

# A study of ancient glass based on a modeling perspective

Xiaohan Wang<sup>\*</sup>, Jingwen Li, Lingong Dai

Department of Stony Brook Institute, Anhui University, Hefei, China, 230039

<sup>\*</sup> Corresponding author: r22014052@stu.ahu.edu.cn

**Abstract.** As a valuable physical evidence of the cultural exchange between China and the West in ancient times, ancient glass is of great research significance. However, ancient glass is highly susceptible to weathering by the burial environment, and previous studies have only applied statistical methods for simple classification and generalization. In order to further study the composition analysis and prediction of ancient glass, glass subclass classification and identification, this paper constructs a data prediction model based on the theory of chi-square test, decision tree classification, and BP neural network prediction, so that the research on ancient glass can broaden the current research progress from a new perspective.

**Keywords:** Chi-square test, Decision tree classification, BP neural network prediction.

## 1. Introduction

### 1.1. Background of the study

The true origin of glass-making was introduced to China along with imported glass eye beads, which were frequently traded along the Silk Road in ancient times and are important physical evidence of early trade [1]. In the firing process of glass, combustion aids and stabilizers are added, and the addition of combustion aids and stabilizers introduces other chemical components in the glass. Among them, lead barium glass varieties have high content of lead oxide and barium oxide, and potassium glass has high potassium content. Affected by the burial environment, ancient glass is highly susceptible to changes in its composition ratio due to weathering, which further affects the judgment of its type.

### 1.2. Purpose of the study

Ancient glass is highly susceptible to weathering due to storage conditions, relative humidity and concentrations of pollutant gases [2]. From the perspective of mathematical modeling and machine learning, we analyzed and predicted the composition of ancient glass, classified subsets of glass, analyze its chemical composition content through chi-square test and interval mean difference model, then analyzed the relationship between surface weathering and other properties of glass artifacts and the related statistical law, and predicted the chemical composition content before weathering. We also used decision tree classification model and BP neural network classification model to classify different subsets of ancient glass data sets, and predict the type of glass. It enables the study of ancient glass to broaden the current research progress from a new perspective.

### 1.3. Problem Study

The data in this study are derived from the forms and data provided by 2022 Higher Education Cup National College Students Mathematical Contest in Modeling.

## 2. Methods and Principles

### 2.1. Data pre-processing

First completed the missing value processing. After screening the valid data, the two types of glass are introduced into the virtual variable function (1) and converted into Quantitative variable.

$$D(t) = \begin{cases} 0, & t \text{ is the surface unweathered type} \\ 1, & t \text{ is the surface weathered type} \end{cases} \quad (1)$$

In the practical operation of the neural network algorithm, the standardized processing of the original data is usually carried out to eliminate the impact of the dimension and order of magnitude of each index data on the algorithm [3]. The heritage glass is divided into four types: high potassium unweathered, lead-barium weathered, high potassium weathered, and high potassium unweathered, and the dummy variable function (2) is introduced to represent the above four types with 0, 1, 2, and 3, respectively.

$$D(t) = \begin{cases} 0, & t \text{ is high potassium unweathered glass} \\ 1, & t \text{ is lead-barium unweathered glass} \\ 2, & t \text{ is high potassium weathered glass} \\ 3, & t \text{ is lead-barium weathered glass} \end{cases} \quad (2)$$

## 2.2. Introduction to the method

To determine the correlation between the surface of glass and three features, we used Chi-square tests which are commonly used for tests of fit of measurement models [4] mainly to analyze the correlation between two fixed categories of variables. Secondly, the glass artifacts were further classified based on the above results using the gray correlation analysis step whose analysis for the purpose of objectiveness and embodying be individual advantage of each alternative [5].

Once the classification results are obtained, there is a correlation between the weathering of the surface and their chemical type, the glass artifacts can be subdivided into subclasses according to their chemical content.

