Artificial Intelligence Technologies and Applications
C. Chen (Ed.)

doi:10.3233/FAIA231389

An Ancient Glass Component Analysis Identification Model Based on Correlation Analysis and Neural Network Algorithm

Jiamin JIANG[1], Jinchen YANG, Xiao CHEN
College of Internet of Things Engineering Hohai University
Nanjing, Jiangsu, China

Abstract. Identification models for glassware are important because the influence of the burial environment on the degree of weathering of old glass can lead to an inaccurate assessment of its category. This paper explores the relationship between glass surface weathering and decoration, color and type by **Chi-square test** and **Fisher's exact test**. The experiment results show that the surface weathering degree of glass is not related to its pattern and color, but is significantly related to its type. This paper has introduced a new method to obtain important characteristic factors for distinguishing high potash glass and barium lead glass and for subclassification by **Random Forest Model**, so as to carry out **R-type clustering** to obtain classification results with 95% accuracy and high sensitivity. To identify the unknown category of cultural relics, the method of applying the **BP neural network** with the Bayesian regularization algorithm for training is proposed to improve the generalization ability of the model. In addition, based on the **Spearman correlation coefficient**, the differences in chemical composition correlation between different types of glass are compared in this paper. In conclusion, this paper explores a method for modeling and validation of glass component data and constructs an identification model that takes the accuracy of classification and prediction results and high robustness into account.

Keywords. BP neural network, Random Forest Model, Spearman correlation coefficient

1. Introduction

Different fluxes added to ancient refined glass in different regions would result in different chemical compositions and types [1]. The chemical composition and colour of glass products will also change as a result of weathering of the external environment. Therefore, it is important to study Chinese Silk Road culture in order to analyse the composition and identify the types of ancient glassware.

According to the data of a batch of ancient glass products collected by archeologists, the samples were divided into lead barium glass and high potassium glass from the aspects of chemical composition. Based on the above background and data, this paper establishes mathematical models to solve the following problems. **I.** It is necessary to study the relationship between the surface weathering of ancient glass

[1] Corresponding author: Jiamin JIANG, College of Internet of Things Engineering Hohai University; Nanjing, Jiangsu, China; email: 2062910102@hhu.edu.cn

products and their types, colors and patterns, to find out the classification basis of lead barium glass and high potassium glass, and to further classify by different chemical composition content. **II.** The sensitivity of the type and classification results is determined by analyzing the component content of unknown categories of cultural property. **III.** It is essential to analyze the correlation between the chemical composition content of ancient glass objects of different categories and to compare the differences in the correlation between the chemical composition of glass objects of different categories.

2. Materials and Method

The sum of the proportions of the components and data between 85 % and 105 % are considered reliable data. In addition, the weathered glass sampling points are divided into two categories: weathered and unweathered, and the unweathered points of weathered glass are included in the unweathered situation for subsequent analysis.

2.1 Classification Law of Glass

Firstly, we check whether the sample size is large enough. If so, Chi-square test can be applied to compare the degree of correlation between the surface weathering of cultural relics and the type, pattern and colour of glass; otherwise, Fisher's exact test should be adopted.

This paper takes the strategy of discussing unweathered and weathered glass respectively, and draws three-dimensional scatter plots for the contents of the three most important chemical components affecting high potash glass and barium lead glass under two conditions, so as to obtain different clustering centers, and further analyzes the classification rules of high potash glass and barium lead glass combined with the images. As for the selection of chemical component indexes affecting the classification of high potash glass and barium lead glass [2], this paper chooses to establish a random forest classification model [3, 4] to calculate the characteristic importance of different chemical components. Similar to the above, the random forest model is established to select 5 chemical composition characteristics indicators that affect glass color, texture and weathering, and then the R-type clustering of Ward minimum variance method is carried out to obtain the classification results. The sum of squares of deviation for i-th glass relics specimens is as follows:

$$S_i = \sum_{j=1}^{n_i} \left(E_i^{(j)} - \bar{E}_i\right)' \left(E_i^{(j)} - \bar{E}_i\right). \tag{1}$$

At last, it is judged whether there is a strong correlation between other characteristic indicators except for these 5 variables. If so, it can explain the rationality of the classification results. For the analysis of the sensitivity of the model, the random function was used to randomly select 10 values and offset the original data by 0.5%~10% to test the similarity between the model and the initial classification result. The higher the similarity of the result, the lower the sensitivity and the higher the stability of the model.

