# Research on the Composition Analysis and Identification of Ancient Glass Products

Xin Li *, Zheng Zhang

College of Science, North China University of Technology, Beijing, China

* Corresponding author: lixin20201118@163.com

**Abstract.** Chinese glass has a history of thousands of years, and ancient glass is valuable material evidence of early cultural exchanges and trade between China and the West. With the continuous development of economy and technology, Chinese scholars have made deeper research on glass composition, but at present, most of the works in China focus on the chemical point of view of ancient glass composition, detection technology and cultural exchanges, and less on the statistical law of chemical composition, identification and classification. In this paper, the relationship between weathering, ornamentation, colour and type of glass surface is analyzed by correspondence analysis. Then the Mann-Whitney test is used to test the content of each chemical substance of different types of glass before and after weathering, and the statistical law of the chemical composition content of the surface of cultural relics samples with or without weathering is expounded. Then, the classification criteria of high-potassium glass and lead-barium glass were obtained by using a decision tree and random forest model, and the sub-classification method of four types of glass was given by using K-means clustering algorithm based on standardized data, and each type of glass was divided into two categories. Finally, the neural network is used to analyze and identify the unknown glass. It provides a reference for the further study and identification of the chemical composition of ancient glass.

**Keywords:** Ancient glass, Correspondence analysis, Decision tree, Clustering analysis, Neural network.

## 1. Introduction

Glass is an important material evidence of ancient trade and cultural exchanges between China and the West on the Silk Road, and its appearance characteristics such as decoration, colour and chemical composition are the basis for cultural relics research [1]. The study of ancient glass relics can deepen the understanding of other ancient cultural relics and the ancient Silk Road, and is also of great significance to the production and research of modern glass products in China. A Jiayao and others elaborated on the raw materials and production process of ancient glass [2]. In the previous study of ancient glass relics, Cheng Qian et al. Made a simple description of the appearance and chemical composition statistics of the samples when studying glass relics [3], but did not deeply analyze the correlation between glass types, decorations, colours and other characteristics, nor did they pay attention to the changes of chemical composition before and after glass weathering; Wen Rui et al. Made a simple division of the composition system of ancient glass [4]. This paper explores the classification rules of high-potassium glass and lead-barium glass according to their chemical composition [5]. The sub-classification of glass cultural relics can simplify the research work of cultural relics, but there is no clear research method on the sub-classification of ancient glass at home and abroad; the identification of the type of glass cultural relics is the focus of the research work. Wang Chengyu and others have conducted in-depth research on the principle of glass weathering, which has reference value for the identification of ancient glass [6].

Given the above problems, this paper first uses the chi-square test to analyze the correlation of variables, uses multiple correspondence analysis to describe the relationship between variables, and then uses the Mann-Whitney U test to test the content of each chemical substance of different types of glass before and after weathering, and analyzes the influencing factors of glass surface weathering; Then, the decision tree model and random forest model were established to analyze the classification rules of high-potassium glass and lead-barium glass. Then K-means clustering algorithm is used to sub-classify the four kinds of glass. To analyze the chemical composition of the unknown glass relics and identify their types, MATLAB is used to fit the BP neural network, and the quantitative conjugate gradient method is used to train and optimize the model.

## 2. Statistical Analysis of Appearance and Chemical Composition of Glass Cultural Relics

### 2.1. Analysis of Surface Weathering Factors of Glass Cultural Relics

According to the analysis of the data, the surface weathering, glass type, ornamentation and colour of glass cultural relics are classified variables. Referring to the number of samples, this paper uses the method of multiple correspondence analysis to study the correlation of classification variables. The basic idea of multiple correspondence analysis is to represent the proportional structure of the elements in the rows and columns of a contingency table in a relatively low-dimensional space in the form of points. Before doing correspondence analysis, this paper first uses the chi-square test to test the correlation between variables and then reveals the relationship between variables through correspondence analysis.

**Table 1.** Chi-square test table

| Variable | $X^2$ | Corrected $X^2$ | P |
|---|---|---|---|
| Color | 6.287 | 6.287 | 0.507 |
| Glass type | 5.400 | 4.134 | 0.020** |
| Texturization | 5.747 | 5.747 | 0.056* |
| Note : * * *, * *, * represents the significance level of 1 %, 5 %, 10 % respectively | | | |

The results of the chi-square test are shown in Table 1. The two groups of variables, colour and texture, are not significant at a high level. The original hypothesis is accepted, that is, colour and texture are not related to surface weathering. Based on surface weathering and type, the significance P value is 0.020 *. The original hypothesis is rejected, that is, glass type is related to weathering.

