A research on subclassification of ancient glass based on FLDA and systematic cluster analysis

Boshang Hu*, Bolun Fu, Zhaokang Han

Guangdong University of Technology, Guangzhou, China, 510006

*Corresponding author: 13733993818@163.com

Abstract. For the purpose of subclassification of ancient glass, this study first used FLDA (Fisher Linear Discriminant Analysis) to classify the glass, in this study, unknown glasses of known composition were reasonably classified by model analysis with a high accuracy of 92.5%, and the classification pattern of the two types of glass was analysed statistically in conjunction with the data, and the results showed that the high potassium type contained high levels of K and Na chemical components; the lead-barium type, on the other hand, contains a high content of Pb and Ba chemical components. Next, a systematic clustering model was used for cluster analysis, avoiding the disadvantages of the K-means clustering model which is sensitive to outliers, and further analysed by a combination of clustering dendrograms and the law of the elbow in the line graph of aggregation coefficient. The analysis was further Integrated analysed by the law of the Elbow in the clustering dendrograms and the line graphs of aggregation factors. The final subclasses were concluded as follows: high potassium glasses can be divided into two subclasses, K_2O-CaO (~10wt%)-SiO_2 and K_2O-SiO_2; lead-barium glasses can be divided into three subclasses, PbO-BaO-SiO_2, PbO (~25wt%)-BaO-SiO_2 and CaO-PbO (~40wt%)-BaO-SiO_2. The study of the subclassification of ancient glass is very helpful and influential for industrial applications. Through the construction of the classification and clustering model in this study, the subclassification division of ancient glass is well carried out, which has an important theoretical basis to support the work of identification and detection, and is of great significance for the study of the origin of ancient glass.

Keywords: Glass, Chemical components, Subclassification, FLDA, Systematic clustering.

1. Introduction

Ancient glass was one of the first man-made materials and carries a wealth of cultural and technological information [1]. According to archaeological research, glass products were introduced to China early from the West Asian and Egyptian regions. We have absorbed its technology and localised its production. Ancient glasswork is meanwhile an important physical evidence of the Silk Road and earlier trade exchanges [2-4]. Therefore, the study of ancient glass subclassification is important for the study of the origin of ancient glass. We can gain a better understanding of the ancient glass-making process, the sources of raw materials and production patterns, and thus a better understanding of ancient civilisations [5]. Ancient glass is divided into two main categories, high potassium glass and lead-barium glass, which have undergone weathering effects that have altered some of their chemical composition and have important implications for identification and testing work.

An extensive literature review shows that FLDA models and systematic clustering models are widely used in various fields. Mohcene Bessaoudi's approach to improve the accuracy of multimodal 2D and 3D face verification based on multilinear enhanced Fisher discriminant analysis [6]. Pasadas Dario J. Improved performance of a debonding classifier based on improved Fisher's discriminatory criteria for recognition and ranking of damage sensitive features [7]. Koren Oded solves the small document issues with a systematic clustering algorithms for efficient file management [8]. Salgado-Hernández J E Non-linear correlation analysis in financial markets using systematic clustering, providing an overview of the stochastic evolution of the state of financial markets [9].

In summary, although both methods are used in different areas, the combination of the two is less commonly used for classification problems. In this paper, we innovatively combine two analytical models to conduct a comprehensive study on the subclassification of ancient glass and finally obtain a more reasonable classification result.

In this study, the FLDA model was used to classify unknown glass samples and analyse the classification pattern based on the basic conditions of ancient glass samples. A systematic clustering model was then used to rationalise the choice of chemical composition and to sub-classify each of these two glass types.

2. Materials and methods

2.1. Data acquisition and preprocessing

The data for this study were obtained from Question C of the 2022 National Student Mathematical Modelling Competition (http://www.mcm.edu.cn). The analysis of the study revealed that there were still missing values and outliers in the data, so we performed data pre-processing.

Outlier handling: Since the sum of the component proportions is between 85% and 105%, the data are considered valid. However, the sum of the compositional proportions of the high potassium samples with artefact numbers 15 and 17 was calculated to be 79.47% and 71.89%, so the data for these two samples were discarded as anomalous.

Zero replenishment processing: As can be seen from the data description, the proportion of the corresponding major component given in Form 2 is blank to indicate that the component was not detected, therefore for the blank the proportion of the major component is 0, fill in the table with 0 to complete the gap.

