

Glass classification and identification based on systematic clustering and BP neural network algorithm

Ruoxin Lu ^{1,*}, Luyao Wang ¹, Hanbo Wang ²

¹ School of Economics and Management, Beijing Jiaotong University, Beijing, China, 100044

² School of Computer and Information Technology, Beijing Jiaotong University, Beijing, China, 100044

* Corresponding Author Email: 20725017@bjtu.edu.cn

Abstract. As glass is buried in the ground for years, its surface has been eroded by weathering, thus changing its chemical composition and structure. In this paper, we search for the basis for the classification of high potassium glass and lead-barium glass, and on this basis, we further carry out subclass classification, and finally perform sensitivity analysis on the classification results. After establishing the glass classification model, the artifacts with unknown glass types are predicted, and the data are input into BP neural network for training and prediction to determine their artifact glass types, and sensitivity analysis is performed on the prediction results.

Keywords: Systematic Clustering, BP Neural Network, Principal Component Analysis, Glass Classification.

1. Introduction

Glass is one of the ancient crystallization of the wisdom of the working people, it is valuable either as a vessel or ornament, sought after and loved by the nobility and merchants [1]. The ancient aristocracy in China was fond of the crystal clear glass products, especially the transparent texture of glass and the sense of luster refracted by light was the reason why the ancients regarded glass as an honored treasure, therefore, glass and its production methods were introduced to China along the ancient Silk Road [2-3], and the glass production methods of China and the West promoted each other in trade and exchange. China gradually formed a production chain of high-potassium glass with Na₂O-CaO-SiO₂, K₂O-SiO₂ and PbO-BaO-SiO₂ as the main components [4], as well as lead-barium glass, by adding materials such as lead pellets and saltpeter in the glass-making process [5-6]. In recent years, glass artifacts have been unearthed throughout the country, and the study of glass image characteristics and chemical composition is important for restoring historical glass production methods, cultural customs, and trade exchanges. In the process of studying the composition of ancient glass, etc., archaeologists have found that because glass was buried underground for many years, its surface was subjected to weathering erosion, which led to changes in its chemical composition and structure. The degree of weathering corrosion is related to the burial environment and glass composition, and weathering corrosion can even cover the original appearance of the glass when it is severe, which will seriously affect our analysis and research of cultural relics [7-8]. However, processing the weathered data and using them to infer the true parameters of the past glass can help us to recreate the full picture of ancient glass [9-10].

2. Build a glass classification model

2.1. Principal component and glass classification law analysis method

To investigate the classification laws of high potassium glass and lead-barium glass, firstly, considering that the severely weathered glass cannot accurately reflect the classification laws of the two types of glass, the data of severely weathered detection points in Form II: 08, 26 and 54 severely weathered points were excluded. Second, the data were further processed by Matlab to obtain the covariance matrix, eigenvalues and eigenvectors, principal component contribution rate and

cumulative contribution rate of high potassium and lead-barium glass. Finally, the expressions of principal components were obtained. The principal component contributions of high potassium and lead-barium glass are shown in Table 1 and Table 2.

Table 1. Table of contribution of high potassium glass factor

Principal components	f1	f2	f3	f4	f5
Contribution rate	0.402978	0.18766	0.124897	0.116515	0.068542
Cumulative contribution rate	0.402978	0.590638	0.715535	0.832049	0.900591

Table 2. Table of contribution of lead barium glass factor

Principal components	f1	f2	f3	f4	f5	f6	f7
Contribution rate	0.25539928	0.210623	0.118901	0.077482	0.064827	0.060043	0.053378
Cumulative contribution rate	0.25539928	0.466022	0.584923	0.662405	0.727232	0.787275	0.840653

Here, the factors with a cumulative contribution of 85% were selected as principal components, i.e., these principal components can basically contain most of the information of the data. Therefore, f1, f2, f3, f4 and f5 were selected as principal components for high potassium glass, and the cumulative contribution of its four principal components reached 90.0%; f1 to f7 were selected as principal components for lead barium glass, and the cumulative contribution of its seven principal components reached 84.1%. Meanwhile, the linear combination of high potassium and lead-barium glass about each chemical component can be obtained based on the eigenvalues and eigenvectors. Due to the limitation of space, only the expressions of principal components with the contribution rate greater than 20% are shown, and the rest of the expressions are shown in the Appendix.

