Mechatronics and Automation Technology
J. Xu (Ed.)

doi:10.3233/ATDE221213

Composition Analysis and Identification of Ancient Glass Products Based on Linear SVM Algorithm

Xiantong SHI[1,#], Jianfei XU[#], Yanping LIU, Shengbo CHEN
Xiangtan University, School of Mechanical Engineering and Mechanics, China.
#contribute equally to the work

Abstract. The weathering of ancient glass changes the amount of its chemical components, influencing the classification. The impact of several elements on weathering is investigated in this research using controlled variables. The composition of numerous glass-like cultural artifacts before and after weathering was examined and forecasted using applicable mathematical statistical information. The Linear SVM method is used to classify and identify glass artifacts. To address the following problems, it is important to analyze and model using the appropriate data provided in the attachment:1: Examine the link between the glass decoration, kind, and color of these glass objects and their surface deterioration. In addition to the glass type, the statistical law of the content of chemical components that have weathered and not weathered on the surface of samples of cultural artifacts was examined separately, their relationship was discussed, and the chemical composition content prior to weathering was predicted using the data from weathering point detection. 2: The categorization law of lead-barium glass and high-potassium glass is examined in light of the attachment dat. The proper chemical components are chosen to split each category into its subcategories, and particular classification findings and procedures must be provided. The logic and sensitivity of the classification results are also examined.3: In Annex Form 3, analyze the chemical composition of unknown glass artifacts, identify the kind of glass to which they belong, and do sensitivity analysis on the classification results. 4: Analyze the correlation between the chemical compositions of different types of glass cultural relics samples, and examine the differences in the correlation between different categories. Use the strategy of controlling variables to start with for issue 1, merely changing the items in the ornamentation, type, and color each time. Then, consider how these changes affect weathering and determine the connection between the three and the surface of the cultural artifacts. The samples are then separated into four groups according to the sampling site and the category of cultural relics. The chemical components of each type of sample are used as a box-type diagram to eliminate the group value, and then the average value is utilized as the proportional law of this type of sample. Based on the data change before and after we weathering, the effect of the weathering composition of the weathered, it can forecast the content of each component before the sampling point based on this foundation to identify the data based on the realization point. So the solution to issue 1 is as follows. The following factors have an impact on weathering (beginning with the easiest): lines B> C> A; lead barium> high potassium; black = blue green> others in color. For the chemical composition rules of cultural relics and the preliminary forecast of weathering spots, the K2O concentration of high potassium glass is normally approximately 9.3% before weathering, and the weathering has reduced its content

[1] Corresponding Author, Xiantong SHI, Xiangtan University, School of Mechanical Engineering and Mechanics, China; E-mail: wenfordir1@163.com.

to 0.55%. The prediction effect is achieved by 8.8%, and the entire findings are included in the model solution. Based on the successful prediction results of problem 1, the weathered data is replaced with prediction data plus uncontrolled data for problem 2 and the Linear SVM method is used. There are two types of categorization, with 97.2% accuracy in the training set and 100% accuracy in the test set. The same process is used to separate the two types of glass into two sub-categories: magnesium and silicon. MGO has a total training set and an accuracy rate of 96.6%, a total test set and an accuracy rate of 100%, and a decent classification. The data is then disrupted by the sensitivity test. When the disturbance range approaches 20%, the overall accuracy rate is 88%, and the sensitivity is low, indicating that the stability is improved.In response to problem 3, the model of problem 2 is categorized, and the data is disturbed based on the classification findings, progressively increasing the range of disturbances to 20%. There is no change, suggesting that the sensitivity is low and the stability is high. The categorization findings are shown, such as polarized high potassium glass in the A1 cultural relics category and silicon lead glass in the A2 cultural relics category. The model solution contains the remaining outcomes. In response to problem 4, we use the Spearman correlation coefficient to examine the relationship between the chemical components of high potassium and lead-barium, and we calculate the P value using the p value. Different studies of the two varieties of glass of the same component P value differences between the two. The particular correlation coefficient findings are presented in the model solution, and the difference between the P values is that the results of the K2O and CAO in the high potassium glass are the strongest, with a P value of 0.000496, and the MGO and AL2O3 are significant in the lead glass glass. The highest P value is 0.00000189. Finally, we assessed the model's merits and weaknesses, and the model was promoted to other fields of archeology and history on the basis of suggesting comparable improvement solutions for the flaws.

Keywords. Classification variable mathematical statistics, linearism spearman coefficient.

Symbol Description

Symbol	Description
v	SVM algorithm ultra -plane legal vector
a	SVM algorithm ultra -plane offset
p	Two variables unrelated probability
t_a	Spellman related coefficient
R^2	Error analysis determination coefficient
L	Ragone Japanese
(x_i, y_i)	SVM algorithm sample collection
d_i	Spirman coefficient level poor

1. Introduction

The fabrication of glass goods requires melting agents since the principal component quartz sand has an excessively high melting point. Variations in the melting agent will result in variations in the chemical composition of the glass. Leading glass glass, for example, is used as a melting agent during the fire process, increasing the lead concentration and oxidation of lead and oxide. The climate readily weathers ancient glass, causing variations in the proportions of each component, affecting the accurate classification. The color and ornamentation are both affected by the glass' surface, and various locations may have varying degrees of weathering. Data about ancient Chinese glassware, such as two varieties of leading glass and high potassium glass, are now available. The percentage of the respective primary elements is provided in Annex Form 2, and Annex Form 1 contains the categorization information for these cultural artifacts. The total of each component ratio should add up to 100%, however due to detection or other circumstances, the total may only add up to a fraction of the composition.

1.1. Analysis of Problem 1

The first question is split into two sections. The link between ornamentation, type, and color and its surface weathering is examined in the first section. Deck tests are appropriate for application in this non-parameter classification variable data correlation analysis. The derivative test cannot completely rule out the influence of the other two variables and can only be attained if the two or two variables are interrelated. As a consequence, we simply utilize it to evaluate our findings. To determine the link between the three and the surface weathering essence, we use the control variable approach to vary just one of the decorations, kinds, and colors at a time. We then assess how its changes affect the weathering condition. In order to push the data before the weathering data after weathering, the second portion must examine the statistical laws of two types of glass and whether we are weathering. In all, four types of cultural relic sample chemical composition ratios must be considered. It is first necessary to pre-process the data in accordance with the topic's requirements, removing any erroneous data and replacing any missing values with 0. It is important to note that the difference between weathering and unsatisfactory is based on a sample point rather than the entire population of cultural artifacts. The statistical laws we must obtain in order to meet the requirements of these parts must not only include the average proportion of different chemical components present in the four different types of cultural relics, but also take into account how the two different types of glass weather both before and after the change in chemical composition. [1]. Depending on the box type, it is important to delete the aberrant value. [2]. Based on this change law, you may identify information based on the weathered point and the chemical composition prior to the weathering of cultural artifacts.

