Glass classification and identification based on multiple clustering model

WenJun Ding [1], Yichen Bao [2], Yitong Tao [1, *]

[1] School of Control and Computer Engineering, North China Electric Power University, BaoDing, HeBei, 071000

[2] School of International Education, North China Electric Power University, BaoDing, HeBei, 071000

* Corresponding author: TYT@ncepu.edu.cn

Abstract. Ancient glass is extremely susceptible to weathering by the influence of the buried environment, resulting in a change in the proportion of glass components. In order to classify different kinds of glass, this paper uses k-means++ clustering method to analyze the classification law, and summarizes the major and minor chemical composition data of the cluster center, so as to establish a glass classification model. In order to identify unknown types of glass, this paper mixes the data of eight groups of unknown types of glass into the known data, outputs the types of eight groups of glass through cluster analysis, calculates the center distance of the lead barium glass group to the subclass of the type in these eight groups, and then divides the subclass, and at the same time divides the subclass of the high potassium glass in the eight groups.

Keywords: k-means++ clustering, clustering, sensitivity analysis.

1. Introduction

As a famous overland trade route connecting Asia and Eurasia, the Silk Road has been an important economic and cultural bridge between East and West since ancient times [1-2]. Today, when diverse cultures coexist, the Silk Road has also been endowed with richer connotations [3].

Glass was one of the items transported to China by the Silk Road, and its own characteristics also made glass a precious material evidence of the early trade of the Silk Road [4]. On the basis of Western glass craftsmanship, ancient Chinese craftsmen chose Chinese local materials to innovate and make Chinese glass, so although China's ancient glass and foreign glass are more similar in appearance, they are different in chemical composition [5-6].

The main raw material for making glass is quartz sand, and its main chemical component is silica (SiO2). In order to facilitate firing, flux needs to be added to reduce the melting temperature, and the main chemical composition of the glass will also change depending on the type of flux. For example, lead barium glass with lead ore as a flux has a high content of lead oxide (PbO) and barium oxide (BaO); The flux used in potassium glass is a substance with a high potassium content, such as grass and wood ash [7-8].

Because the manufacturing process is not perfect, ancient glass is extremely easy to weather under the influence of the buried environment. In the process of weathering, the elements inside the glass are exchanged with the elements in the environment, resulting in a change in the proportion of the components of the glass, which in turn affects the correct judgment of the glass category in today's research. Therefore, it is of great significance for archaeological research to carry out appropriate modeling and analysis of the relationship between the composition category of glass products and weathering [9-10].

2. Establish a glass classification model

2.1. Analysis of classification law of high potassium glass and lead barium glass

Using the K-means++ clustering model, we clustered 67 groups of glass artifacts in clusters of 3. The flowchart is shown in Figure 1 below:

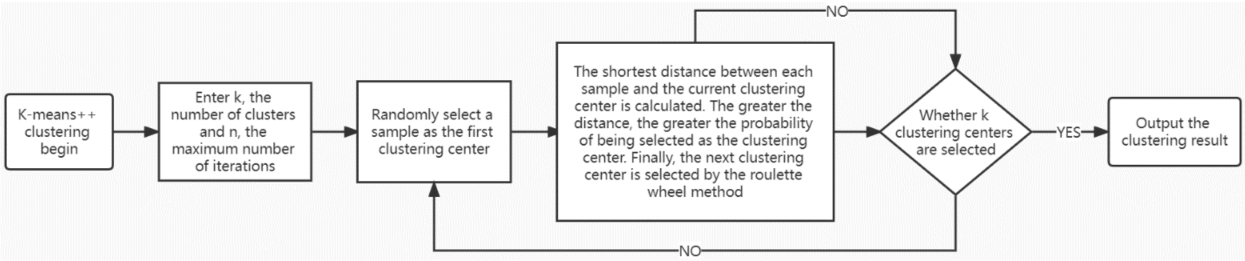

Figure 1. Clustering flowchart

The 67 sample points were divided into 3 clusters. Analysis of statistical data shows that cluster 1 is mainly the weathered part of lead-barium glass, cluster 2 is mainly the unweathered part of lead-barium glass, and cluster 3 is mainly high-potassium glass.

