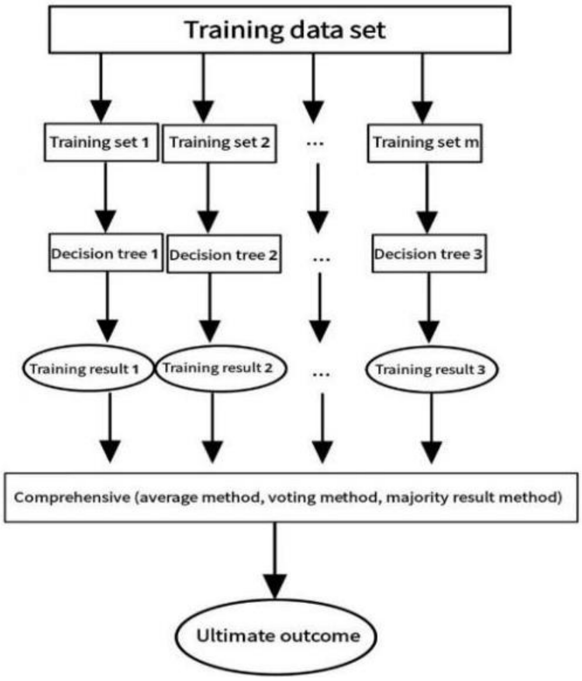

Figure 1. Principle of random forest algorithm

We independently extracted the training set from random vector X and Y distributions and averaged the predicted values from the regression of the decision tree to obtain the predicted values from random forest regression. The formula was shown as follows:

$$M(X) = \frac{1}{m}\sum_{i=1}^{n} h_i(x) \tag{2}$$

2.4. Assumption

We made the following reasonable assumptions and conditional constraints based on the actual situation to construct a more accurate mathematical model.

Hypothesis 1: The sum of the chemical composition ratios of all cultural relics is between 85% and 105%, and any data outside this range will be considered invalid and excluded.

Hypothesis 2: If there is no significant correlation between chemical composition and surface weathering, we assume that the content ratio of chemical composition remains unchanged before and after weathering in the prediction.

Hypothesis 3: We assume that all glass relics are in the same reaction environment before and after weathering.

Hypothesis 4: We ignore the influence of other chemical compositions on the prediction result when conducting a sensitivity analysis to study the prediction of a fixed chemical composition on the glass type,

3. Analysis of glass classification based on decision tree model

We applied the decision tree model to calculate the results shown in Figure 2.

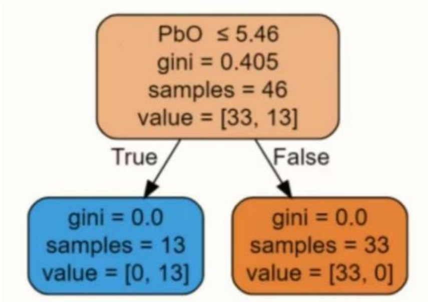

Figure 2. Decision tree structure

As can be seen from Figure 2 of the decision tree, the classification rules of high-potassium glass and lead-barium glass were shown as follows. When the content of lead oxide (PbO) in the chemical composition of glass relics was less than 5.46%, it was classified as high-potassium glass. When the content of the chemical component lead oxide (PbO) in the glass relics was greater than 5.46%, it was classified as lead-barium glass.

Combined with the accessory data, the proportion of lead oxide (PbO) in high-potassium glass was between 0 and 1.626%, which was much lower than the standard of 5.46%. It can be seen that the standard was reasonable and realistic.

4. Subclass analysis based on cluster analysis

Ancient glass cultural relics can be divided into high potassium glass and lead barium glass according to type. We discussed high-potassium glass and lead-barium glass separately, and clustered the two types of glass according to the chemical composition in the cultural relics.

