基于高光谱成像技术的亚麻籽粒品种快速分类

李 怡*,韩俊英#

甘肃农业大学信息科学技术学院, 甘肃 兰州

收稿日期: 2024年3月27日; 录用日期: 2024年10月18日; 发布日期: 2024年10月30日

摘要

亚麻是世界上一种重要的经济作物和油料作物,快速筛选不同品种的亚麻籽对于亚麻育种以及农业种植 具有重要意义。市场上亚麻籽品种繁多,如何快速、准确、高效地鉴别亚麻籽的品种,从而实现亚麻品 种优育以及高产高收亟待解决。本文提出基于高光谱图像技术的K近邻(KNN)分类模型,以期实现不同品 种亚麻籽的快速分类。本研究以甘肃省农业科学院提供的10个不同品种的亚麻籽作为本次试验的样本, 每个品种随机选取50粒亚麻籽,利用高光谱成像系统采集870.07~1709.45 nm范围内的亚麻籽光谱图像。 将采集到的图像黑白校正之后,以50粒亚麻籽为感兴趣区域,求取该区域内平均光谱作为原始光谱数据。 由于原始光谱数据两端噪声较大,有效信息会受到噪声干扰,为了增强信噪比,本研究截取950~1680 nm 范围内的亚麻籽的光谱波段作为有效波段进行分析。为避免数据在采集过程中受无关信息的干扰较强从 而影响建模效果,因此对去噪后的光谱波段信息进行SG (Savitzky Golay)平滑预处理,并在SG平滑预处 理的基础上分别进行最大归一化(MN)和二阶求导预处理(2ndDer)。数据预处理后分别采用竞争性自适应 重加权算法(CARS)、连续投影算法(SPA)单一提取和CARS + SPA、CARS-SPA组合筛选方法提取特征波长, 利用K近邻算法建立CARS-KNN、SPA-KNN、CARS + SPA-KNN、CARS-SPA组合筛选方法提取特征波长, 利用K近邻算法建立CARS-KNN、SPA-KNN、CARS + SPA-KNN、(SG-2ndDer)-CARS + SPA-KNN、 (SG-2ndDer)-CARS-SPA-KNN这4种分类模型对亚麻籽的分类准确率最高可以达到100%。故利用近红外 高光谱成像技术结合KNN算法对亚麻籽品种进行快速无损鉴别是优异和可靠的方案。

关键词

高光谱成像技术,亚麻,品种分类,K近邻算法,组合筛选

Rapid Classification of Flax Seed Grain Varieties Based on Hyperspectral Imaging Technique

Yi Li*, Junying Han#

*第一作者。 #通讯作者。 College of Information Science and Technology, Gansu Agricultural University, Lanzhou Gansu

Received: Mar. 27th, 2024; accepted: Oct. 18th, 2024; published: Oct. 30th, 2024

Abstract

Flax is an important economic and oil crop globally; rapid screening of different flax seed varieties is essential for flax breeding and agricultural cultivation. There are many flax seed varieties on the market; how to quickly, accurately, and efficiently identify the flax seed varieties to realize the superiority of flax varieties and high-vield vield needs to be resolved. This paper proposes the K-Nearest Neighbor (KNN) classification model based on the hyperspectral image technique to classify flax seed varieties quickly. In this study, ten different varieties of flax seeds provided by the Gansu Provincial Academy of Agricultural Sciences were used as the samples for this experiment, and 50 flax seeds of each variety were randomly selected. The spectral images of flax seeds in the range of 870.07~1709.45 nm were collected by using a hyperspectral imaging system. After the black-and-white correction of the acquired images, the 50 flax seeds were the region of interest. and the average spectrum in the region was the original spectral data. Due to the significant noise at both ends of the raw spectral data, the practical information will be interfered with by noise; in order to enhance the signal-to-noise ratio, this study intercepted the spectral bands of linseed in the range of 950~1680 nm as the influential bands for analysis. In order to avoid the data in the acquisition process by irrelevant information interference is more muscular, thus affecting the modeling effect, the de-noised spectral band information is pre-processed by SG (Savitzky Golay) smoothing, and based on SG smoothing, respectively, the Maximum Normalization (MN) and second-order derivation pre-processing (2ndDer). After data preprocessing, the feature wavelengths were extracted by Competitive Adaptive Re-weighting Algorithm (CARS). Successive Projection Algorithm (SPA) single extraction and CARS + SPA, CARS-SPA combined screening methods, respectively, and the K-Nearest Neighbor algorithm was utilized to establish the identification models for the four flax seed varieties, namely, CARS-KNN, SPA-KNN, CARS + SPA-KNN, and CARS-SPA-KNN. The experimental results show that the classification accuracy of flax seed based on the four classification models (SG-2ndDer)-CARS-KNN, (SG-2ndDer)-SPA-KNN, (SG-2ndDer)-CARS + SPA-KNN, and (SG-2ndDer)-CARS-SPA-KNN) can reach up to 100%. Therefore, using near-infrared hyperspectral imaging technology combined with the KNN algorithm for rapid and non-destructive identification of linseed varieties is an excellent and reliable program.

