Volume 9, Issue 9, pp. 61-63, 2025.
International Journal of Scientific Engineering and Science
ISSN (Online): 2456-7361

A Study on the Classification of Ancient Glass Artifacts Based on Statistical Models

Xunru Yin[1], Shan Jiang[1], Xueyong Chen[1], Dan Sun[1], Weiyan Liu[1]
[1] School of Mathematics and Statistics, Taishan University, Tai'an 271000, China

Abstract**—***This paper analyzes the composition of ancient glass and establishes a classification model. First, the data was cleaned and transformed. Then, contingency tables were used to perform chi-square tests and Fisher's exact tests to identify the factors influencing surface weathering of glass. Finally, two clustering methods—hierarchical clustering and k-means clustering—were applied to classify glass types. By comparing the results of the two clustering methods, the hierarchical clustering model based on the sum of squared deviations was ultimately selected.*

Keywords**—** *Statistical model, Cluster analysis, Correlation analysis.*

I. INTRODUCTION

During the weathering process of ancient glass, internal elements exchange extensively with external environmental elements, leading to changes in the proportions of chemical components in the glass. This, in turn, affects the classification of glass types. Many scholars have studied this issue. Chen et al. [1] constructed a classification model for ancient glass based on PCA-BP neural networks, where the principal components obtained from principal component analysis (PCA) were used as inputs to a backpropagation (BP) neural network. Gao et al. [2] applied grey relational analysis to examine the correlations and differences among chemical components, and then established a support vector machine (SVM) model for the classification of ancient glass artifacts. Wang et al. [3] analyzed the composition of ancient glass artifacts using methods such as grey relational analysis, fuzzy clustering, Fisher's discriminant analysis, and correlation analysis. Lou et al. [4] explored classification rules using supervised feature selection methods and further investigated subclassification using unsupervised feature selection methods. Yang et al. [5] built a support vector regression (SVR) model based on training data to predict the chemical composition of glass artifacts prior to weathering. In addition to these approaches, several other analytical methods have also been proposed [6,7].

Based on the collected data, this paper mainly studies the factors influencing the surface weathering of glass, and through cumulative variance contribution, selects the component variables for category division, ultimately proposing classification models for different types of glass.

II. DATA PROCESSING AND ANALYSIS

The original dataset collected information on each glass artifact, including pattern, type, color, and surface weathering. The chemical composition of each ancient glass sample belongs to compositional data, meaning that the proportions of all chemical components for a given sample should sum to 100%. However, in practice, the sums of these chemical components were not consistent. After handling missing values and zero values, the chemical composition data were normalized. The purpose of normalization was to transform these compositional data into constant-sum data. However, due to the inherent constraints of constant-sum data, multicollinearity exists among the component variables, making them unsuitable for conventional statistical analysis. To address this issue, this paper applies the Centered Log-Ratio (CLR) transformation. For a sample with observed data $(\boldsymbol{x_1}, \boldsymbol{x_2}, \cdots, \boldsymbol{x_n})$, the transformation formula is:

$$\boldsymbol{CLR_{data}} = \ln\frac{x_i}{g(x)}, \quad \boldsymbol{g(x)} = (\boldsymbol{x_1 x_2} \cdots \boldsymbol{x_n})^{1/n}$$

The numbers of glass type classifications, pattern categories, and color categories were separately counted under both weathered and non-weathered conditions, and contingency table analysis was conducted. The results are shown in Table I.

TABLE I. Contingency table analysis for surface weathering

	Surface Weathering	
	p-value	Test Method
Type	0.0087	Chi-square test
Pattern	0.8361	Fisher's exact test
Color	0.4713	Fisher's exact test

At a significance level of 0.05, the results in Table I lead to the following conclusion: surface weathering is related to glass type, but unrelated to pattern and color. Based on this conclusion, it can be inferred that in clustering analysis of ancient glass, only glass type needs to be considered as the influencing factor.