4.1. Classification of high potassium glass

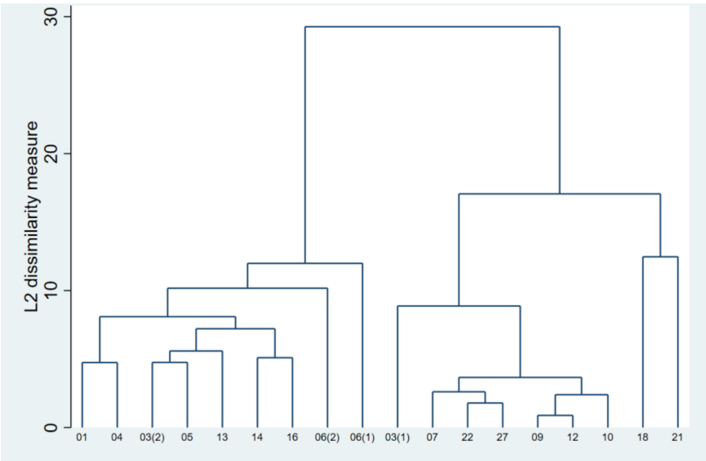

Figure 3. Classification pedigree of high-potassium glass relics

It can be seen that high-potassium glass can be roughly divided into two categories, one was high content of silica (SiO_2) and the other was low content of silica (SiO_2) through Figure 3 of the cluster plot. The classification table was shown as Table 1.

Table 1. Classification of high-potassium glass relics

Category	Cultural relic number	SiO_2 content
Category I	03(1) 07 22 27 09 12 10 18 21	Relatively high
Category II	01 04 03(2) 05 13 14 16 06(2) 06(1)	Relatively low

We analyzed the sensitivity of the change in the number of high-potassium glass classifications on its classification results. The sensitivity analysis table was shown as Table 2.

Table 2. High potassium glass cultural relic sensitivity analysis table

Number of categories	Category	Cultural relic number	Chemical composition content
3	Category I	01 04 03(2) 05 13 14 16 06(2) 06(1)	Low SiO_2 content
	Category II	03(1) 07 22 27 09 12 10	High SiO_2 content
	Category III	18 21	High P_2O_5 content
4	Category I	01 04 03(2) 05 13 14 16 06(2) 06(1)	Low SiO_2 content
	Category II	03(1) 07 22 27 09 12 10	High SiO_2 content
	Category III	18	High K_2O content
	Category IV	21	K_2O content is 0

4.2. Classification of lead barium glass:

Since the content of lead-barium glass silica (SiO_2) before weathering accounted for the largest proportion on average, lead oxide (PbO) was the second. The content of lead oxide (PbO) after weathering accounted for the largest proportion, followed by silicon dioxide (SiO_2). The proportion of silica (SiO_2) content generally decreased during the weathering process, while the proportion of lead oxide (PbO) content generally increased during the weathering process. Based on this, we firstly divided lead barium glass into unweathered lead barium glass and weathered lead barium glass. Then the cluster analysis method was used to cluster separately.