1.2. Analysis of Question 2

There are two sections to Question 2, as well. The categorization guidelines for high potassium glass and lead glass are examined in the first section. In general, it can be inferred from the name of the glass and relevant literature [3] that the two forms of glass are based on potassium oxide and lead and pyrine, respectively, as constituents. If the content is categorized, a specific analysis is necessary to meet the requirements of the applicable classification standard, which is the determination value of the component content ratio. It is important to remember that the topic said that weathering will impact the category judgment; therefore the data after weathering is absurd. The problem 1's solution for weather prediction causes the data before weather 1 to replace the data after the attachment form 2's corresponding data. The LINEAR SVM in machine learning is categorized by the content of the glass cultural relics based on the content of potassium oxide and lead, as well as the elements of the pupae, in accordance with the changed data. The proper chemical component of two different varieties of glass must be broken down into sub-categories in the second section. The classification rules of high potassium and lead in the analysis of the high potassium and lead in the first part may be referred to as the division of rules. Its sensitivity and logic are examined. The categorization approach is appropriate if nature is better.

1.3. Question 3 Analysis

In order to answer Question 3 and evaluate Question 2's analysis, you must classify the unknown category of glass cultural relics in Annex Form 3, determine whether or not they are windy, and examine each cultural relic's composition category. To see if the categorization situation changes, add a certain type of chemical element content (the absolute value is 5%, 10%, and 20%), and then examine its sensitivity.

1.4. Question 4 Analysis

For Problem 4, samples of various glass cultural relics are needed to examine the link between their chemical compositions. It is appropriate for the relationship between the associated coefficients of Pilson for the relationship between numerous sets of data. The data must be large enough and conform to the normal distribution. The two requirements stated by the subjects are obviously not satisfied. Each component of each ingredient, the relationship between Spellman's relationships, may gain its distinctions by providing a representative indicator analysis of various categories of cultural relics.

2. The Establishment and Solution of the Model

There are three varieties of decorations among the 58 cultural relics: A, B, and C, as well as two types of high potassium and lead. They come in the following colors: purple, light green, dark green, light blue, and blue-green. two different weathering kinds. There are four missing data points for the color, and there are complete data for the ornamentation, kind, and surface weathering. The ensuing procedure of solving the problem will be hampered by these four lacking variables. Essence Then classify all 54 of the data's patterns, colors, surface weathering, and types, simplify the data of cultural artifacts, and make it easier to compare them later. Each component of the cultural relic sample point's data and processing was done to an excel level. The complete data of these 58 sample points is obtained from the 58 sampling points. It is discovered that two data are less than 85% and do not match the topic's standards. As a result, we require these two pieces of information. Remove the 54 cultural relic detection spots' data. Then, for Forms 2 and 3, set the empty value to 0, so that the tables corresponding to the composition on the table have values.

2.1. Question 1: Analysis Model Based on Control Variable Method

2.1.1. Control Variable

The three variable aspects in this topic are the ornamentation of the cultural relics, the type of cultural relics, and the color of the cultural relics. The surface weathering of the cultural artifacts is an independent variable component. 17 independent variables are retrieved after pre-processing. The fraction of each cultural relic weather is derived by data processing in the Excel table, as shown in table 1:

Table 1. 17 types of cultural relics after a pre –processing

Pattern	Type	Color	Number of cultural relics	Weathering proportion
A	High K	Blue -green	5.00	0.00
A	High K	Deep blue	1.00	0.00
A	Ba-Pb	Black	2.00	1.00
A	Ba-Pb	Blue -green	1.00	1.00
A	Ba-Pb	Light blue	10.00	0.60
A	Ba-Pb	Deep blue	1.00	0.00
B	High K	Blue -green	6.00	1.00
C	High K	Blue -green	1.00	0.00
C	High K	Light blue	4.00	0.00
C	High K	Dark green	1.00	0.00
C	Ba-Pb	Blue -green	2.00	1.00
C	Ba-Pb	Green	1.00	0.00
C	Ba-Pb	Light blue	6.00	1.00
C	Ba-Pb	Light green	3.00	0.33
C	Ba-Pb	Dark green	6.00	0.67
C	Ba-Pb	Purple	4.00	0.50

To study the aforementioned theories and facts, we will conduct the following groups of experiments: (a) when the texture and color are the same, the relationship between the textures on the surface of the cultural relics is analyzed; (b) when the texture and color are the same, the relationship between the types of analysis on the surface of the cultural relics is analyzed; and (c) when the texture and color are the same, the relationship between the colors on the surface of the cultural relics is analyzed. We utilize code to study the relationship of several specific scenarios, and the findings are displayed in table 2.

Table 2. Comparison of various factors on the weathering relationship when maintaining other factors

Influential factors	The degree of influence on weathering
Texture	B> c> A
type	Ba-Pb > High K
color	Black = Blue and Green> Other

The preceding table shows that when all other things remain constant, B texture is simpler to weather than A and C. Leading kinds are more prone to weather, while black, blue, and green are easier to weather than other colors.

2.1.2. Based on the Establishment of Four Types of Cultural Relic Samples and Detection Data Prediction Models

According to the results of the control variable technique, the statistical law of the surface of glass cultural relic samples has the statistical law of weathered chemical content when combined with the types of glass cultural relics. The emphasis here is on investigating the statistical law of the type and whether or not there is weathering. As a result, we may here The statistical model of two or two combinations is used to determine the statistical likelihood of each combination, and an average diagram of each component ratio of the sampling points before and after weathering is created to determine its statistical law. Then, using the differential point's detection data and statistical model data, first use the box-type diagram to identify the reasonable interval, and then delete the abnormal values that do not correspond to the interval, and compute the average value except the abnormal value. Calculate the average of weathering and weatherlessness to anticipate the chemical composition content prior to the weathering of glass cultural relics.

(1)Statistics of cultural relic samples

There are two types of glass: high potassium and lead barium. The surface of the sample is divided into two types: weathered and non-weathered. This combination contains high potassium, high potassium non-weathering, lead melody, and lead barium. There are four sorts of no weather combinations. We sort the 54 sampling points that we acquire in order to generate four sorts of data that belong to the relevant categories. We cannot use the Rida standard to eliminate these four forms of data since they do not satisfy the normal distribution, and the box-type diagram does not need data to obey any distribution, and there are no further data requirements. Originally, the identification of the abnormal value is more actual and objective; therefore the box-type map is used to eliminate the abnormal value in the data (figure 1).

