

# 古代玻璃制品的成分分析与鉴别

项辉<sup>1</sup>, 王可<sup>2</sup>, 金乐<sup>1</sup>, 李霁昊<sup>2</sup>, 李响<sup>3</sup>

<sup>1</sup>上海理工大学机械工程学院, 上海

<sup>2</sup>上海理工大学光电信息与计算机工程学院, 上海

<sup>3</sup>上海理工大学出版印刷与艺术设计学院, 上海

收稿日期: 2022年11月4日; 录用日期: 2023年1月3日; 发布日期: 2023年1月11日

## 摘要

在经过初步分析后, 使用SPSS中的Fisher判别分析对附件1的样本数据整体进行两次Fisher判别分析, 得出化学成分中对于分类有明显影响的主要变量和高钾玻璃、铅钡玻璃的分类规律。分析得知风化文物的未风化采样点与未风化文物的采样点性质相似, 故将其视为未分化组样本再使用excel进行统计性分析。首先使用SPSS分别画出高钾玻璃化学成分和铅钡玻璃化学成分的散点图, 若判断具有线性关系, 再使用皮尔逊相关系数进行不同类型玻璃文物化学成分相关性分析。考虑到风化作用的影响, 所以将高钾玻璃和铅钡玻璃分为风化高钾、未风化高钾、风化铅钡、未风化铅钡4类进行分析。在进行未知文物的类型鉴别分析时, 使用RUSBoost算法和决策树模型分别进行鉴别, 将两种方法得到的结果使用ROC曲线进行比较得出类型鉴别结果更好的模型, 并使用网格搜索法进行模型的敏感性分析。本文对于探究高钾、铅钡玻璃的分类规律以及玻璃文物的分类有一定的应用价值。

## 关键词

Fisher判别分析, 皮尔逊相关系数, RUSBoost算法, 决策树模型, ROC曲线, 网格搜索法

# Composition Analysis and Identification of Ancient Glass Products

Hui Xiang<sup>1</sup>, Ke Wang<sup>2</sup>, Le Jin<sup>1</sup>, Jihao Li<sup>2</sup>, Xiang Li<sup>3</sup>

<sup>1</sup>School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai

<sup>2</sup>School of Optoelectronic Information and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

<sup>3</sup>College of Communication and Art Design, University of Shanghai for Science and Technology, Shanghai

## Abstract

After the preliminary analysis, the Fisher discriminant analysis in SPSS was used to conduct two Fisher discriminant analyses on the whole sample data of Attachment 1, and the main variables in the chemical composition that had a significant impact on the classification and the classification rules of high-potassium glass and lead-barium glass were obtained. After analysis, it is found that the unweathered sampling points of weathered cultural relics have similar properties to those of unweathered cultural relics, so they are regarded as undifferentiated group samples and then used excel for statistical analysis. Firstly, SPSS was used to draw scatter plots of chemical compositions of high-potassium glass and lead-barium glass respectively. If it is judged that there is a linear relationship, then Pearson correlation coefficient was used to analyze the chemical composition correlation of different types of glass relics. Considering the influence of weathering, high-potassium glass and lead-barium glass are divided into four categories: weathered high-potassium, unweathered high-potassium, weathered lead-barium and unweathered lead-barium for analysis. RUSBoost algorithm and decision tree model were respectively used for the type identification analysis of unknown cultural relics. ROC curve was used to compare the results of the two methods to obtain a model with better type identification results, and grid search method was used to analyze the sensitivity of the model. This paper has certain application value to explore the classification rules of high potassium, lead barium glass and the classification of glass relics.

## Keywords

Fisher Discriminant Analysis, Pearson Correlation Coefficient, RUSBoost Algorithm, Decision Tree Model, ROC Curve, Grid Search Method

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

我国古代吸收来自西亚和埃及地区的玻璃技术后, 在本土就地取材, 为了降低炼制玻璃时的熔化温度, 需要在石英砂中添加助熔剂和作为稳定剂的石灰石。由于添加的助熔剂的不同, 导致了最终炼制出的玻璃的化学成分也不相同。常常作为助熔剂的物质有铅矿石和钾含量高的草木灰, 前者形成的玻璃被称为铅钡玻璃是我国发明的品种, 后者形成流行于我国岭南、东南亚及印度等范围内的钾玻璃。

受环境影响, 古代玻璃极易被风化, 使其内部元素与环境元素大量交换, 成分比例产生变化。

使用 SPSS 对附件 1 的样本数据整体进行两次 Fisher 判别分析, 分别得出化学成分中对于分类有明显影响的主要变量和玻璃文物的分类规律。将风化文物的未风化采样点视为未分化组样本, 使用 excel 进行统计性分析。分别画出高钾玻璃化学成分和铅钡玻璃化学成分的散点图, 判断线性关系再使用皮尔逊相关系数进行不同类型玻璃文物化学成分相关性分析。考虑到风化作用的影响, 所以将高钾玻璃和铅钡玻璃分为 4 类进行分析。在进行未知文物的类型鉴别分析时, 使用 RUSBoost 算法和决策树模型分别进行鉴别, 将两种方法得到的结果使用 ROC 曲线进行比较得出类型鉴别结果更好的模型, 并使用网格搜索法进行模型的敏感性分析。