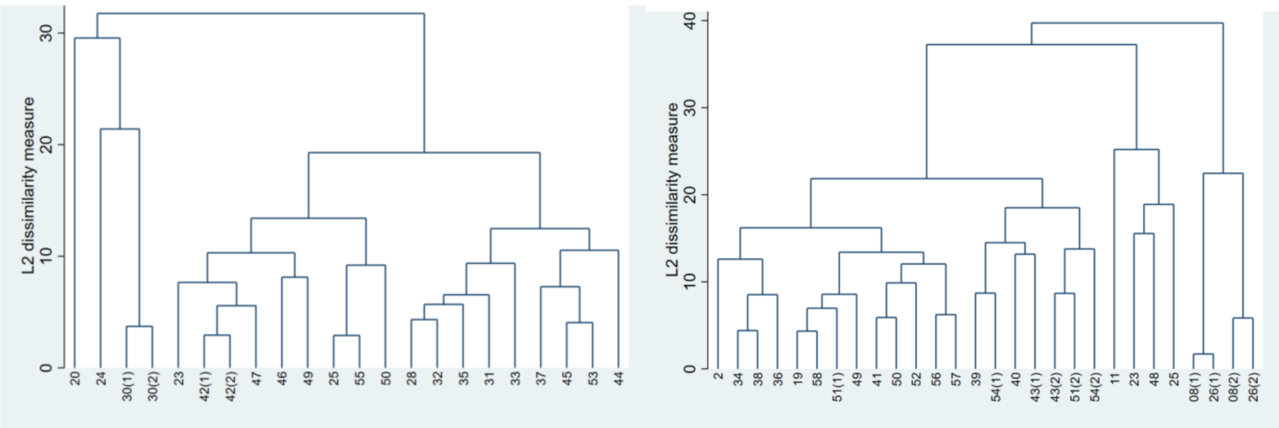

Figure 4. Classification of unweathered and weathered lead barium glass relics

4.2.1. Unweathered lead barium glass relics

It can be concluded that the unweathered lead barium glass cultural relics can be roughly divided into two categories through the cluster plot Figure 4, one was high content of silica (SiO_2) and the other was low content of silica (SiO_2). The classification table was shown as Table 3.

Table 3. Classification of unweathered lead barium glass relics

Category	Cultural relic number	SiO_2 content
Category I	23 42(1) 42(2) 47 46 49 25 55 50 28 32 35 31 33 37 45 53 44	Relatively high
Category II	20 24 30（1）30（2）	Relatively low

We analyzed the sensitivity of the change of the classification number of unweathered lead-barium glass cultural relics on the classification results. Sensitivity analysis was shown in Table 4.

Table 4. Cultural relic sensitivity analysis table of unweathered lead barium glass

Number of categories	Category	Cultural relic number	Chemical composition content
3	Category I	20	Low SiO_2 content Low **PbO** content
	Category II	24 30(1)30(2)	Low SiO_2 content
	Category III	23 42(1) 42(2) 47 46 49 25 55 50 28 32 35 31 33 37 45 53 44	High SiO_2 content
4	Category I	20	Low **PbO** content
	Category II	24	High **BaO** content
	Category III	30(1) 30(2)	Low SiO_2 content
	Category IV	23 42(1) 42(2) 47 46 49 25 55 50 28 32 35 31 33 37 45 53 44	High SiO_2 content

4.2.2. Weathered lead barium glass relics

It can be seen from the cluster map that weathered lead barium glass relics can be roughly divided into two categories, one was the high content of lead oxide (PbO), the other was the low content of lead oxide (PbO). The classification table was shown as Table 5:

Table 5. Classification table of weathered lead barium glass relics

Category	Cultural relic number	PbO content
Category I	02 34 38 36 19 58 51(1) 49 50 52 56 57 39 54(1) 40 43(1) 43(2) 51(2) 54(2) 11 23 48 25	Relatively high
Category II	08(1) 26(1) 08(2) 26(2)	Relatively low

We analyzed the sensitivity of the change of the classification number of weathered lead-barium glass cultural relics on the classification results. Sensitivity analysis was shown in Table 6.

Table 6. Weathered lead barium glass cultural relic sensitivity analysis table

Number of categories	Category	Cultural relic number	Chemical composition content
3	Category I	02 34 38 36 19 58 51(1) 49 50 52 56 57 39 54(1) 40 43(1) 43(2) 51(2) 54(2)	High **PbO** content
	Category II	11 23 48 25	High **PbO** content High SiO_2 content
	Category III	08(1) 26(1) 08(2) 26(2)	Low **PbO** content
4	Category I	02 34 38 36 19 58 51(1) 49 50 52 56 57 39 54(1) 40 43(1) 43(2) 51(2) 54(2)	High **PbO** content
	Category II	11	High **BaO** content
	Category III	23 48 25	Low **BaO** content
	Category IV	08(1) 26(1) 08(2) 26(2)	Low **PbO** content

4.3. Rationality and sensitivity analysis of high potassium glass

Rationality analysis. Based on the attachment data and the previous analysis, we can see that silicon dioxide (SiO_2) was the main component of high-potassium glass, with a proportion range of [59.01%, 96.77%]. The weathered high-potassium glass had a higher level of silica (SiO_2) in its chemical composition than the unweathered high-potassium glass. Therefore, it was reasonable to use the level of silicon dioxide (SiO_2) as the criterion for the classification.

Sensitivity analysis. As shown in Table 1 and Table 2, when the number of classifications was three, cultural relics 18 and 21 were assigned to different categories than when the number of classifications was two. This was because artifacts 18 and 21 had a high level of phosphorus pentoxide (P_2O5), which can form a separate category. From this, we can infer that when the number of classifications was 3, the chemical composition of the subcategories was determined by the level of silica (SiO_2) first, followed by the level of phosphorus pentoxide (P_2O_5).