Find the three cluster centers, as shown in Table 1 below:

Table 1. summarizes the data clustering center results table

	clustering		
	1(Barium weathering)	2(Lead barium not weathered)	3(High potassium)
Silica (SiO_2)	25.735	57.875	77.490
Sodium oxide (Na_2O)	0.178	2.023	0.445
Potassium oxide (K_2O)	0.166	0.196	6.189
Calcium oxide (CaO)	2.837	1.187	3.746
Magnesium oxide (MgO)	0.628	0.700	0.811
Alumina (Al_2O_3)	2.709	5.305	4.992
Iron oxides (Fe_2O_3)	0.702	0.675	1.327
Copper oxide (CuO)	2.469	1.251	2.099
Lead oxide (PbO)	44.964	20.559	1.114
Barium oxide (BaO)	12.727	8.496	0.569
Phosphorus pentoxide (P_2O_5)	5.106	1.234	0.996
Strontium oxide (SrO)	0.455	0.238	0.027
Tin oxide (SnO_2)	0.048	0.078	0.128
Sulfur dioxide (SO_2)	1.279	0.183	0.066

Analyze the proportions of each chemical component in the cluster center of the three data.

SiO_2 is the main chemical component of glass, according to the classification law of the chemical component content, it is judged that the glass cultural relics with high SiO_2 content are high potassium glass, the glass cultural relics with moderate SiO_2 content are the unweathered parts of lead barium glass, and the glass cultural relics with low SiO_2 content are the weathered parts of lead barium glass.

K_2O, as a chemical component of high-potassium glass, has a high content in high-potassium glass and almost zero in lead-barium glass.

PbO and BaO, as secondary chemical components of lead barium glass, have higher content in lead barium glass, higher in the weathered part of lead barium glass than in the unweathered part, and almost zero in high potassium glass.

Therefore, according to the comprehensive analysis of chemical composition of lead-barium glass and the classification law of high potassium glass, the glass cultural relics are classified as shown in Table 2 as follows:

Table 2. Basis for classifying glass artifacts

SiO2	BaO	PbO	Glass composition
high	low	low	High potassium glass
middle	middle	middle	Unweathered part of lead barium glass
low	high	high	Weathering part of lead barium glass

Select obvious indicators (SiO2, K2O, PbO) for visualization and plot for example Figure 2 and Figure 3 show a three-dimensional scatterplot.

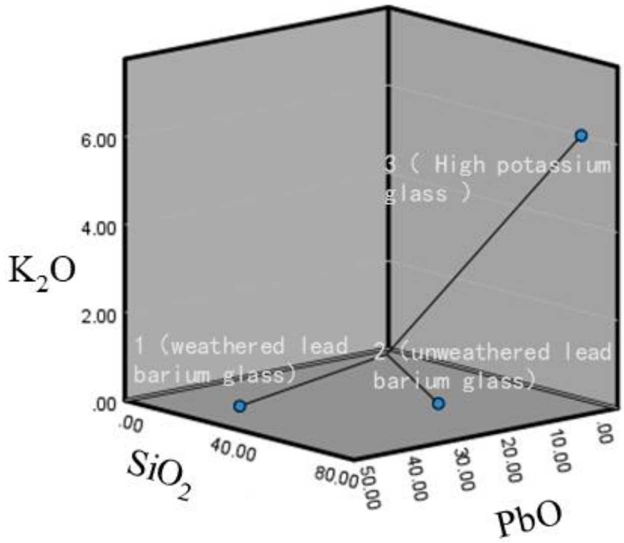

Figure 2. Visualization 3D plot of characteristic chemical composition clustering results