2.2 Glass type identification

In order to better dig and analyze the implicit relationship in the data, this paper adopts the One-Hot Encoding to encode the variables, and adopts the BP neural network model [5, 6] for learning. The loss function and cost function are defined as follows:

$$L(y,y) = -y\log y - (1-y)\log(1-y) \tag{2}$$

$$J(w,b) = \frac{1}{m}\sum_{i}^{m} L(y,y) \tag{3}$$

in the formula, y represents sample label values, and w and b represent model construction as input parameters.

Then, the sensitivity analysis of the model was carried out, and 30 data were randomly selected from 15*8 data, and the prediction test was performed again by changing certain values. When the values changed from 0.5% to 50%, the accuracy rate remained above 80% all the time, and most of them were 87.5% or above, which proved that the model had extremely high stability.

2.3 Chemical Composition Difference

It was found that the chemical composition content data of the glass cultural relics did not conform to a normal distribution and the relationship between the two variables was not linearly correlated, so Spearman correlation coefficient [7] was used for research. The greater the absolute value of the correlation coefficient, the stronger the correlation.

3. Discussion and Results

3.1 Classification Rule Description

By chi-square test on the correlation between glass type and surface weathering, the following results are obtained in Table 1:

Table 1. Surface Weathering * Type Crosstab.

Type	Weathered Count	Weathered Proportion	Unweathered Count	Unweathered Proportion
High potassium glass	6a	33.3%	12a	66.7%
Lead barium glass	28b	70.0%	12b	30.0%

The experiment results showed that among the 34 groups weathered on the surface, 28 groups are barium lead glass, and 6 groups are high potash glass. Therefore, it can be assumed barium lead glass is more easily weathered.

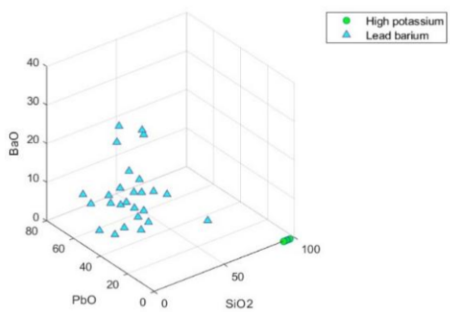
Then, Fisher's exact test was carried out on the correlation between glass pattern, color and surface weathering, and it was concluded that surface weathering is related to the type of glass. The surface of lead barium glass is easy to weather, while the surface of high potassium glass is not. By referring to relevant research, it is found that the surface weathering of glass relics is determined by glass composition, burial humidity, burial temperature, burial time, and atmosphere, but has little correlation with color, which supports the conclusion of this paper [8].

Three important features of weathering and non-weathering glass surfaces selected in this paper are as follows in Table 2:

Table 2. Important factors of weathering chemical characteristics on surfaces.

Weathering surface		Unweathered surface	
Feature name	Feature importance	Feature name	Feature importance
PbO	30.10%	PbO	27.70%
SiO_2	23.50%	K_2O	14.80%
BaO	14.30%	SiO_2	11.20%

Then, the contents of the three most important chemical components affecting the classification of high potash glass and barium lead glass under the two conditions of weathering and unweathering of the glass relics surface are drawn as follows in Fig.1 and Fig.2.



