Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 7, No. 1, 2023

Study of Ancient Glass Classification and Subclassification Based on Systematic Clustering Models

Jun Luo[1, #, *], Peiyuan Jiao[2, #], Keyan Zeng[3, #], Yixin Zhang[4, #]

[1]School of Water Conservancy and Civil Engineering, Zhengzhou University, Henan, Zhengzhou, 450000, China
[2]College of Mechanical and Power Engineering, Zhengzhou University, Henan, Zhengzhou, 450000, China
[3]School of Management, Zhengzhou University, Henan, Zhengzhou, 450000, China
[4]School of Water Conservancy and Civil Engineering, Zhengzhou University, Henan, Zhengzhou, 450000, China

[*]Corresponding author: lj0204062023@163.com

[#]These authors contributed equally.

Abstract: Ancient glass is highly susceptible to weathering by environmental influences, resulting in changes to its internal chemical composition, which can affect the correct determination of its category. For the purpose of classifying ancient glass, three indicators of classification were chosen based on the magnitude of the mean value of each chemical composition of the two main types of glass. The σ principle and the lower quartile principle are chosen, respectively, to determine the critical values according to whether the data follow a normal distribution or not, and the final classification results are obtained by hard voting. The hard voting model was trained with an accuracy of 97%. The two main classes of glass were then subclassed according to systematic class clustering, and by making the real data fluctuate within ±1%, the clustering was performed again under the same conditions. The same results were found as for the stable data clustering, indicating that the systematic class clustering model is stable for subclassification of each glass category.

Keywords: Systematic clustering, Hard voting, Glass classification.

1. Introduction

The main raw materials and chemical composition of glass are quartz sand and SiO2.Pure quartz sand has a relatively high melting point and refining requires the addition of a flux to lower the melting temperature. Depending on the flux added and the main chemical composition, the glass is divided into two main categories: high potassium glass and lead-barium glass. In order to classify ancient glass, the literature that has been obtained shows that the theoretical circle has carried out relevant studies mainly from the following aspects: Yang J et al. obtained the composition of glass products by analysing the correlation of factors external to glass and conducting an independent sample Mann-Whitney[1] test on glass, building a model based on random forest[2] and constructing different classification criteria[3]; Zhilin Huang et al. used grey correlation analysis for different types of ancient glass to investigate the correlation between the chemical components of ancient glass[4], and used overall average and k-mean[5], [6] clustering analysis to identify the differences between the chemical components of glass; And scholar Jiajia Lu et al. has incorporated feature selection methods[7] in machine learning and classification algorithms into the study of the problem of compositional analysis and category identification of ancient glass objects[8]. Using accuracy and AUC as classification performance metrics, an attempt was made to construct an integration feature selection model for chemical composition selection of ancient glassware and a random forest model[9] for identification and classification. In this paper, key chemical components are extracted as classification criteria, three indicators are used to judge the type, and an innovative approach is proposed to judge the category based on the critical value, combined with hard voting (minority-majority mechanism).In this paper, for the chemical composition indicators that obey normal distribution, the principle is used to take the threshold; for the indicators that do not obey normal distribution, the lower quartile is used to take the threshold; For indicator data that do not follow a normal distribution, this paper uses the lower quartile to take the critical. The system clustering model[10] makes full use of data information in the selection of classification indicators, and the indicators selected correspond to known information and have a strong ability to be combined in practice. It was found that using the above method, not only is the identification fast and accurate, but it is an important guide for the differentiation of high potassium glass from lead-barium glass, as well as for further subclassification.

2. Establishment and Solution of Ancient Glass Classification and Subclassification Models

2.1. Classification Rules of High Potassium Glass and Lead Barium Glass

Based on the type and weathering degree of glass obtained from on-site investigation data, this article divides all glass into five categories: high potassium weathering, High potassium without weathering, lead barium weathering, lead barium without weathering, and severe lead barium weathering. Statistically describe various chemical composition data, extract various means and standard deviations, and the results are shown in the figure1:

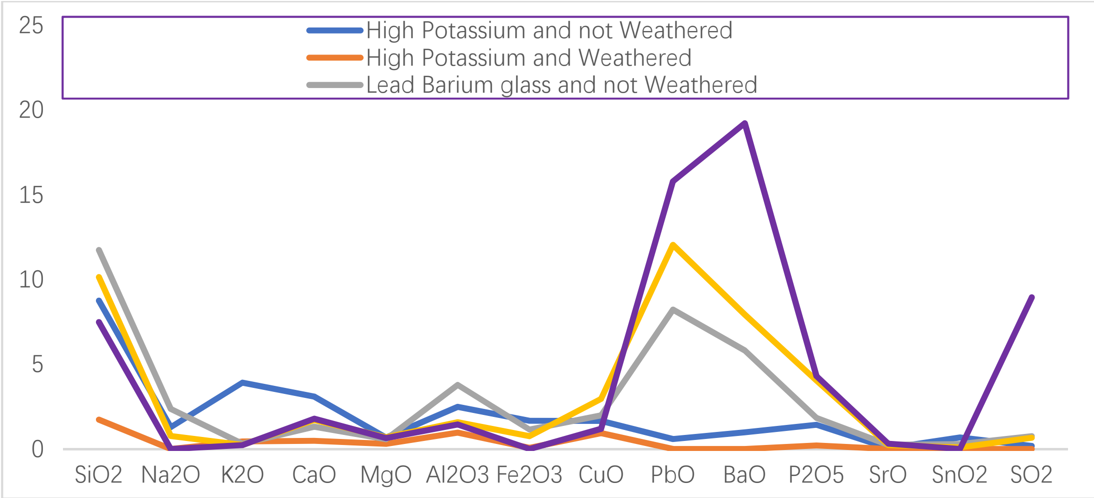

Figure 1. Mean value of various chemical components in different types of glass

From the table, it can be found that the chemical composition of silicon dioxide (SiO2) and potassium oxide (K2O) in high potassium glass is significantly higher than that in lead barium glass, and the content of lead oxide (PbO) and barium oxide (BaO) is significantly lower than that in lead barium glass. However, in the indicator of potassium oxide (K2O), high potassium without weathering glass is significantly higher than that in lead barium glass, but high potassium weathered glass is not significant. Therefore, silicon dioxide (SiO2) is chosen. The three indicators of lead oxide (PbO) and barium oxide (BaO) serve as the basis for determining the type of glass. In order to judge the glass type, it is necessary to determine the critical value. For the index data subject to normal distribution, this paper adopts the 1 sigma principle; For the index data that do not conform to the normal distribution, this paper takes the quartile as the critical value. And this article has three indicators for the type of judgment. In order to reduce errors and improve the accuracy of judgment, this article uses hard voting, which is a method of combining the three indicators to make judgments based on the mechanism of minority obeying majority. The flowchart is shown in the figure 2:

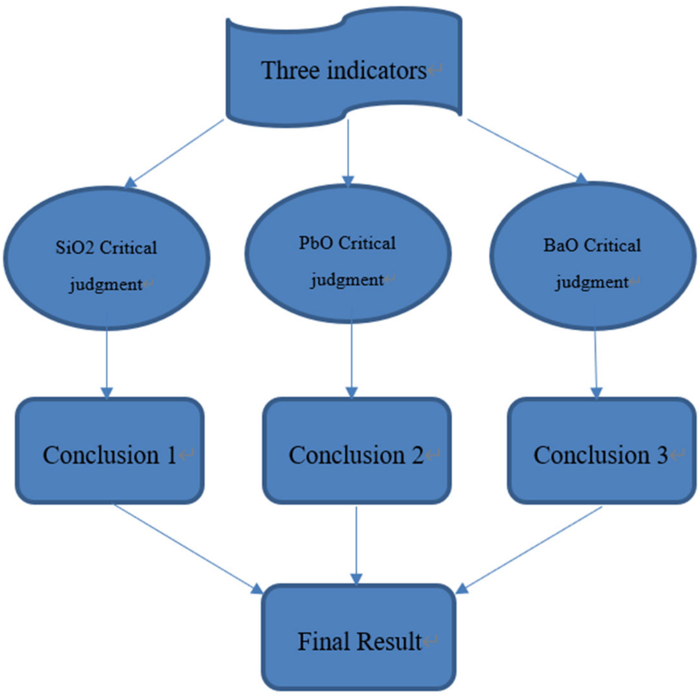

Figure 2. Hard Voting Process

For the indicator of silicon dioxide (SiO2), it can be found that the mean value of high potassium without weathering is smaller than the mean value of high potassium weathering. Therefore, this article chooses high potassium without weathering data to solve the critical value. Firstly, the normal distribution test is carried out, which is implemented by JB test. The other three indicators were tested using the same method, and the results are shown in Table.1:

Tables 1. Jarque-Bera Test

Indicators	Test Object	P
SiO_2	High Potassium and not Weathered	0.0842
PbO	Lead Barium glass and not Weathered	0.5
BaO	Lead Barium glass and not Weathered	0.0022

It can be found that two indicators of silicon dioxide (SiO2) and barium oxide (BaO) reject the original hypothesis under the 90% confidence interval, indicating that the data of two indicators obey the normal distribution. However, the lead oxide (PbO) index accepts the original hypothesis under the 90% confidence interval, indicating that the data of the two indexes do not obey the normal distribution. Obtain the critical point by removing the quartile. The critical values can be obtained from the previous description statistics, and the comprehensive results are shown in Table.2:

Tables 2. Critical value

Indicators	Principle	Critical Value
SiO_2	1σ principle	59.24
PbO	Lower Quartile principle	16.26
BaO	1σ principle	3.02

For silicon dioxide (SiO2), if it is below the critical value, it indicates that the glass is lead barium glass; For barium oxide (BaO), if it is below the critical value, it indicates that the glass is high potassium glass; For barium oxide (BaO), if it is below the critical value, it indicates that the glass is high potassium glass.

Conduct a rationality check on it, program it using MATLAB based on the method used to determine the type of glass. And calculate the accuracy based on the processed data to analyze the rationality of the model. The results obtained are shown in Table.3:

Table 3. Accuracy Results

	Prediction		
Actual Measurement	High Potassium	Lead Barium	Accuracy
High Potassium	17	1	94.4
Lead Barium	1	48	98.0
Total Precent	26.9	73.1	97.0

From the table, it can be seen that the accuracy of the model for determining the glass category is as high as 97%, thus meeting the rationality.

2.2. Establishment and Solution of System Clustering Model

This article is based on two major categories of lead barium glass and high potassium glass to subdivide their subcategories, and uses systematic clustering method to simulate and solve. Cluster analysis was conducted respectively on the chemical composition of lead barium glass and high potassium glass.

According to the distance between different data, variables with similar distances are first clustered into classes, and variables with farther distances are then clustered into classes. And processed in sequence until each variable is classified into a suitable class. According to the above description, the standard deviations of different types of glass were statistically extracted, and the extracted results are shown in Figure3:

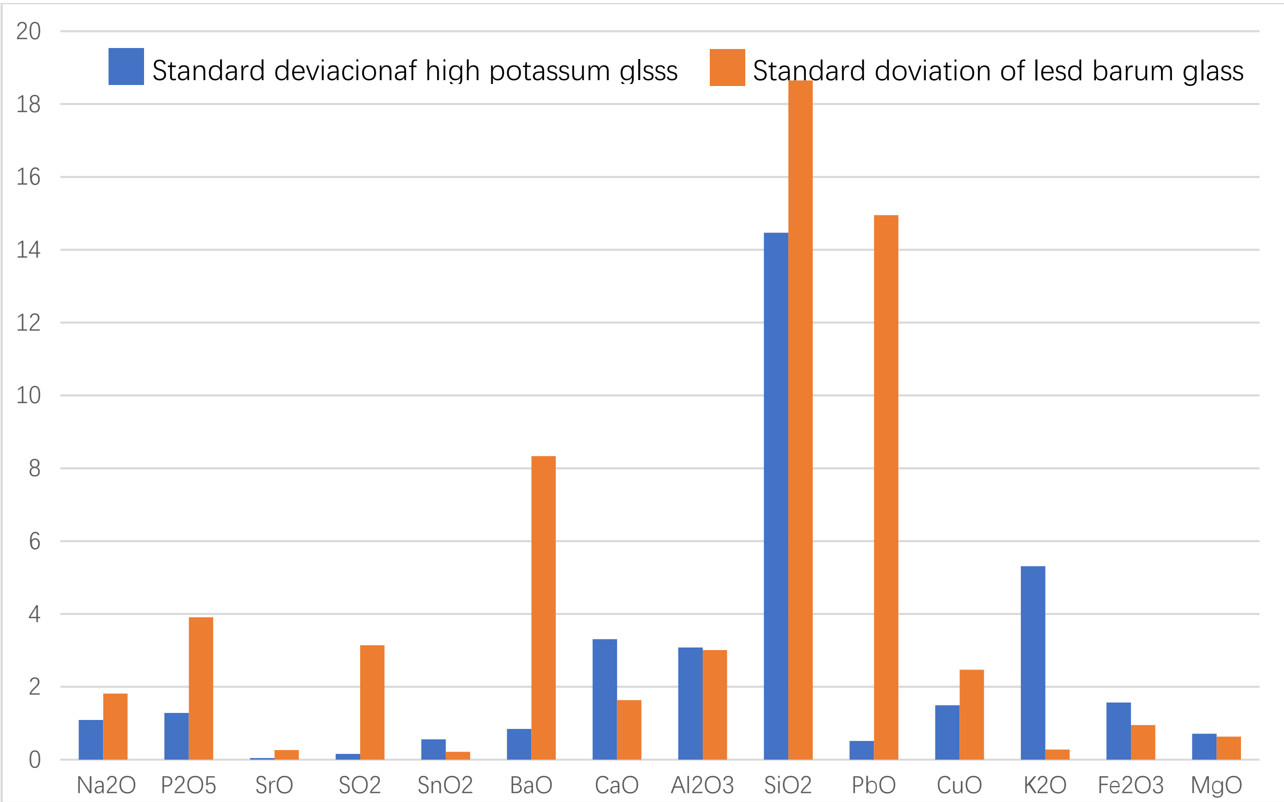

Figure 3. Standard deviation of various chemical components

From the figure, it can be seen that in lead barium glass, the standard deviation of silicon dioxide, lead oxide, and barium oxide is significantly higher than the other indicator values; In high potassium glass, the standard deviation of silicon dioxide, potassium oxide, and calcium oxide is significantly greater than the other indicator values. For the standard deviation, the larger the standard deviation of the indicator, the greater the amount of information contained in the indicator. Therefore, this article selects appropriate chemical indicators based on the value of standard deviation.

Finally, for lead barium glass, this article selects silicon dioxide, lead oxide, and barium oxide as subcategory classification indicators; For high potassium glass, this article selects silicon dioxide, potassium oxide, and calcium oxide as subcategory classification indicators.

In this paper, spss is used for systematic clustering. And the clustering pedigree chart and the line chart of polymerization coefficient obtained. The results are shown in Figure4:

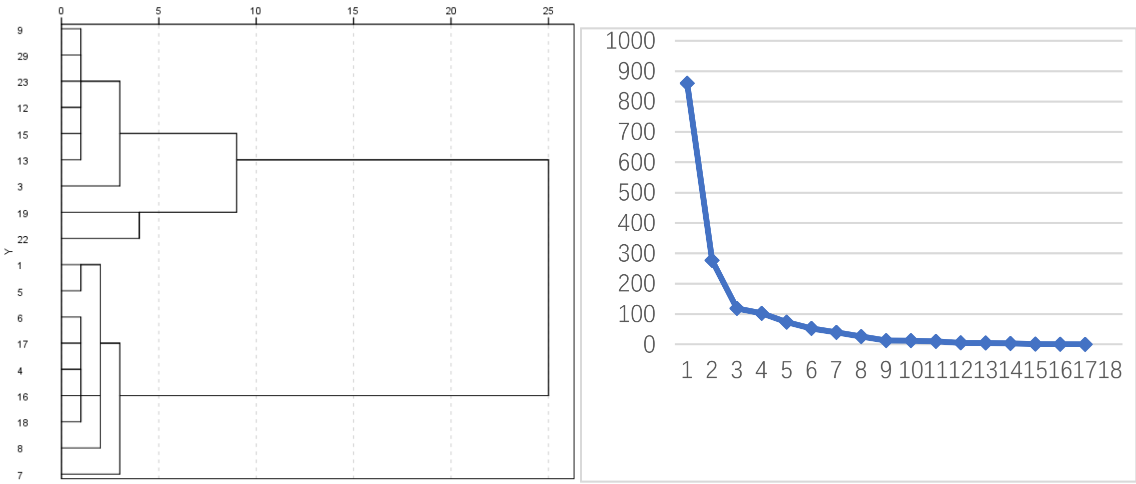

Figure 4. Cluster pedigree diagram and line chart of polymerization coefficient of high potassium glass

According to the line chart of polymerization coefficient, for high potassium glass, when the number of categories is 3, the downward trend of the broken line becomes slow-er, and when the K value is from 1 to 3, the distortion degree changes most. After exceed-ing 3, the degree of distortion significantly decreases. Therefore, the elbow is K=3, so the number of categories can be set to 3. Therefore, this article sets the number of subcatego-ries for high potassium glass classification to 3 categories. the visualization diagram is shown in Figure 5:

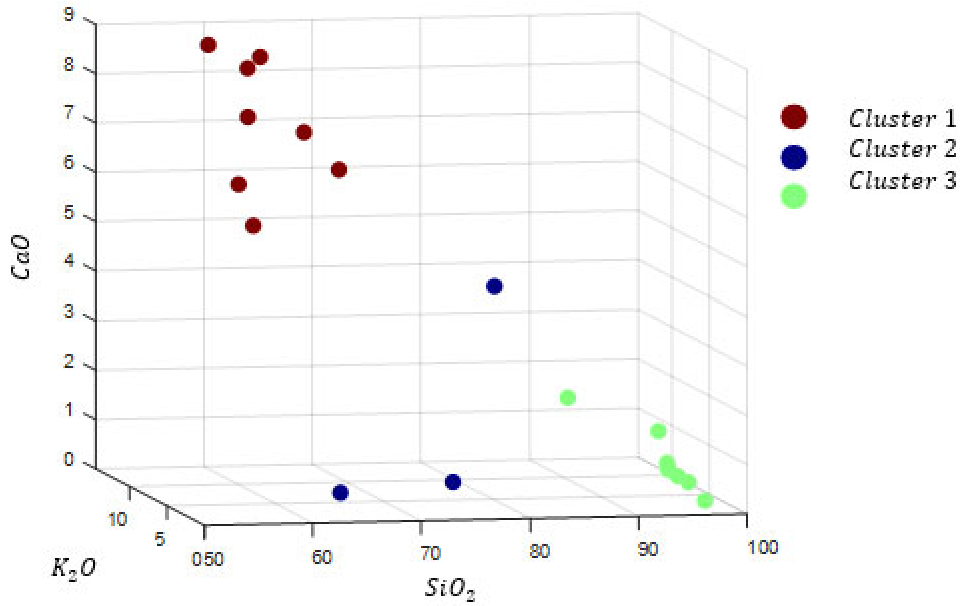

Figure 5. Cluster grouping of high potassium glass

According to the three-dimensional scatter plot, it can be found that the content of calcium oxide in group one is significantly higher than that in the other two groups, which can be defined as the high calcium oxide group; The content of silica in group three is significantly higher than that in the other two groups, which can be defined as the high silica group; Group 2 can be defined as the low silica and low calcium oxide group based on the other two groups.

Similarly, the cluster pedigree diagram and line chart of polymerization coefficient of lead barium glass are shown in Figure6:

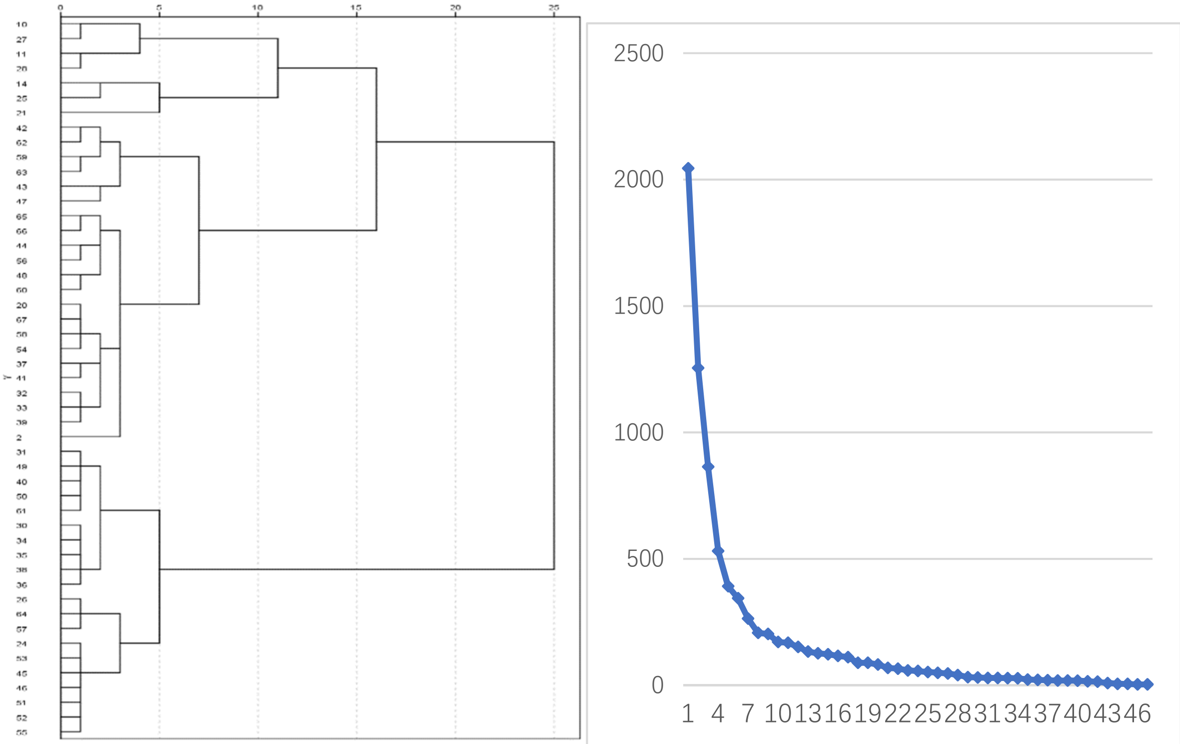

Figure 6. Clustering pedigree diagram and line chart of polymerization coefficient of lead barium glass

According to the line chart of polymerization coefficient, for lead barium glass, when the number of categories is 5, the downward trend of the broken line becomes slower, and when the K value is from 1 to 5, the distortion degree changes most. After exceeding 5, the degree of distortion significantly decreases. Therefore, the elbow is K=5, so the num-ber of categories can be set to 5. Therefore, this article sets the number of subcategories for lead barium glass classification to 5 categories. The visualization diagram is shown in Figure 6:

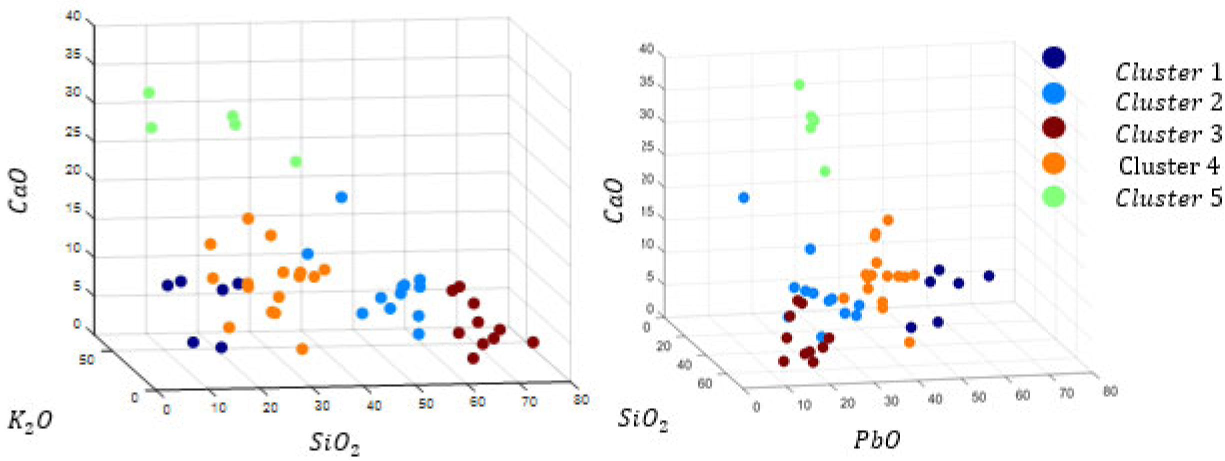

Figure 6. Cluster grouping of lead barium glass

According to the three-dimensional scatter plot on the left, it can be found that the content of silicon dioxide in group three is significantly higher than that of other groups, which can be defined as the high silicon dioxide group; The content of barium oxide in group five is significantly higher than that in other groups, which can be defined as the high barium oxide group; The content of silicon dioxide in group two is only lower than that in group three and significantly higher than that of other groups, which can be defined as the higher silicon dioxide group; According to the three-dimensional scatter plot on the right, it can be found that the content of lead oxide in group one is significantly higher than that in other groups, which can be defined as the high lead oxide group; The content of lead oxide in group four is only lower than group one and significantly higher than the other groups, which can be defined as the higher lead oxide group.

In summary, the subclass classification results of the two major types of glass are shown in Table.4:

Table 4. Subclass Classification Table

Grouping of subclasses of high potassium glass	Names
1	High CaO Class
2	Low SiO2 and CaO Class
3	High SiO_2 Class

Grouping of subclasses of lead barium glass	Names
1	High PbO Class
2	Relatively High SiO_2 Class
3	High SiO_2 Class
4	Relatively High PbO Class
5	High BaO Class

3. Stability and Sensitivity Analysis

In this paper, the method of data fluctuation is used for sensitivity analysis. Accord-ing to the data processed here, each of the data fluctuates within 1%. The random number generated by matlab is used to simulate the size of the data fluctuation. Generate the data after fluctuation, and process the data after fluctuation with the same systematic cluster-ing method as above to obtain the clustering pedigree diagram and line chart of polymer-ization coefficient of high potassium glass, as shown in Figure7:

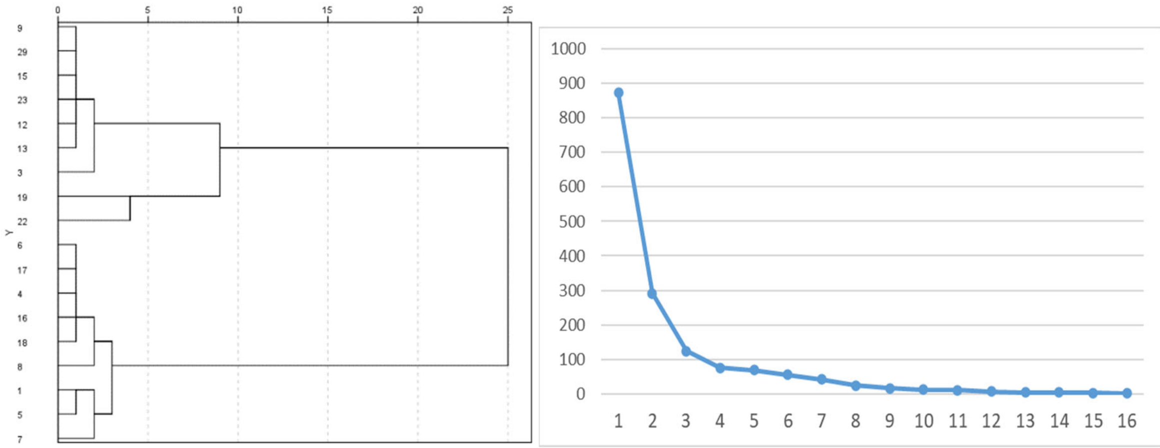

Figure 7. Sensitivity Analysis

Comparing this clustering pedigree with the real cluster analysis clustering pedigree, it is found that there is a great similarity, and three subcategories are also selected ac-cording to the line chart graph of the polymerization coefficient. Comparing the data be-fore and after fluctuations, it is found that the members of each subcategory are the same, indicating that this model has good stability and low sensitivity for subcategories of high potassium glass.