In this model, its goal is to learn how to classify objects or situations by analyzing a set of instances whose classes are known [6]. A decision tree is a tree-like structure 'which starts from the root node' and tests a data sample (consisting of a set of instances 'which have several attributes') and divides the data sample into different subsets of data samples depending on the results ' Each subset of data samples constitutes a sub-node. It is the process of classifying the data by a set of rules.

Decision tree is essentially a top-down stepwise construction method, which generally uses information gain metric in the construction process. The maximum information gain indicates that the dataset is able to minimize its uncertainty during the classification process, so the features selected by ID3 in the process of constructing the algorithm have better classification results. The information first (H) and information gain (G) can be defined as follows:

$$H(p) = -\sum p \times \log_{10} p \quad (3)$$

$$H(Y|X) = \sum_{i=1}^n p_i H(Y|X = X_i) \quad (4)$$

$$G(D, A) = H(D) - H(D|A) \quad (5)$$

Where p denotes the probability of a random variable, A denotes the feature, D represents the data set, and H(D) is defined as the empirical entropy. H(Y|X) Defined as the bar sh that H(D|A) denotes the empirical condition of feature A under the bars of data set D.

BP neural network is a kind of typical forward network, composed of input layer, hidden layer and output layer [7]. The prediction accuracy of BP neural network is greatly affected by the number of

neurons in the hidden layer. The empirical formula for selecting the number of neurons in the best hidden layer is as follows:

$$l < \sqrt{(m+n)} + a \tag{6}$$

Where:  $l$  is the number of nodes in the hidden layer;  $m, n$  are the number of nodes in the input layer and output layer;  $a \in (0,10)$ .

After network is decided, BP network studies and modifies the connecting weight and threshold value among neural units, according to the input and output of example.

The output-input relationship of the neuron  $h$  is expressed as (7).

$$net_h = \sum_{h=1}^n w_{ih} x_h \tag{7}$$

$$y_h = f(net_h) \tag{8}$$

Where  $X = (x_1, x_2, \dots, x_n)^T$  and  $Y = (y_1, y_2, \dots, y_n)^T$  are the input variables and output results, respectively,  $w_{ih}$  are the weights between the input and implied layers, and  $f$  is the activation function. In this model, softmax function (9) is selected as the activation function and cross entropy (10) is the mathematical logic of loss function to evaluate the prediction results.

$$y_k = \frac{e^{a_k}}{\sum_{i=1}^n e^{a_i}} \tag{9}$$

$$Loss = -\sum_i^k y_i \log_e S_i \tag{10}$$

### 3. Modeling and Analysis of Results

First, the preliminary analysis of the data can determine whether the surface weathered, and its type may have a certain correlation. The cardinality test in the nonparametric test was chosen to further determine the correlation among them, and the test results are shown in Table 1.

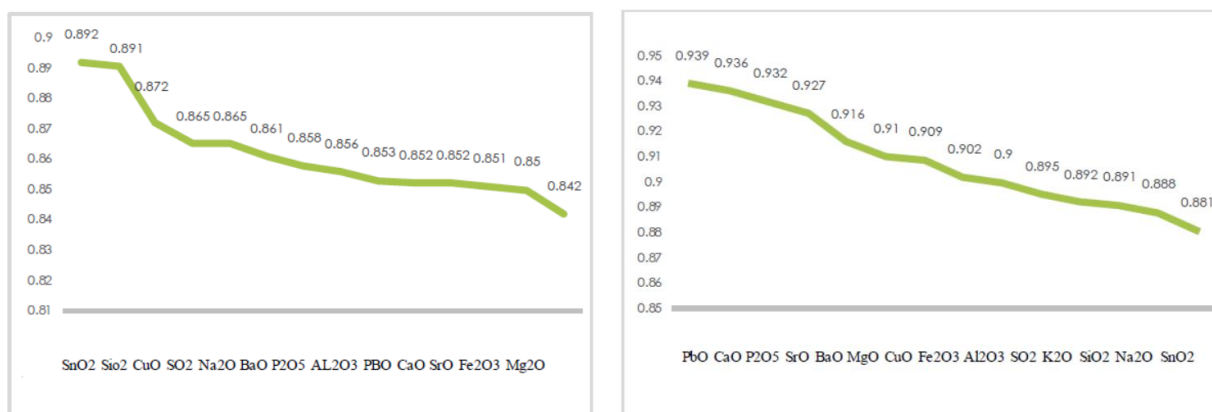
**Table 1.** Results of the cardinality test analysis of glass color, type and weathering

