



南方科技大学  
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

# 本科生毕业设计（论文）

题    目： 基于红外光谱的玉器类型识别

姓    名： 孙含曦

学    号： 11910838

系    别： 计算机科学与工程

专    业： 智能科学与技术

指导教师： 刘  江

2023 年 5 月 8 日

# 基于红外光谱的玉器类型识别

孙含曦

(计算机科学与工程系 指导教师：刘江)

**[摘要]**：玉器作为传统文化中广为流传的标识，在中国古代历史上有着非常重要的地位。软玉的特殊显微结构和高韧性决定了玉器使用的生命较多数器物更长，在流出和交易过程中不易损坏，因此玉器可以世代相传而物性基本保持不变。这使软玉成为了考古学的重要研究内容，帮助考古学家分析区域文化交流交融情况。本文基于玉器光谱数据，通过对文物进行是否为软玉的鉴定，提供可靠的玉器鉴定服务，辅助考古学研究。本文基于玉器的近红外光谱数据，提出了一个智能软玉类型识别框架，包括光谱特征提取，特征重要性分析和玉器类型识别。本文所采用的光谱数据集来自新出土的 M54 号亚长墓 200 多件玉器，实验结果表明本文提出的框架能达到不错的识别性能，为考古研究提供重要衡量标准，为考古学提供新的分类思路，确定软玉的数据特点。

**[关键词]**：近红外光谱数据；古玉分类；机器学习；特征选取；可解释性

**[ABSTRACT]:** As a widely recognized symbol in traditional culture, jade has played a very important role in ancient Chinese history. The special microstructure and high toughness of soft jade determine that it has a longer lifespan compared to most other utensils, making it less susceptible to damage during circulation and trade. Therefore, jade can be passed down through generations while its physical properties remain basically unchanged. This has made soft jade an important research topic in archaeology, studying the reasons for its circulation. Based on jade spectral data, this paper provides reliable jade identification services by identifying whether the cultural relics are made of soft jade, assisting archaeological research. Based on near infrared spectral data of jade, an intelligent jade type recognition framework is proposed in this paper, including spectral feature extraction, feature importance analysis and jade type recognition. The spectral data set used in this paper comes from more than 200 jade objects from the newly excavated Tomb of M54. The experimental results show that the framework proposed in this paper can achieve good identification performance, provide an important measurement standard for archaeological research as well as a new classification idea for archaeology, and determine the data characteristics of jade.

**[Keywords]:** Near infrared spectroscopy analysis; ancient jade classification; machine learning; feature selection; interpretability

# 目录

1. 绪论.. . . . .	页码
1.1 引言.. . . . .	页码
1.2 研究现状.. . . . .	页码
1.1.1 传统玉器分类方法.. . . . .	页码
1.1.2 基于机器学习的玉器分类方法.. . . . .	页码
2. 方法.. . . . .	页码
2.1 近红外光谱无损矿物学分析.. . . . .	页码
2.2 频域特征参数.. . . . .	页码
2.3 特征选取.. . . . .	页码
2.4 分类器.. . . . .	页码
3. 数据集及实验设置.. . . . .	页码
3.1 数据集介绍和预处理.. . . . .	页码
3.1.1 数据集.. . . . .	页码
3.1.2 数据标记.. . . . .	页码
3.1.3 数据预处理.. . . . .	页码
3.2 实验评估方法.. . . . .	页码
3.3 基线.. . . . .	页码
3.4 实验设置.. . . . .	页码
4. 实验结果及分析.. . . . .	页码
5. 总结.. . . . .	页码

参考文献..... 页码

致谢..... 页码

# 1. 绪论

## 1.1 引言

在多数与中国古代器物和社会关联性的研究中<sup>[1]</sup>，都会强调软玉在古代中国的社会地位与在文化关系中起到的重要象征作用。Rawson 等人在研究软玉由新石器时期起到清王朝和现代中国发展变迁情况中收获颇多<sup>[2]</sup>，从完璧归赵的故事代代相传，而后有蔺相如以和氏璧作为与秦国交换城池的条件颇为美谈，可见在古代中国软玉在文化价值与社会关系方面居于重要地位。玉的传承也一直流传至今，玉经常作为传家宝见证了一个家族的荣辱兴衰，寄托了家族成员的精神与情感，凝聚了家族的力量<sup>[3]</sup>。玉石的传承往往承载着祖先对于后代的期待，希望后人能够成为一个有信仰，有追求，温润如玉的人<sup>[4]</sup>。当然具体的用玉方式、社会意义在不同的文化、社会条件和历史阶段有着很大的差异，先民们对玉的意识、认同感、使用和尊崇<sup>[5]</sup>，直接与玉的物性以及物性所赋予的社会价值和文化内涵有关。



图1 常见的玉器形制图

软玉特有的高韧性可能用来体现和建构具有生命活力，持久，永恒，不可破碎的性格与社会关系。高韧性决定了玉器使用的生命可以很长，因此一件玉器可以世代相传，而其物性基本保持不变<sup>[6]</sup>。这也是软玉作为考古学研究的重要原因。软玉的韧性是由其纤维束本身的性状和纤维束之间的相互关系共同决定的，组成纤维束的单晶体越细小，各晶体间的错位程度越大，纤维束本身的强度就越大<sup>[7]</sup>。另一方面组成纤维束的单晶体越小，由它们构成的纤维束也越小；纤维束越细小，交织程度越高，软玉的韧性就越高，致密度也越高，其稳定性就越好。正是软玉纤维结构的特殊性决定了软玉特有的高韧性，高致密度，进而有了软玉特有的温润光泽<sup>[8]</sup>，这些也共同构成了与特有的质感，古人所谓玉德。因此，为了辅助考