When the number of classifications was four, cultural relics 18 and 21 were each assigned to a separate category. This was because artifact 18 had a high level of potassium oxide (K_2O) and artifact 21 had no potassium oxide (K_2O), so they were classified differently. We can infer that when the number of classifications is 4, the chemical composition of the subcategories was determined by the level of silicon dioxide (SiO_2) first, followed by the level of phosphorus pentoxide (P_2O_5), and then the level of potassium oxide (K_2O).

As the analysis above showed, the number of categories had a negligible effect on the classification results, so changing the classification number did not make much difference.

4.4. Rationality and sensitivity analysis of lead barium glass:

4.4.1. Unweathered lead barium glass relics

Rationality analysis. Based on the attached data and the previous analysis, we found that silicon dioxide (SiO_2) was the major component of lead barium glass, with a proportion range of [31.94%, 75.51%]. Hence, it was reasonable to use the chemical composition of silica (SiO_2) as the basis for classification.

Sensitivity analysis. Table 3 and Table 4 showed that when the number of classifications was 3, artifact 20 was classified separately from the other artifacts, which were grouped into two categories. This was because artifact 20 had a low lead oxide (PbO) content, which made it distinct from the others. Therefore, we inferred that when the number of classifications was 3, the chemical composition of each category was determined by the proportion of silica (SiO_2) first and then lead oxide (PbO).

When the number of classifications was 4, artifact 20 and artifact 24 were each assigned to a separate category, while the rest of the artifacts were divided into two groups. This was because artifact 24 had a high barium oxide (BaO) content, which distinguished it from the others. Therefore, we inferred that when the number of classifications was 4, the chemical composition of each category was determined by the proportion of silica (SiO_2) first, then lead oxide (PbO), and finally barium oxide (BaO)

4.4.2. Weathered lead barium glass relics

Rationality analysis. Based on the supplementary data and the previous analysis, we observed that lead oxide (PbO) was the most abundant component in weathered lead barium glass, and its proportion varied widely among different artifacts. Therefore, it was reasonable to use the lead oxide (PbO) content as the criterion for classification.

Sensitivity analysis. As shown in Table 5 and Table 6, when the number of categories was 3, artifacts 11, 2, 48 and 25 were each allocated to a different category than when the number of categories was 2. This was because these four artifacts had a high lead oxide (PbO) content, which set them apart from the others. We deduced that when the number of categories was 3, the chemical composition of each category was determined by the proportion of silica (SiO_2) first, then lead oxide (PbO).

When the number of categories was 4, artifact 11 was grouped with artifacts 23, 48 and 25, whereas they belonged to different categories when the number of categories was 2. This was because artifact 11 had a high barium oxide (BaO) content, which distinguished it from the others. We inferred that when the number of categories was 4, the chemical composition of each category was determined by the proportion of silica (SiO_2) first, then lead oxide (PbO), and finally barium oxide (BaO).

It can be seen that the number of categories did not significantly affect the classification result, indicating that the classification result was robust to the change of the number of categories.

5. Glass cultural relic identification model based on random forest regression

5.1. Modeling steps

Step 1: The type of high-potassium glass was represented by the number 1, and the type of lead-barium glass was represented by the number 2.

Step 2: The random forest model for glass type identification was a set of N trees $\{S_1(X), S_2(X), S_3(X), \ldots, S_N(X)\}$, where $X = \{x_1, x_2, x_3, \ldots, x_M\}$ was the M-dimension feature vector of glass type.

Step 3: N results were obtained through tree building set operation, among which was the predicted value of the nth tree for glass type.

Step 4: When training on the training set sample data given a range of feature variables, the feature variables of the glass type were trained with $D = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}, X_i (i = 1, \ldots, n)$to denote, and Y_i refered to the representative numerical value of the glass type.

5.2. Classification and identification results of glass cultural relics and sensitivity analysis

We applied the random forest model for the calculation and the results obtained were shown in Table 7.