After performing descriptive statistics on the data, for high potassium glass and lead-barium glass the chemical composition box line diagram are shown in Figures 1 and 2.

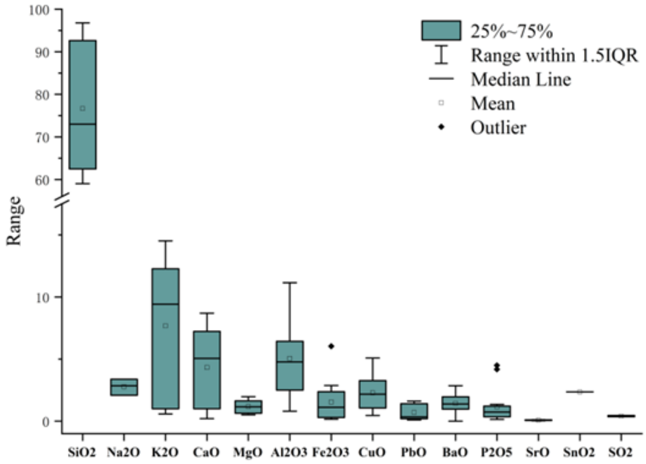

Figure 1. Box line diagram of the chemical composition of high potassium glass

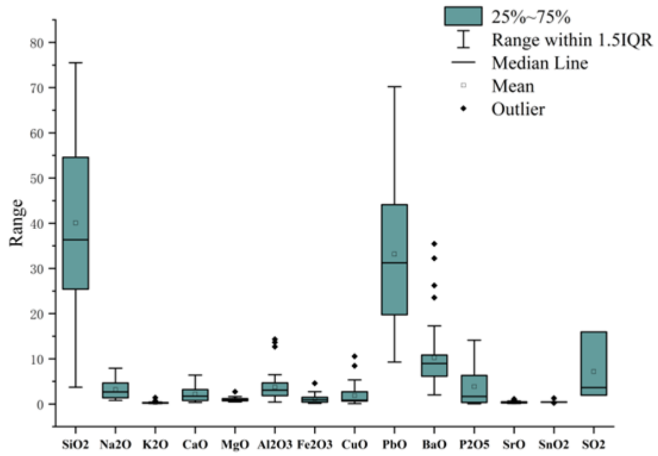

Figure 2. Box line diagram of the chemical composition of lead barium glass

The box line diagram shows that the high potassium glass has a high SiO_2 content, between 60% and 90%, and a high content of K_2O; Lead-barium glass has a lower SiO_2 content, between 30% and 60%, and a higher content of PbO. This gives a general classification of high potassium and lead-barium glass.

2.2. Methodology

2.2.1 Principles and advantages of the FLDA model

FLDA (Fisher Linear Discriminant Analysis), also known as Fisher discriminant analysis, is a typical discriminant method. By projecting the sample data on a straight line, given the sample training set data, combined with Analysis of variance (ANOVA), so that like data are as close as possible and dissimilar data are as far away as possible, the line can be understood as a hyperplane. That is, we need to find a hyperplane with a linear coefficient vector of ω. As in Figure 3, under a two-dimensional plot of x_1 and x_2 variables, two categories can be distinguished according to the discriminant line $y = w^T x$.

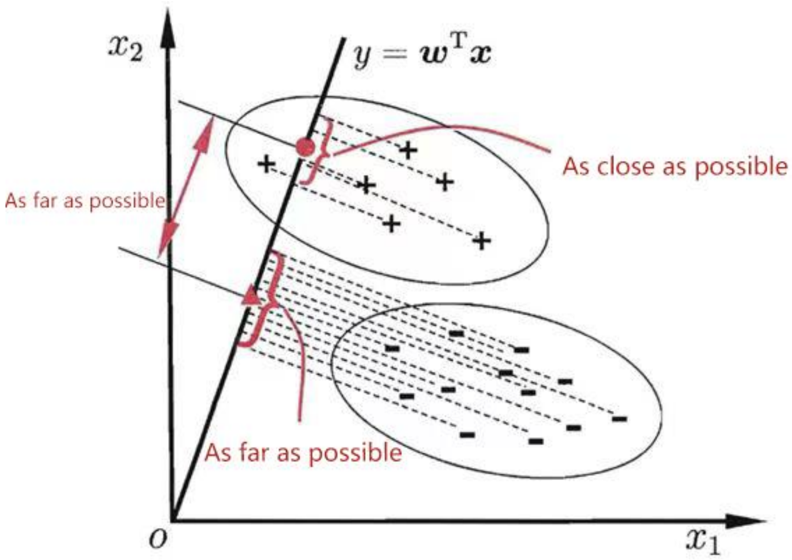

Figure 3. Schematic diagram of FLDA

Benefits of the FLDA model:

1) FLDA is a supervised learning method that can be trained using samples from known categories to improve classification accuracy and maximise the retention of classification information in a supervised situation, this classification information is measured by a non-reference indicator, the Fisher metric.