**Table 2.** Effect quantitative analysis table

| Analysis item | Phi | Crammer's V | coefficient of contingency | lambda |
|---|---|---|---|---|
| Color | 0.341 | 0.341 | 0.323 | 0.000 |
| Type | 0.316 | 0.316 | 0.302 | 0.000 |
| Texturization | 0.326 | 0.326 | 0.310 | 0.000 |

It can be seen from the results of the effect quantification analysis in Table 2 above that the Phi coefficient and Crammer's V coefficient of surface weathering and analysis variables colour, type and ornamentation are between 0.3 and 0.6, so the different degree of surface weathering and colour, ornamentation and type is moderate.

Based on the above analysis of the samples, it can be seen that there is a significant correlation between the surface weathering of glass relics and the type of glass, and there is no significant relationship between the surface weathering of glass relics and the colour and decoration.

### 2.2. Analysis of chemical composition of glass before and after weathering

To analyze the difference in chemical composition between the two kinds of glass before and after weathering, the Mann-Whitney U test was carried out on the contents of various chemical substances of high-potassium glass and lead-barium glass before and after weathering.

**Table 3.** Mann-Whitney U test results

| chemical materials | | $SiO_2$ | $Na_2O$ | $K_2O$ | CaO | MgO | $Al_2O_3$ | $Fe_2O_3$ |
|---|---|---|---|---|---|---|---|---|
| Sig. | High potassium glass | 0.001 | 0.194 | 0.004 | 0.025 | 0.013 | 0.001 | 0.025 |
| | lead barium glass | 0.000 | 0.011 | 0.143 | 0.002 | 0.853 | 0.024 | 0.745 |
| chemical materials | | CuO | PbO | BaO | $P_2O_5$ | SrO | $SnO_2$ | $SO_2$ |
| Sig. | High potassium glass | 0.261 | 0.025 | 0.123 | 0.013 | 0.045 | 0.480 | 0.194 |
| | lead barium glass | 0.052 | 0.000 | 0.521 | 0.000 | 0.031 | 0.620 | 0.208 |

The test results are shown in Table 3. It is considered that there is a significant difference in the content of chemical components before and after glass weathering if the significance level is less than 0.05. For high-K glass, the chemical substances with different contents before and after weathering are $SiO_2$, $K_2O$, CaO, MgO, $Al_2O_3$, $Fe_2O_3$, PbO, $P_2O_5$ and SrO; For lead-barium glass, the chemical substances with different contents before and after weathering are $SiO_2$, $Na_2O$, CaO, $Al_2O_3$, PbO, $P_2O_5$ and SrO.The statistical analysis of the chemical composition content of the two kinds of glass before and after weathering is shown in Table 4 and Table 5:

**Table 4.** Comparison of average chemical composition of high potassium glass before and after weathering

| chemical composition | $SiO_2$ | $K_2O$ | CaO | MgO | $Al_2O_3$ | $Fe_2O_3$ | PbO | $P_2O_5$ | SrO |
|---|---|---|---|---|---|---|---|---|---|
| No weathering | 67.98 | 9.33 | 5.33 | 1.08 | 6.62 | 1.93 | 0.41 | 1.40 | 0.04 |
| weathering | 93.96 | 0.54 | 0.87 | 0.20 | 1.93 | 0.27 | 0.00 | 0.28 | 0.00 |

**Table 5.** Comparison of average chemical composition of lead-barium glass before and after weathering

| chemical composition | $SiO_2$ | $Na_2O$ | CaO | $Al_2O_3$ | PbO | $P_2O_5$ | SrO |
|---|---|---|---|---|---|---|---|
| No weathering | 54.66 | 1.68 | 1.32 | 4.46 | 22.09 | 1.05 | 0.27 |
| weathering | 24.91 | 0.22 | 2.70 | 2.97 | 43.31 | 5.28 | 0.42 |

It can be seen from the above table that after weathering, the $SiO_2$ content of high-potassium glass increases significantly, and other chemical components decrease; after weathering, the $SiO_2$ content of lead-barium glass decreases significantly, the PbO content increases significantly, the $Na_2O$ and $Al_2O_3$ contents decrease, and the CaO, $P_2O_5$ and SrO contents increase.