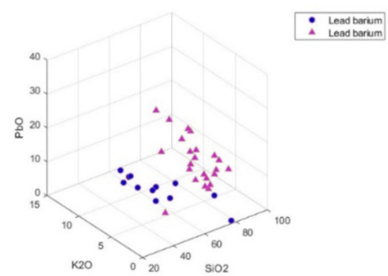


Fig.1 Weathering samples' chemical composition **Fig.2** Unweathered samples' chemical composition

The classification rules of high potash glass and barium lead glass can be analyzed from the above two figures.

For the glass relics samples weathered on the surface, compared with barium lead glass, high potash glass has significantly lower lead oxide content (high potash glass 0%, barium lead glass 40%), significantly higher silica content (high potash glass 90%, barium lead glass 35%), and significantly lower barium oxide content (high potash glass 0%, barium lead glass 10%).

For the glass relics samples without weathering on the surface, the content of lead oxide in high potash glass is significantly lower than that in barium lead glass (0% in high potash glass, 20% in barium lead glass), the content of potassium oxide is higher (10% in high potash glass, 4% in barium lead glass), and the content of silica is similar to that in barium lead glass.

According to Wang's work [9], the firing of ancient Fionse can be divided into PbO-BaO-SiO2 (lead barium), K2OCaO-SiO2 (high potassium) and other systems. Lead barium glass with lead oxide content is much higher than high potassium glass, and high potassium glass basically does not contain lead oxide. Potassium glass with high potassium oxide content is much higher than barium lead glass, and barium lead glass basically does not contain potassium oxide, which also proves that the conclusion of this paper is highly accurate.

All the glass cultural relics samples are divided into high potash glass and barium lead glass and renumbered. R-clustering method is used to divide the sub-categories. The clustering results are as follows in Fig.3 and Fig.4.

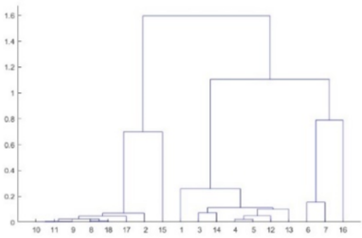


Fig.3 Clustering results of high potassium

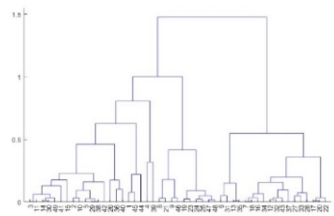


Fig.4 Clustering results of lead barium

The horizontal axis represents the renumbering of the glass cultural relic samples.

For all the sample data of high potash glass and barium lead glass, 10 values are randomly selected to increase or decrease by a certain multiple. In this paper, 0.005 is taken as the gradient to draw the following line graph (Fig.5 and Fig.6) where the vertical axis represents the accuracy rate and the horizontal axis represents the shrinkage multiple.

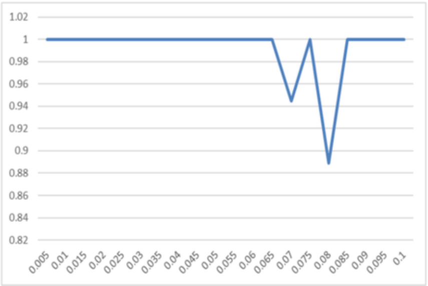


Fig.5 Sensitivity analysis of high potassium glass model

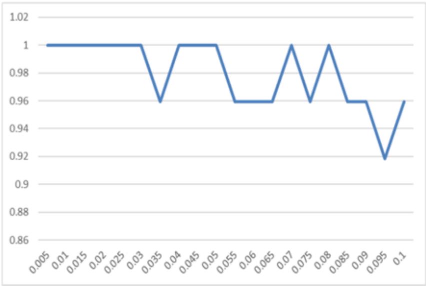


Fig.6 Sensitivity analysis of lead barium glass model

From the above two figures, it can be seen that for high potash glass, when the increase or decrease multiple of 10 values is randomly selected from 0.005 to 0.1, the accuracy of classification results remains above 88%, and most of the accuracy can reach 100%. For barium lead glass, the accuracy of classification results remains above 90%, and most of the accuracy is above 95%, when the increase or decrease ratio of ten random values is increased from 0.005 to 0.1. Combined with the above sensitivity analysis results, the conclusion can be drawn: the classification results obtained by principal component analysis with the R clustering method in this paper show that the sensitivity of the model is weak and the model stability is high.