Title	Name	Surface weathering		Total	X <sup>2</sup>	Correction X <sup>2</sup>	P
		No weathering	Weathering				
Ornamentation	A	11	9	20	5.747	5.747	0.056*
	B	0	6	6			
	C	13	15	28			
Type	Lead Barium	12	24	36	5.400	4.134	0.020**
	High Potassium	12	6	18			
Color	light green	2	1	3	6.287	6.287	0.507
	Pale Blue	8	12	20			
	Dark Green	3	4	7			
	Deep Blue	2	0	2			
	Purple	2	2	4			
	Green	1	0	1			
	Blue-Green	6	9	15			
Black	0	2	2				

Note: \*\*\*, \*\*, \* represent 1%, 5%, 10% significance levels, respectively

The p-value of significance of whether the surface of glass artifacts is weathered or not with ornamentation and color is 0.056 and 0.507 respectively, so there is no significant relationship between ornamentation and color of glass artifacts and weathering. The p-value of significance for weathering and type of glass artifacts is 0.020, which is significant at the level, thus indicating that there is a correlation between weathering and type of glass artifacts.

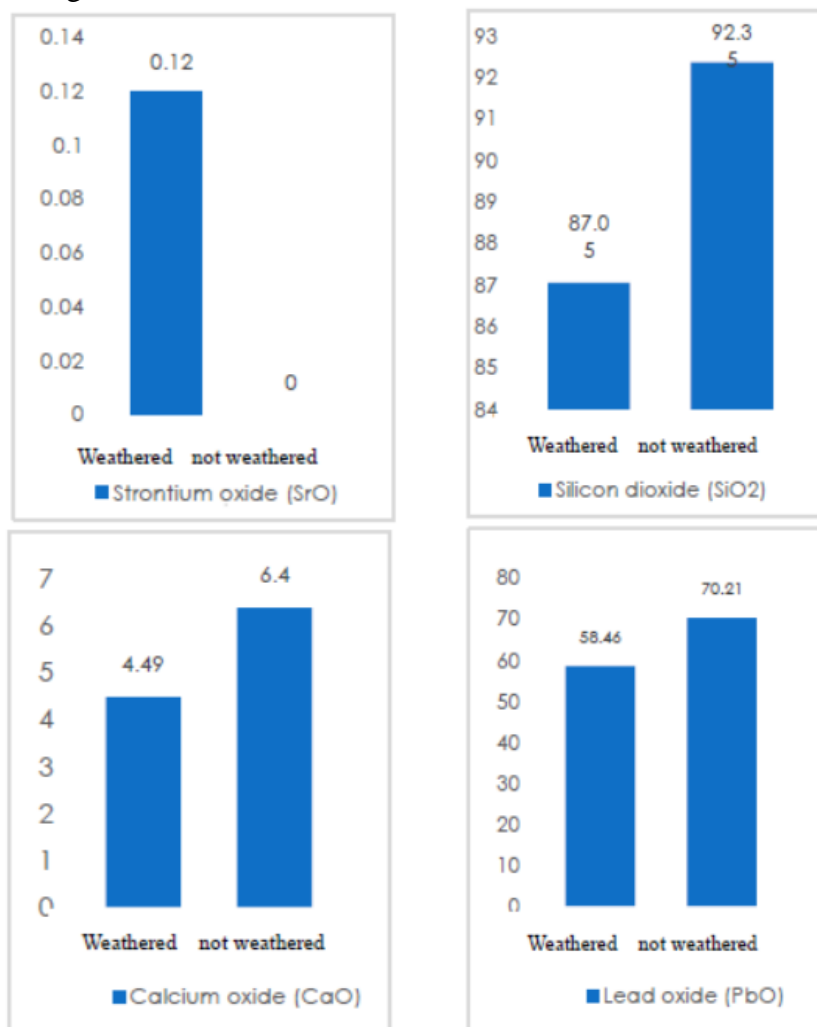
To further explore the statistical patterns of the chemical composition content of glass artifacts whether weathered or not, the glass artifacts were first divided into high-potassium glass and lead-barium glass according to the glass type. Gray correlation analysis was used to assess the degree of association between the chemical composition of glass artifact samples and whether the surface of glass artifacts is weathered or not. The results are shown in Figure 1.



