Wu *et al. Heritage Science* (2023) 11:160
https://doi.org/10.1186/s40494-023-00999-0

Heritage Science

RESEARCH Open Access

Identification of ancient glass categories based on distance discriminant analysis

Shuyu Wu[1†], Jingyang Zhong[2,4†], Hui Ye[3] and Xusheng Kang[1,4*]

Abstract

It is crucial for archaeological investigations to identify the category of cultural relics by analyzing their chemical composition. This study analyzed the chemical composition distribution of glass cultural relics and applied distance discriminant analysis methods to classify them into two categories. Through stepwise regression, four key feature factors (SiO_2, K_2O, PbO, and the presence of weathering on the artifact's surface) were selected from a total of 15 features, including surface weathering. Aside from using columnar table analysis to determine weathering on the surface of the artifact and correlations between categories, and using Spearman correlation coefficients to select key feature factors such as SiO_2, K_2O, PbO, BaO, *and* SrO from 14 total feature factors (excluding weathering on the surface), we established a Mahalanobis distance discriminant model to differentiate unknown glass artifacts. Results indicate that Spearman-Mahalanobis distance discrimination outperformed stepwise regression-Mahalanobis distance discrimination, with an overall accuracy of 99.10% for the former and 98.69% for the latter in identifying high-potassium glass or lead-barium glass.

Keywords Mahalanobis distance discriminant, Stepwise regression analysis, Glass category identification, Spearman correlation coefficient

Introduction

When glass is buried in soil or immersed in water, it is weathered due to the external environment and internal chemical composition [1, 2]. Ancient glass provides valuable physical evidence of early trade on the Silk Road [3], but it is easily weathered by the influence of the buried environment [4]. A large amount of exchange between its internal substances and environmental substances has led to a significant change in the ratio of its internal chemical composition, which in turn has led to a qualitative change in its chemical composition [5], seriously affecting the correct judgment of modern restorers on the category to which ancient glass belongs. Therefore, the timely detection of the internal chemical compositions of glass products and the accurate determination of the state of the glass belonging to, the conservation and restoration of ancient glass relics has positive significance.

Various techniques have been developed to detect the chemical composition of artifacts, including laser ablation inductively coupled plasma mass spectrometry (LA-ICPMS)[6], proton-excited X-ray fluorescence (PIXE) techniques [7–9], particle-induced gamma-ray emission (PIGE) techniques [8, 9], a combination of electron microprobe analysis (EPMA) and thermal ion mass spectrometry (TIMS) [10]. Among these, the PIXE technique is a non-destructive, multi-element quantitative nuclear technique [11] that has proven useful in scientific and technological archaeology [8, 9]. It is commonly applied

†Shuyu Wu and Jingyang Zhong are contributed equally to this work.

*Correspondence:
Xusheng Kang
kangxs@zucc.edu.cn
[1] School of Computer and Computing Science, Hangzhou City University, Huzhou Street, Hangzhou 310015, Zhejiang, China
[2] School of Business, Hangzhou City University, Huzhou Street, Hangzhou 310015, Zhejiang, China
[3] School of Information and Electrical Engineering, Hangzhou City University, Huzhou Street, Hangzhou 310015, Zhejiang, China
[4] Institute of Digital Finance, Hangzhou City University, Huzhou Street, Hangzhou 310015, Zhejiang, China

Table 1 Sample data characteristics

Glass artifacts	Glass types	SiO_2	Na_2O	K_2O	…	Surface weathering of cultural relics
01	0	69.33	0	9.99	…	0
02	1	36.28	0	1.05	…	1
Part 1 of 03	0	87.05	0	5.19	…	0
Part 2 of 03	0	61.71	0	12.37	…	0

to identify ancient glass by measuring the content of its internal chemical components. However, the chemical composition of glass is complex, and there are cases in which the sum of the component contents is lower or higher than non-percentage values, making it difficult to determine the category to which ancient glass belongs. For this reason, this paper proposes a method to establish a distance discrimination analysis model among chemical components to identify the category of ancient glass.

Discriminant analysis is a multivariate statistical method to identify the class to which a newly acquired sample belongs based on several quantifiable characteristics of the observed sample [12]. Distance discriminant analysis is a kind of discriminant analysis [13], which identifies a newly obtained sample's category by calculating the distance between samples. Distance discriminant analysis has been applied in slope stability [14], and the classification of balanced common spherical meteorites [15]; however, it has been less frequently used in identifying categories of cultural heritage. In this paper, we apply stepwise regression analysis and Spearman's correlation coefficient to select key feature factors based on the determined chemical composition of ancient glass [16, 17]. To obtain efficient category identification of ancient glass, we need to select the appropriate combination of feature factors and establish a distance discrimination model. This will ensure that the category to which the glass belongs can be accurately mapped.