# 3. Study on Classification of High Potassium Glass and Pb-Ba Glass

## 3.1. Preliminary Classification of High Potassium Glass and Lead-Barium Glass

The decision tree divides the samples arriving at a node according to a specific attribute, and each subsequent branch of the node corresponds to a possible value of the attribute. The mode of the output variable of the sample contained in the leaf node of the classification decision is the result of the classification. According to the variables of silicon dioxide, lead oxide, potassium oxide and the like of the glass relics, a decision tree which can distinguish the glass relics of high-potassium glass and lead-barium glass is generated, and the classification rules of the two types of glass relics are visually processed.

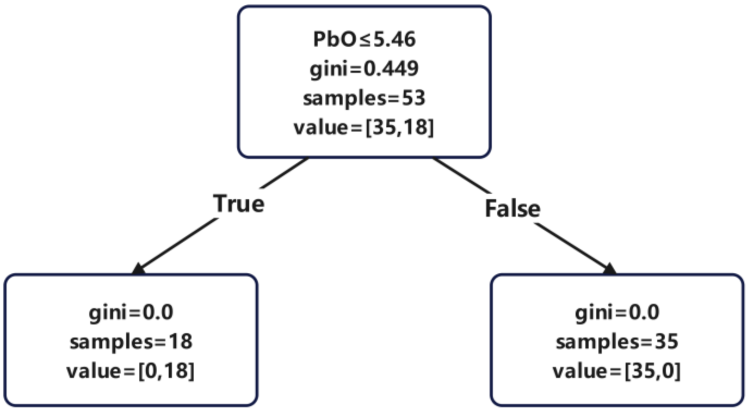


**Figure 1.** Decision tree structure diagram

It can be seen from the above Figure 1 that high-potassium glass and lead-barium glass are classified according to the chemical composition of lead oxide, which is an important characteristic. When the content of lead oxide is less than or equal to 5.46, the glass can be classified as high-potassium glass.

Because the decision tree is prone to overfitting and ignoring the correlation of attributes in the data set, the random forest model is used to optimize the model. In the process of generating many decision trees, random forest classification can overcome the overfitting problem by randomly sampling the sample observation and feature variables of the modelling data set, and randomly selecting the training set and feature attributes of the random forest model [7].

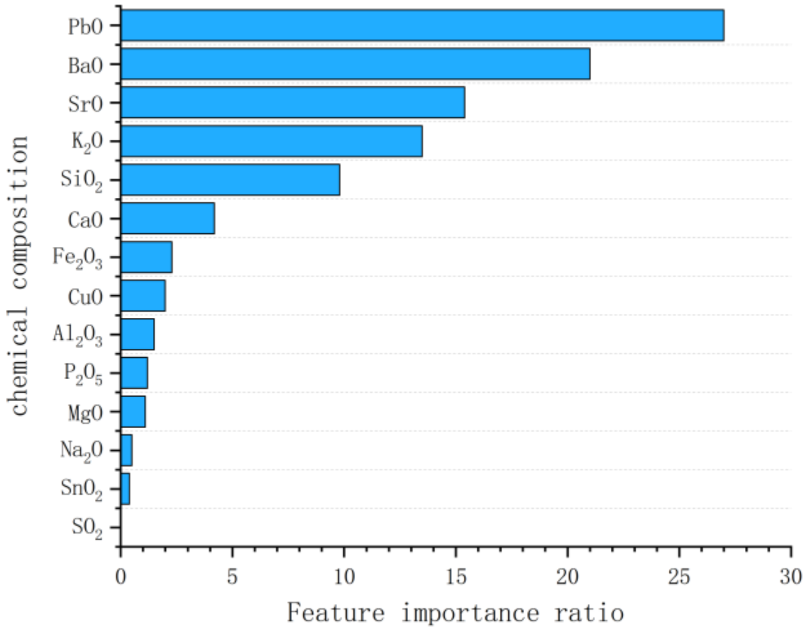


**Figure 2.** Feature importance ratio graph

Fig. 2 is the characteristic importance ratio chart of the random forest model, which shows that the importance of PbO is the greatest and is much higher than other chemical composition variables, which is consistent with the decision tree analysis results, indicating that the classification rule of high-potassium glass and lead-barium glass is based on PbO content. At the same time, the training set and test set of the confusion matrix perform very well, which shows that the classification of glass relics based on PbO composition can classify high-potassium glass and lead-barium glass very well, and has a strong rationality.

### 3.2. Subclassification of Glass Cultural Relics

To further study the classification of glass, this paper uses the K-means clustering algorithm to divide the glass into sub-categories according to the key chemical composition content of glass relics based on four categories of high-potassium weathered glass, high-potassium unweathered glass, lead-barium weathered glass and lead-barium unweathered glass.