古学研究，探索文化起源与人类发展，增强文化认同感，研究人员常使用近红外光谱的计算机方法提取和分析玉背后的文化信息<sup>[9]</sup>。

本文基于玉器的近红外光谱数据，提出了一个智能软玉类型识别框架，包括光谱特征提取，特征重要性分析和玉器类型识别。在近红外光谱领域的应用中，监督学习已被证明是一种可行的方法，它可以根据光谱数据对玉进行分类从而判断其真实性。本文使用四种监督学习方法对数据分别进行训练，利用特征筛选的技术为考古学提供化学与材料层面的光谱关键频域分析，使用融入频域特征参数的方法改进了方法同时比较，了不同特征数量下模型的性能，确定软玉的数据特点。实验结果表明本文提出的框架能达到不错的识别性能，证明了机器学习应用于光谱数据来辅助考古学研究的可行性与可拓展性。将机器学习融入考古学的研究，与领域知识配合分析，为考古研究提供重要衡量标准，为考古学提供新的分类思路，是跨学科交叉研究的重要突破。

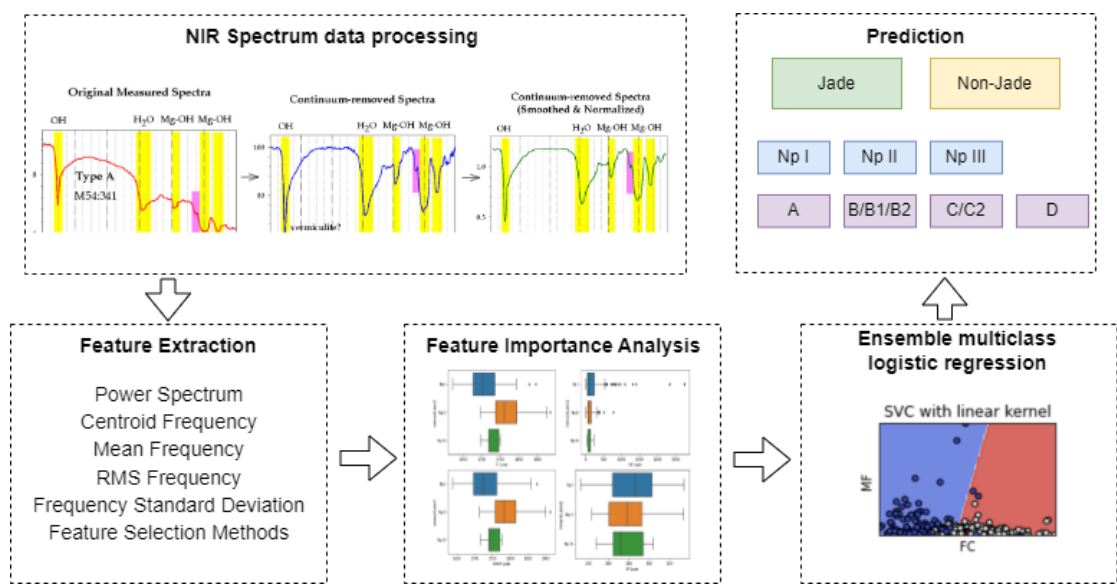


图2 常见的玉器形制图

## 1.2 研究现状

### 1.2.1 传统玉器分类方法

刘继富等人<sup>[10]</sup>在湖北随州曾侯乙墓出土玉器材质分析与产源初探一文中通过表面特征即其风化程度和断裂面颜色判断产源。山东大学的姜亚飞<sup>[11]</sup>在关于西周时期玉器流通状况的初步研究中通过纹饰和文献记载来辨别玉器产地。还有各类射线与光谱分析方法进行分类<sup>[12]</sup>，比较出名的是通过透闪石-阳起石的铁含量来细分玉器种类，例如成小林等人<sup>[13]</sup>利用红外光谱和扫描电镜对新干商墓出土玉