Keywords

Hyperspectral Imaging Technique, Flax, Classification of Varieties, K-Nearest Neighbor Algorithm, **Combined Screening**

Copyright © 2024 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/ (cc)

۲ **Open Access**

1. 引言

亚麻属亚麻科亚麻属一年生草本植物。生产上应用的为栽培亚麻,简称亚麻,分为纤用、油用、油 纤兼用三种。油用亚麻俗称胡麻。亚麻是世界上最古老的作物之一, 原产于波斯湾、黑海及里海[1]。亚 麻籽含油率能达到 34%~45%,其中胚芽含油量约为 45%,子叶含油量约为 51%,种皮和胚乳含油量约为 23% [2],属于高油料作物。亚麻籽中包含丰富的木酚素、油酸、亚油酸和亚麻酸、蛋白质和矿物质等营 养成分[3] [4],其中α-亚麻酸具有降血脂、调节心脏功能、改善血液循环、增强记忆力等功效。亚麻籽中 α-亚麻酸含量在58%左右,是人体所需靠外界获取的不饱和脂肪酸主要来源之一,可媲美"深海鱼油" [5]。研究表明,α-亚麻酸可在人体肝脏内生成二十碳五烯酸(EPA)和二十二碳六烯酸(DHA) [6],而 EPA 和 DHA 就是人们所谓的"脑黄金",是保证人体和大脑正常工作的重要不饱和脂肪酸[7]。此外,用亚 麻籽榨取的食用油中还含有丰富的生育酚和类胡萝卜素,是一种功能性保健食用油[8]。

近年来,快速无损鉴别亚麻籽品种已成为亚麻育种的一个重要研究课题。随着高光谱成像技术的迅速发展,为作物快速无损监测提供了一种便捷的技术手段[9] [10]。利用高光谱技术对目标物进行无损检测的方法已经在农作物种子品质、活力检测、品种识别等方面有着广泛应用。使用高光谱技术对农作物或者农产品的品种分类检测的研究较多,国内外专家学者开展了大量的研究,例如: Zhu 等采集了 10 个品种大豆的高光谱数据,使用卷积神经网络建立了大豆品种识别模型,测试集最高精度达到了 97.2% [11]; Moola等(2021)则基于 Sentinel 1A SAR 图像的时间序列,采用加权的动态时间规整算法和模糊分类方法,实现了蔬菜作物的有效分类[12];曹宇佳等采用 PCA 算法对高光谱数据进行了降维并采用支持向量机分类方法对水稻、玉米、大豆、烤烟等典型农作物高光谱数据进行了分类,分类总体精度达到 87.08% [13]; 袁建清等建立了基于 CARS-PCs 特征的 SVM 模型和 LDA 模型,总体分类精度分别为 95%和 97%,实现 了田间水稻冠层穗颈瘟的早期识别[14]; Zhang 等建立了基于近红外高光谱成像技术的黄秋葵种子品种鉴 定模型,结果表明,基于 SVM 建立的模型取得了较好的结果,为杂交黄秋葵种子的快速筛选提供了新方 法[15];孙婷提出了一种快速无损的高粱品种分类方法和混合高粱品种识别方法,使用 PCA + MD 异常 数据剔除方法结合 MSC 预处理方法,再进行 SPA 特征波长筛选建立的 SVM 分类模型,训练集、测试集 精度达到 100%,外部验证集精度达到 90.91% [16]。