$$\text{High Potassium: } f_1 = -0.40x_1 + 0.10x_2 + 0.30x_3 + 0.28x_4 + 0.31x_5 + 0.37x_6 + 0.35x_7 + 0.21x_8 + 0.18x_9 + 0.23x_{10} + 0.27x_{11} + 0.29x_{12} - 0.04x_{13} + 0.12x_{14} \quad (1)$$

$$\text{Lead Barium: } f_1 = 0.24x_1 + 0.02x_2 + 0.22x_3 + 0.17x_4 + 0.35x_5 + 0.37x_6 + 0.27x_7 - 0.41x_8 - 0.13x_9 - 0.42x_{10} + 0.33x_{11} - 0.14x_{12} + 0.24x_{13} - 0.29x_{14} \quad (2)$$

$$\text{Lead Barium: } f_2 = -0.46x_1 - 0.31x_2 - 0.00x_3 + 0.47x_4 + 0.17x_5 - 0.10x_6 + 0.18x_7 + 0.02x_8 + 0.41x_9 - 0.05x_{10} + 0.41x_{11} + 0.25x_{12} + 0.03x_{13} - 0.06x_{14} \quad (3)$$

* Definition of principal component factors

Since the larger coefficients in front of an indicator represent the greater influence of that indicator on that principal component, combined with the expressions, the principal component factors for high potassium glass are defined as the first principal component to the fifth principal component, and the principal component factors for lead-barium glass are defined as the first principal component to the seventh principal component, respectively. Each principal component has its own characteristics, for example, the acidic oxides silicon oxide (SiO₂) and strontium oxide (SrO₂) in the expression f1 of the first principal component of high potassium contain large negative loadings, and the remaining alkaline oxides all contain similar positive loadings, which can be presumed to be related to the acidity and alkalinity of the glass matrix, and therefore, the principal component is called the matrix environmental alkaline component.

2.2. Systematic Clustering

Systematic clustering principles and steps

Based on the principal components of high potassium and lead-barium glass obtained from the above analysis, considering the small amount of data and the absence of inclusion and nested relationships, this paper uses SPSS 27 to systematically cluster each of the principal components of high potassium and lead-barium glass, respectively, so as to classify the two classes of glass into subclasses.

Systematic clustering is performed by calculating the distance between the data points of two classes, combining the two closest data points, and iterating this process repeatedly until all data points are combined into one class and generating a clustering spectrum graph, and finally obtaining the class values of the clusters. Next, the aggregation coefficients were used to determine the class values of the clusters k . Finally, the clustering of the four principal components of high potassium glass and the seven principal components of lead-barium glass were obtained based on the spectrograms. Figure 1 shows the flow chart of the algorithm for systematic clustering.