Similarly, the clustering pedigree and line chart diagram of polymerization coeffi-cient of lead barium glass are obtained. By comparing with the real cluster analysis clus-tering pedigree, it is found that there is also a great similarity, which indicates that this model has good stability for the subclassification of lead barium glass, that is, it is less sensitive.

4. Conclusions

In order to carry out the subclassification of each glass category, we selected three indicators for the classification according to the magnitude of the mean values of each chemical composition of the two major glass categories: silica, lead oxide, and barium oxide.

According to whether the data obeyed normal distribution, the sigma principle and the lower quartile principle were chosen to determine the critical values, and the critical values of 59.24 for silica, 16.26 for lead oxide, and 3.02 for barium oxide were obtained to judge the categories, and the final classification results were obtained by hard voting, and the hard voting model was trained with an accuracy of 97%. Then, the two major glass categories were subclassified according to the systematic class clustering, and the clustering was performed again under the same conditions by making the real data fluctuate within 1%, and it was found that the clustering results were the same as those of the stable data, indicating that the systematic class clustering model has good stability for the subclassification of each glass category.

References

[1] Ahmad Zahoor,Nguyen Tuan-Khai,Rai Akhand,Kim Jong-Myon. Industrial fluid pipeline leak detection and localization based on a multiscale Mann-Whitney test and acoustic emission event tracking[J]. Mechanical Systems and Signal Processing,2023,189.

[2] Wall Emerson Robert. Mann-Whitney U test and t-test[J]. Journal of Visual Impairment & Blindness,2023,117(1).

[3] Yang J,Wang Z,Gao S. Component Analysis and Classification Model of Ancient Glassware Based on K-means Clustering and Random Forest[C]//Wuhan Zhicheng Times Cultural Development Co., Ltd..Proceedings of 2023 International

Conference on Mathematical Modeling, Algorithm and Computer Simulation (MMACS 2023).2023:398-406.

[4] Zhilin Huang,Yali Xiong,Yifeng Fu,Feng Li. A study of the problem of compositional analysis of ancient glassware based on grey models[J]. Academic Journal of Materials & Chemistry, 2023, 4(2).

[5] Xu K,Chen Y,Nie J. Subclass classification of ancient glassware based on K-Means and GMM[C]//Wuhan Zhicheng Times Cultural Development Co., Ltd..Proceedings of 2023 International Conference on Mathematical Modeling, Algorithm and Computer Simulation (MMACS 2023). 2023:278-285.

[6] Dong K. Composition analysis and identification of ancient glass based on K-Means clustering[C]//Wuhan Zhicheng Times Cultural Development Co., Ltd..Proceedings of 2023 International Conference on Mathematical Modeling, Algorithm and Computer Simulation (MMACS 2023). 2023: 347-356.

[7] Du Y,Zhou J,Li J. Classification and Identification of Ancient Glass Objects based on K-means++ and Fisher Discriminant [C]//Wuhan Zhicheng Times Cultural Development Co., Ltd..Proceedings of 2023 International Conference on Mathematical Modeling, Algorithm and Computer Simulation (MMACS 2023).2023:357-365.

[8] Lu Jiajia. Ancient glass classification model based on integrated feature selection and random forest[J]. Journal of Silicates, 2023,51(04):1060-1065.

[9] McIntyre Amanda,Saikaley Marcus,Janzen Shannon,Blackport Daymon,Singh Jaswinder,Baker Carmen,Miller Tom,Teasell Robert. A Random Forest Classification Model for Spasticity after Stroke[J]. Archives of Physical Medicine and Rehabilitation,2022,103(12).

[10] ChávezDurán Álvaro Agustín, OlveraVargas Miguel, Figueroa Rangel Blanca, García Mariano, Aguado Inmaculada, RuizCorral José Ariel. Mapping Homogeneous Response Areas for Forest Fuel Management Using Geospatial Data, K-Means, and Random Forest Classification[J]. Forests, 2022, 13(12).