III. CLASSIFICATION OF GLASS TYPES

For the convenience of subsequent statistical analysis, the transformed data was divided into two datasets: high-potassium glass and lead-barium glass. Since each glass artifact contains 14 chemical components, there may be correlations among these component variables. For example, in the correlation matrix of high-potassium glass components, the largest correlation coefficient is -0.84, which is the correlation between CuO and SnO_2. Therefore, it would be unreasonable to use all chemical components for category division.

Here, we use the cumulative variance contribution ratio to filter out a subset of component variables. When the cumulative variance contribution ratio exceeds 80%, it can be

International Journal of Scientific Engineering and Science
Volume 9, Issue 9, pp. 61-63, 2025. ISSN (Online): 2456-7361

considered that most of the features of the original data are retained. Based on this principle, the selected chemical components for high-potassium glass are: Na_2O, CaO, BaO, K_2O, PbO, SnO_2, MgO, Fe_2O_3, and SO_2. For lead-barium glass, the selected components are: Na_2O, Fe_2O_3, P_2O_5, MgO, CuO, CaO, and K_2O. These chemical components serve as the data variables for glass classification.

Currently, physicochemical methods generally divide high-potassium glass into two categories and lead-barium glass into three categories. Following this approach, hierarchical clustering was applied to subdivide high-potassium glass. Using both the shortest distance method and the sum of squared deviations method, the clustering results were obtained, as shown in Figures 1 and 2.

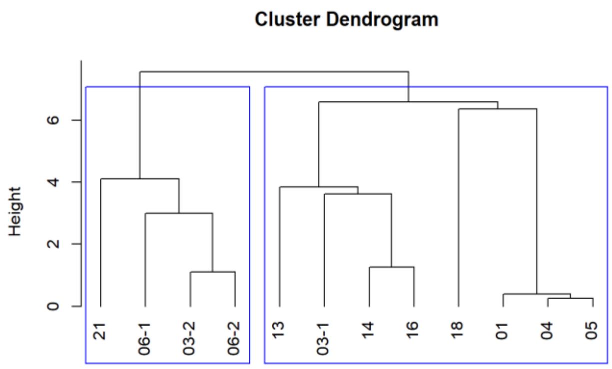

Fig. 1. Clustering Based on Sum of Squared Deviations

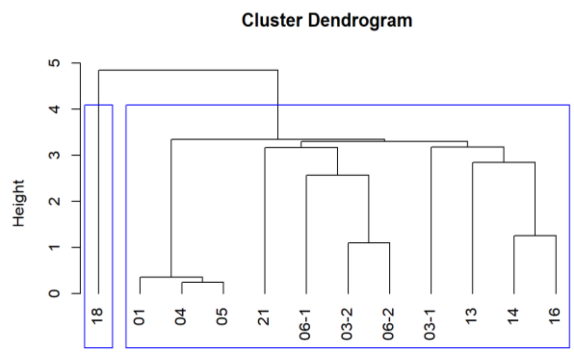

Fig. 2. Single-Linkage Clustering

To further verify the rationality of the classification results, the K-means clustering method was applied. The clustering results for high-potassium glass are shown in Table II.

TABLE II. K-means Clustering Results for High-Potassium Glass

Artifact	1	03-1	4	5	13	14
Category	1	1	1	1	1	1
Artifact	16	18	03-2	06-1	06-2	21
Category	1	1	2	2	2	2

It can be observed that the K-means clustering results are identical to those of the hierarchical clustering method using the sum of squared deviations.

Next, clustering was performed for lead-barium glass. Based on the hierarchical clustering criterion of the sum of squared deviations, the results are shown in Figure 3.