Models and algorithms

Data sources

Chemical composition data of glass cultural relics, whether the surface of cultural relics is weathered or not, and the glass category data used in this paper are all from the 2022 Mathematical Modeling Competition for Chinese College Students (http://www.mcm.edu.cn/). There are 77 glass cultural relics samples (including 69 known category samples and 8 unknown category samples). In order to produce glass, quartz sand—with a high melting point—is used as a raw material. However, to ensure efficient production, various fluxes such as lead ore and grass ash are added to reduce the melting point. Glass samples can be classified into two types based on the cosolvents added during the refining process high-potassium glass and lead-barium glass. These types are quantified with values of 0 and 1, respectively. Each glass sample provided data on 15 characteristic factors, including 14 factors that represented the chemical composition percentages of glass (e.g.$SiO_2, Na_2O, K_2O, and CaO$) and 1 factor that represented whether the surface of a cultural artifact was weathered or not (with the values of 1 and 0 representing weathered and unweathered surfaces, respectively). During data detection, errors may occur due to improper manual operation, testing equipment, or environmental temperature. In order to achieve precise calculations, we exclude abnormal data and set the effective data set to be between 85 and 105% of the cumulative percentage of chemical components. Finally, two sets of anomalous data are excluded, and the remaining 75 glass sample data are used for the establishment and inspection of the category identification model, and 8 unknown glass sample data are used for the prediction of glass cultural relic samples. The processing results of some data are shown in Table 1. Complete data can be found in Additional file 1.

Stepwise regression analyses

Stepwise regression [18] is a method of factor screening, which progressively eliminates independent variables in the model that does not have significant predictors, which better avoids multicollinearity [19] among factors and throws off the heavy factor screening work (Additional file 1).

The full-factor model of the regression equation is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_i X_i + \varepsilon \tag{1}$$

where, the predictor $X_i(\mathrm{i} = 1, 2, \ldots, 15)$ is the percentage of composition of the glass artifacts, as shown in Table 2, the response variable Y is the category to which the glass artifact belongs, β_0 is a constant term, $\beta_i(\mathrm{i} = 1, 2, \ldots, 15)$ is the partial regression coefficient of the response variable Y on the predictor X_i, and ε is the regression residual.

According to the effect size value of each predictor on the response variable, the predictors with significant effects were included in the regression equation one by one from the largest to the smallest, and an F-test [20] was conducted to test the significance size, while the predictors with small significant effects in the current regression equation were excluded. The test statistics in the F-test are:

$$F = \frac{ESSP_k}{\frac{RSS}{n-k-1}} \sim F(1, n-k-1) \tag{2}$$

Table 2 Percentage of glass cultural relics

Components	Content ratio of each component
SiO_2	X_1
Na_2O	X_2
K_2O	X_3
CaO	X_4
MgO	X_5
Al_2O_3	X_6
Fe_2O_3	X_7
CuO	X_8
PbO	X_9
BaO	X_{10}
P_2O_5	X_{11}
SrO	X_{12}
SnO_2	X_{13}
SO_2	X_{14}
Whether the surface of the glass relic is weathered	X_{15}

Table 3 Percentage of glass cultural relics

	Unweathered	Weathering	Total
High-potassium glass	12	6	18
Lead-barium glass	20	29	49
Total	32	35	67

where, n is the glass artifact sample data; k is the number of predictors; $ESSP_k$ is the partial regression sum of squares of X_k, and RSS is the residual sum of squares.

Contingency table analysis

To investigate whether there is a correlation between whether the surface of glass artifacts is weathered and the type of glass artifacts, a 2×2 contingency table [21] analysis was implemented for glass type Y and whether the surface of artifacts is weathered X_{15}, as shown in Table 3.

The computing method of the test statistic is shown in the Eq. (3):

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\left(n_{ij} - n\widehat{p}_{ij}\right)^2}{n\widehat{p}_{ij}} \tag{3}$$

where, $\widehat{p}_{ij} = \frac{n_{i,}}{n} \times \frac{n_{,j}}{n}$, n is the overall capacity of the sample, $n_{i,}$ denotes the number of each glass category in the sample, n_{ij} denotes the number of each weathering condition in the sample, and $n_{,j}$ denotes the number of each weathering condition in each glass sample, such as n_{11} represents the number of unweathered glass in the high-potassium glass.