2) FLDA is essentially a projection technique for dimensionality reduction. For linearly divisible samples, it is always possible to find a projection direction that makes the reduced samples still linearly divisible and better divisible, i.e. "low coupling and high cohesion", with minimum intra-class distance and maximum inter-class distance.

3) FLDA can solve directly for normal vectors $y = w^T x$.

4) FLDA is not only suitable for training deterministic pattern classifiers, but also for stochastic modal machines, and FLDA models can be extended to multi-class problems.

2.2.2 Principles and advantages of systematic clustering models

Systematic clustering, also called hierarchical clustering, is a merging algorithm that combines the two closest classes of data points by calculating the distance between them, clustering variables that are close together into classes first and variables that are farther apart into classes later, and iterating this process until all data points are combined into one class and a clustering spectrum chart is generated. As shown in Figure 4, it presents a remarkable hierarchical structure that is intuitive and clear.

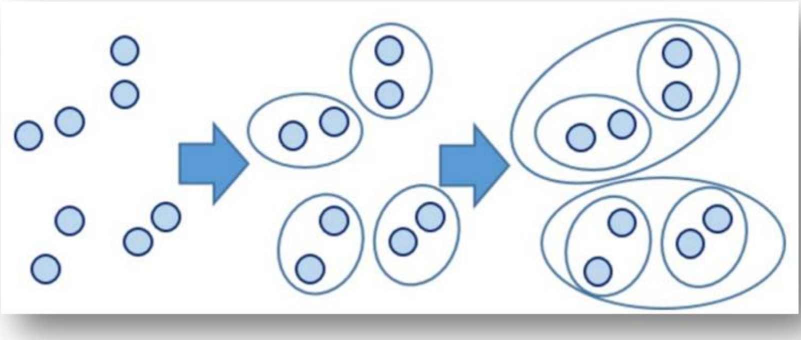

Figure 4. Schematic of system clustering

Advantages of systematic clustering models:

1) No need to pre-determine the number of clusters and avoid the disadvantages of the K-means model which is sensitive to the initial cluster centres and outliers.

2) A wealth of methods for calculating interclass distances and representing results, and the possibility of transformation and standardisation of data for processing.

3) The similarity of distances and rules is easy to define and less restrictive.

4) The results are presented in an intuitive hierarchy and the conclusions are expressed in a clear and concise form.

3. Model building and solving

3.1. FLDA model building

Firstly, the training high potassium glass samples and lead barium glass samples are divided into two subsets X_1 and X_2, and the dummy variables are set to 0 and 1 respectively.

Next, calculate the mean vector for each category$u_i, i = 1,2,\ u_i = \frac{1}{N_i}\sum_{x \in X_i} x$;

Further calculate the intra-class dispersion matrix for each class$S_{\omega i},\ i = 1,2,\ S_{\omega i} = \sum_{x \in X_i}(x - u_i)(x - u_i)^T$;

Then calculate the total dispersion matrix within class$S_\omega = S_{\omega 1} + S_{\omega 2}$;

Then calculate the inverse matrix${S_\omega}^{-1}$ of the matrixS_ω and find the vector $\omega = {S_\omega}^{-1}(u_2 - u_1)$;

Further find the discriminant function$y = \omega^T x$ and the threshold value of the discriminant function$\omega_0,\ \omega_0 = \frac{N_1\omega^T u_2 + N_2\omega^T u_2}{N_1 + N_1}$;

Finally, the classification rules are based on: comparing the magnitude of the y-value with the threshold ω_0 to arrive at its classification.

The modelling was done in a discriminant analysis in SPSS.

3.2. Solving the model

The typical discriminant function coefficients are the linear coefficient vectors ω. Through SPSS discriminant analysis, the individual portions of the linear coefficient vector ω on the 15 indicators can be obtained, on the basis of which the Bayesian discriminant function coefficients ω_0 can be further derived, which in turn can be classified by comparison.