## 2. 高钾玻璃、铅钡玻璃的分类规律

LDA (Liner Discriminant Analysis)是一种十分经典的线性判别方法,又称为 Fisher 判别分析。其目标为求出一个最优的投影空间,使得异类点之间尽可能的远离,同类点之间尽可能的接近和密集,使得在完成投影后,可以将两类数据点有效的分开[1]。

在使用 SPSS 进行线性判别分析时,将附件 1 中的数据根据玻璃文物是否风化分成两组分别进行判别分析和整体一起判别分析得到的结果进行比较,得到的结果相差不大,故此处为了简化模型的求解,因此采取整体一起判别分析的方法进行该部分的分析。由于玻璃文物的成分变量一共有 14 个,无法准确判断哪几个为主要变量,所以先采用一次判别分析,得到标准化典型判别函数系数如表 1 所示。

**Table 1.** Coefficients of full-component standardized canonical discriminant functions  
**表 1.** 全成分标准化典型判别函数系数

成分	系数
二氧化硅	0.753
氧化钠	0.440
氧化钾	-0.595
氧化钙	0.095
氧化镁	0.164
氧化铝	0.516
氧化铁	0.033
氧化铜	-0.320
氧化铅	1.341
氧化钡	1.296
五氧化二磷	0.255
氧化锶	-0.131
氧化锡	-0.008
二氧化硫	-0.202

现对上表进行进一步的分析,由于玻璃的主要成分均为二氧化硅对玻璃类型的划分影响较小,所以在选取主要变量时剔除二氧化硅。在剩下的 13 个变量中,可以由初步分析的系数得知氧化铅和氧化钡这两种成分对于玻璃类型的相对影响较大,故接下去的第二次线性判别分析只选用这两个成分作为变量,得到表 2。

**Table 2.** Normalized canonical discriminant function coefficients of some components  
**表 2.** 部分成分标准化典型判别函数系数

类型分类	铅钡玻璃	高钾玻璃
氧化铅(PbO)	0.196	0.002
氧化钡(BaO)	0.228	0.008
常量	-5.098	-0.695

在第二次判别分析中，对于玻璃类型的判别成功率为 91.4%，达到了较高的成功率。分析表 2 可以得知相较于高钾玻璃，氧化铅和氧化钡对铅钡玻璃的相对贡献率更加的明显。得到如下结论，氧化铅和氧化钡含量高的玻璃有较高的概率是铅钡玻璃，而氧化铅和氧化钡含量低的玻璃有较高的概率是高钾玻璃。

### 3. 风化前后化学成分的统计性分析

由分析得知风化文物的未风化采样点与未风化文物的采样点数据差别不大，所以在进行风化前后化学成分统计性分析时，把附件 1 中表面风化但采集点是未风化的文物的数据样本归属于未风化组。

由图 1、图 2 可知对于高钾玻璃，氧化钠、氧化钾、氧化钙、氧化镁、氧化铝、氧化铁、氧化铜、氧化铅、氧化钡、五氧化二磷、氧化锶、氧化锡、二氧化硫这十三个成分风化后的平均占比相较于风化前均降低，只有二氧化硅风化后的均值占比较于风化前大幅度提高。由此可以看出风化对于各成分的相对作用都比较明显，且相较于二氧化硅对于其余十三个成分的相对作用率更高。

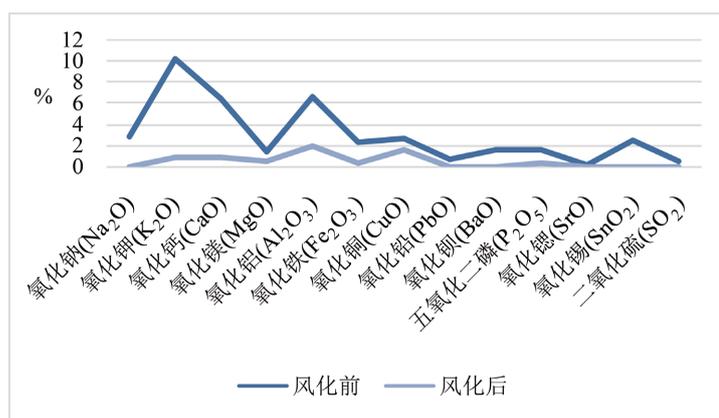


Figure 1. Changes in the mean proportion of 13 components of high potassium glass before and after weathering

图 1. 高钾玻璃十三个成分风化前后占比均值变化

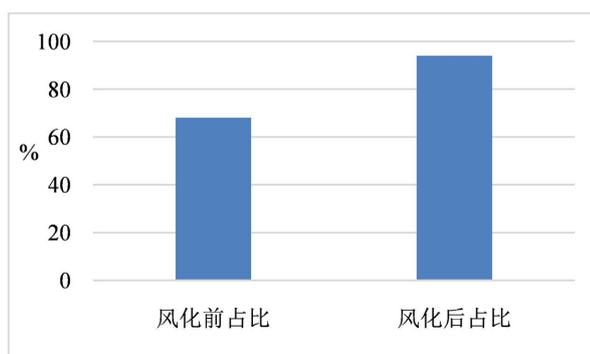


Figure 2. Comparison of the mean proportion of high potassium glass silica before and after weathering

图 2. 高钾玻璃二氧化硅风化前后占比均值对比

由图 3 可知铅钡玻璃中二氧化硅的占比均值降低较明显，说明风化作用对于二氧化硅的作用量和作用率均较明显。氧化钠、氧化钾、氧化铁、氧化锶这四种成分在风化作用后占比均值减小不明显，说明