When the original hypothesis holds, Eq. (3) approximately obeys a chi-square distribution with degrees of freedom of $(2-1) \times (2-1) = 1$. For a given significance level α, a judgment can be made to accept or reject the original hypothesis that the two are independent based on the corresponding test statistic and the calculated test statistic.

Correlation analysis of Spearman coefficient

For correlation analysis between variable data, Pearson's correlation coefficient [22] or Spearman's correlation coefficient [23] is generally used. After testing, it was found that the data in this paper did not conform to a normal distribution, so Spearman's correlation coefficient was chosen for the correlation analysis of the glass data.

Assuming that X and Y denote two aggregates with the number of elements N, then the Spearman correlation coefficient ρ between X and Y is calculated by Eq. (4).

$$\rho = \frac{\sum_{i=1}^{N}(x_i - \overline{x})\left(y_i - \overline{y}\right)}{\sqrt{\sum_{i=1}^{N}(x_i - \overline{x})^2 \sum_{i=1}^{N}\left(y_i - \overline{y}\right)^2}} \tag{4}$$

where, x_i is the ratio of the ith chemical composition, y_i is the type of the artifact glass, $\overline{x}$ and $\overline{y}$ is the mean value of different chemical compositions and the mean value of the type, respectively.

The value of Spearman's correlation coefficient ranges from $[-1, 1]$, and the larger the absolute value of ρ, the stronger the correlation. When $\rho > 0$, the two groups of variables under discussion are positively correlated; when $\rho < 0$, the two groups of variables under discussion are negatively correlated.

In general, the significance test for the Spearman correlation coefficient uses a P value test when the sample size is greater than 30. Since the sample size in this study is much larger than 30, we perform the significance test of Spearman's correlation coefficient based on the size of the P value at a 0.05 significance level. We consider a correlation between variables to be significant when the P value is less than 0.05 [24].

Distance discriminant model predicts unknown sample categories

This paper uses the Mahalanobis distance [25] to calculate the shortest distance between each heritage glass sample point and the overall glass of the two categories. Assuming that high-potassium glass is set as the overall A and lead-barium glass is set as the overall B, the Mahalanobis distance from the unknown category as the sample point x of the heritage glass to each overall $M(A\text{or}B)$ is:

$$d_M(x) = \sqrt{(x-\mu)^T \Sigma^{-1}(x-\mu)} \tag{5}$$

where, μ is the sample mean and $\sum$ is the covariance matrix of multidimensional random variables. The discriminant function used in this paper is:

$$w(x) = {d_B}^2(x) - {d_A}^2(x) \tag{6}$$

In summary, the discriminant criterion for the sample x to be tested belongs to:

$$x \in \begin{cases} A, \ w(x) \geq 0 \\ B, \ w(x) < 0 \end{cases} \tag{7}$$

Results and discussion

Descriptive analyses

To better investigate the quantitative characteristics of the chemical composition content factors of glass artifacts, we investigate the concentration trends and dispersion of the chemical composition content factor data for a sample of 67 known categories of glass artifacts.

The concentration trend is assessed by two measures: the sample mean and the sample median. The median is not affected by extreme values, making it useful in joint analysis with the sample mean to determine the importance of chemical component content factors. Dispersion is measured by four indicators: sample standard deviation, coefficient of variation [26], skewness, and kurtosis. The coefficient of variation effectively eliminates the impact of data error caused by dimensionality. Therefore, a comprehensive analysis of all four indicators can better identify the degree of fluctuation in each chemical component content factor. Table 4 also shows the correlation analysis of the four main components.

Figures 1 and 2 are the index charts of the concentration trend and dispersion trend of high-potassium glass respectively. The graphs show that the concentration degree is relatively high in high-potassium glass, so it can be explained that this chemical composition has an important influence on the composition of high-potassium glass; In the weathered state, the contents of SiO_2, PbO, BaO, SrO and is close to 0, which shows that the content of these four components will change with the weathered degree of cultural relics.

The dispersion of chemical composition content in non-weathered samples is higher than in weathered samples. This suggests that the chemical substances in weathered high-potassium glass are more stable and uniform. However, for this type of glass, the fluctuations in the contents of PbO, BaO, SiO_2 differ significantly between weathered and non-weathered samples, reflecting the correlation between the weathering degree of