Table 7. Identification results of random forest model

Cultural relic number	Forecast results
A1、A6、A7	Hyperkalemia
A2、A3、A4、A5、A8	Lead barium

We correlated all the chemical components in the data with their predicted glass types and obtained the correlation coefficients as shown in Table 8.

Table 8. Correlation coefficients of each chemical composition with the predicted results

SiO_2	Na_2O	K_2O	CaO	MgO	Al_2O_3	Fe_2O_3
-0.898 (0.002***)	0.293 (0.482)	-0.208 (0.621)	0.250 (0.551)	0.085 (0.841)	0.114 (0.788)	0.347 (0.399)
CuO	**PbO**	**BaO**	**P_2O_5**	**SrO**	**SnO_2**	**SO_2**
0.098 (0.817)	0.857 (0.006***)	0.621 (0.100)	0.498 (0.209)	0.679 (0.064*)	0.293 (0.482)	0.161 (0.703)

Note: * * *, * * and * represent the significance level of 1%, 5% and 10% respectively

The correlation coefficient table indicated that silica (SiO_2) and lead oxide (PbO) had significant effects at the 1% level and strontium oxide (SrO) at the 10% level. The other chemical compositions did not show any remarkable influence. This suggested that the predicted results were strongly correlated with silica, lead oxide, and strontium oxide, while other chemical components had weaker correlations. Therefore, we examined how the changes in the contents of these three chemical components affected the prediction results by adjusting their levels of silica, lead oxide and strontium oxide.

The data showed that the cultural relics had silica, lead oxide and strontium oxide contents within the effective range of the total content (85%-105%). These contents increased or decreased proportionally (except for zero values, which only increased) while keeping other chemical contents unchanged. Based on this, we can identify the glass type of the cultural relics. We substituted the sorted data into the random forest regression model above, and the calculated partial results were shown in Table 9, 10, and 11.

Table 9. Sensitivity analysis table of silica to predicted results

Cultural relic number	SiO_2	Na_2O	…	PbO	…	SO_2	Forecast results
A1	78.45	0		0		0.51	Hyperkalemia
A1	78.45	0	…	1	…	0.51	Hyperkalemia
A1	78.45	0		2		0.51	Hyperkalemia
	…				…		
A8	51.12	0		24.24		2.26	Lead barium
A8	51.12	0	…	25.24	…	2.26	Lead barium
A8	51.12	0		26.24		2.26	Lead barium

Table 10. Sensitivity analysis table of lead oxide to predicted results

Cultural relic number	SiO_2	Na_2O	…	PbO	…	SO_2	Forecast results
A1	78.45	0		0		0.51	Hyperkalemia
A1	78.45	0	…	1	…	0.51	Hyperkalemia
A1	78.45	0		2		0.51	Hyperkalemia
	…				…		
A8	51.12	0		24.24		2.26	Lead barium
A8	51.12	0	…	25.24	…	2.26	Lead barium
A8	51.12	0		26.24		2.26	Lead barium

Table 11. Analysis table of sensitivity of strontium oxide to prediction results

Cultural relic number	SiO_2	Na_2O	…	SrO	SnO_2	SO_2	Forecast results
A1	78.45	0		0.03	0	0.51	Hyperkalemia
A1	78.45	0	…	1.03	0	0.51	Hyperkalemia
A1	78.45	0		2.03	0	0.51	Hyperkalemia
	…					…	
A8	51.12	0		3.31	0	2.26	Lead barium
A8	51.12	0	…	4.31	0	2.26	Lead barium
A8	51.12	0		5.31	0	2.26	Lead barium

The prediction results did not change when the levels of silicon dioxide, lead oxide and strontium oxide in cultural relics varied. This indicated that the prediction results were insensitive to the variations in the levels of silicon dioxide, lead oxide and strontium oxide.

We discovered that when the content of compounds increased or decreased within the effective range of total content, the predicted results were not sensitive to their changes. This indicated that the model was stable.