**Figure 1.** Correlation results between the chemical composition content of high potassium glass and whether the surface is weathered or not and correlation results of chemical composition content of lead barium glass and whether the surface is weathered or not

From Figure 1, the correlation between strontium oxide and silica in high potassium glass is closest to 1, and the decreasing trend of chemical composition correlation after silica gradually tends to be smooth, that is, it shows that the correlation between them and whether the surface of high potassium glass artifacts is weathered or not is the strongest, so strontium oxide and silica are selected as the statistical law of chemical composition content of high potassium glass with or without weathering. Similarly, in lead-barium glass, lead oxide and calcium oxide were selected as the statistical law of the chemical composition content of lead-barium glass with and without weathering. To further summarize the specific statistical laws based on the most correlated chemical composition content

changes before and after weathering of the two glass types, the maximum values of strontium oxide and silica when the high potassium glass was unweathered, the minimum values of strontium oxide and silica after weathering of the high potassium type glass, the maximum values of calcium oxide and lead oxide when the lead barium glass was unweathered and the minimum values of calcium oxide and lead oxide after weathering of the lead barium glass were calculated respectively. The results are shown in Figure 2.



**Figure 2.** Comparison of strontium oxide and silica threshold content before and after weathering of high potassium glass and comparison of the threshold content of calcium oxide and silica before and after weathering of lead-barium glass

From the result comparison of in Figure 2, we knew that if the content of strontium oxide or silica in high potassium glass is less than or equal to 87.05, it is a glass sample with not weathered high potassium surface, and if the content of strontium oxide or silica in high potassium glass is not present or greater than or equal to 92.35, it is a glass sample with unweathered high potassium surface. If the content of calcium oxide is less than or equal to 4.49 or the content of lead oxide is less than or equal to 58.46, then the glass sample is of the unweathered type of lead-barium surface, and if the content of calcium oxide is greater than or equal to 6.4 or lead oxide is greater than or equal to 70.21, then the glass sample is of the weathered type of lead-barium surface.

In the study of subcategory classification of glass products, the above idea of classifying glass products according to their chemical content is continued. We used the method of Cluster analysis whose aim of clustering is to find structure in data and is therefore exploratory in nature [8] to analyze.