器进行分析研究。

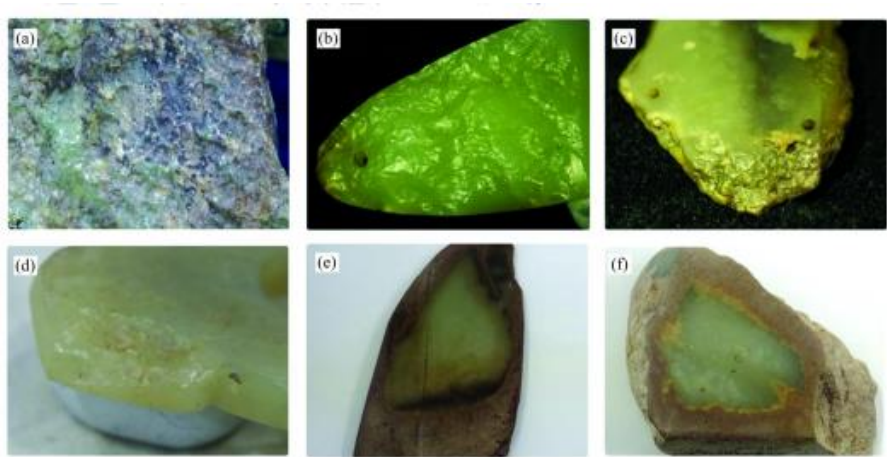


图3 部分具有产地信息特征的样品

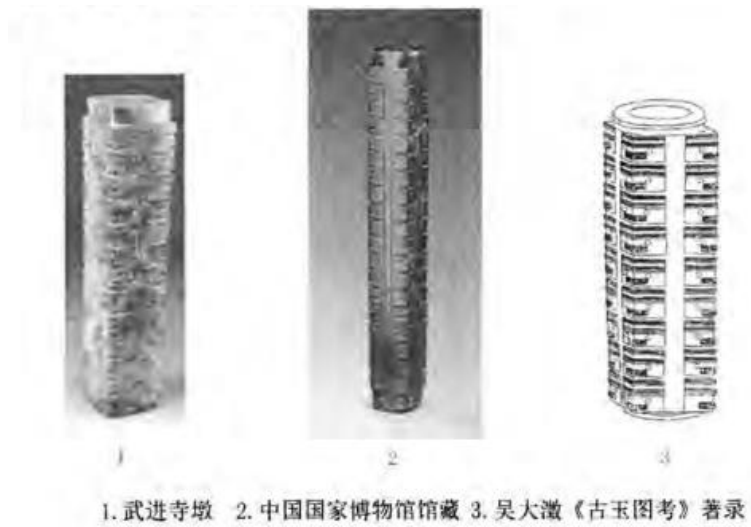


图4 良渚文化晚期玉琮

与此同时，国外的同期研究也采用了许多受领域知识限制颇多的传统方法<sup>[14]</sup>。由此可见，考古学届目前对于玉器流通的研究还停留在比较经典的传统分析方法上，对于计算机辅助分析有较大的发展空白需要填补，本实验旨在通过机器学习和特征分析的手段进行玉器光谱的分析。

### 1.2.1 基于机器学习的玉器分类方法

在玉器分类方式研究上，国外对于玉器光谱的研究和分析早在 1974 年就有所突破<sup>[15]</sup>，但这项研究只探讨了在激发出电子转移状态下的玉器光谱。彭明生等人对于台湾软玉<sup>[16]</sup>和郭立鹤等人对于和田玉、玛纳斯碧玉和岫岩老玉的研究<sup>[17]</sup>聚焦于红外光谱中的喇曼光谱。Harder 等人<sup>[18]</sup>对于红外光谱的利用停留在微量元素对玉器颜色的影响。Bao 等人<sup>[19]</sup>使用热学测量的方法，但并没有探究获得的晶



体结构、化学成分对于玉器种类细分的帮助。

在特征选取方法上，也有不少成熟的研究提供可选方法。单变量统计检验 (Analysis of Variance, 简称 ANOVA F-value)<sup>[20]</sup> 可使用 `f_classif` 进行单变量特征选择的方法根据与目标变量的相关性进行单变量统计检验, 选择  $k$  个最佳特征。这是一种简单快速的方法, 适用于具有大量特征的数据集。然而, 它可能不适用于具有高度相关特征的数据集, 因为它没有考虑特征之间的相互作用。具有互信息的单变量特征选择 (`mutual_info_classif`, 简称 MIC)<sup>[21]</sup> 来执行单变量统计检验, 但使用互信息而不是方差分析  $f$  值。互信息度量一个特征提供的关于目标变量的信息量, 对于识别特征和目标变量之间的非线性关系很有用。对于具有非线性关系的数据集, 它是一种很好的方法, 但对于大型数据集, 它可能比基于方差分析的方法慢。使用卡方 (`chi2`)<sup>[22]</sup> 进行单变量特征选择则是执行卡方检验, 根据它们与目标变量的关联选择  $k$  个最佳特征。对于具有分类特征的数据集, 它可以捕获分类特征与目标变量之间的非线性关系, 是一种很好的方法。然而, 它可能不适用于具有连续特征的数据集。(分类器) 逻辑回归方法[23]则是根据其系数选择  $k$  个最佳特征。系数测量每个特征对目标变量的影响, 并可以捕获特征和目标变量之间的线性和非线性关系。对于具有线性或非线性关系 数据集, 这是一种很好的方法, 但对于高度相关的特征可能不太有效。

## 2. 方法

### 2.1 近红外光谱无损矿物学分析

近红外光谱 (Near InfraRed spectroscopy , 简称为 NIR) 无损矿物学分析<sup>[24]</sup>是一种先进的非破坏性检测技术, 广泛应用于矿物、宝石和岩石的鉴定和研究。近红外光谱的波长范围为 700-2500 纳米, 它可以通过对矿物样品进行光谱扫描, 收集样品在不同波长下的反射或透射光谱信息, 从而获得样品的化学和结构信息。

玉在矿物学上包括软玉和硬玉两种, 中国古玉特指软玉。软玉 结构成分为钙角闪石 (tremolite-actinolite 透闪石-阳起石), 学名蓝晶石 (Nephrite), 是一种具有良好观赏价值和实用性的矿物。它在化学上属于钙镁硅酸盐矿物, 主要成分是  $\text{Ca}_2(\text{Mg}, \text{Fe})_5\text{Si}_8\text{O}_{22}(\text{OH})_2$ 。软玉的颜色丰富多样, 从乳白色、翠绿色到

深棕色，这主要取决于其内部杂质元素的种类和含量。在物理性质方面，软玉的硬度较低，摩氏硬度为 6-6.5。但是，它具有非常高的韧性和抗压强度，这使得它在古代和现代雕刻中非常受欢迎。此外，软玉的密度为 2.8-3.2 g/cm<sup>3</sup>，这使得它在触感上显得温润光滑。从微观结构特征来看，软玉由纤维状晶体组成，这些晶体相互交织在一起形成了坚固的纤维束结构。这种结构使得软玉具有良好的抗压性和韧性，也是其光滑温润质感的原因。

在软玉的近红外无损矿物学分析中[26]，可以观察到一些特征吸收峰，这些峰通常与软玉中的特定化学键和分子振动模式有关。通过对这些特征峰进行定量和定性分析，可以快速、准确地鉴定软玉的成分、结构和来源等信息。这种方法在保护文物、鉴定真伪和研究矿物成因等方面具有重要应用价值。