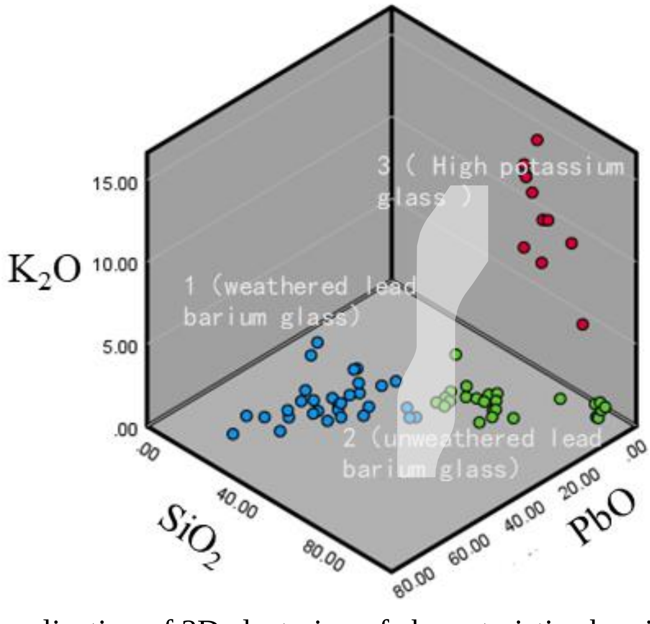

Figure 3. 3D visualization of 3D clustering of characteristic chemical components

The appropriate chemical composition is subclassified for each category

Through the overall cluster analysis, the lead-barium glass is divided into two categories, one is the unweathered part of the lead-barium glass, and the other is the differentiated part of the lead-barium glass. Then, the lead barium glass is clustered separately, and k=2 is set to obtain the clustering result and clustering center.

According to the chemical composition of lead barium glass, it is divided into two subcategories: one is low silicon high phosphorus lead barium glass (cluster center 1), and the other is high silicon low phosphorus lead barium glass (cluster center 2).

Then the subclass analysis of high potassium glass can be carried out, and the high potassium glass can be divided into one class through overall cluster analysis, which does not have obvious characteristics. The high-potassium glass is clustered separately, set k=2, and the clustering results and clustering centers are shown in Table 3 and Table 4 below:

Table 3. High potassium glass clustering results

Item number, type and	clustering	distance	Item number, type and	clustering	distance
Surface weathering	1	7.295	Surface weathering	2	4.866
01High potassium; No weathering	1	5.06	03High potassium; No weathering - part 1	2	3.89
03High potassium; No weathering - part 2	1	4.299	07High potassium weathering	2	5.442
04High potassium; No weathering	1	2.31	09High potassium weathering	2	7.207
05 High potassium; No weathering	1	9.545	10 high potassium; weathering	2	4.809
06 High potassium; No weathering - part 1	1	7.951	12 high potassium; weathering	2	11.981
06 High potassium; No weathering - part 2	1	6.548	18 high potassium; No weathering	2	13.978
13 High potassium without weathering	1	5.582	21 high potassium; No weathering	2	2.775
14 high potassium; No weathering	1	5.682	22 high potassium; weathering	2	4.138

Table 4. High potassium glass clustering results

	1(High potassium without weathering)	2(High potassium weathering)
Silica (SiO2)	64.9	90.29
Sodium oxide (Na2O)	0.94	0
Potassium oxide (K2O)	11	2.016
Calcium oxide (CaO)	6.5	1.337
Magnesium oxide (MgO)	1.16	0.444
Alumina (Al2O3)	7.49	2.79
Iron oxides (Fe2O3)	2.36	0.445
Copper oxide (CuO)	2.88	1.502
Lead oxide (PbO)	0.41	0.141
Barium oxide (BaO)	0.59	0.222
Phosphorus pentoxide (P2O5)	1.55	0.54
Strontium oxide (SrO)	0.05	0.008
Tin oxide (SnO2)	0	0.27
Sulfur dioxide (SO2)	0.14	0

From the above data, the division results of high potassium glass are obtained. According to the chemical composition of high-potassium glass, it is divided into two subcategories: one is high-potassium low-silicon high-potassium glass (cluster center 1), and the other is low-potassium high-silicon high-potassium glass (cluster center 2).