风化作用对于这四种成分的作用率和风化作用对于整体的作用率差不多。其余的九个成分在风化作用后占比均值均升高，说明风化作用对于这九种成分的作用率均低于风化作用对于整体的作用率。

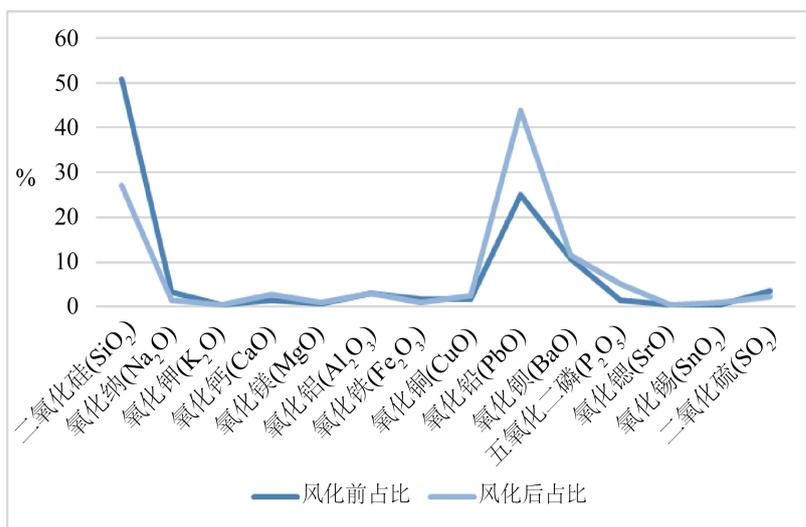


Figure 3. Comparison of mean ratios of 14 components of lead-barium glass before and after weathering

图 3. 铅钡玻璃十四成分风化前后占比均值对比

#### 4. 不同类型玻璃文物化学成分相关性和差异性分析

使用 SPSS 画出附件 1 中样本高钾玻璃化学成分和铅钡玻璃化学成分的散点图如图 4 和图 5 所示。

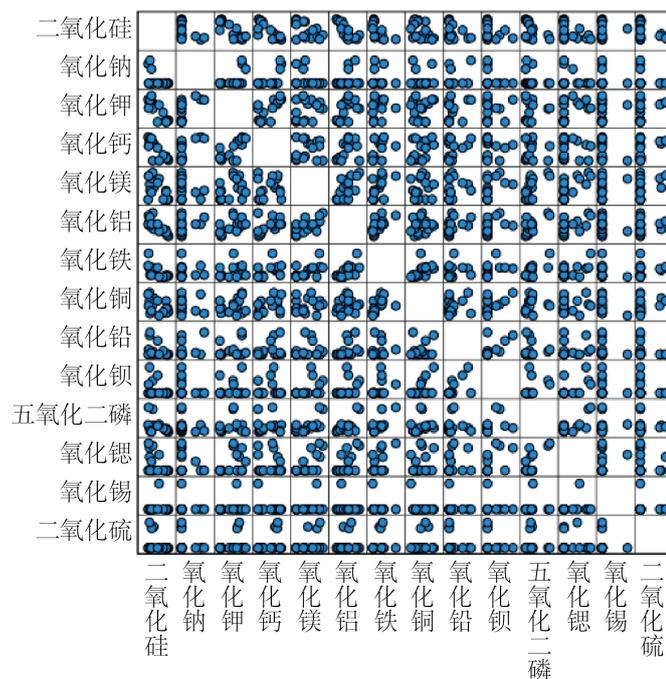
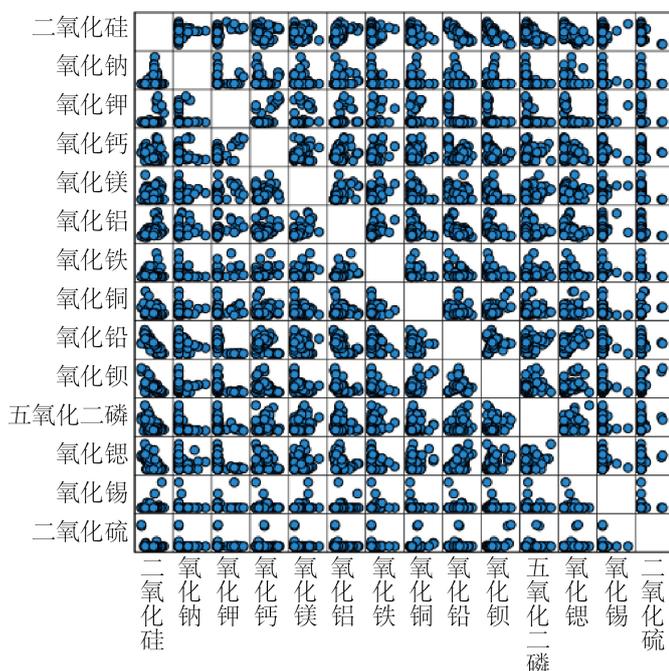