3.2 Identification result

After model training is completed, model parameters are tested, and the results are as follows in Fig.7.

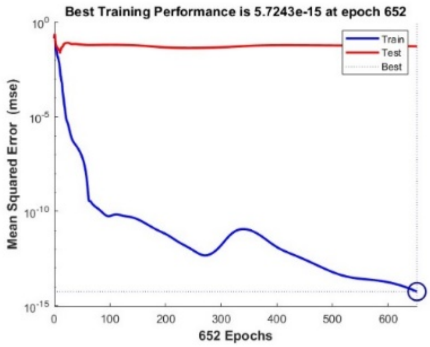


Fig.7 Optimal training results in the 908 round

The training effect reached the optimum in the 908 round, and this generation model will be used as the final training model in this paper. The relationship between the gradient of training figures and the mean square error is shown in the in Fig.8 below.

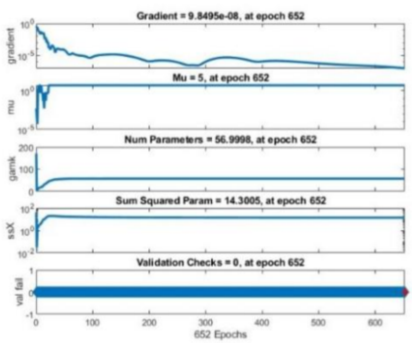


Fig.8 Gradient and mean square error of training figures

It can be seen from the figure above that the gradient and mean square error of training figures are gradually stable with the increase of algebra, and it can be clearly seen that the results of the training model are relatively stable. It can be seen from the results in the following figure that the gradient and mean square error of the training data are gradually stable with the increase of algebra, and it can be clearly seen that the results of the training model are relatively stable.

As a test set, the following historical residuals of the training set, the test set and all data after model training are drawn respectively in Fig.9.

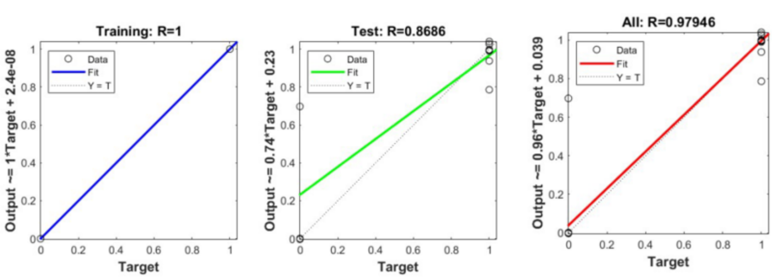


Fig.9 Historical residuals of training data

It can be seen from the figure that the R values of the three groups of data are close to 1, which proves that the goodness of fit of the model is very high.

3.3 Chemical Composition Difference Analysis

The heat map of the correlation coefficient drawn by Spearman correlation coefficient of all chemical components in high potash glass and barium lead glass is shown in the figure below in Fig.10 and Fig.11.