2.2. Plausibility and sensitivity analysis

(1) Plausibility analysis

In this problem, the k-means++ clustering method is used to classify lead-barium glass and high potassium glass according to the chemical composition of glass, and different clustering results are obtained according to the distance between different chemical components. Since the relationship between glass subclass classification and chemical composition is still in the research stage, there is a lack of sufficient professional knowledge for the rationality of classification results, and we judge the rationality of classification results with the help of prediction intensity method.

For the division, the method of cross-validation is used, the data is randomly divided into 4 equal parts, and one of them is taken out each time as the test set, and the rest is the training set to calculate the prediction strength. Repeat this step to get a total of 4 values, and take the average to get the average prediction intensity. The result is shown in the figure 4 to 5.

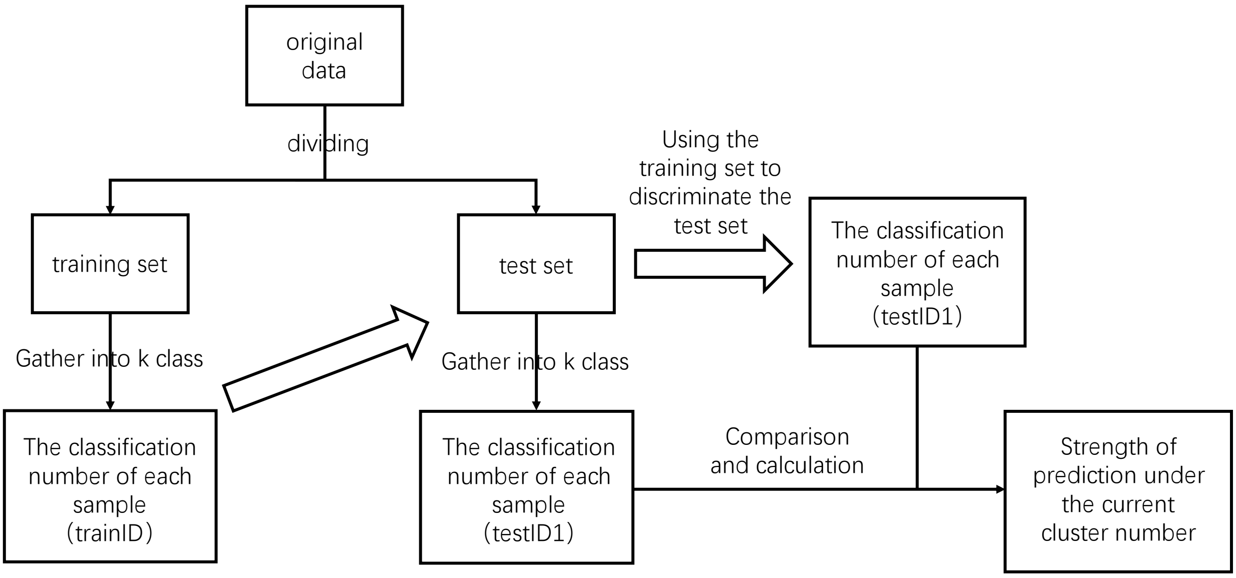

Figure 4. Rationality Analysis Mind Map

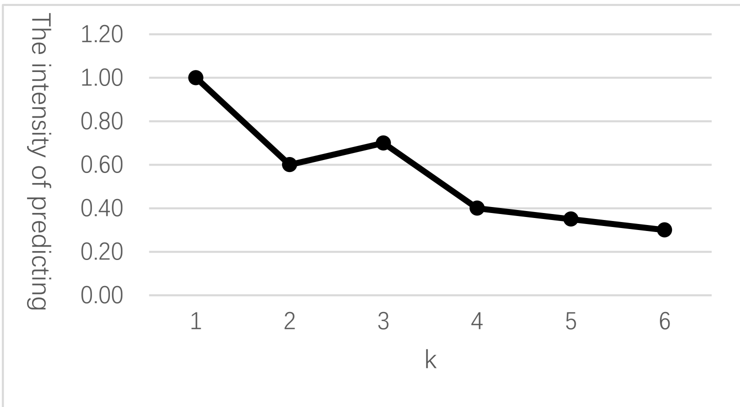

Figure 5. Predicted intensity

The predicted intensity is relatively small in multiple clusters, and the chemical composition data itself is very mixed, peaking at the predicted intensity k=2 and k=3. The chemical components are clustered k=2 and k=3 respectively, and the classes at k=3 have more significant differences while the classes at k=3 do not show significant differences, so the clustering at k=3 is more reasonable.