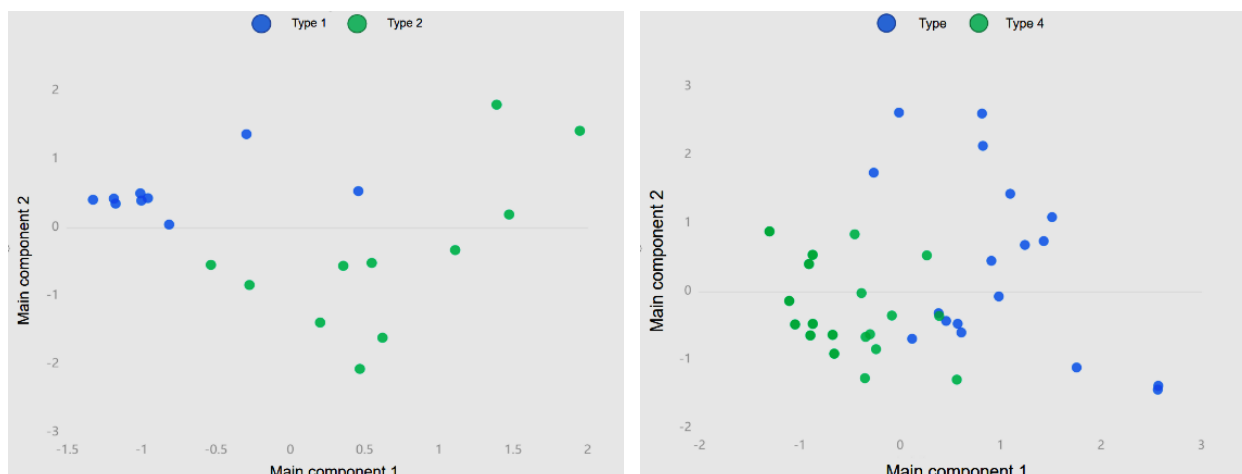
In general, partitional methods suppose that the data set can be represented by finite cluster prototypes with their own objective functions. Therefore, defining the dissimilarity (or distance) between a point and a cluster prototype is essential for partition methods [9].

It is known that the k-means algorithm is the oldest and popular partitional method, by using the fviz\_nbclust function in R language for judgment, as shown in Figure 3, according to the function image interruption point can be known high potassium, lead barium two types of glass sub the classifications are finally selected as 2 classes.



**Figure 3.** Comparison of the number of clusters of high potassium glass and comparison of the number of clusters of lead barium glass

The clustering analysis was performed by K-means with the number of clusters determined to be 2. The results are shown in Figure 4 for high potassium glass and lead-barium glass, respectively.



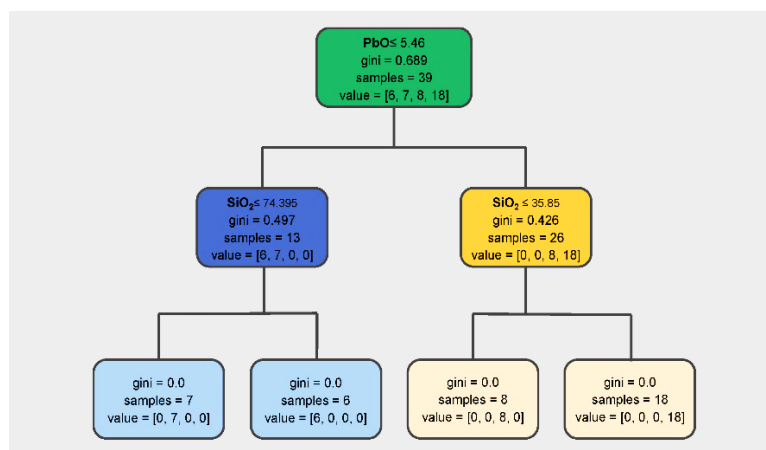
**Figure 4.** Cluster scatter plot of high potassium glass and cluster scatter plot of lead barium glass

According to the results of the cluster analysis, compact clusters indicated good registration [10], so the two categories of the original dataset were expanded into four categories, as shown in Table 2.

**Table 2.** Cluster analysis results of ancient glass

Type of Glass	Subcategories	Glass number
High Potassium Glass	Type_1	13 06Part2 17 05 03Part215 14 16 04 06Part1 01
	Type_2	21 18 03Part1 22 07 27 12 09 10
Lead Barium Glass	Type_3	43part1 40 50 41 26 08 51part2 43part2 54 51part1 57 52 39 49 56 19 58 38
	Type_4	24 11 30part1 34 02 30part2 20 36 50 55 25 421 422 47 48 23 49 46 37 44 45 29 53 31 35 28 32 33

For the ancient glass data set of this experiment according to the results of cluster analysis, 60% of the data were taken as the training set and 40% of the data were taken as the test set to obtain the decision tree shown in Figure 4. The results of fitting the glass subspecies classification with the decision tree classification model are shown in Figure 5.