The k-means clustering algorithm is a simple iterative clustering algorithm, which divides the sample set into K subsets, that is, to form K classes, and divides n samples into K classes. The centre distance between each sample and its class is the smallest, and each sample belongs to only one class [8]. As shown in Figure 3.

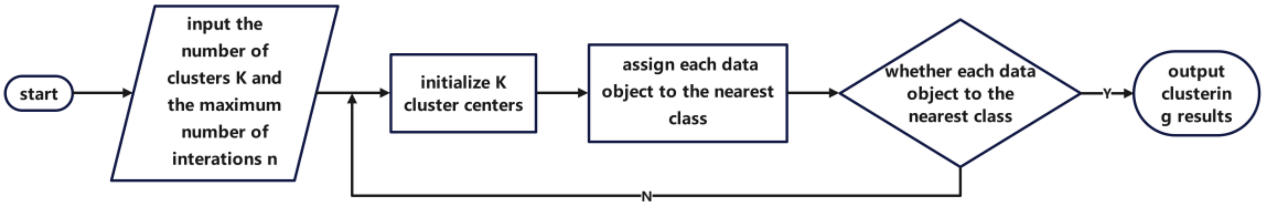


**Figure 3.** K-means clustering flow chart

Before clustering analysis, the form data is processed by Z-score standardization. Direct use of raw data may lead to highlighting the role of indicators with higher values in the analysis and weakening the role of indicators with lower values, so to ensure the reliability of the results, the data are standardized. In this paper, Euclidean distance is used as the similarity measure, and the sum of squared errors (SSE) is used as the objective function to measure the quality of clustering, and the relationship between the number of clusters K and SSE is drawn, and then the elbow method is used to determine the number of clusters K.