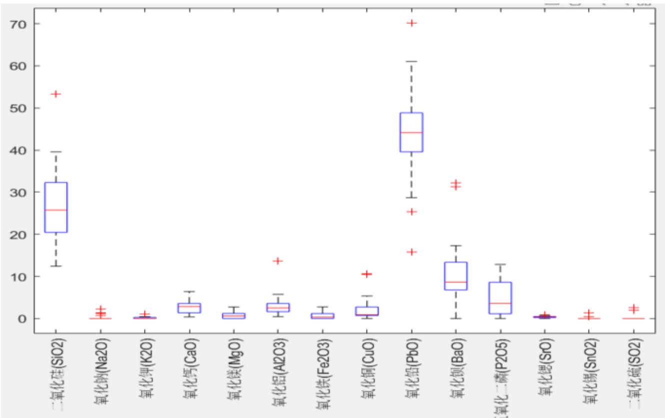

Figure 1. Abnormal value detection ---- box map

A box-type diagram is a statistical graphic that is used to diversify data. The statistical diagram is typically made up of six data nodes: higher quadri-bit, medium, quadrilateral, quadrilateral distance, upper limit, and lower limit. When the selected value is less than or more than the box's lower or higher limits, it is deemed abnormal. The anomalous values in the data are deleted using the relevant box-type figure code [4]. (See particular code appendix). Use Excel table formulae to obtain the average data after removing the four categories of aberrant values: high potassium glass non-windy, high potassium glass, tough and glassy glass, and the average value of each material change. Use the code to compute its average value, and then render figure 2 as follows:

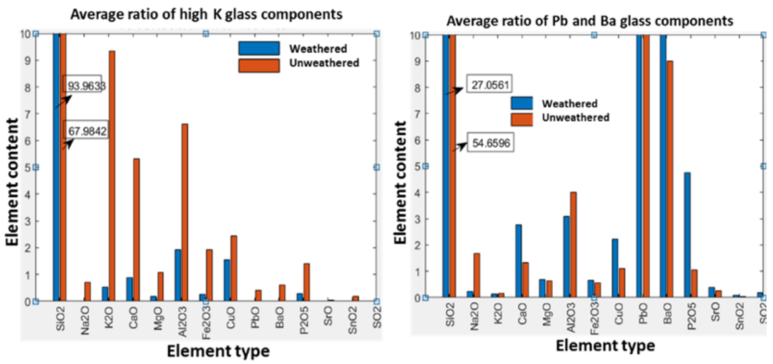

Figure 2. The average component ratio of each component of glass sampling points

The statistical laws of these four combinations may be shown in figure 2. Whether it is worn, windy, or weathered, the silicon dioxide content, which is the major component of the glass, is dominant. The content of silicon dioxide in lead glass is relatively low after weathering, and the content of leading elements is improved; in high potassium glass, the quantity of silicon dioxide accounted for relatively poor content. Other surface components are decreased, but the potassium level remains larger than that of lead-pyramid glass. The content of leading items is nearly zero.

(2) Forecast of chemical composition content before weathering

We will predict using the average value in this case. Before weathering, the predicted value of the chemical composition content should meet the following relationship:

Predictive value = the average value after weathering (the average before weathering-the average after weathering)

Use this connection, combine the discovered statistical laws, and build the relevant code to obtain the following outcomes (table 3):

Table 3. The prediction results of the component of high potassium glass weathering points

Composition / Sample	(SiO_2)	(Na_2O)	(K_2O)	CaO)	(MgO)	(Al_2O_3)	(Fe_2O_3)
07	66.65	0.70	8.79	5.53	0.88	6.67	1.84
09	69.04	0.70	9.38	5.08	0.88	6.01	1.99
10	70.79	0.70	9.71	4.67	0.88	5.50	1.93
12	68.31	0.70	9.80	5.18	0.88	6.15	1.96
22	66.37	0.70	9.53	6.12	1.52	8.19	2.02
27	66.74	0.70	8.79	5.40	1.42	7.20	1.87

Composition / Sample	(CuO)	(PbO)	(BaO)	(P_2O_5)	(SrO)	(SnO_2)	(SO_2)
07	0.45	23.02	4.19	4.32	0.30	0.00	0.00
09	4.13	0.41	0.60	1.73	0.04	0.20	0.10
10	2.44	0.41	0.60	1.47	0.04	0.20	0.10
12	1.73	0.41	0.60	1.12	0.04	0.20	0.10
22	2.54	0.41	0.60	1.27	0.04	0.20	0.10
27	1.44	0.41	0.60	1.33	0.04	0.20	0.10

Through this result, we can see that the predictions can see that each prediction result is within the range of 85%to 105%, indicating that the high accuracy we obtain is obtained.

As shown in table 4 and table 5, we choose the total value of the total value of the 49, 50 uncomfortable sampling point and the two predicted values for error analysis and calculation. The analysis calculation formula of the error is as follows:

$$R^2 = 1 - \frac{\sum_i (\hat{y} - y_i)^2}{\sum_i (\bar{y} - y_i)^2} \tag{1}$$

Table 4. Predicted component content at 49 un-weathered and 49 weathered points

Composition	(SiO_2)	(Na_2O)	(K_2O)	(CaO)	(MgO)	(Al_2O_3)	(Fe_2O_3)
49 Unweathered point	54.61	0.00	0.30	2.08	1.20	6.50	1.27
49 Weathering point prediction	57.95	1.40	0.19	0.00	0.00	2.82	0.00
Composition	**(CuO)**	**(PbO)**	**(BaO)**	**(P_2O_5)**	**(SrO)**	**(SnO_2)**	**(SO_2)**
49 Unweathered point	0.45	23.02	4.19	4.32	0.30	0.00	0.00
49 Weathering point prediction	0.47	18.63	14.49	0.00	0.00	0.00	0.00

Table 5. Predicted component content at 50 unweathered and 50 weathered points

Composition	(SiO_2)	(Na_2O)	(K_2O)	(CaO)	(MgO)	(Al_2O_3)	(Fe_2O_3)
50 Unweathered point	45.02	0.00	0.00	3.12	0.54	4.16	0.00
50 Weathering point prediction	54.22	1.40	0.19	0.00	0.00	3.15	0.00
Composition	**(CuO)**	**(PbO)**	**(BaO)**	**(P_2O_5)**	**(SrO)**	**(SnO_2)**	**(SO_2)**
50 Unweathered point	0.70	30.61	6.22	6.34	0.23	0.00	0.00
50 Weathering point prediction	0.84	22.48	16.34	0.00	0.00	0.00	0.00

With the value of the above formula's content, you can compute the decision coefficient of 49 points = 0.9455 and the decision coefficient of 50 points = 0.9346. These two decision coefficients are near to one, showing that there is a fitting effect between the two. The model built is sufficient if the mistake is modest.

2.2. Question 2: Classification Model Based on the LINEAR SVM Algorithm

By checking the appropriate literature, we can determine that the elements contained in the glass name are essentially high content in each of the glass's components, so we can know If the ingredient content is categorized, the proportion of the ingredient content is the primary classification basis. Weathering impacts the judgment of the category before classification because the backdrop of the issue addressed, therefore the data after weathering is evidently inappropriate. To maintain the data's integrity, the data before to the forecast of the first question replaces the weathered data. Following these steps, we may create a classification model to summarize the problem, and then utilize machine learning's Linear SVM technique to solve the model.