Table 1. Typical discriminant function coefficients

SiO_2	Na_2O	K_2O	CaO	MgO	Al_2O3	Fe_2O_3	Constants
0.054	-0.150	0.247	0.012	0.044	-0.046	0.056	
CuO	**PbO**	**BaO**	**P_2O_5**	**SrO**	**SnO_2**	**SO_2**	-3.629
0.343	0.021	-0.075	0.033	-0.661	0.104	0.316	

The classification function coefficients are the Bayesian discriminant function coefficientsω_0, and the corresponding Bayesian discriminant function coefficients are tabulated as shown in Table 2. The 14 components of the sample are substituted into the 2 Bayesian discriminant functions and the larger function values indicate the corresponding sample classification.

Table 2. Classification function coefficients

High Potassium Glass	**SiO_2**	**Na_2O**	**K_2O**	**CaO**	**MgO**	**Al_2O_3**	**Fe_2O_3**
TRUE	32.630	34.713	37.811	38.867	48.065	26.298	32.485
FALSE	32.443	35.228	36.961	38.825	47.914	26.455	32.294
CuO	**PbO**	**BaO**	**P_2O_5**	**SrO**	**SnO_2**	**SO_2**	**Constants**
18.459	32.962	40.582	42.392	-76.615	24.718	16.108	-1620.408
17.281	32.891	40.840	42.279	-74.343	24.360	15.021	-1605.728

The 14 chemical compositions attributed to high potassium glass and lead-barium glass can be classified from Table 3.

The corresponding ancient glass samples were judged on the following basis:

Table 3. Analysis of classification results

Glass type			**Forecast Team Member Information**		**Total**
			High Potassium	**Lead Barium**	
Original	**Counting**	**High Potassium**	19	2	21
		Lead Barium	3	43	46
		Ungrouped cases	4	4	8
	%	**High Potassium**	90.5	9.5	100.0
		Lead Barium	6.5	93.5	100.0
		Ungrouped case	50.0	50.0	100.0
a. 92.5% of the original grouped cases were correctly classified.					

The analysis of the classification results shows that after classifying 67 samples, 19 out of 21 high potassium glass samples were correctly classified by the FLDA classification model, with a correct rate of 90.5%; The number of correct classifications obtained by the FLDA classification model was 43 out of 46 lead-barium glass samples, with a correct rate of 93.5%.

The above shows that the FLDA has correctly classified 90.5% of the high potassium type and 93.5% of the lead-barium type for unknown samples, and 92.5% of the overall classification.

Therefore, eight unknown sample artefacts were classified based on this FLDA model and the classification results were obtained by SPSS solution as shown in Table 4.

Table 4. Classification results

Artifact number	**Type**
A1	High Potassium
A2	Lead Barium
A3	Lead Barium
A4	Lead Barium
A5	Lead Barium
A6	High Potassium
A7	High Potassium
A8	High Potassium

To sum up: ancient glass can be divided into two categories: high potassium and lead-barium, with the high potassium type containing a high content of K and Na chemical components; The lead-barium type, on the other hand, contains a high content of Pb and Ba chemical components. For the eight unknown sample artefacts, the classification model yielded that A1, A6, A7 and A8 were high potassium glass samples; A2, A3, A4 and A5 were lead-barium glass samples and were correctly classified at 92.5%.

3.2.1 Systematic clustering model development

ReferencesIn this article, a systematic clustering model was selected and SPSS was used to carry out systematic clustering, generating the corresponding clustering dendrogram shown in Figure 5, which gives feedback on the class merging in each iteration of the process, and the systematic clustering model building flow chart shown in Figure 6.