Figure 4. Dot plot of chemical dispersion of high potassium glass

图 4. 高钾玻璃化学成分散点图



**Figure 5.** Dot plot of chemical dispersion of lead-barium glass  
**图 5.** 铅钡玻璃化学成分散点图

由散点图矩阵可知各化学成分之间具有线性关系，因此可以使用皮尔逊相关系数进行进一步的分析 [2]。由于风化作用会影响玻璃制品化学成分的含量，所以将高钾玻璃和铅钡玻璃分为 4 类，风化高钾、未风化高钾、风化铅钡、未风化铅钡进行分析。通过 SPSS 的皮尔逊模型利用附件 1 中的样本数据分析这 4 类玻璃制品化学成分之间的相关关系，得到的具有显著性关系的化学成分类型与其相关系数如表 3~6 所示。

**Table 3.** Relationship between weathering high potassium chemical components  
**表 3.** 风化高钾化学成分之间的关系

正相关	相关性系数	双尾	负相关	相关性系数	双尾
氧化钙与氧化镁	1	0	二氧化硅与氧化钙	-0.897	0.015
氧化钙与氧化铁	0.958	0.003	二氧化硅与氧化镁	-1	0
氧化镁与氧化铝	1	0	二氧化硅与氧化铝	-0.869	0.025
氧化镁与氧化铁	1	0	氧化镁与氧化铜	-1	0
			氧化镁与五氧化二磷	-1	0

**Table 4.** Relationship between chemical components of unweathered high potassium  
**表 4.** 未风化高钾化学成分之间的关系

正相关	相关性系数	双尾	负相关	相关性系数	双尾
氧化钠与氧化镁	1	0	二氧化硅与氧化钾	-0.612	0.046
氧化钠与氧化铅	1	0	二氧化硅与氧化钙	-0.813	0.004
氧化钾与氧化钙	0.865	0.003	氧化钠与五氧化二磷	-1	0

## Continued

氧化钾与二氧化硫	1	0.016	氧化钾与氧化镁	-0.819	0.007
氧化镁与氧化铁	0.681	0.044	氧化钙与氧化锶	-0.995	0.005
氧化镁与五氧化二磷	0.67	0.048			
氧化铝与五氧化二磷	0.623	0.041			
氧化铁与五氧化二磷	0.738	0.023			
氧化铜与氧化钡	0.984	0.002			
氧化铅与氧化钡	0.912	0.031			

Table 5. Relationship between weathering lead barium chemical compositions

表 5. 风化铅钡化学成分之间的关系

正相关	相关性系数	双尾	负相关	相关性系数	双尾
二氧化硅与氧化铝	0.612	0.001	二氧化硅与氧化钡	-0.603	0.004
二氧化硅与氧化锡	1	0	二氧化硅与氧化锶	-0.600	0.004
氧化钙与氧化镁	0.626	0.029	二氧化硅与氧化硫	-0.999	0.001
氧化钙与氧化铁	0.538	0.047	氧化钠与氧化钾	-1	0
氧化钙与五氧化二磷	0.853	0	氧化钠与氧化镁	-1	0
氧化钙与二氧化硫	0.992	0.008	氧化钙与氧化锡	-1	0
氧化镁与氧化锡	1	0	氧化铝与氧化铅	-0.405	0.05
氧化铝与氧化锡	1	0	氧化铁与氧化锡	-1	0
氧化铁与五氧化二磷	0.534	0.049	氧化铜与氧化铅	-0.654	0.001
氧化铜与氧化钡	0.692	0.001	氧化铜与二氧化硫	-0.995	0.005
氧化锶与氧化铅	0.475	0.03	氧化铅与氧化钡	-0.474	0.03
			氧化铅与氧化锡	-1	0
			氧化钡与氧化锡	-1	0
			五氧化二磷与氧化锡	-1	0
			氧化锶与氧化锡	-1	0

Table 6. Relationship between chemical compositions of unweathered lead and barium

表 6. 未风化铅钡化学成分之间的关系

正相关	相关性系数	双尾	负相关	相关性系数	双尾
氧化钠与氧化铜	0.833	0.003	二氧化硅与氧化钾	-0.765	0.001
氧化钠与氧化钡	0.769	0.009	二氧化硅与氧化钙	-0.471	0.036
氧化钾与氧化钙	0.833	0.001	二氧化硅与氧化铜	-0.737	0
氧化钾与氧化铁	0.888	0.008	二氧化硅与氧化铅	-0.613	0.002
氧化钾与氧化铜	0.679	0.008	二氧化硅与氧化钡	-0.706	0
氧化钾与氧化铅	0.604	0.017			
氧化钾与五氧化二磷	0.818	0.004			

Continued

氧化钾与氧化锶	1	0
氧化钙与氧化铅	0.466	0.038
氧化钙与五氧化二磷	0.550	0.034
氧化镁与氧化铝	0.565	0.028
氧化铜与氧化钡	0.901	0
氧化铜与氧化锶	0.629	0.012
氧化钡与氧化锶	0.710	0.001
五氧化二磷与氧化锡	1	0

高钾玻璃与铅钡玻璃的化学成分有不同种类的差异性，如化学成分成正相关的种类、个数不同等，接下去将从差异性较明显的方面对上述表 3~6 及附件 1 进行分析。

由表 6 中的相关系数可知，铅钡玻璃中与氧化钾成正相关性的其他化学成分有 6 个，与二氧化硅成负相关性的其他化学成分有 6 个；而表 4 表明，高钾玻璃中与氧化钾成正相关性的其他化学成分只有 2 个，与二氧化硅成负相关性的其他化学成分只有 2 个。