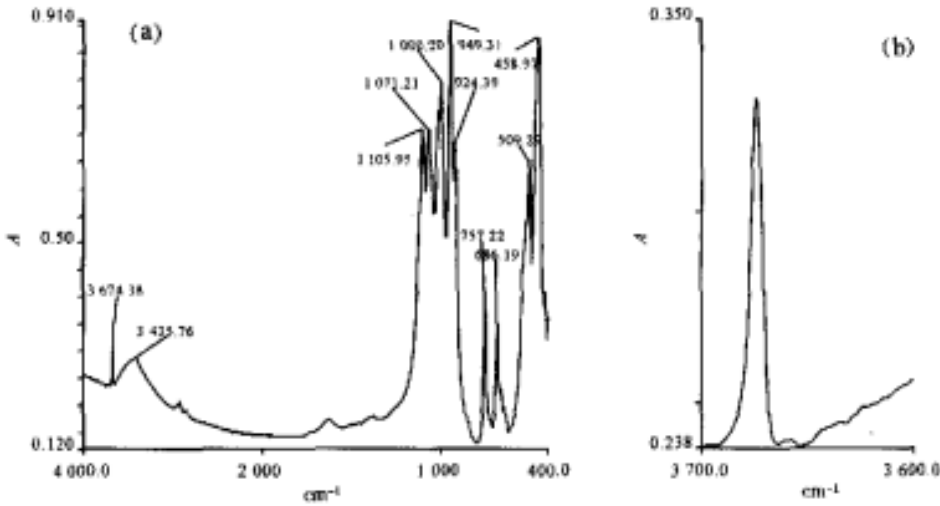


图5 无损矿物学分析下的软玉光谱图

## 2.2 频域特征参数

频谱分析包括幅度频谱和相位频谱。当玉器材质不同时，样本数据信号频谱中的频率分量会相应地不同。故可以先通过分析振动信号的频域特征准确地表征信号频谱信息，然后获取在不同材质器型下近红外光谱频率的特征来帮助频率的判断。研究者常用的频域特征参数包括重心频率、平均频率、均方根频率以及频率标准差等。这些频域特征参数可以用于分析信号的频谱内容，并可以提供有关能量在频率范围内分布的信息。它们通常用于语音和音频处理应用程序，用于说话人识别和情感识别等任务。

1. 重心频率：重心频率，也称为 FC 或中心频率，是信号频率分量的加权平

- 均值。它是测量信号频率含量的“重心”，可以用来表征整体频率分布。
2. 平均频率:平均频率，也称为 MF，是信号频率分量的算术平均值。它是信号频率内容集中趋势的另一种度量。
  3. 均方根频率:均方根频率，或 RMSF，是一个信号的平方频率分量的平均值的平方根。它是信号能量含量的度量，RMSF 值越高，表示在越高的频率范围内能量越多。
  4. 频率标准差:频率标准偏差，或 VF，是对信号频率分量的传播或可变性的度量。它是信号中频率内容多样性的度量，可以用来表征频率分布。
  5. 幅度谱:幅度谱是信号幅度作为频率函数的图。它显示了能量在信号频率范围内的分布。
  6. 频率响应:频率响应是将系统的输入信号与输出信号作为频率函数联系起来的复函数。它通常用于设计滤波器和其他信号处理系统。
  7. 传递函数:传递函数是系统在频域中输出信号与输入信号的比值。它是分析线性定常系统时常用的特征参数。
  8. 功率谱密度:功率谱密度是对信号单位频率功率的度量。它常用于分析随机信号和噪声。
  9. 相干性:相干性是两个信号之间作为频率函数的相关性的度量。它通常用于分析系统中两个信号之间的关系。相干函数是以频率为函数来衡量两个信号之间的相关性，但它既考虑了信号之间的相位关系，也考虑了信号之间的幅度关系。
  10. 交叉谱密度:交叉谱密度是两个信号在频域中关系的度量。它描述了一个信号在每个频率上有多少功率被另一个信号共享。
  11. 频谱平坦度:频谱平坦度是衡量信号的能量在其频率范围内的均匀程度。频谱平坦度高的信号能量在其频率范围内均匀分布，而频谱平坦度低的信号则被少数频率分量所支配。

## 2.3 特征选取

本文采用了递归特征消除(Recursive feature elimination, 简称 RFE)的方法来筛选玉器光谱特征，使用机器学习模型递归地删除最不重要的特征，直到达到所需的特征数量。RFE 的工作原理是，首先在完整的特征集上训练机器学习模

型，然后计算每个特征的重要性。然后删除最不重要的特征，并重复该过程，直到达到所需的特征数量，得到的特征子集用于训练最终的线性 svc 模型。

2.4 分类器

本文采用了随机森林的分类器来进行玉器类型分类。基于模型的特征选择方法使用机器学习模型(随机森林)根据它们的重要性得分选择 10 个最佳特征。重要性分数衡量每个特征对模型准确性的贡献，并可以捕捉特征和目标变量之间的非线性关系。对于特征和目标变量之间关系复杂的数据集，它是一种很好的方法，但对于大型数据集，它可能比单变量方法慢。

3. 数据集及实验设置

3.1 数据集介绍和预处理

3.1.1 数据集

为了在保护文物本身的基础上对玉器结构成分进行分析，本文采用了玉器的无损矿物学分析方法，即通过光谱扫描仪对玉器进行全方位的扫描，得到玉器的一组光谱数据。采集数据所需的玉器由南方科技大学社科中心文化遗产实验室提供。器型丰富，大致可分为玉璧（环）、玉璜（珩）、玉佩、玉玦、玉管、玉饰等。样品制备所有样品都经过目视检查，并根据其颜色、透明度和质地进行初步分类。然后用软毛刷轻轻地清洁样品，去除表面的污染物，如灰尘和土壤。由于分析是非破坏性的，因此不需要进一步的样品制备。

表 1 数据集概况表

Data Sets	spectra number	catagories
M54 一号墓	449	8
亚长墓	231	8

采用傅里叶变换近红外光谱仪 (Fourier transform near infrared spectrometer, 简称 FT-NIR) 对样品进行了近红外光谱分析。光谱范围为 1300 ~ 2500 nm，光谱分辨率为 4 cm<sup>-1</sup>。每个样品放置在积分球窗口上，采用漫反射模式采集光谱。为了确保测量结果的可靠性和可重复性，每个样品的光谱记录为三份。然后计算三个光谱的平均值，并用于后续的数据分析。