Table 4 Descriptive statistical patterns of heritage glass elements

Glass type	Surface weathering	Chemical compositions	Average%	Median%	Standard deviation	Coefficient of variation	Skewness	Kurtosis
High potassium glass	Unweathered	SiO_2	67.98	65.53	8.38	8.76	1.16	0.54
		K_2O	9.33	9.83	3.75	3.92	− 1.20	1.90
		PbO	0.41	0.16	0.56	0.59	1.37	0.42
		BaO	0.60	0	0.94	0.98	1.49	1.24
		SrO	0.04	0.02	0.05	0.05	0.57	− 1.45
	Weathering	SiO_2	93.96	93.51	1.58	1.73	0.85	− 0.39
		K_2O	0.54	0.55	0.41	0.45	− 0.54	− 1.91
		PbO	0	0	0	0		
		BaO	0	0	0	0		
		SrO	0	0	0	0		
Lead barium glass	Unweathered	SiO_2	54.66	54.61	11.57	11.83	− 0.37	− 0.54
		K_2O	0.22	0.15	0.30	0.31	2.88	10.06
		PbO	22.08	20.12	8.03	8.22	0.62	− 0.46
		BaO	9.00	8.99	5.70	5.83	1.80	3.76
		SrO	0.27	0.26	0.24	0.24	1.23	1.96
	Weathering	SiO_2	24.91	25.02	10.40	10.61	0.31	1.23
		K_2O	0.13	0	0.24	0.24	2.51	7.79
		PbO	43.31	44.06	11.99	12.23	− 0.03	0.23
		BaO	11.80	8.79	9.78	9.98	1.28	0.84
		SrO	0.42	0.43	0.26	0.26	0.54	0.78

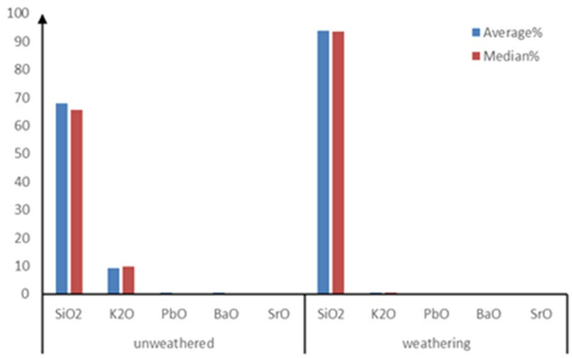

Fig. 1 Concentration trend index of high-potassium glass

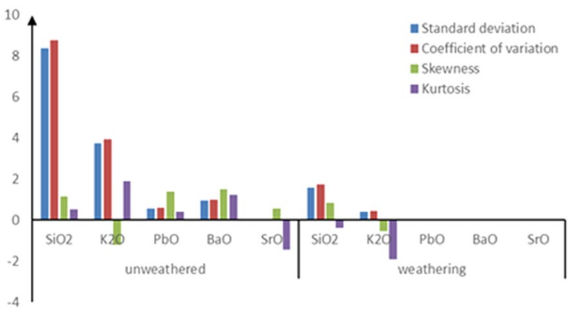

Fig. 2 Dispersion index of high-potassium glass

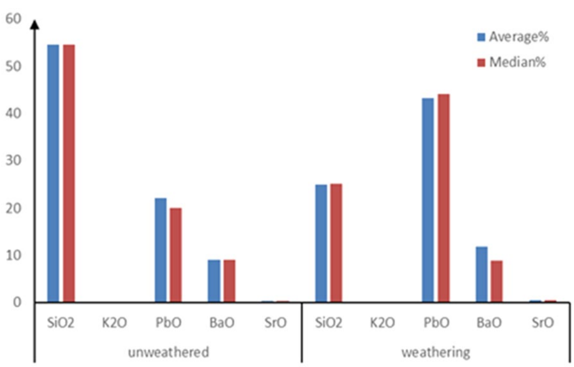

Fig. 3 Lead-barium glass concentration trend index

high-potassium glass and its internal chemical makeup, particularly with respect to PbO, BaO, SiO_2.

Figures 3 and 4 display the concentration and dispersion trends of lead-barium glass, respectively. The data reveals that the concentration of PbO and SiO_2 in lead-barium glass is high, with mean and median values exceeding 40%. These components play a crucial role in the composition of lead-barium glass. Conversely, the indicators for $SrOandK_2O$, which reflect the dispersion

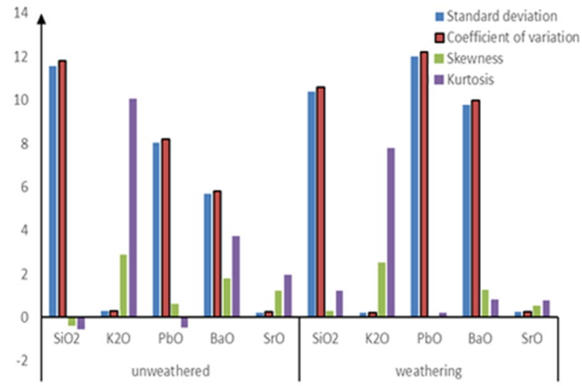

Fig. 4 Lead-barium glass dispersion index

degree, are relatively low, indicating that these chemical components are stable and exhibit minimal fluctuations.