2.2.1. Linear SVM Algorithm [5]

SVM is a wide linear classifier that allows vector computers to do dual classification of data chosen through supervision and learning. Its decision-making boundary is the greatest superplane boundary for solving the total number of data samples. SVM can calculate the risk of experience by using hinge loss functions. Regularization items are inserted into the solution system to address structural hazards. This is a classifier that is both loose and stable. SVM may also employ nuclear methods for non-linear or linear classification, which is another popular nuclear learning approach. Under certain limitations, linear SVM seeks an ultra-plane classification to differentiate positive and negative samples. Among the selected sample capacity, the division of ultra -plane can be used to use the following linear equations to express: $v^T x + a = 0$

Here, $v = (v_1, v_2, \cdots v_n)$ The direction of the super plane is represented by the law vector. The displacement item, A, represents the distance between the original point and the ultra-plane. The split of the supercopic displacement A and the French vector V is obvious.

The basic model of LINEAR SVM is a convex two -time plan:

$$\begin{cases} \min\limits_{v,a} \ \dfrac{1}{2}\|\mathrm{v}\|^2 \\ s.t.\ \ \mathrm{y}_i(v^T x_i + a) \geq 1 \quad \mathrm{i} = 1, 2, \cdots, \mathrm{m} \end{cases} \tag{2}$$

However, it usually uses the Lagrangri multiplier to answer instead of using the existing optimized computing package:

$$L(v, a, \varepsilon) = \frac{1}{2}\|\mathrm{v}\|^2 + \sum_{i=1}^{m} \varepsilon_i (1 - y_i (v^T x_i + a)) \tag{3}$$

where, $\varepsilon = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_m)$,making $L(v, a, \varepsilon)$ and a the number of partial guidance is zero can be obtained

$$\begin{cases} 0 = \sum\limits_{i=1}^{m} \varepsilon_i y_i \\ v = \sum\limits_{i=1}^{m} \varepsilon_i y_i x_i \end{cases} \tag{4}$$

According to the above formula $L(v, a, \varepsilon)$, v and a in can be eliminated, and then combined with the constraints of the above equation to obtain the dual form of the most primitive problem

$$\begin{cases} \max \ \sum\limits_{i=1}^{\mathrm{m}} \varepsilon_i - \dfrac{1}{2} \sum\limits_{i=1}^{m} \sum\limits_{j=1}^{m} \varepsilon_i \varepsilon_j y_i y_j x_i^T x_j \\ s.t.\ \sum\limits_{i=1}^{\mathrm{m}} \varepsilon_i y_i = 0 \quad \varepsilon_i \geq 0 \ \ \mathrm{i} = 1, 2, \cdots, \mathrm{m} \end{cases} \tag{5}$$

After solving ε , just ask for v and a, to get the model:

$$f(x)=v^T x+a=\sum_{i=1}^{m}\varepsilon_i y_i x_i^T x+a \tag{6}$$

Obtaining ε by inserting $v=\sum_{i=1}^{m}\varepsilon_i y_i x_i$ 中,we can get v.

For arbitrary vectors (x_s, y_s) all contains $y_d f(x_d)=1$, that is ,

$$y_d(\sum_{i\in D}\varepsilon_i y_i x_i^T x_d+a)=1 \tag{7}$$

where, $D=\{i \mid \varepsilon_i>0, i=1,2,\cdots,m\}$ is a collection of subscripts for all support vectors. According to the theoretical knowledge, any vector can be selected to solve the above equation to obtain "a" or it can be obtained by solving the mean of the entire support vector:

$$a=\frac{1}{|D|}\sum_{d\in D}(y_d-\sum_{i\in D}\varepsilon_i y_i x_i^T x_d) \tag{8}$$

The above introduction, however, is merely a hard side distance Linear SVM. Even if no sample data violates the upper formula's requirements, that is, the data sample is linearly divided in the sample space; there is a super flat plane that can separate the different sorts of samples. In reality, though, this arrangement is far too good. As a result, the more realistic way is to allow some sample data to break the restrictions of the issue, resulting in the soft edge line Linear SVM. Here is the optimized target function:

$$\min_{v,a,\lambda_i}\ \frac{1}{2}\|\mathrm{v}\|^2+C\sum_{i=1}^{m}\lambda_i \tag{9}$$

This problem can usually still be solved using the Lagrangian multiplier method:

$$L(v,a,\varepsilon,\lambda,\mu)=\frac{1}{2}\|\mathrm{v}\|^2+C\sum_{i=1}^{m}\lambda_i+\sum_{i=1}^{m}\varepsilon_i(1-\lambda_i-y_i(v^T x_i+a))-\sum_{i=1}^{m}\mu_i\lambda_i \tag{10}$$

Among them, the Lagrange multiplier is $\varepsilon_i\ge 0, \mu_i\ge 0$

$$v=\sum_{i=1}^{m}\varepsilon_i y_i x_i,\ 0=\sum_{i=1}^{m}\varepsilon_i y_i,\ C=\varepsilon_i+\mu_i \tag{11}$$

The same can be done to obtain the duality of the original problem

$$\max\ \sum_{i=1}^{\mathrm{m}}\varepsilon_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\varepsilon_i\varepsilon_j y_i y_j x_i^T x_j \tag{12}$$

$$s.t.\ \sum_{i=1}^{\mathrm{m}}\varepsilon_i y_i=0\ \ 0\le\varepsilon_i\le \mathrm{C}\ \mathrm{i}=1,2,\cdots,\mathrm{m} \tag{13}$$

Therefore, the data is arbitrarily trained on samples (x_i, y_i), containing $\varepsilon_i = 0$ or $y_i f(x_i) = 1 - \lambda_i$.So, $\varepsilon_i = 0$, there might be incorrect ion in f(x)effecting while $\varepsilon_i > 0$, Hence $y_i f(x_i) = 1 - \lambda_i$, That is, the sample is a support vector;as $\varepsilon_i < C$,Ba $\mu_i > 0$,Therefore $\lambda_i = 0$, This sample can be obtained at the maximum interval boundary.

2.2.2. Solution of Classified Models

(1) Classification of high potassium glass and lead glass

We consider non-weathering data and anticipated data that hasn't yet been weathered as the general unwavering data in accordance with the guiding principles and procedures of the aforementioned LINEAR SVM method. As illustrated in figure 3, the SVM toolbox executes the nuclear function of the high potassium glass, the lead glass, and the categorized scattered point picture.

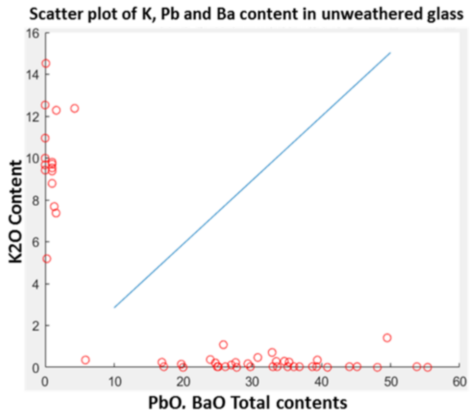

Figure 3. Scatterplot of potassium and barium lead content of unweathered glass

The nuclear function of the high potassium glass and leading glass is y = 6.095X-0.2 as a consequence of the toolbox's running result. High potassium glass is in the function's upper right corner. It is 97.2%, while the test set's accuracy rate is 100%, demonstrating great accuracy and supporting the validity of the categorization.