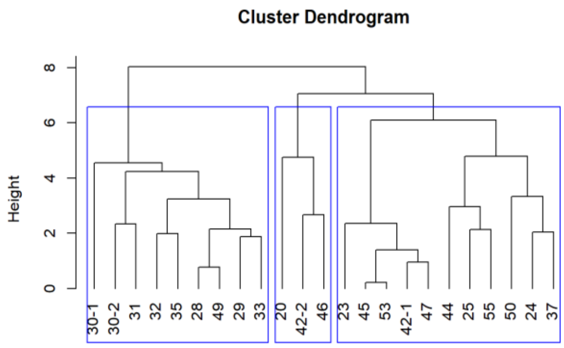

Fig. 3. Hierarchical Clustering of Lead-Barium Glass

K-means clustering was also applied to examine the classification results, as shown in Table III.

TABLE III. K-means Clustering Results for Lead-Barium Glass

Artifact	20	24	33	37	46	50
Category	1	1	1	1	1	1
Artifact	23	25	29	42-1	42-2	44
Category	2	2	2	2	2	2
Artifact	45	47	53	55	28	30-1
Category	2	2	2	2	3	3
Artifact	30-2	31	32	35	49	
Category	3	3	3	3	3	

It can be seen that the results of these two clustering models differ. As an example, one of the categories was compared, as shown in Table IV.

TABLE IV. Comparison Between Two Clustering Models

System Clustering	28 *29* 30-1 30-2 31 32 *33* 35 49
K-means Clustering	28 30-1 30-2 31 32 35 49

The differences in the classification results are minor, with only two glass samples being classified differently. Considering the clustering results of high-potassium glass, the hierarchical clustering method based on the sum of squared deviations was ultimately adopted.

IV. CONCLUSION

The statistical analysis methods employed in this paper, such as the chi-square test, Fisher's exact test, and cluster analysis, demonstrate strong practicality and general applicability, providing valuable guidance in real-world contexts. When performing supervised learning for sample classification, models such as the Naive Bayes classifier can also be considered. Therefore, when analyzing chemical products in practical situations, such statistical methods can be effectively applied to solve related problems.

ACKNOWLEDGMENT

This work was supported by grants from the 16th Teaching Reform Project of Taishan University (Grant Nos. JG202424, JG202407, and JG202425).

REFERENCES

[1] Chen, S.H., Wang, Y.K., Li, X.B., et al. "An Ancient Glass Classification Model Based on PCA-BP Neural Network," *Journal of Guangxi University (Natural Science Edition)*, vol. 49, issue 05, pp.1088-1098, 2024.

[2] Gao, G.Y., Wang, Q.Y. "Classification of Ancient Glass Products Based on Support Vector Machine (SVM)," *Journal of Gannan Normal University*, vol. 45, issue 03, pp. 19-22, 2024.

[3] Wang, C., Han, Y., Yang, L.L., et al. "Study of Glass Artifact Composition Based on Fisher's Discriminant Method," *Journal of Beijing Institute of Petrochemical Technology*, vol. 31, issue 04, pp. 49-55, 2023.

[4] Lou, Y., Dou, L., Zhuo, C.Y., et al. "Research on Ancient Glass Composition Analysis and Subclassification Methods," *Mathematical Modelling and its Applications*, vol. 12, issue 04, pp. 73-83, 2023.

[5] Yang, L., Wei, T., Zong, Y.H. "Mathematical Model for the Composition Analysis and Identification of Ancient Glass Products," *Journal of Beijing Institute of Printing*, vol. 31, issue 09, pp. 61-66, 2023.

[6] Xu, H., Hu, S., Yao, X.M., et al. "Study of Glass Artifact Composition Using K-Means Clustering and Grey Relational Analysis," *Journal of Xinjiang Normal University (Natural Science Edition)*, vol. 42, issue 03, pp. 66-73, 2023.

[7] Chen, X.J., Ye, Z., Shi, H.W. "Classification and Identification of Ancient Glass Artifacts Based on K-Means Clustering and SVM Algorithm," *Journal of Harbin Normal University (Natural Science Edition)*, vol. 39, issue 04, pp. 70-79, 2023.