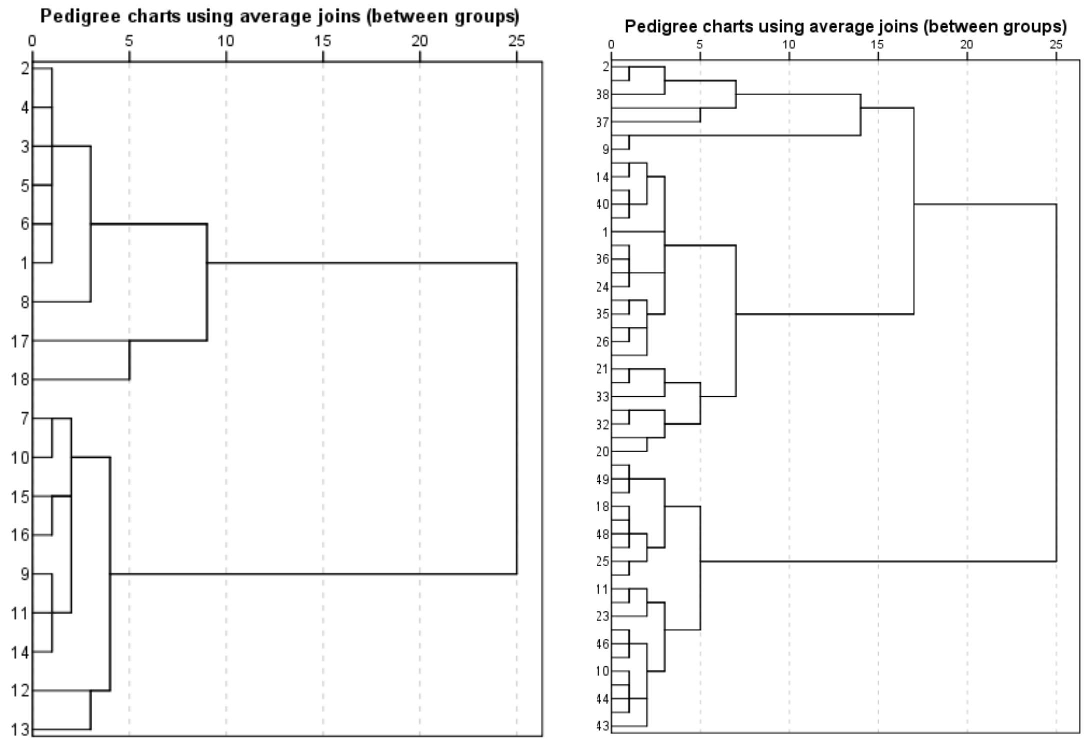

Figure 5. Cluster dendrograms for high potassium glass (left) and lead-barium glass (right)

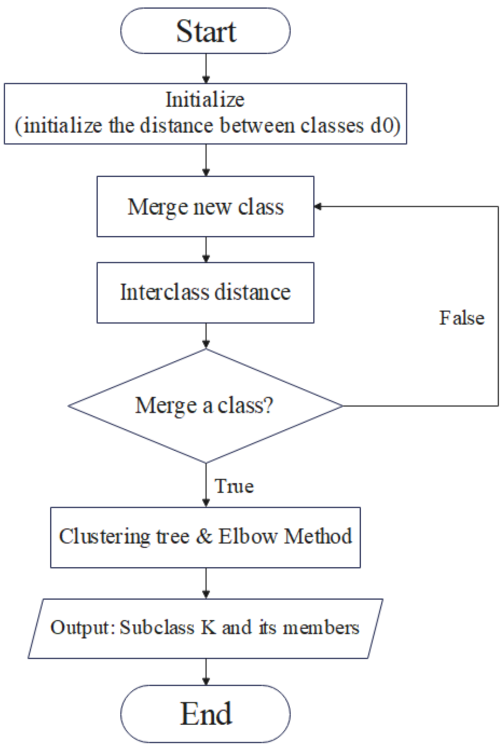

Figure 6. Flowchart of system clustering model building

Law of the Elbow:

The elbow method is a way of roughly estimating the optimal number of clusters from a graph. Let n samples be divided into K classes, each with at least two chemical components, and the Kth class is denoted by C_k. Define the degree of distortion of each class j_k as the sum of the squares of the distances between the centre of gravity of the class u_k and the positions of its internal members $x_i \cdot \sum |x_i - u_k|^2$, then the total degree of distortion of all classes, also known as the coefficient of aggregation, has a value $J = \sum_{k=1}^{K} j_k$.

A line graph of the aggregation coefficient J versus the number of clustering categories K is plotted as shown in Figure 7:

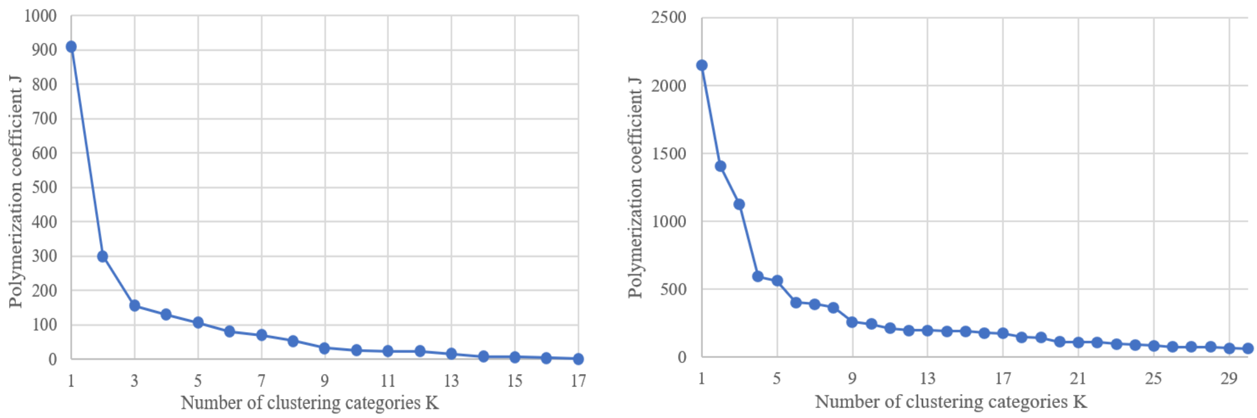

Figure 7. Line graph of aggregation coefficient of high potassium glass (left) and line graph of aggregation coefficient lead-barium glass (right)

The graph is shaped like an elbow and the gently sloping points are shaped like elbows, so the analytical method for roughly estimating the optimal number of clusters based on the elbow graph and elbow points can be called the elbow rule.

As shown in the figure, in the case of high potassium glass, when the number of categories K = 2, the decreasing trend of the fold tends to level off from the next point in the category, i.e. Total distortion level J plummets before this, and after this total distortion level J at the next point decreases gently, so that the number of categories K = 2 is the optimal number of clusters. Similarly, the optimal number of clusters K = 3 for lead barium glass.

3.2.2 Solving the model

The corresponding samples for each subclass are found in the clustering dendrogram based on the number of subclasses determined by the elbow rule.

The subclasses for high potassium glass and lead-barium glass are divided as shown in Figure 8:

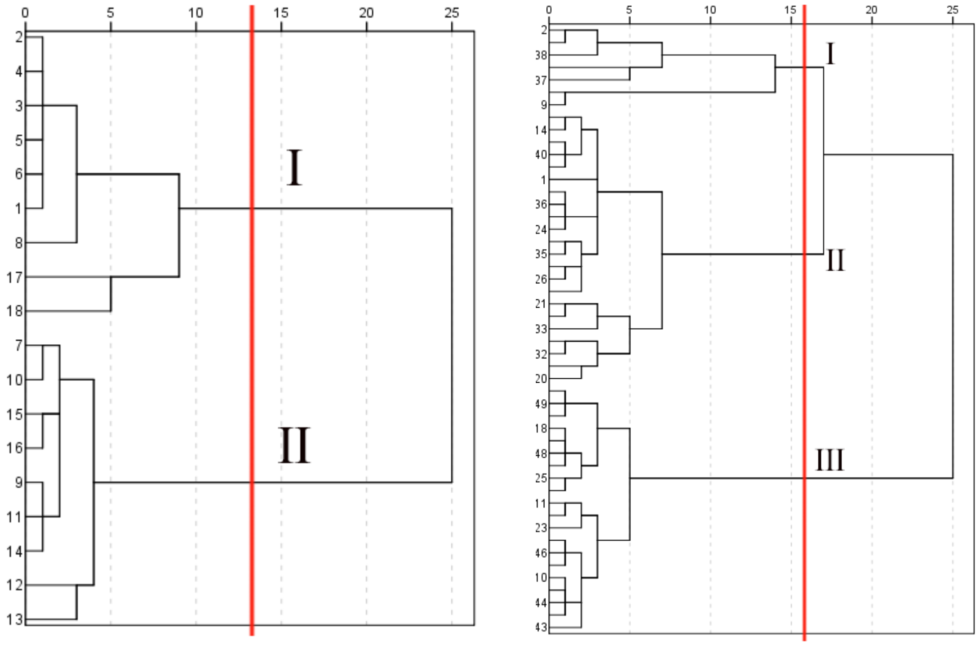

Figure 8. Subclasses of high potassium glass (left and lead-barium glass (right)

From the research documentation [10] it is clear that: glass can be divided into the following five categories, as shown in Table 5:

Table 5. Classification of glass

Class number	Category name	Category characteristics
G1	K_2O-CaO(~10wt%)-SiO_2	Contains high levels of CaO and K_2O
G2	K_2O-SiO_2	Significantly lower CaO content and increased (SiO_2+K_2O) content compared to G1
G3	PbO-BaO-SiO_2	High SiO_2 content in lead barium glass
G4	PbO(~25wt%)-BaO-SiO_2	It clearly shows a decrease in SiO_2 content and an increase in PbO content
G5	CaO-PbO(~40wt%)-BaO-SiO_2	The PbO content is already significantly higher than the SiO_2 content

The high potassium glass is divided into subclasses I and II, where subclass I has samples 1, 2, 3, 4, 5, 6, 8, 17 and 18; subclass II has samples 7, 9, 10, 11, 12, 13, 14, 15 and 16; the value is the number of rows of the filtered data table, and the correct classification result is obtained by querying the corresponding artefact number.

A descriptive statistical analysis of the content of the key chemical components of the two subclasses of high potassium glass is shown in Figure 9.

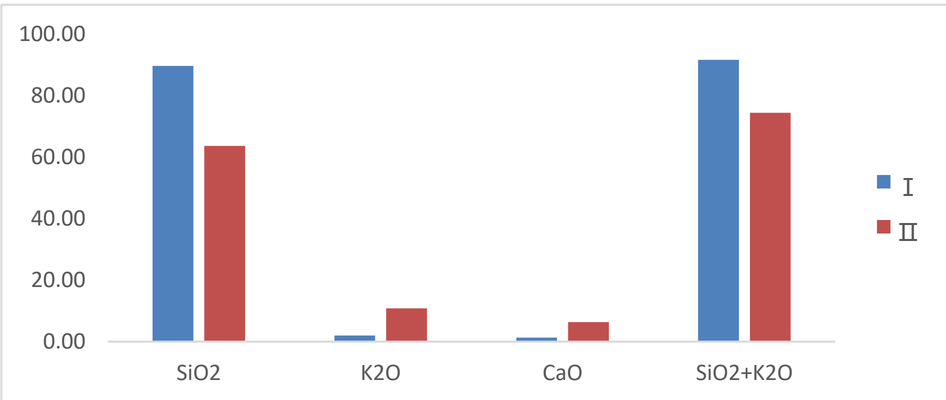

Figure 9. Comparative analysis of the content of the key chemical components of the two subclasses of high potassium glass

As can be seen from the above diagram, subclass I conforms to the category characteristics of G2 and subclass II conforms to the category characteristics of G1. Therefore, a query of the corresponding artefact numbers leads to the classification results in Figure 10:

SubclassI (K_2O-SiO_2)

- Artifact number： 3, 7, 9, 10, 12, 22, 27, 18, 21

SubclassII (K_2O-CaO (~10wt%)-SiO_2)

- Artifact number： 1, 3, 4, 5, 6, 13, 14, 16

Figure 10. Subclassification results for high potassium glass

The barium lead glass is divided into subclass I, subclass II and subclass III, where subclass I has samples 2, 3, 4, 8, 9, 37 and 38; subclass II has samples 12, 14, 39, 40, 13, 1, 5, 36, 28, 24, 34, 35, 17, 26, 30, 21, 29, 33, 15, 32, 16 and 20; subclass III has samples 7, 49, 27, 18, 19, 48, 6, 25, 47, 11, 22, 23, 31, 46, 45, 10, 42, 44, 41 and 43; Similarly, the corresponding artefact numbers are queried to obtain the correct classification results.

A descriptive statistical analysis of the content of the key chemical components of the Subcategory III of lead barium glass is shown in Figure 11:

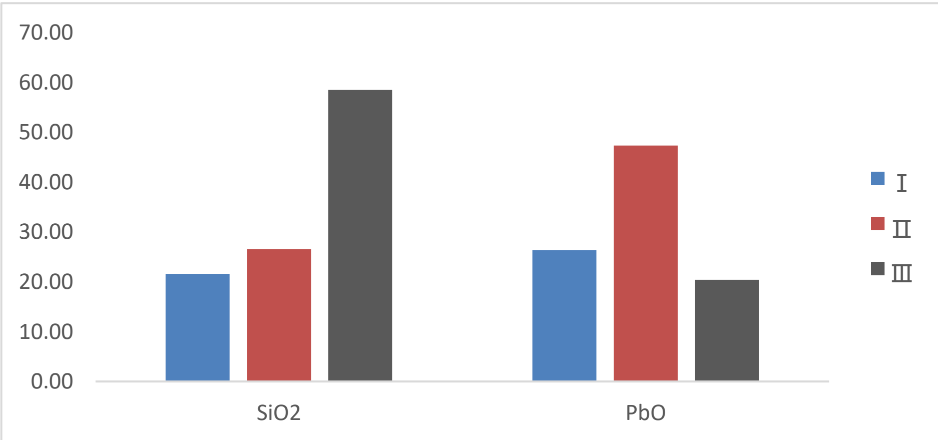

Figure 11. Comparative analysis of the content of the key chemical components of the Subcategory III of lead and barium glass