(2) The sub -category division of each category

Types of high potassium glass and leading glass are discussed in this article. There are 14 chemical components in each kind. To choose the proper chemical components in this situation, we shall employ enumeration methods. Each glass has two chosen chemical components that are divided. Select the combination with the highest priority using two or more comparisons to compare the division impact of each combination. The fundamental idea behind the enumeration technique is to make a list of every scenario that may arise, examine which circumstance best fits the problem's requirements, and determine which solutions are viable and should be eliminated. The fundamental concept behind employing enumeration techniques:

(1) Determine the topic's enumeration objects, scope, and judgment criteria

(2) All feasible solutions, then confirm whether it is a problem solution.

The enumeration object is a mixture of two or two chemical components, with 14 components of each kind performing the range. The determining condition is whether or not it is the best division combination. In the SVM toolbox, enter each enumeration technique combination in turn, and run the relevant scatter distribution map through the code of the preceding classification. Essence we chose 20% of the non-weathering data from the silicon magnesium composition combination and the projected before weathering, and the remaining 80% as the test set. 1) A nuclear functions and categorization distributed points according to the principles and code (figure 4 and figure 5). The kernel functions of high potassium glass, lead-barium-silicon, and magnesium content are y=0.05454x-2.5, y=0.03962x-1.265, respectively, the upper right of the function is magnesium glass, the bottom is partial silica glass, and the test set and training set are also 100%, with extremely high accuracy. As a result, high-potassium glass is subdivided into meta-magnesium and high-potassium glass and meta-silicon high-potassium glass; lead barium glass is subdivided into meta-silicate lead silicate glass and meta-magnesium lead silicate glass.

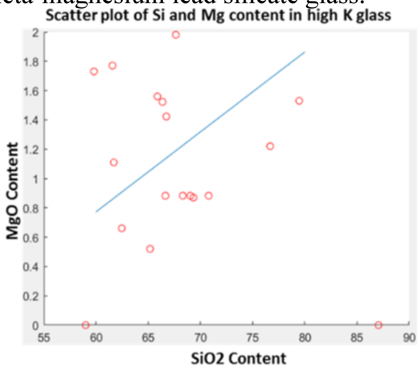

Figure 4. Scatter plot of high potassium glass silicon and magnesium content

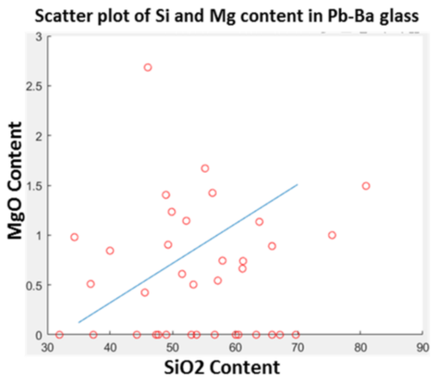

Figure 5. Scatter plot of lead-barium glass silicon and magnesium content

(3) Rationality analysis of subclass classification results

To validate the rationality of the subclass classification results, the model must be reasonable, that is, it must demonstrate that the kernel function established in the scatter plot is always correct, and the data used for subclassification are non-weathering components, including the original non-weathering and alternative weathering predictions. When the silicon magnesium content data in the weathered glass composition is extracted, through the SVM toolbox to represent its data, after testing, the kernel function in the above figure is still applicable, indicating that our classification method and classification results are very reasonable (figure 6 and figure 7).

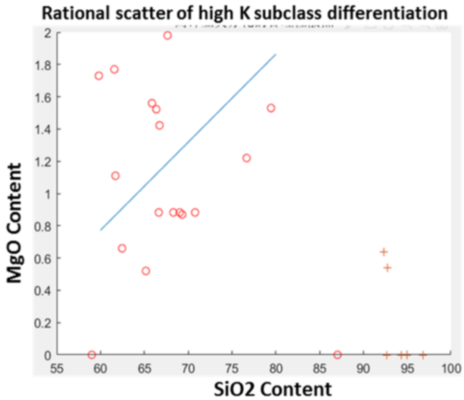

Figure 6. Test of the rationality of the classification of high-potassium glass subclass

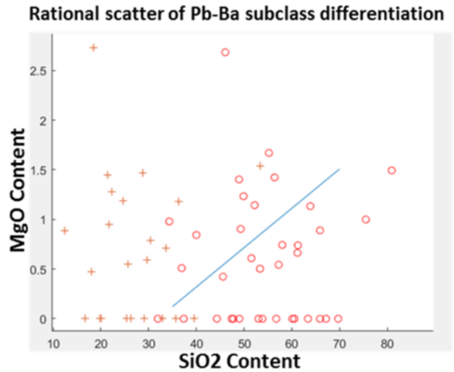

Figure 7. Scatter plot of the rationality of the classification of the subclass of lead barium glass

(4) Sensitivity analysis of substrate classification results

Using the unweathered component data and prediction data as the whole training set and the unweathered component data as the test set, we have chosen a reasonable subclass partition based on the prior data. Create a line chart showing the change in percentage (absolute value) in these situations, as shown in figure 8, and then set the test set to have perturbations of 5%, 10%, 15%, and 20%. Then, see how much the test set influences the classification outcomes after each perturbation.

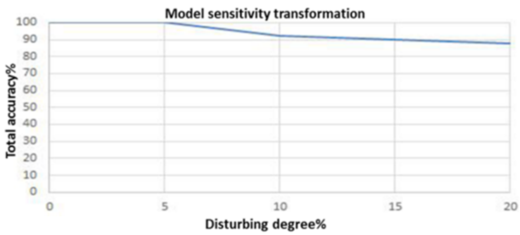

Figure 8. Sensitivity test plot of subclass segmentation model

The change in the polyline in the figure shows that when the disturbance size (absolute value) is 5%, there is no effect on the sum change rate, and when the disturbance gradually increases, the sum accuracy rate gradually decreases until the disturbance reaches 20%, at which point the sum accuracy rate is 88%, indicating that the model's sensitivity is relatively low and that the model has high stability.

2.3. Problem 3: Identification of Cultural Relics of Unknown Category Based on the Model of Problem 2

2.3.1. Identification of Cultural Relics of Unknown Category

The model and procedure from issue two are used to answer this question, and the overall solution concept is as follows:

1) The eight categories are first divided into their respective categories, and they are first divided into high-potassium glass and lead-barium glass. The potassium oxide content, barium oxide, and lead oxide content after pretreatment are first extracted, identified through the SVM toolbox, and then they are first divided into their respective categories.

2) The silica and magnesium oxide content are then extracted, identified using the SVM toolkit, and classified into the appropriate classes.