Moreover, the overall dispersion of $PbO, BaO, andSiO_2$ is large, suggesting that the content of these components varies significantly and has a considerable impact on whether the sample is in a weathered state.

Stepwise regression analysis-Mahalanobis distance discriminant model

Stepwise regression analyses

Stepwise regression was carried out on 15 predictors, and the regression results were as shown in Table 5:

From the models in Tables 5, 6, and 7, it can be seen that the regression equation gradually introduces new predictors from PbO predictors for stepwise regression testing.

The adjusted R^2 of the model changed from 0.566 to 0.867, so the final stepwise regression obtained the linear regression equation for the glass category Y as:

$$Y = 1.329 + 0.006x_9 - 0.069x_3 - 0.363x_{15} - 0.009x_1 \tag{8}$$

The PbO, K_2O, SiO_2 and the weathering condition of the glass surface in the equation will be used as the key characteristic factors of the distance discrimination model for the next step of prediction.

Distance discriminant analysis

We randomly select 50 data from the sample data as training sets for training, and test the overall sample, repeated 50 times. The model's overall average rate of correctness is 98.69%, with an estimated misjudgment rate of 2.46%. Based on this, the simulation accurately identifies seven out of eight unknown glass cultural relics, resulting in an 87.5% success rate. Table 8 also shows the predicted and actual classes of these eight unknown

Table 5 Table of statistics in Regression process

Model		Unstandardized Coefficient		Standard Coefficient	t	Sig
		B	Standard error	trial version		
1	(Constant)	0.307	0.058		5.313	0.000
	PbO	0.017	0.002	0.757	9.341	0.000
2	(Constant)	0.532	0.064		8.289	0.000
	PbO	0.012	0.002	0.521	6.428	0.000
	K_2O	− 0.050	0.009	− 0.433	− 5.342	0.000
3	(Constant)	0.616	0.053		11.600	0.000
	PbO	0.016	0.002	0.694	9.812	0.000
	K_2O	− 0.057	0.008	− 0.499	− 7.597	0.000
	Weathering	− 0.351	0.058	− 0.396	− 6.093	0.000
4	(Constant)	1.329	0.137		9.668	0.000
	PbO	0.006	0.002	0.243	2.399	0.019
	K_2O	− 0.069	0.007	− 0.599	− 10.440	0.000
	Weathering	− 0.363	0.048	− 0.409	− 7.609	0.000
	SiO_2	− 0.009	0.002	− 0.481	− 5.472	0.000

Table 6 Model ANOVA table

Model		Sum of squares	df	Mean square	F	Sig.
1	Regression	7.544	1	7.544	87.247	0.000
	Residual	5.620	65	0.086		
	Total	13.164	66			
2	Regression	9.277	2	4.639	76.378	0.000
	Residual	3.887	64	0.061		
	Total	13.164	66			
3	Regression	10.719	3	3.573	92.041	0.000
	Residual	2.446	63	0.039		
	Total	13.164	66			
4	Regression	11.515	4	2.879	108.223	0.000
	Residual	1.649	62	0.027		
	Total	13.164	66			

Table 7 Model Goodness of Fit Table

Model	R	R Square	Adjust R side	The error of standard estimation	Durbin-Watson
1	0.757	0.573	0.566	0.29405	
2	0.839	0.705	0.696	0.24644	
3	0.902	0.814	0.805	0.19702	
4	0.935	0.875	0.867	0.16310	1.561

Table 8 Need to predict the category of cultural relics

Numbering	Judgment category	Actual category
A1	1	0
A2	1	1
A3	1	1
A4	1	1
A5	1	1
A6	0	0
A7	0	0
A8	1	1

categories of glass using a Stepwise regression analysis-Mahalanobis distance discriminant model.

Sensitivity analysis

The analysis in this paper involves numerical data that was subjected to a certain degree of disturbance. For category data, any non-weathered artifacts were directly classified as "weathered." This allowed us to study the sensitivity of glass cultural relics to changes in weathering on their surface.

Table 9 illustrates that with each 5% increase in a numerical characteristic factor, the model's prediction results vary, and the degree of change becomes more significant with an increase in error. Among the numerical data, SiO_2 content is crucial for the model, with its impact exceeding that of the other two numerical factors. The result is affected only when the value of each numerical feature is reduced by 20%.