由表 4 和表 6 可知，铅钡玻璃的化学成分中成正相关关系的个数比高钾玻璃的多，如图 6 所示。

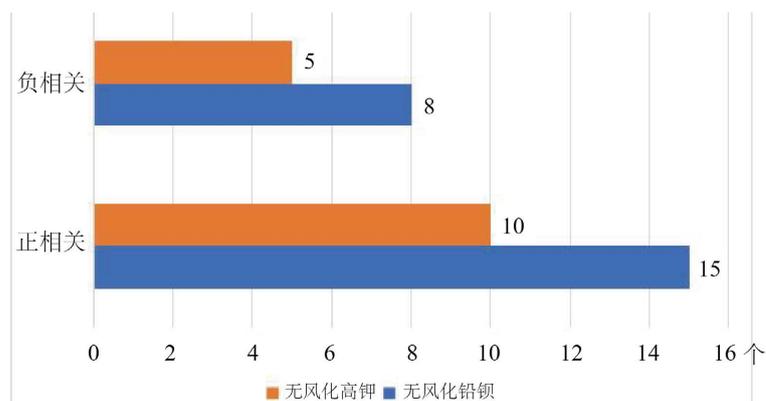


Figure 6. The number of types of chemical composition correlation of unweathered glass

图 6. 无风化玻璃化学成分相关性的种类个数

由表 3 和表 5 可知，铅钡玻璃的各化学成分中有正相关性的比较多，共有 11 组，而高钾玻璃有正相关性的化学成分只有 4 组。在铅钡玻璃中对氧化锡成负相关性的其他化学成分个数比较多，有 6 种(氧化钙、氧化铁、氧化铅、氧化钡、五氧化二磷、氧化锶)，而在高钾玻璃中，没有对氧化锡成负相关性的化学成分。

铅钡玻璃的化学成分中成负相关关系的个数比高钾玻璃的多，是高钾玻璃的 3 倍，如图 7 所示。

## 5. 未知类型玻璃文物的类型鉴别

### 5.1. RUSBoost 预测模型的建立

RUSBoost 算法是一种结合欠采样方法与 Boosting 的混合算法。将 Boost 算法的每轮迭代，在训练弱

分类器之前，使用随机欠抽样方法抽取一定量的多数类样本和少数类组成平衡分布的训练数据集，用于弱分类器训练，RUSBoost 运算步骤如下：

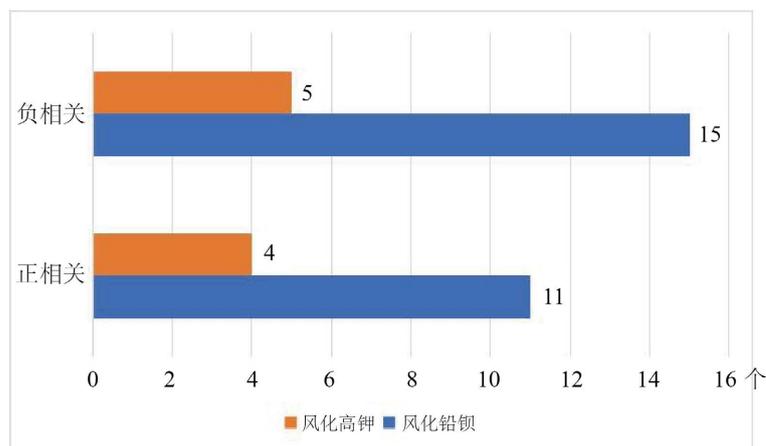


Figure 7. Number of types of correlations between weathering glass chemical compositions

图 7. 风化玻璃化学成分相关性的种类个数

$$e = \sum D(i)[1 - p1(i) + p2(i)] \quad (1)$$

$$\alpha(t) = e/(1 - e) \quad (2)$$

$$D(i) = D(i) \times \alpha(t)^{0.5 \times [1 + p1(i) - p2(i)]} \quad (3)$$

首先，所有样本初始化为  $1/m$ ；其次，利用欠采样方法生成训练集  $S$ ，并且通过对  $S$  中各样本的原有权重进行归一化处理，以获得一个弱分类器  $h(t)$  然后根据  $h(t)$  判别输出概率高为  $p1$ ，低位  $p2$ ，根据公式(1)计算误差  $e$ ，根据公式(2)计算  $\alpha(t)$  根据公式(3)更新权重并归一化。最终通过  $T$  次迭代，输出全部集成分类器：

$$H = \sum h(t) * \log(1/\alpha(t))$$

在迭代过程中，RUSBoost 算法通过调整样本权重的分布来提升后续基学习器的准确率[3]。

由图 8 可知  $TP = 49$ ， $FN = 0$ ， $FP = 0$ ， $TN = 18$  得出分类准确率  $Accuracy = 1$  查全率和查准率的调和平均数  $F1 = 1$ ，可知预测结果较为准确。