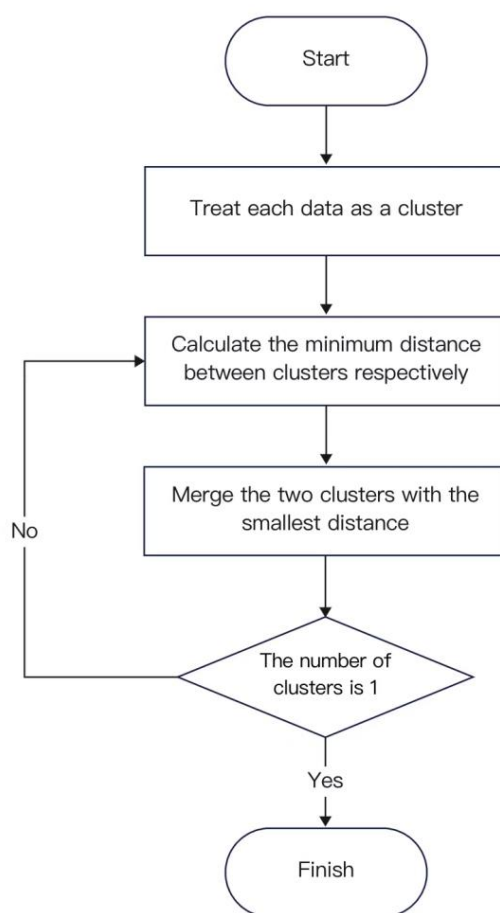


Figure 1. Algorithm flow chart of system clustering

In order to estimate the optimal number of clusters, the Elbow Method (EM) is used in this paper. The core idea is to determine the optimal number of clusters by calculating the sum of various types of distortion degrees, also called aggregation coefficients, and observing their trends. The location where the improvement in the degree of aberration decreases the most is the elbow, and the degree of aberration is generally used to determine the best value. The aggregation coefficients of high potassium glass and lead-barium glass are obtained separately using systematic clustering as line graphs, and Figure 2 and Figure 3 show the line graphs of the aggregation coefficients of high potassium and lead-barium glass, respectively.

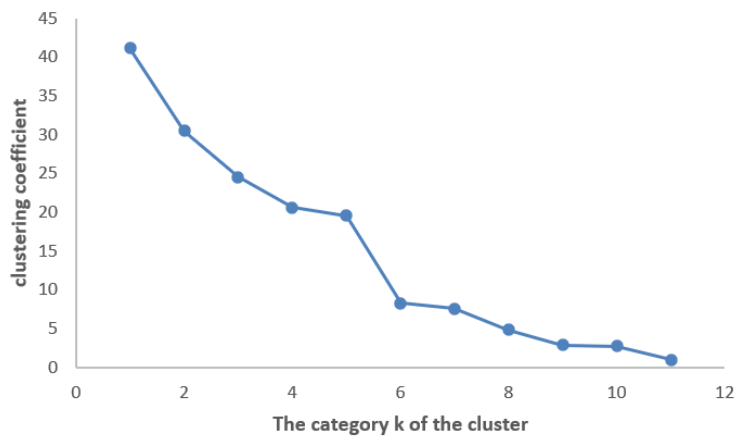


Figure 2. Folding graph of clustering coefficient of high potassium glass

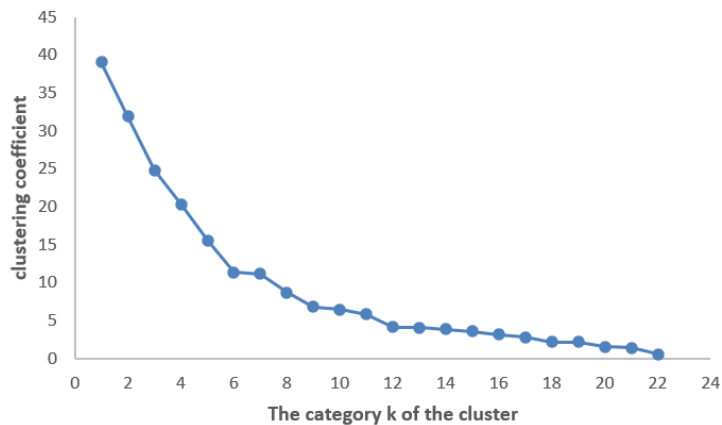


Figure 3. Folding graph of clustering coefficient of lead barium glass

After determining the number of clusters, a new clustering was conducted under the determined number of clusters, and the clustering of high potassium glass and lead barium glass was obtained. The spectral diagrams are shown in Figure 4. Due to the limitation of space, the clustering spectra of lead-barium glass are only partially shown in the paper.