Furthermore, exchanging categories for surface weathering in category-type data cultural relics did not impact the model's results. To summarize, our analysis suggests that SiO_2 content is highly sensitive, whereas K_2O *and* PbO content exhibit weaker sensitivity compared to SiO_2 content. The factor of surface weathering of cultural relics was found to have weak sensitivity.

Table 11 Chi-square test table

	Value	df	Asymp. sig
Pearson Chi-square	3.526	1	0.097
Likelihood ratio	3.567	1	0.097
Linear-by-linear association	3.473	1	0.097
N of valid cases	67		

Spearman-Mahalanobis distance discriminant model

Correlation analysis

To simplify the model calculation, this study will reduce the dimensionality of the sample data. Since the eigenvector data comprises two distinct types of data, the Spearman correlation coefficient is used to select eigenvectors for numerical data. For categorical data, the contingency table analysis method is used to determine whether the surface of glass cultural relics is weathered and whether the type of glass cultural relics is related.

Table 10 indicates the correlation coefficient, the P value, and correlation coefficient strength of each compound composition of glass cultural relics. Based on this analysis, we selected five characteristic factors for the numerical data: SiO_2, K_2O, PaO, BaO *and* SrO.

The correlation coefficient suggests that the content of SiO_2 *and* K_2O is negatively correlated with the glass type, whereas the content of PbO, BaO, *and* SrO is positively correlated with the glass type. In other words, higher SiO_2 *and* K_2O content increases the likelihood

Table 9 Category identification change table after adding disturbance to numerical data

Characteristic	− 20%	− 10%	− 5%	1.1%	5%	10%	15%	20%
SiO_2	0	0	0	1	1	2	3	3
K_2O	1	0	0	1	1	2	2	2
PbO	1	0	0	1	1	2	2	2

Table 10 Correlation coefficient and the P value of cultural relic glass characteristics

	SiO_2	Na_2O	K_2O	CaO	MgO	Al_2O_3	Fe_2O_3
Coefficient	− 0.6703	0.1173	− 0.5832	− 0.1933	− 0.0948	− 0.2411	− 0.2888
Pvalue	5.47e−10	0.3445	2.23e−10	0.1170	0.4455	0.0493	0.0177
	CuO	**PaO**	**BaO**	P_2O_5	**SrO**	SnO_2	SO_2
Coefficient	− 0.2273	0.7695	0.7145	0.1570	0.5881	0.0633	− 0.0649
Pvalue	0.0644	2.78e−14	1.13e−11	0.2046	1.67e−7	0.6110	0.6017

Table 12 Need to predict the category of cultural relics

Numbering	Judgment category	Actual category
A1	0	0
A2	1	1
A3	1	1
A4	1	1
A5	1	1
A6	0	0
A7	0	0
A8	1	1

Table 13 The number of changes in the location artifact type when increasing the error

Error	SiO_2	K_2O	PbO	BaO	SrO
0.01	0	0	0	0	0
0.02	0	1	1	1	1
0.03	1	1	1	1	1
0.04	1	1	1	1	1
0.05	1	1	1	1	1
0.06	1	2	2	2	2
0.07	2	2	2	2	2
0.08	2	2	3	3	3
0.09	3	3	3	3	3
0.10	3	3	3	3	3

of the glass type being high-potassium, whereas higher $PbO, BaO, and\ SrO$ content increases the probability of the glass type being lead-barium.

Table 11 displays a chi-square value of 3.526 and a companion probability of 0.097, which is higher than the significance level of 0.05. Hence, the original assumption that the surface weathering of cultural relics and the type of glass are independent of each other is acceptable. Therefore, no correlation exists between them.

Distance discriminant analysis

The Mahalanobis distance discriminant method was utilized for discriminant analysis based on the selected characteristic factors. For this study, 50 data were randomly selected from the sample data as the training set, and the entire sample was tested 50 times. The test was repeated 50 times, resulting in an overall average correct rate of 99.10% and an estimated misjudgment rate of 1.31%.

The simulation results of the 8 cultural relics are all judged correctly, with a correct rate of 100%. Table 12 also shows the predicted and actual classes of these eight unknown categories of glass using a Spearman-Mahalanobis distance discriminant model.

Sensitivity analysis

Table 13 illustrates that the number of SiO_2 changes slower than other factors as errors increase (ranging from 0.01 to 0.10). When the error is 0.02, the change in the number of SiO_2 is 0. Even when the error is 0.6, the number of cultural relics changes to 2 due to other factors, but the number of SiO_2 changes is still only 1. Therefore, SiO_2 has the lowest sensitivity. The correlation coefficients among the five elements are $-0.6703, -0.5832, 0.7695, 0.7145, and\ 0.5881,$ respectively. The sensitivity of $SiO_2\ and\ K_2O$, with negative correlation coefficients, is low, and the smaller the absolute value, the less sensitive they are. Similarly, we found that the sensitivity increases with the absolute value of the correlation coefficient. It can be seen that the sensitivity of $PbO, BaO, and\ SrO$, which are positively correlated, is ranked from weak to strong as $SrO, BaO, and\ PbO$. In summary, the sensitivity from weak to strong is $SrO, BaO, PbO, SiO_2, and\ K_2O$.