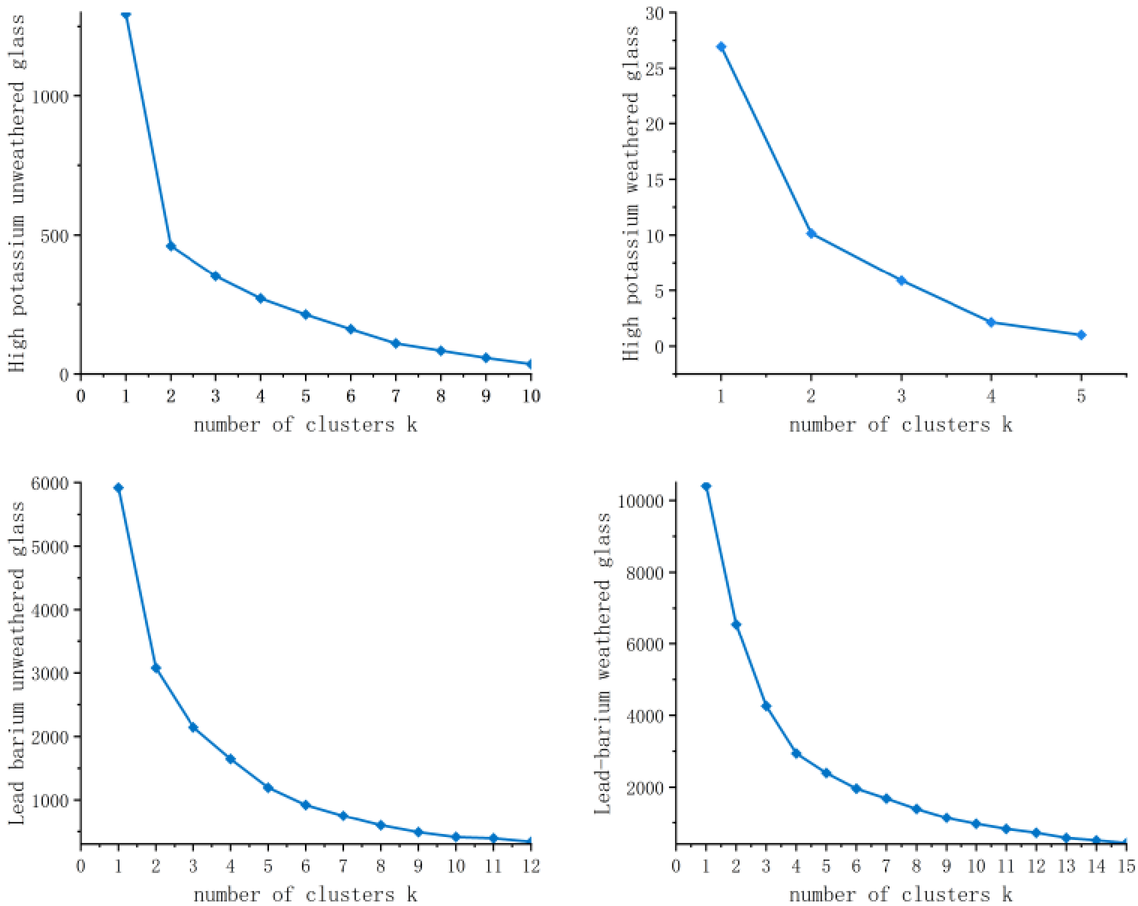


**Figure 4.** determination of glass classification numbers

As shown in Figure 4, the optimal cluster number of high-potassium unweathered glass, high-potassium weathered glass and lead-barium unweathered glass is 2; the optimal cluster number of lead-barium weathered glass should be 3. Considering that the sample data is small and the number of subclassifications is too large to be conducive to research, 2 is selected as the cluster number in this paper. The standardized chemical components were used as variables for K-means clustering in SPSS. The final cluster centres for the subclasses of the four glasses are shown in Figure 5 below:

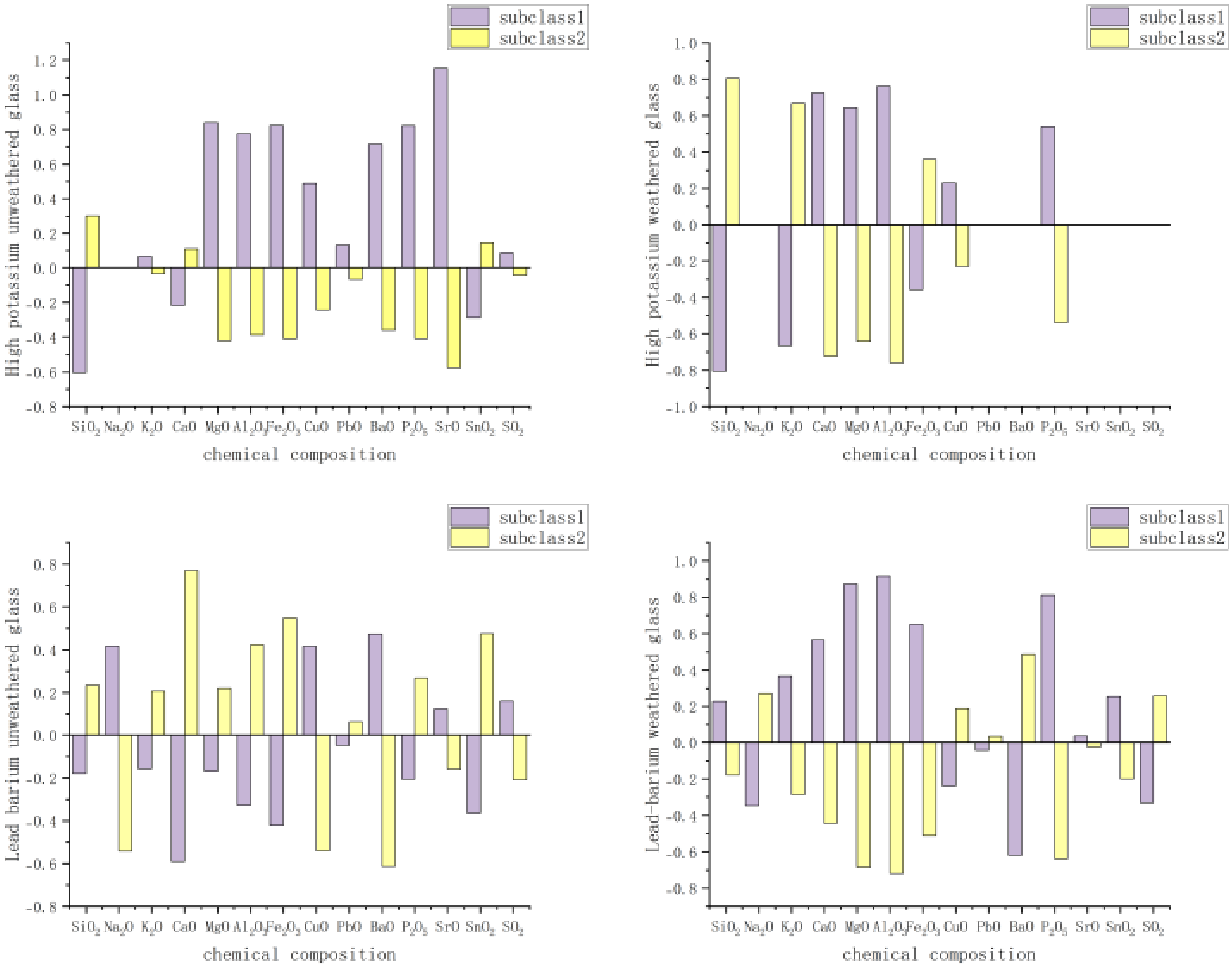


**Figure 5.** The final clustering centers

The final cluster centre is the average value of the standardized chemical composition contained in each category, which can reflect the composition differences between subcategories. From the overall performance of the above figure, it can be seen that the classification results of subclasses are reasonable, the difference in chemical composition between subclasses is large, and there are obvious signs of chemical composition distinction between different subclasses of the same glass type.

## 4. BP neural network model

### 4.1. Establishment of the model

In this paper, a BP neural network model will be established to analyze the chemical composition of eight glass cultural relics of unknown categories, identify their types, and analyze the sensitivity of the classification results. Before the training of the BP neural network, the data is normalized. Then it enters the hidden layer through the input layer of the network, and then enters the input layer, and finally gets the output of the network, in which the neurons adopt the following activation function:

$$f(x) = \frac{1}{(1+e^{-(W^T x+B)})} \tag{1}$$

Where, $W = \left|w_{ji}^{(p)}\right|$ is a weight matrix, and $w_{ji}^{(p)}$ is a weight between the ith neuron in the p-1th layer and the jth neuron in the PTH layer, x is input data, $B = \left|b_j^{(p)}\right|$ is a threshold matrix, and $b_j^{(p)}$ is a threshold of the jth neuron in the PTH layer. The formula is as follows:

$$W(t+1) = W(\mathrm{t}) \text{-} \eta \frac{\partial E(\mathrm{t})}{\partial W(t)} \tag{2}$$

$$B(t+1) = B(\mathrm{t}) \text{-} \eta \frac{\partial E(\mathrm{t})}{\partial B(t)} \tag{3}$$

Where, $E(t)$ is the network error, $\eta \in (0,1)$ is the step factor, and t is the number of training times.

When the error $E(t)$ of the network does not reach the expected target, it will enter the process of error reverse propagation, and the total output error of the network will be determined by the training pair according to the number of output layer-hidden layer-output layer nodes, which are 22 and 5 respectively. Through the empirical formula and continuous testing, the hidden layer setting with the minimum error is selected, and the number of hidden layer nodes is 10. 70% of the data was used as the training set, 15% as the validation set, and 15% as the test set.

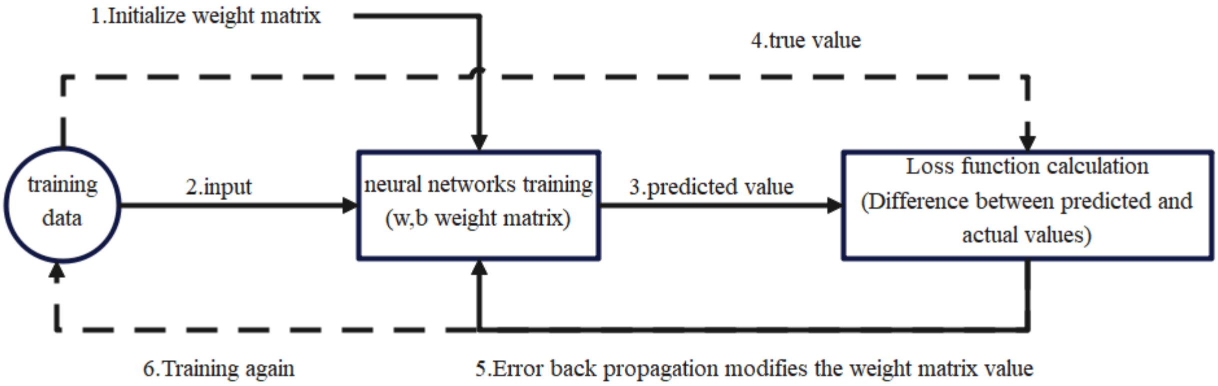


**Figure 6.** Neural network flow diagram

According to Machine Learning-Formula Derivation and Code Implementation [9], the training set, validation set and test set are reasonably divided into 70%, 15% and 15% for data analysis. The training set is the sample data used for model fitting, and the validation set is the known sample data reserved in the model training process. The ability of the existing model is evaluated and the model parameters are adjusted. The function of the test set is to evaluate the generalization ability of the final model, and finally, achieve the optimization of the model through continuous iterative training. Through the theoretical analysis of "Pattern Recognition and Intelligent Computing MATLAB Technology Implementation"[10], this paper uses the quantitative conjugate gradient method to train the model. As shown in Figure 6.