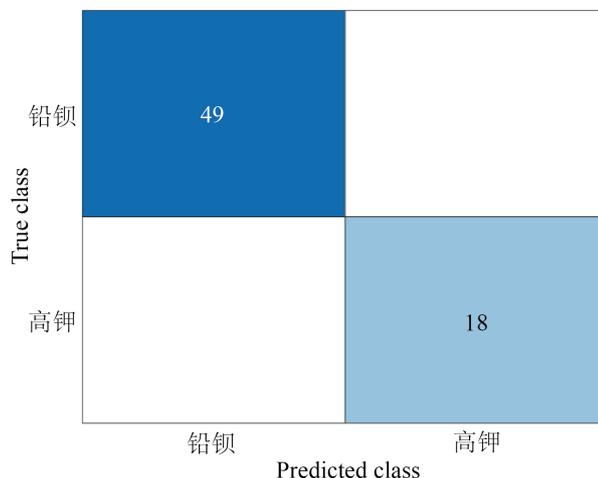
$$F1 = \frac{2TP}{2TP + FP + FN}$$

将附件 2 中的未知类别玻璃文物的化学成分进行分析，鉴别其属性，得出结果如表 7 所示。

## 5.2. 决策树模型的建立

决策树(decision tree)是一种基本的分类与回归方法，在分类问题中，表示根据特征对样本进行分类的过程。

将所有训练数据都放在根节点，选取最佳特征，根据特征将训练数据集分割成子集，使各个子集都具有在当前条件下最好的分类。构建叶节点，将分割后的子集分到所对应的叶节点去。递归进行，直至所有训练数据子集被基本正确的分类，或者没有合适的特征为止。



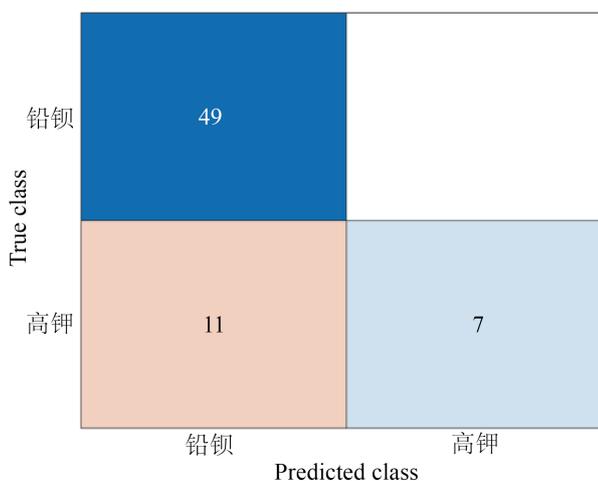
**Figure 8.** RUSBoost confusion matrix  
**图 8.** RUSBoost 混淆矩阵

**Table 7.** Predicted types of unknown glass relics  
**表 7.** 未知玻璃文物预测类型

编号	鉴别类型	编号	鉴别类型
A1	高钾	A5	铅钡
A2	铅钡	A6	高钾
A3	铅钡	A7	高钾
A4	铅钡	A8	铅钡

从根节点开始，对样本的某一特征进行测试，根据测试结果将测试数据分到其子节点，此时每个子节点对应着该特征的一个取值，递归的对要样本进行测试并分配，到达叶节点，最后将实例分到叶节点的类中[4]。

默认决策树最大分裂数为 100，分裂准则为 Gini's diversity index，根据决策树方法可得图 9。



**Figure 9.** Decision tree confusion matrix  
**图 9.** 决策树混淆矩阵

TP = 49, FN = 0, FP = 11, TN = 7, 得出分类准确率 Accuracy = 83.6%, 查全率和查准率的调和平均数 F1 = 0.90。对附件 2 中的未知类别玻璃文物的化学成分进行分析, 鉴别其属性, 得出全为铅钡类型。可知预测结果偏差较大。

### 5.3. 模型的对比与预测结果分析

ROC 曲线是根据一系列不同的分类阈值, 以假正类率(False Positive Rate, FPR)为横坐标, 真正类率(True Postive Rate, TPR)为纵坐标绘制的曲线。

ROC 曲线的用法可以将不同模型的 ROC 曲线绘制在同一张图内, 最靠近左上角的那条曲线代表的模型分类效果最好。

AUC 被定义为 ROC 曲线与下方的坐标轴围成的面积, AUC 的范围位于[0, 1]之间, AUC 越大则模型分类效果越好, 本文将 AUC 小于等于 0.5 的模型视为无效。在本文中当 AUC 大于 0.85 的模型可以认为表现较好[5]。