(2) Sensitivity analysis

Through the k-means++ model, all glass artifacts are divided into three categories, and then the sensitivity of a certain chemical component of the artifact to the cluster to which it belongs is studied, and the traditional sensitivity analysis is to change a parameter and control other parameters unchanged to study the sensitivity of a certain parameter. After data normalization, the total content of chemical components is 100%, changing the content of a chemical component, will inevitably lead to the change of the content of other chemical components, in order to reduce the impact, the amount of change according to the initial proportion of each chemical component affects the chemical ratio. Set the initial values for each chemical component as shown in Table 6:

Table 6. gives preliminary values to each chemical component

chemical composition	content /%	chemical composition	content /%
Silica (SiO2)	57	Copper oxide (CuO)	1.2
Sodium oxide (Na2O)	2	Lead oxide (PbO)	20
Potassium oxide (K2O)	0.2	Barium oxide (BaO)	10
Calcium oxide (CaO)	1	Phosphorus pentoxide (P2O5)	1
Magnesium oxide (MgO)	1	Strontium oxide (SrO)	0.3
Alumina (Al2O3)	5	Tin oxide (SnO2)	0.1
Iron oxides (Fe2O3)	1	Sulfur dioxide (SO2)	0.2

Analyze 14 chemical components in turn, and these variables fluctuate up and down by 5%, with 1% as the interval, calculate the distance to the three cluster centers under different amplitude fluctuations, and determine which cluster center to classify. A line chart plotting 14 sets of data is shown in Figure 5.

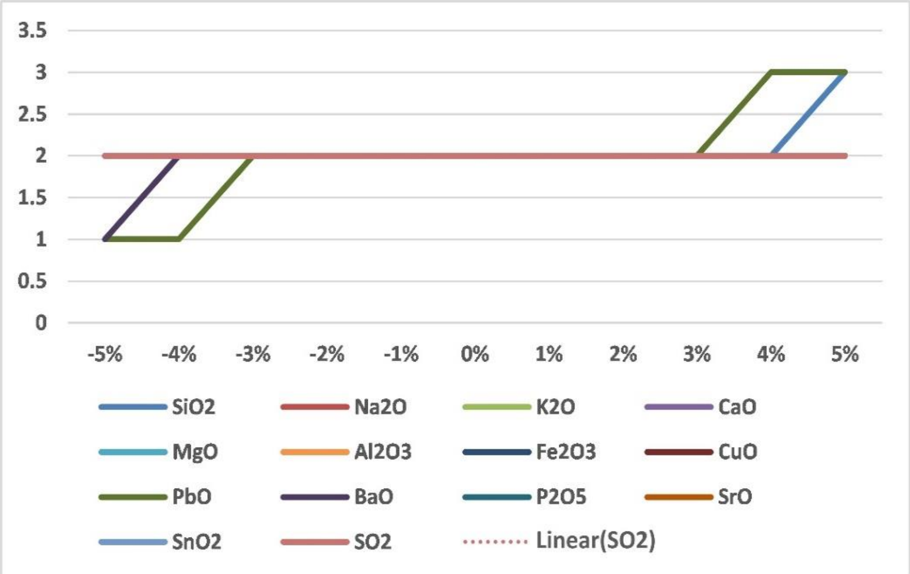

Figure 5. Sensitivity analysis line chart

3. Identify unknown types of glass

There are eight groups of glass cultural relics of unknown categories, in order to save the calculation process, 8 groups of glass cultural relics sampling points of unknown categories can be mixed into about 67 glass cultural relics sampling points of known categories for k-means++ clustering, set k=3, and the clustering results and clustering centers are shown in Table 7 and Table 8 below.