3) Then run through the code to get the final type. The results are shown in table 6:

Table 6. Types of cultural relics A1-A8

Heritage No	Large category (high potassium/barium lead)	Subclass (metasilic/metamagnesium)
A1	High potassium	Metamagnesium
A2	Lead barium	Partial silicon
A3	Lead barium	Metamagnesium
A4	Lead barium	Metamagnesium
A5	Lead barium	Metamagnesium
A6	High potassium	Partial silicon
A7	High potassium	Partial silicon
A8	Lead barium	Partial silicon

2.3.2. Sensitivity Analysis of Categorical Results

The sensitivity analysis technique is consistent with the process described in question 2, in that we first extract data from eight recognized cultural relics and utilize the original unweathered component data and prediction data as the full training set. Set perturbations of -20%, -10%, 10%, and 20% to see how the test set affected the identification results after each perturbation, and create a line chart showing the percentage change in these situations, as shown in figure 9. The change in the polyline in the figure shows that no matter how much disturbance there is, it has no influence on the accuracy of the sum, and its value is 100%, indicating that the model's sensitivity is extremely low, implying that the model's stability and accuracy are very high.

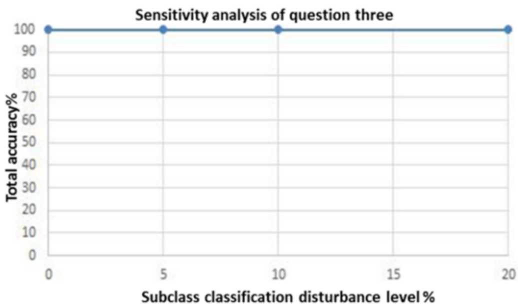

Figure 9. Line diagram of sensitivity transformation of the classification model for unknown categories

2.4. Problem 4：Spearman Correlation Coefficient Solves the Component Association Relationship Model

2.4.1. Spearman Correlation Coefficient

The correlation coefficient is a statistical indicator that represents the degree of connection between two or more variables. This coefficient is determined using the corresponding formula, and it is based on the dispersion between the variables and their respective mean dispersions, with the product of their dispersions reflecting the degree of correlation between the variables chosen. The correlation coefficients we usually use are Pearson correlation coefficient and Spearman correlation coefficient, but Pearson correlation coefficient is used to investigate the linear relationship between variables and requires a large sample size, which this question does not meet, so it is not used. When the selected data is not duplicated and the variables are completely monotonically correlated, then the value of Spearman's correlation coefficient is 1 or -1, showing a strong positive or negative correlation. Spearman's correlation coefficient can measure the correlation between variables, typically using monotonic equations to express the correlation between variables.

In a continuous distribution, Spearman correlation, also known as level correlation, states that the selected data is often more than half the level when the level is substituted for the selected data level. The level correlation coefficient and the rank correlation coefficient, however, are virtually the same in the example we are using. The Spearman correlation coefficient has no strict requirements for the data we choose; it only requires the actual value of the selected variable to be rated or to be converted from the actual value of the variable to obtain the rating; as a result, it can be used to analyze data regardless of sample size or the overall distribution of the selected variable. Spearman correlation is chosen to investigate the link between the chemical compositions of various types of glass artifacts exactly because it has a better advantage over other correlation coefficients. Let X and Y be two sets of data, and the Spearman correlation coefficient formula is:

$$t_a = 1 - \frac{6\sum_{i=1}^{m} d_i^2}{m(m^2-1)}, d_i = X_i - Y_i \tag{14}$$

di is the rank difference, which refers to the rank of the data, that is, the number is arranged from smallest to largest in the column in which it is located, and his position in this column, if there are the same values, then the position of the data is averaged.

2.4.2. Solving of Component Association Relationship Model

The results are as follows when the form's data is combined with the formula mentioned above.

According to the correlation between the values of the correlation coefficient (table 7):

Table 7. Correlation coefficient and correlation

correlation index	Associative relationships
0.0-0.2	Very weakly or no correlation
0.2-0.4	Weak correlation
0.4-0.6	Moderately relevant
0.6-0.8	Strong correlation
0.8-1.0	Ultra-strong correlation

In high-potassium glass, silica and potassium oxide will be chosen for description, along with a few other usual components. Silica, lead oxide, and barium oxide will be chosen for association explanation in lead barium glass, while the remaining constituents will just be briefly mentioned.

In high-potassium glass, silica is negatively correlated with the majority of other substances, and only a small number of substances are positively correlated. As a result, as the silica content rises, the vast majority of the other substances' contents decrease, and only a small portion of the other substances rise. Among them, calcium oxide forms a strong negative correlation with it, sodium oxide and copper oxide form a weak negative correlation, potassium oxide, alumina, and iron oxide form a moderately negative correlation, tin oxide forms a moderately positive correlation, and the rest is an extremely weak correlation. Potassium oxide showed positive correlations with sodium oxide, calcium oxide, iron oxide, copper oxide, and sulfur dioxide, with calcium oxide showing the strongest correlation and sulfur dioxide the weakest. Magnesium oxide, alumina, lead oxide, barium oxide, phosphorus pentoxide, strontium oxide, tin oxide, and others are all negatively connected with it. In particular, potassium oxide and barium oxide have a very weak negative connection, while other components have modest correlations. Silica is positively correlated with four components and negatively correlated with nine components in lead barium glass. Silica is moderately negatively correlated with lead oxide and strontium oxide and weaker negatively correlated with the other seven components, indicating that as silica content increases, these nine components decrease to a corresponding extent. When the quantity of lead oxide increases, the amount of alumina, copper oxide, barium oxide, and phosphorus pentoxide is weakly reduced, the amount of strontium oxide is slightly increased, and the amount of other constituents changes very little or not at all; Alumina increases very weakly, sulfur dioxide reduces dramatically, and other components either grow more noticeably or decrease very weakly as the concentration of barium oxide falls (figure 10).

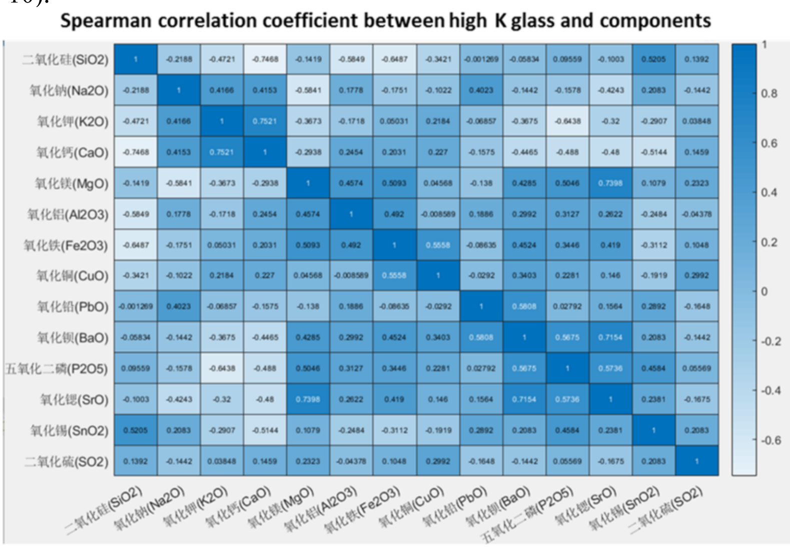

Figure 10. Spearman correlation coefficient between high-potassium glass and components

Based on the correlations between their various chemical constituents shown in figures 11, the following differences between high-potassium glass and lead-barium glass can be made:

1) While the silica content of high potassium glass and the oxygen content of lead barium glass are negatively correlated with the majority of other components, only two of the two have weak negative correlations, compared to seven of the seven in the lead barium glass, suggesting that the silica content of high potassium glass is more important than the oxygen content.