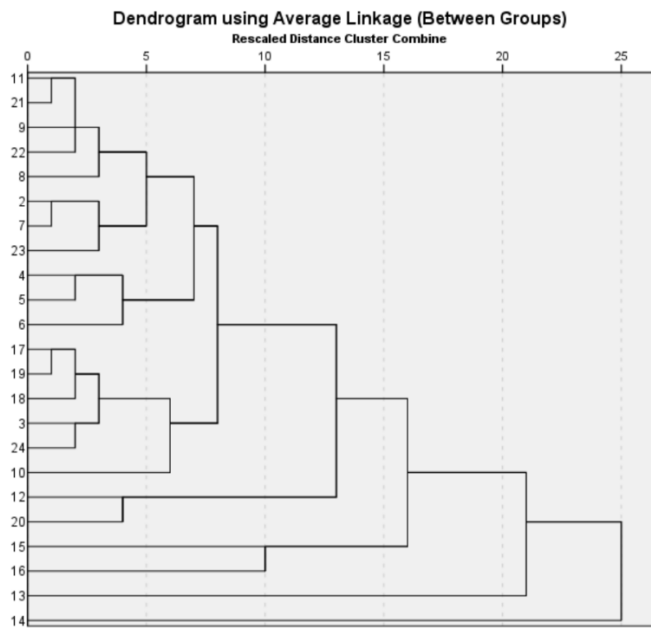


Figure 4. Cluster spectrum of lead-barium glass

The specific classification of high-potassium glass with k value of 4 and lead-barium glass with k value of 6 was determined by cluster spectrum: for high potassium glass, artifacts numbered 01, 03 part1, 04, 05, 13, 14, 16 as the first category, artifacts numbered 03 part 2, 21 as the second category, and artifacts numbered 06 part 1, 06 part 2 as the third category, and artifacts numbered 18 as the fourth category. For lead barium glass, cultural relics numbered 23 unweathered points, 25 unweathered points, 28 unweathered points, 49 unweathered points, 29 unweathered points, 42 unweathered points 1, 42 unweathered points 2, 44 unweathered points, 50 unweathered points, 53 unweathered points, 32, 33, 35, 45, 46, 47, 55 as the first category, cultural relics numbered 20, 37 as the second category, cultural relics numbered 24 as the third category, cultural relics numbered 30 part 1 as the fourth category, cultural relics numbered 30 part 2 as the fifth category, cultural relics numbered 31 as the sixth category.

2.3. Rational and sensitivity analysis

* Rational Analysis:

In order to analyze the rationality of clustering in this question more comprehensively, a combination of macroscopic and microscopic analysis was used. From a macroscopic point of view, two different types of artifacts from the high potassium glass obtained by clustering were selected and their chemical composition contents were analyzed. By comparing the data of the two types of artifacts, it was found that the descriptive values, such as the mean values of their different chemical compositions, were significantly different, which could indicate that the clustering approach was reasonable. Table 3 shows the differences in descriptive values between the first and second categories of artifacts in high-potassium glass.

Table 3. Descriptive numerical differences between Class I and Class II artifacts in high potassium glass

	SiO ₂	Na ₂ O	K ₂ O	CaO	MgO	Al ₂ O ₃	Fe ₂ O ₃
Minimum Difference	-0.8	0	0	0.21	-1.11	-4.69	-2.16
Maximum Difference	20.09	3.38	2.15	2.83	-0.21	-1.92	-3.16
Mean Difference	13.0975	0.6415	-0.825	0.0963	-1.0054	-3.9856	-2.3315
Standard deviation	7.83054	1.24712	0.58193	0.80895	0.19325	-0.14061	-0.84596

From a microscopic perspective, a class of artifacts was selected from high potassium glass and lead-barium glass, respectively. To make the comparison results more intuitive, the third class of high potassium glass (artifact 18) was selected for comparative analysis with the second class of lead-barium glass (artifact 48). Comparison found that: except for copper oxide and sulfur dioxide content is basically zero, there are significant differences in the chemical composition of the two. For example: silica content of cultural relics 18 up to 79.46%, while cultural relics 48 only up to 53.33%. At the same time, due to differences between the two heritage categories, heritage 18 (high potassium glass) potassium oxide content is significantly higher than heritage 48 (lead barium glass), while heritage 48 lead oxide, barium oxide, aluminum oxide content is significantly higher than heritage 18. In addition, there are some differences in content, such as: oxide, iron oxide, tin oxide content also exists a small difference.