本实验中使用的数据集包含了 680 条光谱数据，由无损矿物学分析采集。每

条光谱由 600 个频率组成，每 2nm 采样一次。每个样本都有一个标签，标记为软玉或非软玉。数据集以.mat 格式存储。

3. 1. 2 数据标记

此外，本文通过人文社科中心提供的考古人员编写的玉器特征标签，对照玉石的光谱数据的玉器编号，对数据进行人工标记的工作。其中根据近红外光谱分析矿物成分有：Np I - 软玉类型 I，Np II - 软玉类型 II，Np III 软玉类型 III，At - 叶蛇纹石，Ct - 绿泥石，It - 伊利石，Dc - 迪开石，Qz - 玉髓、隐晶质石英，UN - 未知矿物或岩石。而其他特征、受沁程度、器物类别的质地分类有：A - 矿物组成为 Np I 的器物；B、 B1、 B2 - 矿物组成类同于 A，为 Np I，但含有微量蛭石，B - 未或微弱受沁或次生变化（不包 A 型 I 式玉管），B1 - 中等或强烈受沁或次生变化（不包 A 型 I 式玉管），B2 - A 型 I 式玉管，未或微弱受沁或次生变化；C - A 型 I 式玉管，矿物组成为 Np II；D - 矿物组成为 Np III 之器物；S - 非玉之矿物或岩石。

表 2 样品标记对照表

器物号	器名	本文分类	矿物成分 <sup>1</sup>	质地分类 <sup>2</sup>	Munsell 标准色	光泽 <sup>3</sup>	半透明度 <sup>4</sup>	次生变化 <sup>5</sup>	埋葬受沁 <sup>6</sup>
M54:148	铜戟玉援矛-A	武器、工具形器	Np I	B	10Y5/4 - 5GY 5/4	++	++	+++	++
M54:158	矛	武器、工具形器	Np I	B1	5GY6/2, 10YR6/6	+++	+ / ++	+++	++
M54:159	箍形器	环形器 - 其它	Np I	A	5GY7/2 - 8/2	++++	++++	+	+
M54:317	玉管-A III	装饰器 - 玉管	Np III	D	2. 5Y8/1	++ / +++	+++	+ / ++	+
M54:318	玉管-A I	装饰器 - 玉管	Np II	C2	5Y4/1	+++	+	+	+
M54:322	圭	武器、工具形器	Np7	B1	10Y5/2	+++	+++	+	+
M54:331	玉管-A I	装饰器 - 玉管	Ct	S	5G5/2	++	+ / ++	+	+

3. 1. 3 数据预处理

原始的近红外光谱通常包含噪声和系统变化，可以影响结果的准确性。为了尽量减少这些影响，收集的光谱进行了几个预处理步骤。首先，采用 SG 平滑滤波器(Savitzky-Golay Filter)去除高频噪声;接下来，根据基线失真的性质，使用乘法散射校正(Multivariate scatter correction, 简称 MSC)或标准正态变量(Standard Normal Variate transform, 简称 SNV)变换对光谱基线进行校正。最后，以均值为中心对光谱进行自动缩放，使光谱强度归一化。

本文使用了赫尔商变化 (Hull Quotient Transformation) 获得去掉原始光谱吸收背景值之后的光谱，凸显不同化学键的吸收峰，使用多元散射矫正 MSC 和

小波变换进行去噪，使用了 Savitzky-Golay 平滑滤波和归一化处理。

在数据扩增步骤中，本文通过添加偏移量(Offset)、乘数(Multiplication)和斜率(Slope)的随机变化来扩增。偏移量变化为训练集标准差的 $\pm 0.10$ 倍。乘数是标准差的 $1 \pm 0.10$ 倍。斜率在 $0.95 \sim 1.05$ 之间均匀随机调整。对每个样本执行 9 次，将亚长墓原本的 231 个样本扩增为 2079 个。

### 3.2 实验评估方法

为了评价本文所选的方法效果及模型的整体性能，本文计算了常用的评价指标：准确度(Accuracy, 简称 ACC)、宏观精度(PR)、宏观灵敏度(SE)、宏观 F1 分数。这些评价指标是根据预测标签和真实标签来计算的，这些标签是通过在测试集上应用训练好的线性 svc 模型得到的。这些评价指标可以用以下公式表示，其中 TP、FP、TN、FN 分别表示真正例、假正例、真负例、假负例的个数：

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} , \quad (1)$$

$$SE = \frac{TP}{TP+FN} , \quad (2)$$

$$PR = \frac{TP}{TP+FP} , \quad (3)$$

$$F1 = \frac{2*PR*SE}{PR+SE} , \quad (4)$$

为了评估分类模型的性能，使用了 k 折交叉验证方法，其中数据集被随机分为 k 个子集。每个子集用作验证集，其余子集用于模型训练。此过程重复了 k 次，并对结果进行平均，以获得可靠的性能估计。

### 3.3 基线

在特征选取任务中，本文实现 ANOVA F-value, MIC, Chi2, 随机森林(Random Forest, 简称 RF), 逻辑回归(Logistic Regression, 简称 LR), RFE 六种特征选取方法来选取 10 个关键频率，作为对比来证明本文提出的参数的有效性。本文将实现一个分类器 SVC 来测试 600 个频率特征值下的分类准确性，为后续经过筛选的实验结果。