Model comparison

The stepwise regression analysis-Mahalanobis distance discriminant model includes PbO, SiO_2, K_2O, and glass surface weathering as characteristic factors, with an average correct rate of 98.69%. The Spearman-Mahalanobis distance discriminant model uses $PbO, SiO_2, K_2O, BaO, and\ SrO$ as characteristic factors, with an average correct rate of 99.10%. The latter model performs slightly better than the former, but both are affected by measurement errors in the compound content, which can lead to prediction inaccuracies when the error content exceeds 1.1%. However, the former model is slightly more sensitive than the latter. Despite this, both models demonstrate high accuracy in class prediction.

Conclusions

Both the stepwise regression analysis-Mahalanobis distance discriminant model and the Spearman-Mahalanobis distance discriminant model perform well in identifying glass cultural relics. However, the latter achieves higher accuracy.

The model proposed in this paper offers a more scientifically rigorous analysis and identification of the composition of ancient glass products. It allows for timely identification and prediction of the internal components of glass products in various environments, which can effectively prevent the corrosion and weathering of such cultural relics. With further improvements, this model could provide more effective methods and technologies for the protection and restoration of ancient cultural relics.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40494-023-00999-0.

Additional file 1. Chemical composition ratio data set of different glasses.

Acknowledgements
Not applicable.

Author contributions
SW: conceptualization, methodology design, software implementation, results' visualization and interpretation, writing of the main manuscript, project administration; JZ: revision of the final manuscript, writing of the main manuscript, data curation, supervision, investigation, data analysis; XK: methodology design, revision of the final manuscript, conceptualization; HY: supervision. All authors have read and agreed to the published version of the manuscript.

Funding
This work is supported by the National Undergraduate Training Program for Innovation and Entrepreneurship.

Availability of data and materials
Experimental data from the website [Chinese college students Mathematical Contest in Modeling(mcm.edu.cn)].

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no conflicts of interest in this work. We declare that we do not have any commercial or associative interest that represents a competing interest in connection with the work submitted.