As can be seen from the above figure, subclass I conforms to the category characteristics of G4, subclass II conforms to the category characteristics of G5, and subclass III conforms to the category characteristics of G3. Therefore, a query of the corresponding artefact numbers leads to the classification results in Figure 12:

SubclassI (PbO (~25wt%)-BaO-SiO_2)

- Artifact number： 8, 11, 26 , 20, 24

SubclassII (CaO-PbO (~40wt%)-BaO-SiO_2)

- Artifact number： 2, 19, 34, 36, 38, 39, 40, 41, 43, 49, 50, 51, 52, 54, 56, 57, 58, 30

SubclassIII (PbO-BaO-SiO_2)

- Artifact number： 23, 25, 28, 29, 42, 42, 44, 48, 49, 50, 53, 31, 32, 33, 35, 37, 45, 46, 47, 55

Figure 12. Results of subclassification of lead barium glass

4. Conclusions

In this article, unknown glass samples of known chemical composition are reasonably well classified by using FLDA and systematic cluster analysis, and well further sub-classified. Glass objects were one of the earliest and most important commodities along the ancient Chinese and foreign Silk Roads, and the study of ancient glass sub-classifications can provide a better understanding of ancient glass-making techniques, sources of raw materials and production patterns, and uncover the origins of ancient glass, leading to a better understanding of ancient civilisations. At the same time, the combination of modern methods for modelling and scientifically classifying glass has given a boost to the development and progress of the modern glass industry.

References

[1] Fuxi Gan, Hong Li, Lisong Hou. Development History Of Ancient Chinese Glass Technology [M]. World Scientific Publishing Company: 2021-02-04.

[2] Xu Siwen, Qiao Baotong, Yang Yimin. The rise of the Maritime Silk Road about 2000 years ago: Insights from Indo-Pacific beads in Nanyang, Central China [J]. Journal of Archaeological Science: Reports, 2022, 42.

[3] Siu Ieong, Henderson Julian, Qin Dashu, Ding Yu, Cui Jianfeng, Ma Hongjiao. New light on plant ash glass found in Africa: Evidence for Indian Ocean Silk Road trade using major, minor, trace element and lead isotope analysis of glass from the 15th-16th century AD from Malindi and Mambrui, Kenya. [J]. PloS one, 2020, 15(8).

[4] J. Henderson, J. An, H. Ma. The Archaeometry and Archaeology of Ancient Chinese Glass: a Review [J]. Archaeometry, 2018, 60 (1).

[5] Th. Rehren, M. Brüggler. The Late Antique glass furnaces in the Hambach Forest were working glass - not making it [J]. Journal of Archaeological Science: Reports, 2020, 29(C).

[6] Mohcene Bessaoudi, Mebarka Belahcene, Abdelmalik Ouamane, Ammar Chouchane, Salah Bourennane. Multilinear Enhanced Fisher Discriminant Analysis for robust multimodal 2D and 3D face verification [J]. Applied Intelligence, 2019, 49(4).

[7] Pasadas Dario J., Barzegar Mohsen, Ribeiro Artur L., Ramos Helena G.. Guided waves based debonding classification in lap-joints using modified Fisher discriminant criterion [J]. NDT and E International, 2023, 137.

[8] Koren Oded, Shamalov Aviel, Perel Nir. Small Files Problem Resolution via Hierarchical Clustering Algorithm. [J]. Big data, 2023.

[9] Salgado-Hernández J E, Vyas Manan. Non-linear correlation analysis in financial markets using hierarchical clustering[J]. Journal of Physics Communications, 2023, 7(5).

[10] Fu Xiufeng, Gan Fuxi. A study on the composition of a group of ancient glass from southern and southwestern China based on multivariate statistical analysis [J]. DOI:10.16334/j. cnki. cn31-1652/k. 2006. 04. 002.