### 3.4 实验设置

为了比较不同特征数量下模型的性能，本文使用了不同数量的关键频率进行训练，从 10 到 500 不等，每隔 10 个取一个。本文使用五折交叉验证对每个模型

进行评估。最终，本文使用测试集对每个模型进行测试，并得出了准确率。具体步骤如下：

- (1) 读取.mat 文件，将数据和标签分别存储到变量中。
- (2) 使用多种特征选取算法选择关键频率并通过交叉对比筛选出最有效的 10 个频率。
- (3) 将选定的 10 个频率作为特征，将数据分为训练集和测试集。
- (4) 使用集成学习方法对数据进行训练，获得基线。计算每个模型在测试集上的准确性并比较它们。
- (5) 通过加入频域特征参数优化模型，再次训练获得准确性等特征。
- (6) 使用交叉验证来评估每个模型的性能，确保模型具有良好的泛化能力

### 4. 实验结果及分析

在本实验选取的频域特征参数（Frequency domain characteristic parameter, 简称 FDC）包括重心频率、平均频率、均方根频率、频率响应，功率谱密度，交叉谱密度，频谱平坦度以及频率标准差等。由箱型图获得的如下结果可知，重心频率、均方根频率和频率标准差有肉眼可见较为良好的区分度。

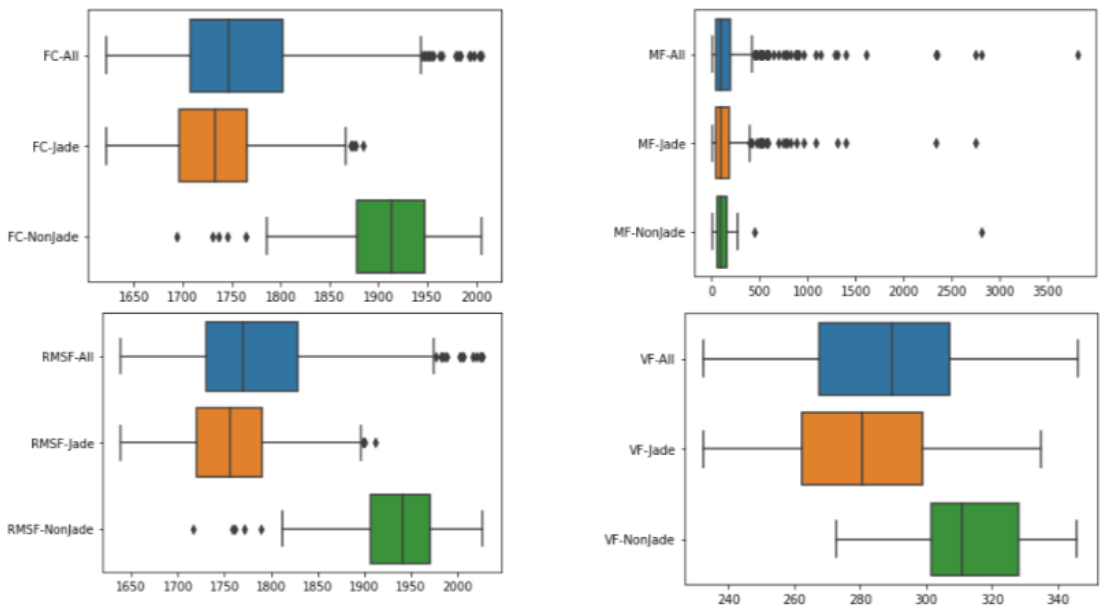


图 8 频域特征在区分玉与非玉材质上的结果

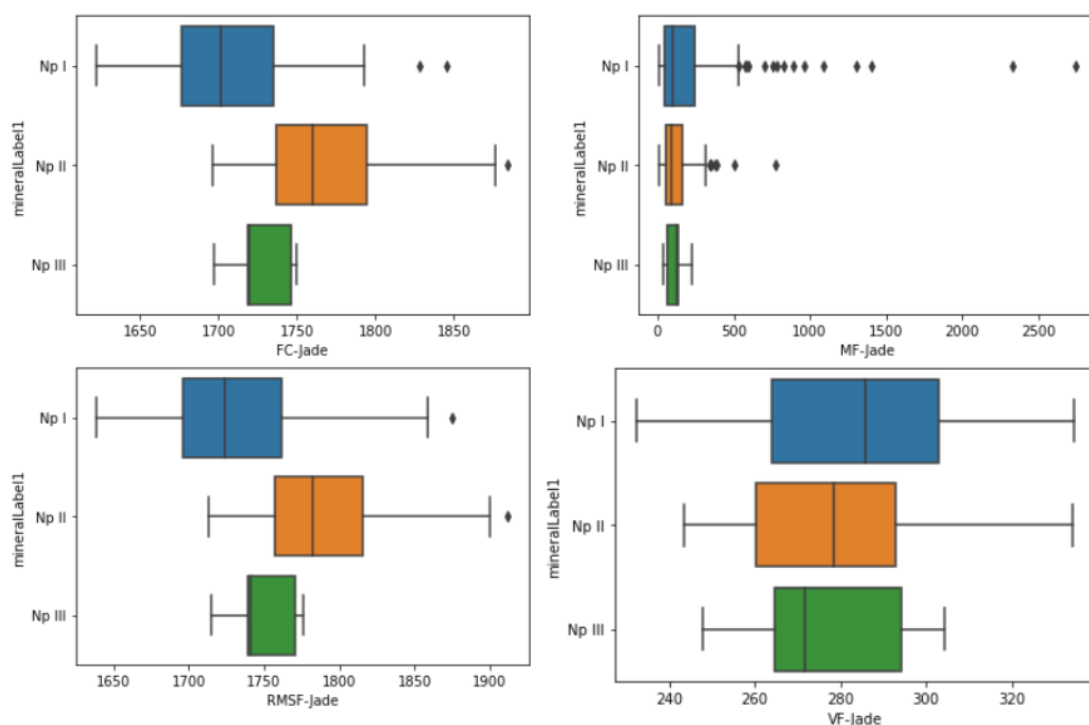


图 9 频域特征在区分玉类器物矿物属性的结果

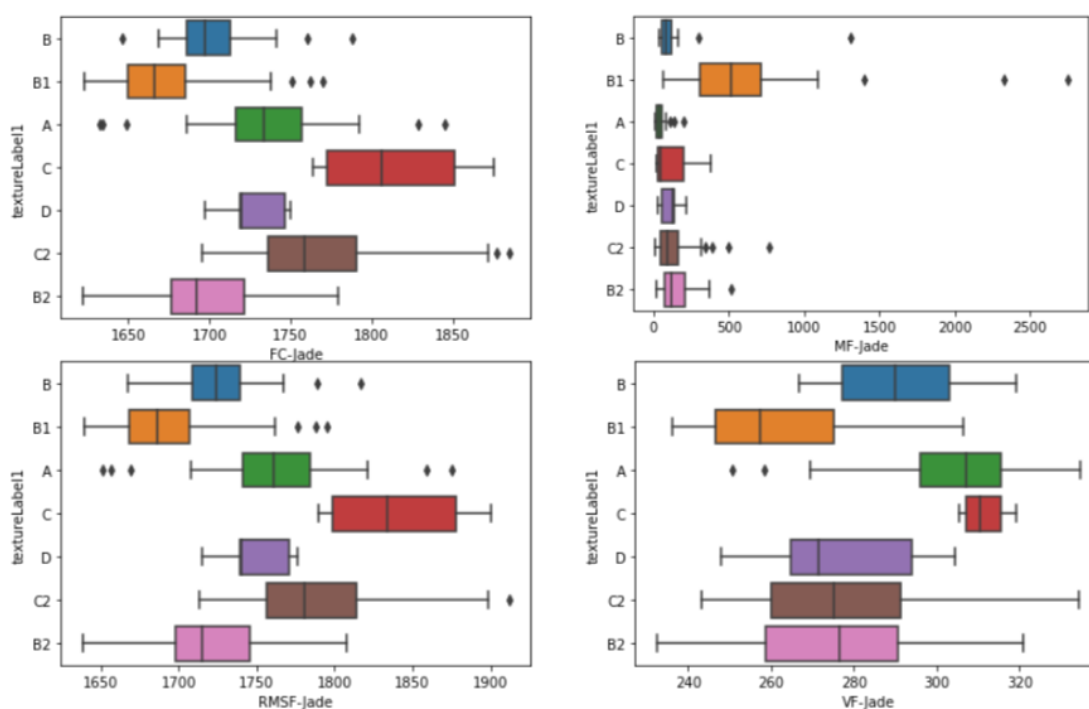


图 9 频域特征在区分玉类器物矿物属性的结果

因此，我们也选择将，重心频率、均方根频率和频率标准差的相应参数加入特征选取的模型，配合 Chi2, RF, LR, RFE 等方法进行分类。结果如表 3 所示。



表 3 特征选取 K=10 时分类结果

Methods	ACC	F1	PR	SE
ANOVA F-value	0.376	0.433	0.429	0.437
MIC	0.433	0.514	0.509	0.518
Chi2	0.393	0.450	0.449	0.453
RF	0.492	0.614	0.645	0.606
LR	0.311	0.493	0.547	0.473
RFE	0.529	0.787	0.746	0.788
ANOVA F-value+FDC	0.494	0.643	0.653	0.636
MIC+FDC	0.506	0.664	0.694	0.652
Chi2+FDC	0.536	0.697	0.761	0.688