2) In contrast to the marker elements in lead-barium glass, which have little to no association with non-metallic elements but clear correlation with metal components, the logo elements in high-potassium glass clearly correlate with non-metallic elements.

3) The sulfur dioxide content of lead barium glass is negatively correlated with the content of the majority of other components, and the correlation with other components is significantly increased. In contrast, the sulfur dioxide content of high-potassium glass has a positive correlation with the majority of other glass components, and these correlations are generally weak.

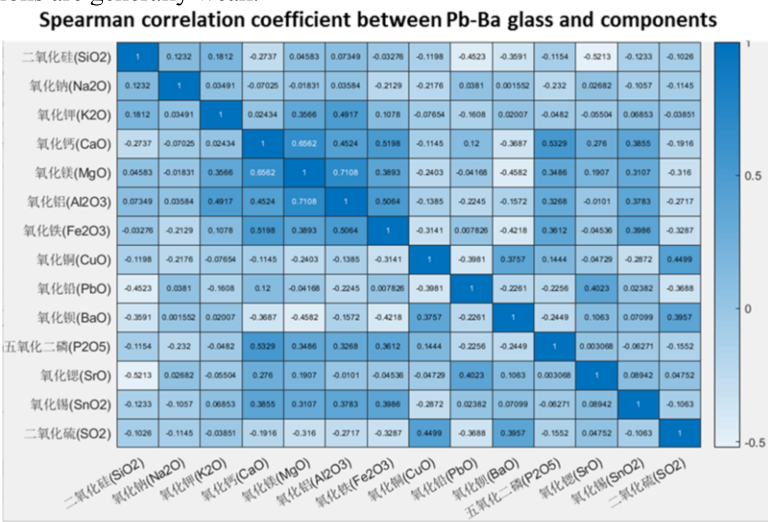

Figure 11. Spearman correlation coefficient between barium lead glass and components

2.4.3. Significance Test (Figures 12 and 13)

In the Spearman correlation coefficient test, when the sample size is greater than 30, it is understood that the sample meets the significance sample test. In this case, we can test the hypothesis by establishing a statistical method, and then found the following method for statistics and calculate the p-value:

$$r_s\sqrt{n-1} \sim N(0,1) \tag{15}$$

The operation's results show that K_2O and CaO have the highest significance in high potassium glass, with a p value of 0.000496, while MgO and Al_2O_3 have the highest significance in lead barium glass, with a p value of 0.00000189, allowing us to determine the difference in the chemical composition associations between the two types of glass. The p-value at this point is substantially lower than 0.05, suggesting a significant difference, and the Spearman correlation coefficient confirms the validity of the correlation.

	二氧化硅 (SiO_2)	氧化钠 (Na_2O)	氧化钾 (K_2O)	氧化钙 (CaO)	氧化镁 (MgO)	氧化铝 (Al_2O_3)	氧化铁 (Fe_2O_3)	氧化铜 (CuO)	氧化铅 (PbO)	氧化钡 (BaO)	五氧化二磷(P_2O_5)	氧化锶 (SrO)	氧化锡 (SnO_2)	二氧化硫 (SO_2)
二氧化硅 (SiO_2)	1	0.398857	0.055695	0.000573	0.586971	0.013651	0.004848	0.178902	0.996143	0.823986	0.715158	0.701847	0.032202	0.594078
氧化钠 (Na_2O)	0.398857	1	0.096203	0.097376	0.013822	0.494806	0.501372	0.696411	0.109395	0.580864	0.545281	0.089626	0.422349	0.580864
氧化钾 (K_2O)	0.055695	0.096203	1	0.000496	0.146976	0.50974	0.847945	0.399703	0.793717	0.146701	0.005292	0.210541	0.257594	0.883436
氧化钙 (CaO)	0.000573	0.097376	0.000496	1	0.252326	0.342436	0.434396	0.380949	0.546156	0.072413	0.046866	0.051191	0.03462	0.57621
氧化镁 (MgO)	0.586971	0.013822	0.146976	0.252326	1	0.064875	0.036799	0.861797	0.597408	0.086111	0.038839	0.000688	0.680118	0.369626
氧化铝 (Al_2O_3)	0.013651	0.494806	0.50974	0.342436	0.064875	1	0.044836	0.973901	0.468584	0.243369	0.22171	0.309295	0.336385	0.867484
氧化铁 (Fe_2O_3)	0.004848	0.501372	0.847945	0.434396	0.036799	0.044836	1	0.02052	0.741771	0.068227	0.175604	0.094096	0.224041	0.688892
氧化铜 (CuO)	0.178902	0.696411	0.399703	0.380949	0.861797	0.973901	0.02052	1	0.911405	0.181345	0.37861	0.576006	0.460502	0.243369
氧化铅 (PbO)	0.996143	0.109395	0.793717	0.546156	0.597408	0.468584	0.741771	0.911405	1	0.014493	0.915296	0.548947	0.260248	0.527418
氧化钡 (BaO)	0.823986	0.580864	0.146701	0.072413	0.086111	0.243369	0.068227	0.181345	0.014493	1	0.017494	0.001245	0.422349	0.580864
五氧化二磷(P_2O_5)	0.715158	0.545281	0.005292	0.046866	0.038839	0.22171	0.175604	0.37861	0.915296	0.017494	1	0.016068	0.064225	0.831877
氧化锶 (SrO)	0.701847	0.089626	0.210541	0.051191	0.000688	0.309295	0.094096	0.576006	0.548947	0.001245	0.016068	1	0.357495	0.520461
氧化锡 (SnO_2)	0.032202	0.422349	0.257594	0.03462	0.680118	0.336385	0.224041	0.460502	0.260248	0.422349	0.064225	0.357495	1	0.422349
二氧化硫 (SO_2)	0.594078	0.580864	0.883436	0.57621	0.369626	0.867484	0.688892	0.243369	0.527418	0.580864	0.831877	0.520461	0.422349	1