* Sensitivity Analysis:

$$F_{\text{High Potassium}} = 0.40 f_1 + 0.19 f_2 + 0.12 f_3 + 0.12 f_4 + 0.07 f_5 \quad (4)$$

$$F_{\text{Lead Barium}} = 0.24 f_1 + 0.23 f_2 + 0.12 f_3 + 0.08 f_4 + 0.07 f_5 + 0.06 f_6 + 0.05 f_7 \quad (5)$$

To investigate the sensitivity and correctness of the model, sensitivity analysis is used in this paper. Sensitivity analysis is a method to study and analyze the sensitivity of a model's state or output change to changes in system parameters or surrounding conditions. The basic step is to change the value of an important parameter in the model while controlling other parameters constant, and then observe

the change of the model's result. By observing Equation (4) and Equation (5), for high potassium glass: the coefficient before f_1 is the largest relative to the other principal components, and the coefficient before f_5 is the smallest relative to the other principal components, representing that the principal component of matrix environmental alkalinity has the largest influence on the classification of high potassium glass, while this principal component has the smallest influence on the classification of modified glass; for lead-barium glass: the coefficient before f_1 is larger relative to the other principal components. The coefficient before f_7 is the smallest relative to the other principal components, representing the greatest influence of the principal component on the classification of lead-barium glass, and this principal component has the least influence on the classification of lead-barium glass.

Based on the above conclusions, the following adjustments were made to the weights of the principal components with the greatest and least influence on the classification of high potassium glass and lead-barium glass, respectively.

Step 1: Keeping the other principal components unchanged, the f_1 weight of high potassium glass and f_1 weight of lead-barium glass were adjusted upward by 20% for re-clustering.

Step 2: Keep the other principal components unchanged, and adjust the weight of f_5 of high potassium glass and f_7 of lead-barium glass upward by 20%, and re-clustering.

The adjusted clustering coefficients of high potassium glass and lead-barium glass are shown in Figure 5 and Figure 6.

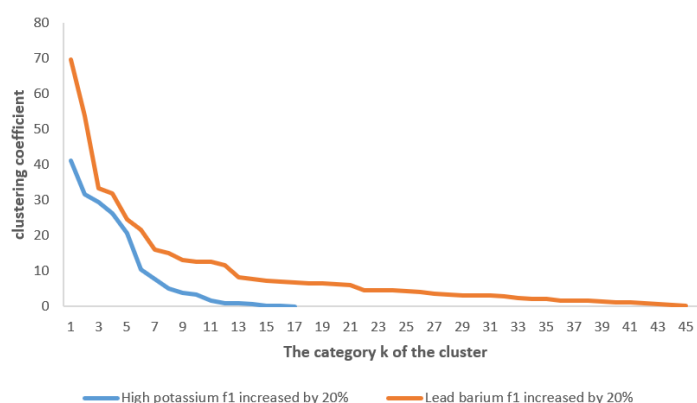


Figure 5. Folding graph of two clustering coefficients with 20% upward adjustment of important principal components

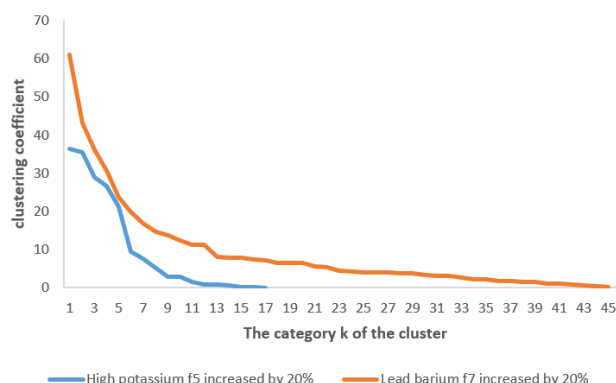


Figure 6. Non-significant principal component upward adjustment of 20% two-class glass clustering coefficient line graph

The graph shows that the clustering result of high potassium glass changed after the important principal component f_1 was adjusted upward by 20%, and the clustering category changed from three categories to two categories, while the clustering result of lead-barium glass pair did not change and still kept $k=3$. When the non-important principal component was adjusted upward by 20%, the

clustering result of high potassium glass did not change and kept $k=3$, while the clustering category of lead-barium glass changed from three categories to two categories.