RF+FDC	0.508	0.688	0.734	0.668
LR+FDC	0.384	0.500	0.513	0.493
RFE+FDC	0.529	0.818	0.877	0.783

本文选取关键频域频率并综合了中心频域和频域标准差后，再次进行分类，由标上结果可知，多数特征选取方法在 k=10 时的准确率都在 0.3-0.4 左右；F1、PR 和 SE 值都在 0.4-0.5 中，Random Forest 能超过 0.6，RDE 能达到 0.7。在增加了频域特征参数后，大多数方法的 ACC 都提升了 6%-12%，F1、PR 和 SE 都提升了 20%-30%达到 0.6-0.7，RFE 的 F1 甚至达到了 0.8。值得一提的是，RFE 模型的 ACC 在加入频域特征参数后没有明显提升，其原因是由于数据类型不均衡和测试集划分的随机性，对于占全数据集不足 1%数据量的类别和占比超过 50%的类别，若一次训练在训练集中划分的比例不理想，模型就会因为对于大比例类别划分错误多而 ACC 表现差，若划分的比例能使大类别得到充分训练，模型就会表现较好，ACC 达到近 0.8 的。对于这种情况，未来可以通过增加部分类别数据，平衡类型差异的方法来解决。

此外，由表 4 的交叉验证评分(Cross\_val\_score)可知，本文使用的改进方法较改进前，不仅有较高的性能也有较高的泛化能力。

表 4 特征选取 k=10 下交叉验证结果表

Methods	Cross_val_score
ANOVA F-value	0.569
MIC	0.671
Chi2	0.562
RF	0.744
LR	0.504
RFE	0.832
ANOVA F-value+FDC	0.771
MIC+FDC	0.790
Chi2+FDC	0.835
RF+FDC	0.779
LR+FDC	0.633
RFE+FDC	0.852

综上所述，本次实验的结果对于玉器矿物类型和材质类型的分类具有重要的实际意义。通过对玉器矿物类型和材质类型的高效、准确的分类，我们可以更好地评估玉器的质量和辅助考古学研究的推进。

## 5. 总结

本文基于 M54 亚长墓出土 200 多件玉器样品的近红外光谱数据，提出了一个智能软玉类型识别框架，包括光谱特征提取，特征重要性分析和玉器类型识别。本文创新性地将频域特征参数用于矿物类型与材质类型的分类对比中，重心频率、平均频率、均方根频率、频率响应和频谱平坦度是对分类性能影响最大的特征，改进后的方法在性能和泛化能力上都有明显提升。实验结果表明本文提出的框架能达到不错的识别性能，为考古研究提供重要衡量标准，为考古学提供新的分类思路，确定软玉的数据特点。为玉器的矿物类型和材质类型的研究和考古学的玉器分类需求提供了有力的技术支持。