Received: 29 March 2023 Accepted: 17 July 2023
Published online: 02 August 2023

References

1. Saijo Y, Suzuki Y, Akiyama R, Shimizu M, Shimotsuma Y, Miura K. Speciation analysis of tin at the tin side of float glass by solvent extraction combined with a stepwise etching technique. J Non-Cryst Solids. 2022;592:121752. https://doi.org/10.1016/j.jnoncrysol.2022.121752.
2. Koob SP, van Giffen NAR, Kunicki-Goldfinger JJ, Brill RH. Caring for glass collections: the importance of maintaining environmental controls. Stud Conserv. 2018;63(S1):146–50. https://doi.org/10.1080/00393630.2018.1492252.
3. Schibille N, Lankton JW, Gratuz B. Compositions of early Islamic glass along the Iranian silk road. Geochemistry. 2022;82(4):125903. https://doi.org/10.1016/j.chemer.2022.125903.
4. Xu H. A study on the composition analysis and identification of ancient glass products based on SVM model. Acad J Comput Inf Sci. 2022;5(13):89–95. https://doi.org/10.25236/JCIS.2022.051314.
5. Tang L, Tang M, Zhang L. Component analysis and identification of glass products based on hierarchical clustering and naive Bayes. Highlights Sci Eng Technol. 2023;41:279–86. https://doi.org/10.54097/hset.v41i.6833.
6. Swan CM, Rehren T, Lankton J, Gratuze B, Brill RH. Compositional observations for Islamic glass from Sīrāf, Iran, in the corning museum of glass collection. J Archaeol Sci Rep. 2017;16:102–16. https://doi.org/10.1016/j.jasrep.2017.08.020.
7. Kaspi O, Girshevitz O, Senderowitz H. Pixe based, machine-learning (pixel) supported workflow for glass fragments classification. Talanta. 2021;234:122608. https://doi.org/10.1016/j.talanta.2021.122608.
8. Bugoi R, Panaite A, Alexandrescu C. Chemical analyses on roman and late antique glass finds from the lower Danube: the case of Tropaeum Traiani. Archaeol Anthropol Sci. 2021;13(9):148. https://doi.org/10.1007/s12520-021-01310-7.
9. Balvanović R, Šmit Ž. Sixth-century ad glassware from Jelica, Serbia—an increasingly complex picture of late antiquity glass composition. Archaeol Anthropol Sci. 2020;12(4):94. https://doi.org/10.1007/s12520-020-01031-3.
10. Henderson J, Ma H, Evans J. Glass production for the silk road? Provenance and trade of Islamic glasses using isotopic and chemical analyses in a geological context. J Archaeol Sci. 2020;119:105164. https://doi.org/10.1016/j.jas.2020.105164.
11. Sottili L, Giuntini L, Mazzinghi A, Massi M, Carraresi L, Castelli L, Czelusniak C, Giambi F, Mandò PA, Manetti M, Ruberto C, Guidorzi L, Re A, Lo Giudice A, Torres R, Arneodo F, Mangani SM, Calusi S, Taccetti F. The role of PIXE and XRF in heritage science: the INFN-CHNet LABEC experience. Appl Sci. 2022. https://doi.org/10.3390/app1213658.
12. He H, An L, Liu W, Zhang J. Prediction model of collapse risk based on information entropy and distance discriminant analysis method. Math Probl Eng. 2017;2017:1–08. https://doi.org/10.1155/2017/8793632.
13. Santos AEM, Lana MS, Cabral IE, Pereira TM, Zare Naghadehi M, de Da Silva DFM, Dos Santos TB. Evaluation of rock slope stability conditions through discriminant analysis. Geotechn Geol Eng. 2019;37(2):775–802. https://doi.org/10.1007/s10706-018-0649-x.
14. Guo J, Wang J, Liu S, Lefik M. Application of an improved cloud model and distance discrimination to evaluate slope stability. Math Probl Eng. 2019;2019:8315894. https://doi.org/10.1155/2019/8315894.
15. Woźniak M, Galazka-Friedman J, Duda P, Jakubowska M, Rzepecka P, Karwowski L. Application of mössbauer spectroscopy, multidimensional discriminant analysis, and mahalanobis distance for classification of equilibrated ordinary chondrites. Meteorit Planet Sci. 2019;54(8):1828–39. https://doi.org/10.1111/maps.13314.
16. Abas MA, Wee ST. Exploring policy governance factors using stepwise multiple regression analysis: a case study of solid waste management policy in Malaysia. Int J Public Sector Perform Manag. 2020;6(6):876–92. https://doi.org/10.1504/IJPSPM.2020.110990.
17. Li H, Cao Y, Su L. Pythagorean fuzzy multi-criteria decision-making approach based on spearman rank correlation coefficient. Soft Comput. 2022;26(6):3001–12. https://doi.org/10.1007/s00500-021-06615-2.
18. Xiao J, Kong J, Leng S. Component analysis and identification of glass artifacts based on logistic regression and k-means clustering. Highlights Sci Eng Technol. 2022;22:333–9. https://doi.org/10.54097/hset.v22i.3399.
19. Abubakar A, Abbas UF, Lasisi KE. Remedying multicollinearity in quantitative analysis: a simulation studies. ATBU J Sci Technol Educ. 2023;10(04):108–16.
20. Sureiman O, Mangera C. F-test of overall significance in regression analysis simplified. J Pract Cardiovasc Sci. 2020;6(2):116. https://doi.org/10.4103/jpcs.jpcs_18_20.
21. Sulewski P. Some contributions to practice of 2×2 contingency tables. J Appl Stat. 2019;46(8):1438–55. https://doi.org/10.1080/02664763.2018.1552665.
22. Cai Z, Zheng Z, Jiang X. Composition analysis and identification of glass products based on Pearson correlation analysis. Highlights Sci Eng Technol. 2022;22:174–86. https://doi.org/10.54097/hset.v22i.3308.
23. Zheng X, Feng Y, Chen H. Analysis of each components of glass samples based on the spearman correlation coefficient model. Highlights Sci Eng Technol. 2022;22:241–5. https://doi.org/10.54097/hset.v22i.3368.

24. Qu Q, Wu W, Guo Y. Study on the classification of glass relics based on spearman correlation coefficient. Acad J Sci Technol. 2023;5(1):147–54. https://doi.org/10.54097/ajst.v5i1.5539.
25. Brereton RG, Lloyd GR. Re-evaluating the role of the mahalanobis distance measure. J Chemometrics. 2016;30(4):134–43. https://doi.org/10.1002/cem.2779.
26. Abdi H. Coefficient of variation. Encycl Res Des. 2010;1(5):170–2. https://doi.org/10.1007/0-387-26336-5_379.

Publisher's Note