**Figure 5.** Decision tree classification results

As shown in the Figure, the classification pattern of ancient glass types is mainly determined by the content of PbO and SiO<sub>2</sub>. According to these two components, this experiment classifies ancient glasses into four subclasses.

- 1) PbO ≤ 5.46% & SiO<sub>2</sub> ≤ 74.395% ---- high potassium weathered glass
- 2) PbO ≤ 5.46% and SiO<sub>2</sub> > 74.395% ---- high potassium unweathered glass
- 3) PbO > 5.46% and SiO<sub>2</sub> ≤ 74.395% ---- lead-barium weathered glass
- 4) PbO > 5.46% and SiO<sub>2</sub> > 74.395% ---- lead-barium unweathered glass

The subclass classification method obtained from the above study allows the prediction of the glass type to which the unknown samples belong.

The input parameters of the BP neural network classification and regression model in this study are the determined values of 14 chemical components such as silica, oxide, etc. The output parameters of the BP neural network regression model are the classification types represented by the dummy variables of the test samples in the test set, i.e., 0, 1, 2, 3.

As it is shown in Table 3, the rate of model classification and the ratio of  $R^2$  to correct rate are all close to one. The model fit effect is excellent, it can be used as a test set of the model for its glass type prediction evaluation.

**Table 3.** Performance evaluation of glass type prediction models for heritage samples

	BP neural network regression $R^2$	BP neural network classification correct rate	$R^2$ /correct rate
Training set	0.816	0.962	0.848
Test set	-0.767	0.862	0.890

Then the learned model is used to data of the content of each chemical component for predicting and classifying. The output of the model prediction is shown in Table 4.

**Table 4.** Predicted results for glass types of unknown artifact samples

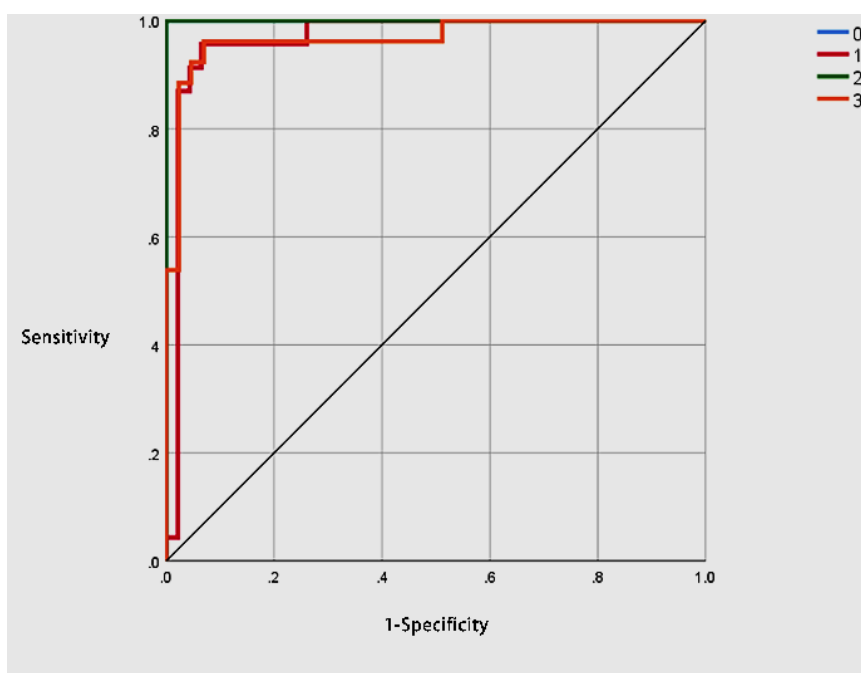
Artifact Number	Prediction Type	Predicted results	Predicted outcome probability_0	Predicted outcome probability_1	Predicted outcome probability_2	Predicted outcome probability_3
A1	High potassium unweathered	0	0.937	0.003	0.060	0.000
A2	Lead-barium weathering	3	0.000	0.082	0.000	0.918
A3	Lead-barium unweathered	1	0.000	0.990	0.000	0.010
A4	Lead-barium unweathered	1	0.000	0.993	0.000	0.007
A5	Lead barium unweathered	1	0.030	0.837	0.001	0.133
A6	High potassium weathering	2	0.039	0.001	0.960	0.000
A7	High potassium weathering	2	0.130	0.003	0.867	0.000
A8	Lead-barium unweathered	1	0.000	0.982	0.000	0.018