Figure 12. Saliency analysis between components of high-potassium glass

	二氧化硅 (SiO_2)	氧化钠 (Na_2O)	氧化钾 (K_2O)	氧化钙 (CaO)	氧化镁 (MgO)	氧化铝 (Al_2O_3)	氧化铁 (Fe_2O_3)	氧化铜 (CuO)	氧化铅 (PbO)	氧化钡 (BaO)	五氧化二磷(P_2O_5)	氧化锶 (SrO)	氧化锡 (SnO_2)	二氧化硫 (SO_2)
二氧化硅 (SiO_2)	1	0.474151	0.290294	0.106236	0.790693	0.67013	0.849543	0.486493	0.005615	0.031476	0.502701	0.001112	0.473546	0.551734
氧化钠 (Na_2O)	0.474151	1	0.839813	0.683918	0.915595	0.835617	0.212547	0.202452	0.825417	0.992832	0.173327	0.876615	0.539348	0.506164
氧化钾 (K_2O)	0.290294	0.839813	1	0.887939	0.032799	0.002321	0.53157	0.657276	0.348744	0.907503	0.780146	0.749872	0.691277	0.823533
氧化钙 (CaO)	0.106236	0.683918	0.887939	1	1.39E-05	0.005596	0.001156	0.506043	0.485784	0.026929	0.00082	0.103177	0.020224	0.263071
氧化镁 (MgO)	0.790693	0.915595	0.032799	1.39E-05	1	1.19E-06	0.018946	0.158042	0.809256	0.004954	0.037164	0.265326	0.065151	0.060436
氧化铝 (Al_2O_3)	0.67013	0.835617	0.002321	0.005596	1.19E-06	1	0.001624	0.420533	0.188155	0.36001	0.051705	0.953399	0.022915	0.108989
氧化铁 (Fe_2O_3)	0.849543	0.212547	0.53157	0.001156	0.018946	0.001624	1	0.062093	0.963869	0.010392	0.030447	0.792807	0.016038	0.050321
氧化铜 (CuO)	0.486493	0.202452	0.657276	0.506043	0.158042	0.420533	0.062093	1	0.016192	0.023943	0.400645	0.784162	0.089444	0.005906
氧化铅 (PbO)	0.005615	0.825417	0.348744	0.485784	0.809256	0.188155	0.963869	0.016192	1	0.184769	0.185799	0.015004	0.890326	0.026854
氧化钡 (BaO)	0.031476	0.992832	0.907503	0.026929	0.004954	0.36001	0.010392	0.023943	0.184769	1	0.149921	0.53731	0.680755	0.0169
五氧化二磷(P_2O_5)	0.502701	0.173327	0.780146	0.00082	0.037164	0.051705	0.030447	0.400645	0.185799	0.149921	1	0.985831	0.716355	0.36606
氧化锶 (SrO)	0.001112	0.876615	0.749872	0.103177	0.265326	0.953399	0.792807	0.784162	0.015004	0.53731	0.985831	1	0.604032	0.783173
氧化锡 (SnO_2)	0.473546	0.539348	0.691277	0.020224	0.065151	0.022915	0.016038	0.089444	0.890326	0.680755	0.716355	0.604032	1	0.537397
二氧化硫 (SO_2)	0.551734	0.506164	0.823533	0.263071	0.060436	0.108989	0.050321	0.005906	0.026854	0.0169	0.36606	0.783173	0.537397	1

Figure 13. Saliency analysis between lead barium glass and composition

3. Model Evaluation and Promotion

3.1. Model Evaluation

3.1.1. Advantages of Model

(1) To make the results more credible, the first section of the first question uses the approach of controlling variables, which eliminates the interference of the other two factors while evaluating the effect of one of the components.

(2) The second half of question 1 employs a box plot to weed out outliers, and the average value calculated on this basis can more accurately depict the statistical rule governing the chemical cost ratio of the two categories of cultural relics and weathering. When the weathering point is not weathered, the percentage of each component is anticipated, and the prediction results are tested, the error is extremely minimal. Questions 2, 3, and 4 successfully employ the forecasting technique and data after the forecasting method's validity has been shown.

(3) Inquiry 2 Potassium and lead-barium are used as the categorization criteria in the archaeological literature, considerably reducing the burden. SVMs are a more reliable method of categorization in machine learning than manually divided thresholds.

(4) In question 3, data perturbations are introduced to the model to increase the amount of data, decrease sensitivity, and enhance the overall classification impact when the model in question 2 is highly accurate.

(5) This study expertly employs programming to apply quantitative statistics, machine learning, feature analysis, and other information and tools for analysis. The results are accurate and trustworthy, the reasoning is rigorous and clear, and the representation is clear and comprehensive.

3.1.2. Shortcomings and Improvements of the Model

(1) The first part of question 1 takes into account the lack of data, and the quantitative analysis error of the relationship between grain, type, and weathering is too large, so only qualitative analysis is done. The effect of some colors on weathering cannot be qualitatively determined due to the lack of samples, which is why the quantitative analysis error of the relationship between grain, type, and weathering cannot be small. If there is enough data, it is possible to expand the parameters further and convert the qualitative analysis into a quantitative study to determine the correlation coefficient.

(2) The average value cannot be used to reflect changes before and after weathering of the same cultural relics because only two cultural relics contain samples of weathering points and unweathered points at the same time, and the effect will be worse.

(3) Although potassium and lead barium are the only criteria used in the first section of question 2, they are still important factors in classifying glass and will have an effect on its kind. In order to strengthen the impact of each component on classification and to offer a more logical regular interpretation of the classification of glass kinds, pertinent literature should be located.

4. Conclusion

Analyzing different types of glass cultural relics is beneficial to determining the type of glass cultural relics and sub-categories. It may assist comprehend the evolution of glass goods in different locations, as well as research local human history, based on the influence of diverse glass production techniques on cultural relics, along with geographical, temporal, and other elements. This question's modeling approaches and analysis can also be encouraged. Outside the glass, do research on the categorization, weathering, and historical evolution of diverse cultural artifacts [6].

References

[1] Zhao Z Q. Study on composition system and manufacturing technology of glass beads unearthed from Shirenzigou Site Group in Balikun, Xinjiang [D]. Northwestern University, 2016.

[2] Wang H T, Li X R, Zhao L Y. Cable Segmentation Method Based on Box Graph and break point Threshold Boundary [J]. Computer Applications and Software, 2021, 38(09):244-249.

[3] Han F, Wang Y Z, Ma H J, Ma Q L. Analysis of glassware excavated from Xinfeng Tomb in Lintong, Shaanxi [J]. Spectroscopy and Spectral Analysis, 2017, 37(05):1546-1552.

[4] Xiong Y Y, Wu X Q. Comparison and Application of four Discriminant Criteria for Gross Error [J]. University Physics Experiment, 2010, 23(01):66-68. DOI:10.14139/j.cnki.cn22-1228.2010.01.008.

[5] Xia J W. Research on application of Support Vector Machine (SVM) in classification of ancient ceramics [D]. Jingdezhen Ceramic University, 2 022. DOI:10.27191/d.cnki.gjdtc.2022.000296.

[6] Liu X L. Comparative Study on Glassware Unearthed from Warring States and Han Dynasty in Hunan, Hubei and Guangxi [D]. Hunan University, 2017.