通过对 Boosted Trees 和 RUSBoost 的 ROC 曲线比较得到图 10 如下所示。

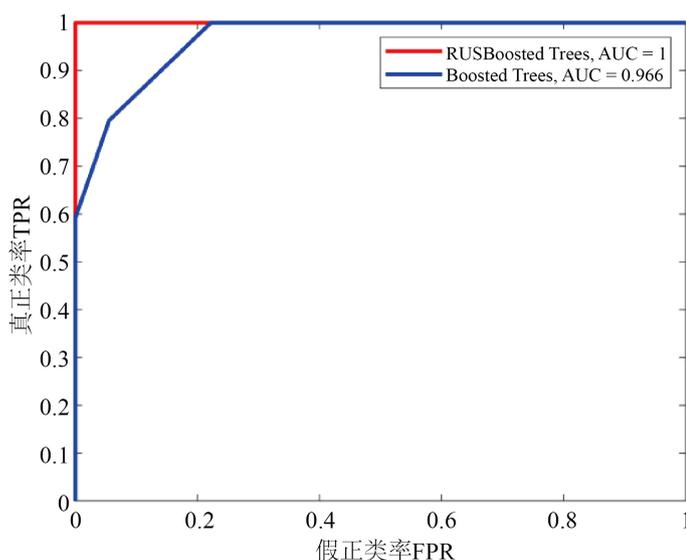


Figure 10. RUS compared with BoostingAUC  
图 10. RUS 与 BoostingAUC 比较

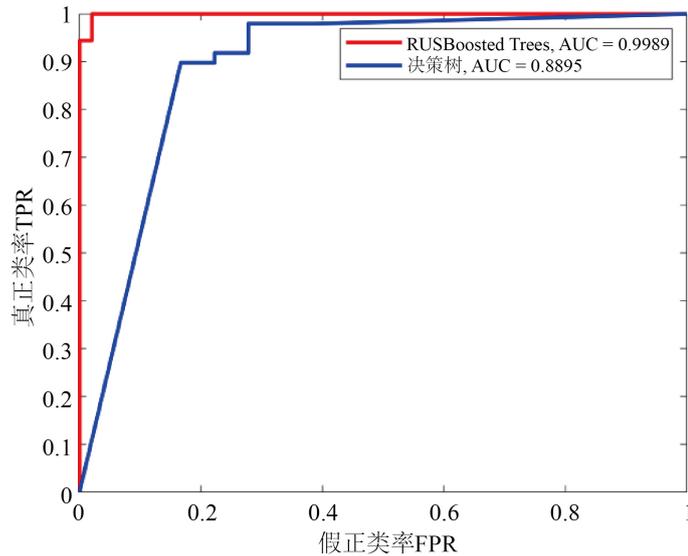
RUSBoosted Trees 的 AUC 值为 1, BoostedTrees 的 AUC 值为 0.966, 可知 RUSBoosted 的分类效果更好。由此得出随机欠采样与标准提升程序 AdaBoost 相结合, 通过删除多数类样本来更好地对少数类进行建模。并且使得算法更简单, 模型训练时间更快。

通过图 11 可以更加直观的进行 RUSBoost 模型和决策树模型比较, 得出预测结果较为准确的值是表 7 所预测的值。

### 5.4. 模型敏感性分析

敏感性分析是研究一个数学模型或数值输出中的不确定性如何被划分和分配到输入中的不确定性的不同来源。通常操作方法为控制其他参数不变的情况下, 改变模型中某个重要参数的值, 然后观察模型的结果的变化情况。选取结果最优的时候的参数。决策树的最大分裂数越大, 树越复杂可能导致过拟合

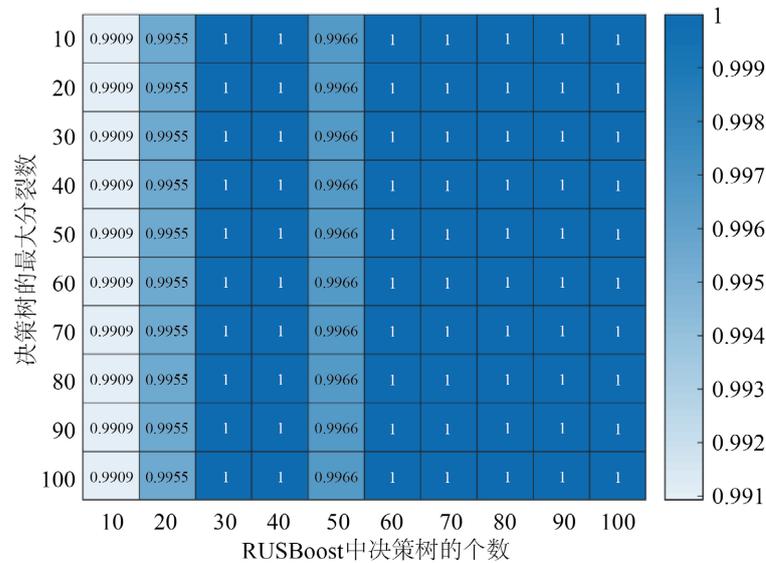
的现象，如果值过小树深度过浅可能导致欠拟合的现象。因此通过敏感性分析获得最佳的最大分裂树。



**Figure 11.** RUS compared to the decision tree AUC  
**图 11.** RUS 与决策树 AUC 比较

网格搜索法是指定参数值的一种穷举搜索方法，通过将估计函数的参数通过交叉验证的方法进行优化来得到最优的学习算法。即将各个参数可能的取值进行排列组合，列出所有可能的组合生成“网络”。然后将各个组合用于 SVM 训练，并使用交叉验证对表现进行评估[6]。

使用网格搜索对 RUSBoost 中决策树的个数和决策树的最大分裂数进行调参得到图 12。



**Figure 12.** RUSBoost thermal map  
**图 12.** RUSBoost 热力图

由图 12 可知决策树的最大分裂数为 10，RUSBoost 中决策树的个数为 30 的时候，AUC 为 1，分类

效果最好。

决策树分裂有三个准则分别为 Gini's diversity index, Twoing rule, Maximum deviance reduction, 通过网络搜索得到图 13。