6. Conclusion

This paper used the decision tree model to explore the classification rules of high-potassium glass and lead-barium glass by studying the chemical composition of ancient glass. The segmentation method of cluster analysis was used to subgroup each category by selecting the appropriate chemical composition. Finally, we analyzed the rationality and sensitivity of the classification results. When the number of classifications changed, the change of classification results was not significant. The change of classification number was not sensitive to the classification result, which indicated that the classification result is reasonable. We analyzed the chemical composition of the unknown categories of glass relics in the data, and the random forest model was established to identify the specific type of glass relics. Through sensitivity analysis, it was found that the predicted results were not sensitive to the change of the content of compounds with strong correlation, indicating that the model was relatively stable.

References

[1] Won-in K, Thongkam Y, Pongkrapan S et al. Raman spectroscopic study on archaeological glasses in Thailand: ancient Thai glass. [J]. Spectrochim Acta A Mol Biomol Spectrosc, 2011, 83: 231-5.

[2] Cheng Qian, Guo Jin-Long, Wang Bo et al. [Characteristics of chemical composition of glass finds from the Qiemo tomb sites on the Silk Road].[J] .Guang Pu Xue Yu Guang Pu Fen Xi, 2012, 32: 1955-60.

[3] Wang Jianlan,Guan Juan,Hawkins Nicholas et al. Analysing the structure and glass transition behaviour of silks for archaeology and conservation.[J] .J R Soc Interface, 2018, 15: undefined.

[4] Henderson Julian,Tradition and experiment in first millennium a.d. glass production--the emergence of early Islamic glass technology in late antiquity.[J] .Acc Chem Res, 2002, 35: 594-602.

[5] Li Qing-lin,Xu Cheng-tai,Ling Xue et al. [Nondestructive analysis of some ancient glass in Jin and Yuan dynasty, by EDXRF probe].[J] .Guang Pu Xue Yu Guang Pu Fen Xi, 2011, 31: 1960-3.

[6] Wen Rui,Zhao Zhi-qiang,Ma Jian et al. [Chemical Analysis of Ancient Glass Beads from Shirenzigou Sites in Balikun County, Xinjiang].[J] .Guang Pu Xue Yu Guang Pu Fen Xi, 2016, 36: 2961-5.

[7] Liu Song,Li Qing-hui,Gan Fu-xi et al. Applacation of portable energy-dispersive X-ray fluorescence spectrometer in chemical composition analysis of Chinese ancient glass].[J] .Guang Pu Xue Yu Guang Pu Fen Xi, 2010, 30: 2576-80.

[8] Melcher Michael,Wiesinger Rita,Schreiner Manfred,Degradation of glass artifacts: application of modern surface analytical techniques.[J] .Acc Chem Res, 2010, 43: 916-26.

[9] Selih Vid S,van Elteren Johannes T,Quantitative multi-element mapping of ancient glass using a simple and robust LA-ICP-MS rastering procedure in combination with image analysis.[J] .Anal Bioanal Chem, 2011, 401: 745-55.

[10] Panighello Serena,Van Elteren Johannes T,Orsega Emilio F et al. Laser ablation-ICP-MS depth profiling to study ancient glass surface degradation.[J] .Anal Bioanal Chem, 2015, 407: 3377-91.

[11] Liu S, Li Q H, Gan F X. Influence of ancient glass samples surface conditions on chemical composition analysis using portable XRF [J]. Guang pu xue yu guang pu fen xi = Guang pu, 2011, 31(7): 1954-9.

[12] Balvanović Roman V,Šmit Žiga,Emerging Glass Industry Patterns in Late Antiquity Balkans and Beyond: New Analytical Findings on Foy 3.2 and Foy 2.1 Glass Types. [J]. Materials (Basel), 2022, 15: undefined.

[13] Zou Ying,Molecular-Composition Analysis of Glass Chemical Composition Based on Time-Series and Clustering Methods.[J]. Molecules, 2023, 28: undefined.

[14] Russell S, Norvig P. Artificial Intelligence. A Modern Approach. Prentice-Hall: Englewood Cliffs, NJ, 1995.

[15] Landau S, Ster I C. Cluster analysis: overview [J]. Á Á, 2010, 11(x12): x1p.

[16] Biau G. Analysis of a random forests model [J]. The Journal of Machine Learning Research, 2012, 13(1): 1063-1095.