The influence of principal component f1 (matrix environmental alkalinity) on the classification of high potassium glass was obtained above by systematic clustering, and the influence of principal component f5 on the classification of high potassium glass was weak. Assuming that the analysis is accurate, then the change of important principal components must affect the clustering results, while the change of non-important principal components will not have a large impact on the clustering results. And in this part, by keeping the other principal components unchanged and adjusting the weights of f1 and f5 respectively through sensitivity analysis, the changes of clustering results were obtained in line with the expectation (f1 change from three categories to two categories, and f5 change still kept three categories), which confirmed the correctness and reasonableness of the above analysis. At the same time, the results for lead-barium glass showed the opposite of the expected results, which may be due to the fact that the principal component expression of lead-barium glass contains seven principal component factors, two more than that of high-potassium glass, and the distribution of each weight is more uniform, the weights of principal component f1 and principal component f2 are basically the same, and the difference between the weights of principal component f3 and f1 and f2 is not large.

3. Identification of glass types using BP neural networks

Firstly, the data were classified according to Form 1 glass type and the invalid sampling points containing severe weathering points in Form 2 data were excluded. Since whether weathering is a categorical variable cannot be directly input into the neural network, this categorical variable was then further quantified, specifying that 0 is unweathered and 1 is weathered. The above processed data were input as the training set into the input layer of the BP neural network for training, and the Matlab calculation results showed that the samples were trained for a total of 45 times, and the output training set goodness of fit (Figure 7.3) and mean square error (Figure 7.4).

As shown in Figure 7.3, the goodness-of-fit of the training set is 99.8%, and the goodness-of-fit of the test set is 96.9%, which proves that the network training is excellent, and the prediction results of this model can fully reflect the glass type situation of the unknown artifacts.

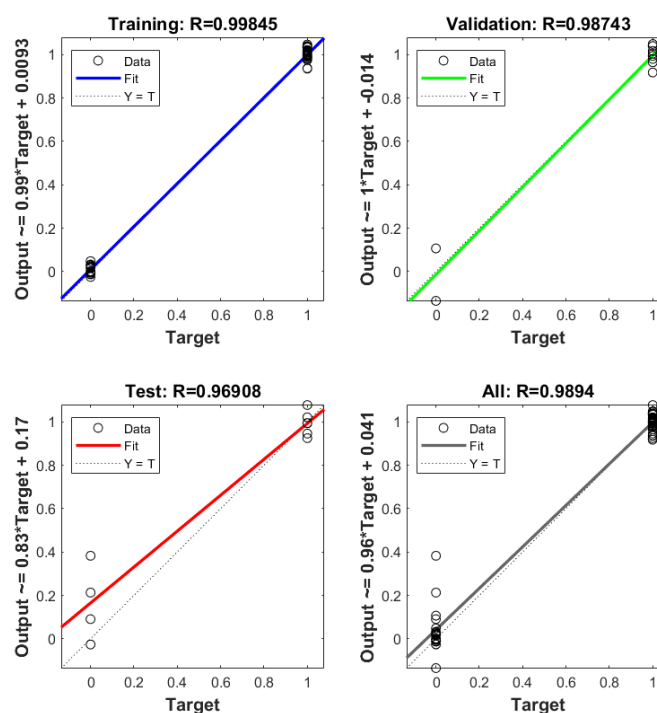


Figure 7. Goodness of fit of the training group

Figure 7. shows the network convergence results of the mean square error of each training result. The network converges after 39 times of training with an error of 0.004, which basically meets the requirements of the network error setting. Meanwhile, it can be seen from Figure 7.4 that with the increase of training times, the training set and test set errors gradually keep fluctuating in a smaller range, and the training set and test set change in the same trend, and when the network is trained to 39 times, the 3 curves gradually approach, the errors gradually stabilize, and the convergence effect is better.