### 4.2. Results and analysis

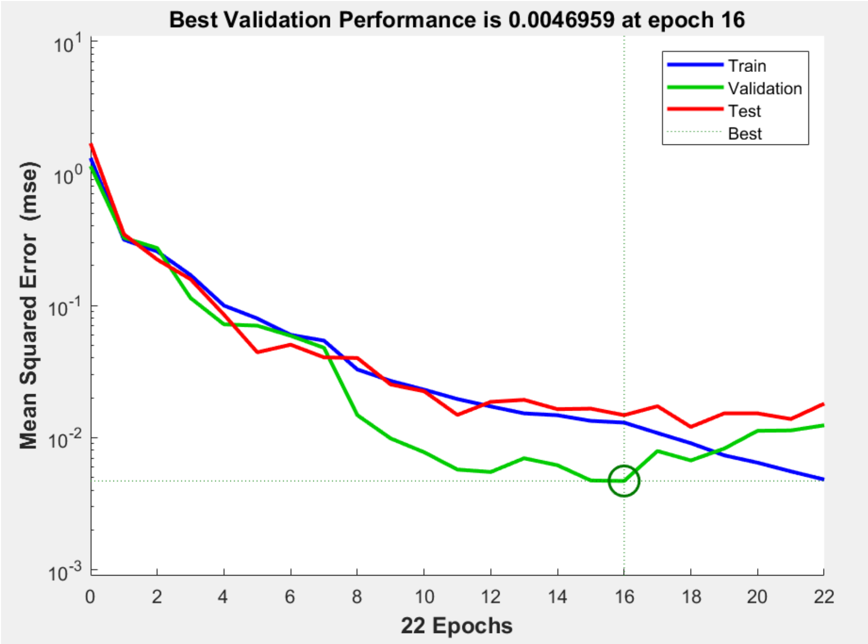


**Figure 7.** Neural network iterative MSE diagram

Bring in the data for training, and the result is shown in the Figure 7. An epoch is equivalent to using all the samples in the training set to train once. After 22 iterations, the optimal MSE is 0.0047 at the 16th iteration, and the model works well.

**Table 6.** Neural network fitting coefficient

| | Training | Validation | Text | All |
|---|---|---|---|---|
| R | 0.9661 | 0.992 | 0.9578 | 0.9696 |

The data are brought in for training, and the results are shown in the above Table 6. The corresponding regression analysis is made on the predicted value and the true value, and the regression coefficients are greater than 0.9 and close to 1, so the goodness of fit is high and the model fitting effect is good. This results in the neural network shown in the Figure 8.

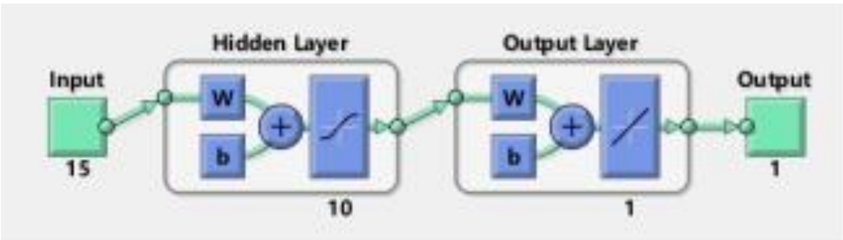


**Figure 8.** Neural network structure diagram

According to the fitted model, the chemical composition of the unknown glass relics was analyzed to identify their types, in which the cultural relics numbered A1, A6 and A7 were judged to be high-potassium glass, and the cultural relics numbered A2, A3, A4, A5 and A8 were judged to be lead-barium glass.