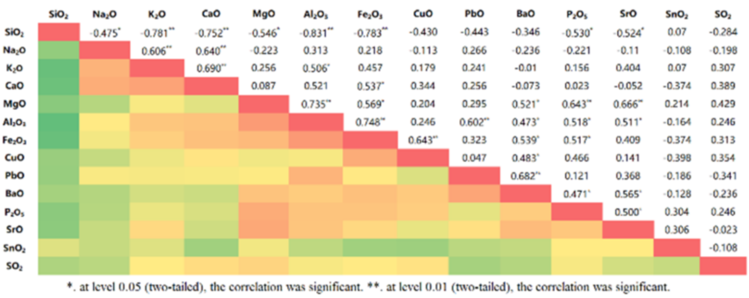


Fig.10 Correlation coefficient of high potassium glass chemical composition

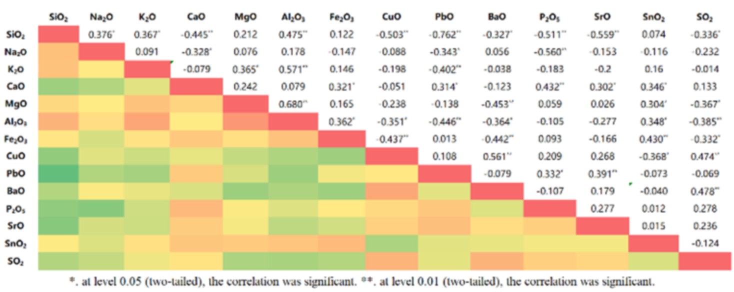


Fig.11 Correlation coefficient of Pb barium glass chemical composition

By comparing the two figures above, this paper summarizes the differences in the following chemical composition correlation. Firstly, the correlation coefficient of the chemical composition content of the high potash glass is usually relatively high on the whole, while the correlation coefficient of the chemical composition content of the barium lead glass is relatively low, which indicates that the chemical composition content of the high potash glass is more closely related. Secondly, the content of silica and sodium oxide in high potassium glass presents a negative correlation, while the content of silica and sodium oxide in lead barium glass presents a positive correlation. Thirdly, potassium oxide content in high potassium glass is strongly correlated with calcium oxide content, while potassium oxide content in lead barium glass is weakly correlated with calcium oxide content [10].

4. Conclusions

This paper proposes an ancient glass component analysis identification model based on correlation analysis and neural network algorithm to identify the type of antique glass accurately according to the chemical composition of cultural relic samples. Through analysis, it is concluded that the weathering degree of glass surface can be identified according to its glass type, while the influence brought by its pattern and color can be ignored. In addition, when using BP neural network, Bayes regularization algorithm is used for training, so as to improve the promotion ability of BP network, so that the model can process the data more accurately and dig out the relationship between the data more deeply. It is practical to analyze and identify the composition of ancient glass.

References

[1] Ma,Q.,Braekmans,D.,Shortland,A.Pollard,M.,2021,The Production and Composition of Chinese Lead-Barium Glass through Experimental Laboratory Replication,Journal of Non-Crystalline Solids,551:0022-3093
[2] Xingling, T.; Jianyu, W.; Yong, C. A study of glass beads recovered from the Ming Dynasty shipwreck of Nan'ao 1. Cult. Relics 2016, 12, 87–92.
[3] Menze, B.H., Kelm, B.M., Masuch, R. et al. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. BMC Bioinformatics 10, 213 (2009).
[4] Zhang Hua, Tao Liyuan, Zhao Yiming. The principle of random forest algorithm and its application in clinical research [J]. Chinese Journal of Pediatrics, 201, 59(09):798-798.
[5] Qin, X.; Liu, Z.; Liu, Y.; Liu, S.; Yang, B.; Yin, L.; Liu, M.; Zheng, W. User OCEAN Personality Model Construction Method Using a BP Neural Network. Electronics 2022, 11, 3022.
[6] Si Shoukui, Sun Xijing. Mathematical Modeling Algorithm and Application [M]. Beijing: National Defense Industry Press, 2016-2.
[7] Chen Zhihao, Ji Jingmin. Based on hierarchy clustering model of ancient glass composition analysis and evaluation [J]. Journal of modern information technology, 2023, 7(8):122-125. The DOI: 10.19850/j.carol carroll nki. 2096-4706.2023.08.031.
[8] Justino de Lima, C., Aldinger, B., de Haan, P. et al. Effects of composition on the durability and weathering of flat glass. Glass Struct Eng 7, 139 – 155 (2022).
[9] Wang Yingzhu. Research on Biweekly Fions technique [D]. University of Science and Technology Beijing, 2019.
[10] Huang Xiaojiao. A Comparative study on ancient glass ware between China and the West in National Museum of China [J]. Cultural Relics World, 2023, 381(03):86-91.