The identification results are: A1-high potassium unweathered glass; A2-lead-barium weathered glass; A3-lead-barium unweathered glass; A4-lead-barium unweathered glass; A5-lead-barium unweathered glass; A6-high potassium weathered glass; A7-high potassium weathered glass; A8-lead-barium unweathered glass.

According to the results in Table 5 and Figure 6, the recall rate of the model classification results is high, indicating that the classification results are highly sensitive, and the prediction ability of the model is good.

**Table 5.** Model evaluation results

	Accuracy	Recall Rate	Accuracy rate	F1
Training set	1	1	1	1
Test set	0.857	0.857	0.864	0.858



**Figure 6.** ROC graph

#### 4. Conclusion

There is a correlation between the surface of glass artifacts is weathered or not and its type. The statistical law of the change of chemical composition content whether weathering is derived from the chemical composition of each of the two types of glass with high potassium and lead-barium corresponding to a greater correlation, and the results are shown in Appendix A. Based on the comparative analysis, it is possible to classify the ancient by introducing dummy variables to classify the glass types, the four well-classified glass types were used as dependent variables and the 14 chemical compositions as covariates, and a BP neural network classification model was used to predict the unknown types of glass.

#### References

- [1] LI H. Origin of the art of glass-making in China [J]. Glass Technology-European Journal of Glass Science and Technology Part A, 2018, 59 (1): 20 - 8.
- [2] CHOPINET M H, VERITA M, FALCONE R, et al. Soda-lime-silica glass containers: chemical durability and weathering products; proceedings of the 9th Conference of the European-Society-of-Glass-Science-and-Technology/Annual Meeting of the International-Commission-on-Glass, Trencin, SLOVAKIA, F Jun 22-26, 2008 [C]. 2008.

- [3] WU J, QI H E, XINGWEI D U, et al. Prediction of crab Yield by BP neural network based on meteorological factors of producing areas in Jiangsu Province [J]. 2020.
- [4] BERGH D. Chi-Squared Test of Fit and Sample Size-A Comparison between a Random Sample Approach and a Chi-Square Value Adjustment Method [J]. Journal of applied measurement, 2015, 16 (2): 204 - 17.
- [5] JING H, ZHAO X A. The application of improved grey multi-hierarchical comprehensive evaluation method to the growth ability evaluation of technological small and medium enterprises; proceedings of the 26th Chinese Control Conference, Zhangjiajie, PEOPLES R CHINA, F Jul 26-31, 2007 [C]. 2007.
- [6] QUINLAN J R. Learning decision tree classifiers [J]. ACM Computing Surveys (CSUR), 1996, 28 (1): 71 - 2.
- [7] LIU L H, CHEN J, XU L X. Realization and application research of BP neural network based on MATLAB; proceedings of the International Seminar on Future BioMedical Information Engineering, Wuhan, PEOPLES R CHINA, F Dec 18, 2008 [C]. 2008.
- [8] JAIN A K. Data clustering: 50 years beyond K-means [J]. Pattern Recognition Letters, 2010, 31(8): 651 - 66.
- [9] KRISHNA K, MURTY M N. Genetic K-means algorithm [J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 1999, 29 (3): 433 - 9.
- [10] ORCHARD J. Multimodal image registration using floating regressors in the joint intensity scatter plot [J]. Medical image analysis, 2008, 12 (4): 385 - 96.