### 4.3. Sensitivity of classification results

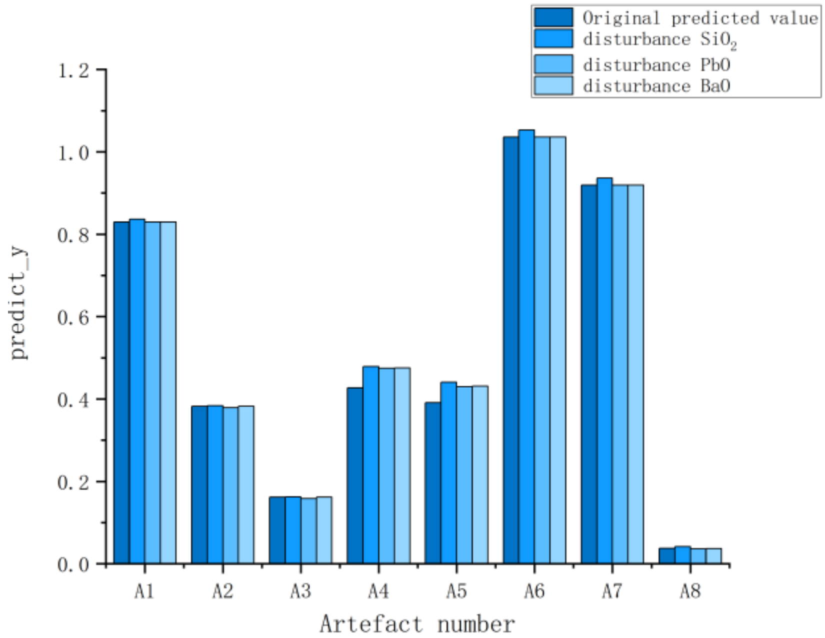


**Figure 9.** System clustering sensitivity analysis results

According to the chemical composition analysis of glass relics, $SiO_2$, PbO and BaO are sensitive parameters with obvious influence. In this paper, the data of these parameters are selected for 1% data perturbation. The sensitivity test of the above discrimination classification analysis is performed, and the results are shown in Figure 9, after the key chemical components are disturbed by 1%, the classification results of the obtained samples are the same as the original classification results, indicating that the sensitivity of the model is good and the model has good stability.

## 5. Conclusion

In this paper, the chi-square test and multiple correspondence analysis are used to analyze the relevant variables of the sample data, and it is concluded that whether the cultural relics are weathered or not has a significant impact on the glass type. Mann-Whitney U was used to test the difference in chemical composition between the two kinds of glass before and after weathering, and the corresponding statistical law of chemical composition was provided, which provided a valuable reference for the follow-up study of cultural relics. Since the 1980s, with the development of materials science in China, significant progress and breakthroughs have been made in the research of glass reference materials in China. Based on the chemical composition of glass relics, this work applies a clustering algorithm to sub-classify glass and uses a neural network model to identify unknown types of glass, which has a good reference value for the composition analysis and identification of glass relics. The purpose of this study is to improve the identification efficiency of glass cultural relics and the production efficiency of modern glass and to improve the economic and social benefits.

## References

[1] Chen Songlin. Research and application of reference materials for glass composition analysis [J]. China Petroleum and Chemical Industry Standards and Quality, 2022, 42 (23): 135 - 137.

[2] An Jiayao. Silk Road and Glassware [J]. Cultural Relics World, 2021, No.366 (12): 87 - 92.

[3] CHENG Qian, GUO Jinlong, Wang Bo, Cui Jianfeng. Analysis of the composition of glassware unearthed at the end of the Silk Road by LA-ICP-AES [J]. Spectroscopy and Spectral Analysis, 2012, 32 (07): 1955 - 1960.

[4] Wen Rui, Zhao Zhiqiang, Ma Jian, Wang Jianxin. Composition analysis of glass beads unearthed from Shirenzigou site group in Balikun, Xinjiang [J]. Spectroscopy and Spectral Analysis, 2016, 36 (09): 2961 - 2965.

[5] Chen Songlin. Research and application of reference materials for glass composition analysis [J]. China Petroleum and Chemical Industry Standards and Quality, 2022, 42 (23): 135 - 137.

[6] Wang Chengyu, Tao Ying. Weathering of silicate glass. Journal of Silicate, 2003 (01): 78 - 85.

[7] Zhang Haiyan, Liu Yan, Ma Limeng, yuan Jinsha, Ju Hanji, Wei Tongjia. Research on comparison and application of decision tree algorithm. North China Electric Power Technology, 2017, No.452 (06): 42 - 47.

[8] Wang Sen, Liu Chen, Xing Shuaijie. Review of K-means clustering algorithm [J]. Journal of East China Jiaotong University, 2022, 39 (05): 119 - 126. DOI: 10. 16749/J. CNKI. Jecjtu. 20220914.

[9] Lu Wei. Machine Learning-Formula Derivation and Code Implementation [M]. Beijing: People's Posts and Telecommunications Press, 2022.

[10] Yang Shuying. Pattern Recognition and Intelligent Computing MATLAB Technology Implementation. Fourth Edition. Beijing: Press of Electronics Industry, 2019.