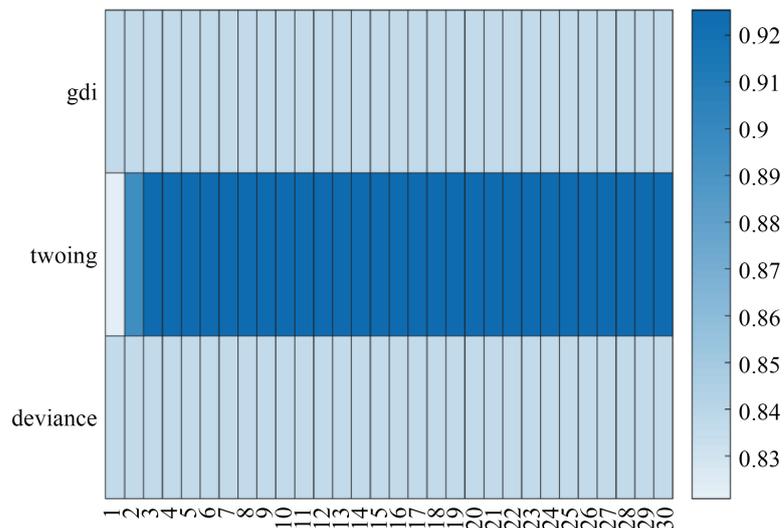


Figure 13. Decision tree thermal map

图 13. 决策树热力图

由图 13 可得当决策树分裂准则为 Twoing rule, 最大分裂数为 3 时, 分类效果最好, 此时 F1 的值为 0.9254, 准确率 Accuracy 提高到 91%, 再次预测结果如表 8 所示。

Table 8. Type prediction scores

表 8. 类型预测分数

模型	铅钡	高钾	预测(调参前)	预测(调参后)
RUS	0.3406	0.6594	高钾	高钾
	1	0	铅钡	铅钡
	1	0	铅钡	铅钡
	1	0	铅钡	铅钡
	0.826	0.1737	铅钡	铅钡
	0.2915	0.7085	高钾	高钾
	0.2915	0.7085	高钾	高钾
	1	0	铅钡	铅钡
决策树	0.3571	0.6429	铅钡	高钾
	1	0	铅钡	铅钡
	1	0	铅钡	铅钡
	1	0	铅钡	铅钡
	1	0	铅钡	铅钡
	0	1	铅钡	高钾
	0	1	铅钡	高钾
	1	0	铅钡	铅钡

## 6. 结语

在经过两次 Fisher 判别分析后, 得出化学成分中对于分类有明显影响的主要变量为氧化铅和氧化钡, 分类规律为氧化铅和氧化钡含量高的玻璃有较高的概率是铅钡玻璃, 而氧化铅和氧化钡含量低的玻璃有较高的概率是高钾玻璃。将附件 1 中表面风化但采集点是未风化的文物的数据样本划归未风化组后, 使用 excel 对附件 1 数据进行统计性分析。得到结论, 风化对高钾玻璃中各成分的相对作用都比较明显, 且相较于二氧化硅对于其余十三个成分的相对作用率更高。风化作用对铅钡玻璃中二氧化硅的作用量和作用率均较明显, 对氧化钠、氧化钾、氧化铁、氧化锶这四种成分的作用率和风化作用对于整体的作用率差不多, 对其余九个化学成分作用率均低于风化作用对于整体的作用率。

使用 SPSS 分别画出高钾玻璃化学成分和铅钡玻璃化学成分的散点图, 在得到其有线性关系后使用皮尔逊相关系数得出 4 类玻璃文物成分的相关系数。分析相关系数表可知未风化组中, 铅钡玻璃的化学成分中与氧化钾成正相关性的其他化学成分有 6 个, 而高钾玻璃只有 2 个; 与二氧化硅成负相关性的其他化学成分种类有 6 个, 而高钾玻璃只有 2 个。在风化组中, 铅钡玻璃的各化学成分中有正相关性的比较多, 共有 11 组, 而高钾玻璃有正相关性的化学成分只有 4 组。在铅钡玻璃中对氧化锡成负相关性的其他化学成分有 6 个, 而在高钾玻璃中, 没有与氧化锡成负相关性的化学成分。

在进行未知文物的类型鉴别分析时, 由 RUSBoost 混淆矩阵和决策树模型预测分别对附件 2 中的未知类别玻璃文物的化学成分进行分析, 鉴别属性。使用 ROC 曲线对比两个模型的预测结果, 可知 RUSBoost 混淆矩阵得出的预测结果更准确。最后使用网格搜索法对 RUSBoost 中决策树的个数和决策树的最大分裂数进行调参, 可知决策树的最大分裂数为 10, RUSBoost 中决策树的个数为 30 的时候, AUC 为 1, 分类效果最好。该模型在实际应用中, 对于玻璃文物的分类有着一定的应用价值。

## 参考文献

- [1] 黄利文. 基于变量择优的 Fisher 逐步判别分析方法[J]. 系统科学与数学, 2021, 41(8): 2338-2348.
- [2] 张景香, 刘亚丹, 杨钊. 基于 SPSS 分析煤灰白度的影响因素[J]. 煤炭技术, 2022, 41(2): 230-231. <https://doi.org/10.13301/j.cnki.ct.2022.02.059>
- [3] 钟华星. 基于 RUSBoost 算法的违约风险预测模型构建与应用[J]. 财会月刊, 2020(10): 74-80. <https://doi.org/10.19641/j.cnki.42-1290/f.2020.10.010>
- [4] 樊良优, 姚小强, 王刚, 刘伟, 夏智权. 基于决策树的指挥控制决策方法[J/OL]. 电光与控制: 1-7. <http://kns.cnki.net/kcms/detail/41.1227.TN.20220913.1940.033.html>, 2022-11-04.
- [5] 宋花玲. ROC 曲线的评价研究及应用[D]: [硕士学位论文]. 上海: 第二军医大学, 2006.
- [6] 马小凤. 机器学习在学生通过率预测中的应用研究[D]: [硕士学位论文]. 重庆: 西南大学, 2019.