## 参考文献

- [1] Chang, K. C. The archaeology of ancient China. Fourth edition. Yale University Press. 1986. 607-609
- [2] Rawson, J. Chinese jade from the neolithic to the qing. British Museum Press. 1995. 298-310
- [3] Mark, Cartwright. Jade in Ancient China. World History Encyclopedia. World History Encyclopedia, 2017. Retrieved from <https://www.worldhistory.org/article/1088/jade-in-ancient-china/>
- [4] 李学勤. 文物中的古文明. 北京: 商务印书馆, 2008.
- [5] Lu, R., & Liu, C. Jade and the Chinese mind. Chinese Sociology & Anthropology. 2002. 35(1), 51-72.
- [6] 许佳君. 廖宗廷. 周征宇; 和田、格尔木与溧阳三地软玉微观结构的对比研究[J]; 上海地质; 2008 年 01 期
- [7] Zhang, R. Y., & Liou, J. G. Formation of jadeite-bearing metasomatic rocks and their country rocks, north of Motagua Fault, Guatemala. International Geology Review. 1998. 40(10), 861-879.
- [8] Wu, X., Fan, X., Qin, F. et al. Compressibility and structural properties of jadeite, NaAlSi<sub>2</sub>O<sub>6</sub> at high pressure. J. Earth Sci. 24, 57–64 (2013). <https://doi.org/10.1007/s12583-013-0311-8>
- [9] Sun, J., Zhou, Q., Shang, J., & Sun, S. Research of machine learning algorithm for broadcasting spectrum signal processing. IOP Conference Series. Materials Science and Engineering. 2019. 569(5) doi:<https://doi.org/10.1088/1757-899X/569/5/052083>
- [10] 刘继富, 杨明星, 苏越, 等. 湖北随州曾侯乙墓出土玉器材质分析与产源初探[J]. 光谱学与光谱分析, 2022, 42(1): 215–221. DOI:10.3964/j.issn.1000-0593(2022)01-0215-07.
- [11] 姜亚飞. 西周时期玉器流通状况的初步研究[D]. 山东: 山东大学, 2021.
- [12] Mei, J., Chen, K., Wang, L., Wang, R., & Wu, J. A study of the elemental composition of ancient Chinese bronze mirrors by synchrotron radiation X-ray fluorescence. Journal of Analytical Atomic Spectrometry. 2015. 30(4), 1024-1028.
- [13] 成小林, 王建平. 一件新干商墓出土玉器的红外光谱和扫描电镜的分析研究[C]. 光谱学与光谱分析. 2006: 57–58.
- [14] Hanus, R., Sobek, K., Souček, K., Staš, L., Georgiou, L., & Selucká, A. A non-destructive analytical study of cultural heritage object from late antiquity: Gold framework and gemstone inlays.

Heritage Science, 2023, 11(1), 33. doi:<https://doi.org/10.1186/s40494-023-00874-y>

- [15] Rossman G R. Lavender Jade. The optical spectrum of  $\text{Fe}^{3+}$  and  $\text{Fe}^{2+} \rightarrow \text{Fe}^{3+}$  intervalence charge transfer in jadeite from Burma[J]. American Mineralogist, 1974, 59(7-8): 868-870
- [16] 彭明生, 李迪恩, 殷小玲; 台湾软玉的喇曼光谱研究[J]; 矿物岩石地球化学通报; 1997 年 S1 期
- [17] 郭立鹤, 韩景仪. 和田玉、玛纳斯碧玉和岫岩老玉中 M1、M3 阳离子占位的红外光谱分析[J]; 岩石矿物学杂志; 2002 年 S1 期
- [18] Harder H. Trace elements as colouring agents in jadeites[J]. The Journal of Gemology, 2012, 48(4) : 273-283.
- [19] Bao, Y., Zhao, C., Li, Y., & Yun, X. A method of determining heated ancient nephrite jades in china. Scientific Reports (Nature Publisher Group), 2018, 8, 1-11. doi:<https://doi.org/10.1038/s41598-018-30564-w>
- [20] Fisher, R.A. Statistical Methods for Research Workers. Oliver and Boyd, Edinburgh, Scotland. Nature 131, 383, 1925. <https://doi.org/10.1038/131383b0>
- [21] Cover, T. M., & Thomas, J. A. Elements of information theory. John Wiley & Sons. 2006. pp. 347-408. <https://doi.org/10.1002/047174882X.ch11>
- [22] Pearson, K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine Series 5, 50(302), 1900, 157-175.
- [23] Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. Applied logistic regression (Vol. 398). 2013. John Wiley & Sons.
- [24] Clark, R. N., & Roush, T. L. Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. Journal of Geophysical Research: Solid Earth, 1984, 89(B7), 6329-6340.
- [25] Zhang, R. Y., & Liou, J. G. Formation of jadeite-bearing metasomatic rocks and their country rocks, north of Motagua Fault, Guatemala. International Geology Review, 1998, 40(10), 861-879.
- [26] Huang, Z., Li, Q., Chen, J., Zhao, Z., Zhao, H., & Zhang, C. Discrimination of green nephrite from different origins by FT-NIR spectroscopy. Guang Pu Xue Yu Guang Pu Fen Xi, 2008, 28(8), 1804-1807.

## 致谢

感谢我的父母孙宁生先生和严军燕女生对我的从小到大的培养，感谢我的导师刘江老师对我的肯定对课题的指导与对我所有想法的支持，感谢章晓庆博士在数个研究项目中给我提供建议与意见，感谢 iMED 实验室里和我一起完成课题研究的小伙伴们，这些年感谢在每个时刻不断努力的自己。