Table 7. Summary clustering results mixed with eight sets of unknown data (Partial display)

Number, type and surface weathering of the artifact	clustering	distance	Number, type and surface weathering of the artifact	clustering	Number, type and surface weathering of the artifact	distance
02Lead barium weathering	1	16.52	56 lead barium weathering	1	A4 is weather-free	2
08 Lead barium weathering	1	27.099	57 lead barium weathering	1	A5 Weathering	2
08 Lead barium weathering - severe weathering point	1	34.316	58 lead barium weathering	1	A8 No weathering	2
11 lead barium weathering	1	20.941	A2 Weathering	1	01 High potassium No weathering	3
19 lead barium weathering	1	8.671	A3 No weathering	1	03 High potassium No weathering - part 1	3
24 lead barium non-weathering	1	23.403	20 lead barium No weathering	2	03 High potassium No weathering - part 2	3
26 lead barium weathering	1	27.541	23 lead barium weathering - unweathered point	2	04High potassium No weathering	3
26 lead barium weathering - severe weathering point	1	38.725	25 lead barium weathering - unweathered point	2	05 High potassium No weathering	3
30 lead barium no weathering - part 1	1	11.872	28 lead barium weathering - unweathered point	2	06 High potassium No weathering - part 1	3

Table 8. Summary cluster centers mixed with eight sets of unknown data

	clustering		
	1	2	3
Silica (SiO2)	26.402	56.962	78.924
Sodium oxide (Na2O)	0.166	1.812	0.385
Potassium oxide (K2O)	0.201	0.233	5.452
Calcium oxide (CaO)	3.154	1.274	3.593
Magnesium oxide (MgO)	0.613	0.759	0.795
Alumina (Al2O3)	2.708	5.582	4.942
Iron oxides (Fe2O3)	0.893	0.914	1.268
Copper oxide (CuO)	2.311	1.564	2.042
Lead oxide (PbO)	44.486	20.434	0.962
Barium oxide (BaO)	12.036	8.352	0.492
Phosphorus pentoxide (P2O5)	5.349	1.527	0.924
Strontium oxide (SrO)	0.442	0.243	0.024
Tin oxide (SnO2)	0.045	0.089	0.110
Sulfur dioxide (SO2)	1.194	0.257	0.086

Analysis of statistical data shows that cluster 1 is mainly the weathered part of lead-barium glass, cluster 2 is mainly the unweathered part of lead-barium glass, and cluster 3 is mainly high-potassium glass. A1, A6, and A7 are cluster 3, so their types are judged to be high-potassium glass; A2, A3 is cluster 1, A4, A5, A8 is cluster 2, because the pretreatment has a certain error, to judge its subclass type is inaccurate, so it is preliminarily judged to be lead barium glass.

Our previous subcategories of lead barium glass were divided into low silicon high phosphorus lead barium glass and high silicon low phosphorus lead barium glass, and the subclass of high potassium glass was divided into high potassium low silicon high potassium glass and low potassium high silicon high potassium glass. Therefore, the coordinates of each component of the lead-barium glass clustering center in Table X can be used to calculate the distances from A1, A2, A3, A6, and A7 to the two cluster centers, and the points closer to a cluster center can be classified as this class. Similarly, using the coordinates of each component of the high-potassium glass cluster center in Table

X before, the distances from A4, A5, and A8 to the two cluster centers are calculated, and the points closer to a cluster center are classified as this class.

Our previous subcategories of lead barium glass were divided into low silicon high phosphorus lead barium glass and high silicon low phosphorus lead barium glass, and the subclass of high potassium glass was divided into high potassium low silicon high potassium glass and low potassium high silicon high potassium glass. Therefore, the coordinates of each component of the lead barium glass clustering center in the previous table X can be used to calculate the distance from A2, A3, A4, A5, and A8 to the two-lead barium glass subclass clustering centers respectively, and the points closer to a cluster center can be classified as this class. Similarly, using the coordinates of each component of the high-potassium glass clustering center in Table X before, the distances from A1, A6, and A7 to the two high-potassium glass subclass clustering centers are calculated, and the points closer to a cluster center are classified as this class.