4. Conclusions

Based on the data processing results, we first removed the data of severe weathering points that could not reflect the chemical characteristics of heritage glass types and did principal component analysis on the removed data to output the main characteristic factors of high potassium glass and lead-barium glass, respectively. Five principal components were obtained for high potassium glass, with a cumulative contribution of 90.0%; seven principal components were obtained for lead-barium glass, with a cumulative contribution of 84.1%. Based on the results, the classification rules of high potassium and lead-barium glass were refined. In the second step, systematic clustering was used to obtain the respective subclasses of high potassium and lead-barium glass. It is worth mentioning that the processing of the input data for the systematic clustering followed closely the output of the principal component analysis in the first step, and the 14 chemical components of each artifact were reassigned weights by the coefficients in the principal component expression, and then the obtained data were systematically clustered. The clustering results showed three subclasses for high potassium and lead-barium glass, respectively. Then, by performing macroscopic and microscopic analysis on the classification results, we proved the rationality of the clustering model. The macroscopic analysis selected different two classes from the high-potassium glass for descriptive statistics and found that the compositions differed greatly; the microscopic analysis selected a subclass from each of the two classes of glass and focused on one artifact in the subclass. In order to observe the differences between the high-potassium and lead-barium glass subclasses more intuitively, the subclass containing only one artifact was selected for display, and the results showed that the two artifacts differed greatly in chemical composition. Finally, the sensitivity of the model was verified to be more sensitive by the sensitivity analysis of the principal components of the two glass types.

References

- [1] Ling Feng,Liu Chengying. Identification of dullness of optical glass grinding wheels based on BP network[J]. Manufacturing Automation,2022,44(06):128-131.
- [2] Chen Z R,Gai L L. Kiln pressure control system of TFT-LCD ultra-thin float glass substrate production process based on intelligent control algorithm[J]. Building Materials World,2022,43(03):58-62.
- [3] Ren Xiaoyu. Semantic segmentation of glass in realistic scenes[D]. Inner Mongolia Normal University, 2022. DOI: 10.27230/d.cnki.gnmsu.2022.000810.
- [4] Lu Mingfeng,Wu Jinmin,Yang Wenming,Zhang Feng,Tao Ran. Newton's ring experiments based on signal processing and deep learning[J]. Physics Experiment,2022,42(04):17-24+40.DOI: 10.19655/j.cnki.1005-4642.2022.04.004.
- [5] Huang X-B, Gao Y-Han, Zhang Y, Zhao L, Wu Y-Q, Sun Su-Zhen. Glass insulator target recognition algorithm based on joint component grayscale algorithm and deep learning[J]. Power Automation Equipment,2022,42(04): 203-209.DOI: 10.16081/j.epae.202201032.
- [6] Wu, B., Yu, D. Y.. Deep convolutional neural network-based classification and detection of cell phone glass cover surface defects[J]. Software Engineering,2021,24(12): 6-10.DOI: 10.19644/j.cnki.issn2096-1472.2021.012.002.
- [7] Shi G. Research on the method of acquiring height information on the surface of glass substrate[D]. Xi'an University of Technology, 2021.DOI: 10.27398/d.cnki.gxalu.2021.000549.

- [8] Hao Shuai, Ma Rize, Zhao Xinsheng, Ma Xu, Wen Hu, An Beiyi. Self-exploding fault detection algorithm for glass insulators based on super-resolution depth residual network[J]. High Voltage Technology,2022,48(05): 1817-1825.DOI: 10.13336/j.1003-6520.hve.20210415.
- [9] Liu JH. Research on the improvement of production efficiency of custom wooden doors in M based on group formation technology[D]. South China University of Technology,2016.
- [10] Yu XL. Novel molecular parameter extraction model and its application in molecular conformational relationships[D]. Hunan University,2016.