The results of Table 9 are as follows:

Table 9. Subclass analysis of cultural objects of unknown categories

numbering	type	extent	Subclass
A1	High potassium glass	No weathering	High potassium low silicon high potassium glass
A2	Lead barium glass	weathering	Low silicon high phosphorus lead barium glass
A3	Lead barium glass	No weathering	High silicon low phosphorus lead barium glass
A4	Lead barium glass	No weathering	High silicon low phosphorus lead barium glass
A5	Lead barium glass	weathering	Low silicon high phosphorus lead barium glass
A6	High potassium glass	weathering	Low potassium high silicon high potassium glass
A7	High potassium glass	weathering	Low potassium high silicon high potassium glass
A8	Lead barium glass	No weathering	High silicon low phosphorus lead barium glass

4. Conclusions

In this paper, the K-means++ clustering method is used to analyze the classification law, and the major and minor chemical component data of the cluster center are summarized. The rationality of the model is verified by the prediction intensity method, and the sensitivity analysis model is innovatively optimized by controlling the clustering to which other parameters remain unchanged and fluctuate by 5% of a parameter, and the sensitivity analysis model is completed by drawing a line chart. Then, eight sets of data are mixed into the known data, and the types of eight groups of glass are output by cluster analysis, and the center distance of the lead barium glass group in these eight groups to the subclass to which they belong is calculated, and then the subclass is divided. At the same time, the eight groups of high-potassium glass were also subclassified.

References

[1] Wang Xiaofu. Journal of Tsinghua University (Philosophy and Social Sciences), 2020, 35 (04): 1 - 14+211. DOI: 10.13613/j.cnki.qhdz.002956.

[2] LIU Song, LV Liangbo, LI Qinghui, XIONG Zhaoming. Glass beads excavated from Han tombs in Lingnan and Sino-foreign exchanges on the Maritime Silk Road of the Han Dynasty [J]. Cultural Relics Conservation and Archaeological Science, 2019, 31 (04): 18 - 29. DOI: 10.16334/j.cnki.cn31-1652/k.2019.04.004.

[3] Cheng Yajuan. Research on Dunhuang Mural Painted Small Glassware: The Evolution of Buddhism in the Eastward Transmission of Glass on the Silk Road [J]. Journal of Nanjing University of the Arts(Art and Design), 2018 (05): 84 - 93+210.)

[4] Wang Feifeng. Glassware excavated from Baekje, Silla and Gaya areas and its related research with the Silk Road [J]. Journal of Yanbian University (Social Science Edition), 2017, 50 (02): 18 - 24. DOI: 10.16154/j.cnki.cn22-1025/c.2017.02.002.

[5] Wei Zhenguo. Shanxi on the Silk Road during the Wei, Jin, Southern and Northern Dynasties [D]. Shanxi University, 2015.

[6] Jiang Feifei. Preliminary study on glass products in the Liao Dynasty [D]. Inner Mongolia University, 2014.

[7] Lou Jianhong. Guangzhou and the Maritime Silk Road in the Han Dynasty: Exploring the Status and Role of Guangzhou in the Maritime Silk Road [J]. People's Forum, 2012 (02): 138 - 139. DOI: 10.16619/j.cnki.rmlt.2012.02.006.

[8] Lu Chi. The influence of the Silk Road on ancient Chinese glass art [J]. Decoration, 2007 (04): 42 - 45. DOI: 10.16272/j.cnki.cn11-1392/j.2007.04.014.

[9] ZHANG Yu. Steppe Silk Road——Khitan and the Western Regions[C]//. Collected Works of Archaeological and Cultural Research in Eastern Inner Mongolia., 1990: 114 - 122.

[10] Zhou Qingji. Glass input and "Maritime Silk Road"[J]. Research on the History of Maritime Communication, 1985 (01): 7 - 11.)