现今已有大量的国内外学者对农作物种子品种鉴别展开了一系列研究,其中多以小麦、玉米、花生、 水稻等大面积种植的粮食作物为研究对象,但对亚麻籽等油料作物的品种识别鲜有文献报道;另外,文 献中对农作物品种鉴定识别多使用 SPA 方法提取特征波段,该方法由于提取的特征波段数目偏少,容易 出现有效信息缺失的情况;而且,文献中报道的品种鉴定识别模型多以 LS-SVM 为主,该方法容易陷入 局部最优,全局搜索能力较差且鲁棒性较低。基于上述原因,本文提出一种基于高光谱成像技术的亚麻 籽品种快速鉴定识别方法,以期达到准确提取亚麻籽特征波段并进行全局搜索建立品种鉴定识别模型。

实验以甘肃省农业科学院作物研究所提供的 10 个不同品种的亚麻籽为试材,每个品种随机选取 50 粒亚麻籽,分别用近红外高光谱成像系统采集亚麻籽的高光谱图像,并以 50 粒亚麻籽为感兴趣区域,求取感兴趣区域平均光谱作为原始数据。为了消除噪声信息对原始数据的影响,首先对原始数据进行 SG 滤波平滑预处理,然后在 SG 滤波平滑预处理基础上再分别进行最大归一化和二阶求导两种处理方式。 采用 CARS 和 SPA 单一方法以及 CARS + SPA、CARS-SPA 组合筛选方法提取特征波长,并利用*K*近邻 算法建模分析,实现对亚麻籽品种的鉴别,以期为近红外高光谱成像技术鉴别油料作物的实际应用提供 一种技术参考。

2. 材料与方法

2.1. 试验材料、设备及软件

2.1.1. 试验材料

田间试验时间是 2022 年 4 月 20 日至 8 月 30 日,试验地点位于甘肃省农业科学院作物研究所景泰县 试验基地内(北纬 37°19',东经 104°08')。试验品种选择中国西北地区广泛种植的 Onyc、张亚 3 号、晋亚 7 号等 10 个品种。为了提高本研究中图像样本的多样性和差异性,从而更好地验证模型的泛化性能和分 类准确性,每个品种共选择代表性植株 100 株;代表性植株在不同地块的不同种植行中尽量选择长势存 在明显差异的植株,并进行挂牌标记。将代表性植株脱粒后称取 75 g 分装进牛皮纸袋,并按照 1~10 的 顺序做好品种标记,运回甘肃农业大学实验室,置于通风干燥阴凉处保存,如图 1 所示。



图 1. 亚麻籽品种展示(共 10 种)

2.1.2. 试验设备与软件

高光谱图像采集系统主要包括 1 台 GaiaField-N17E-GE 推扫型近红外高光谱仪、2 组高稳定性的卤素 灯阵列光源、暗箱和 1 台计算机共同组成,如图 2 所示。近红外高光谱成像系统采集光谱范围是 870.07~1709.45 nm, bands = 512,光谱通道数为 380,光谱分辨率为 2.8 nm,入射光狭缝宽度为 30 μm。 利用 GaiaField 配套的 SpectraVIEW 软件获取亚麻籽高光谱图像。数据处理和绘图软件为 ENVI 5.3、The Unscramber X10.4、Excel 2021、Matlab 2023a、Adobe Photoshop CS6。



Figure 2. Hyperspectral image acquisition system 图 2. 高光谱图像采集系统

2.2. 试验方法

2.2.1. 高光谱图像的采集

将近红外光谱仪启动并预热 30 min,以保证设备能够正常运行。将曝光时间设置为 1.00 ms,物距为 700 mm,帧速设置为 31,近红外光谱仪的光谱图像采集方式为镜头内置推扫,推扫前进速度为 0.04662 cm/s,回退速度为 0.2 cm/s,采集到的高光谱图像块尺寸为 640 × 600 × 512 (像素 × 波段数,即每个样本 采集 512 个波段下像素为 640 × 600 的二维图像)。近红外高光谱仪推扫过程中从右向左扫描样本,每次

扫描得到1行图像的光谱信息,镜头推扫过程中,获取其他位置直至整个样本的光谱信息。为避免杂光的干扰,图像采集的过程全部在暗箱中完成。在收集到的10个品种亚麻籽中,选取籽粒均色质均匀、颗粒饱满、无机械损伤的亚麻籽作为各品种试验样本。采集亚麻籽的光谱图像时,每次选取一个品种,从 牛皮纸袋中倒出亚麻籽,随机选取50粒样本放入暗箱里面的载物台中央部位,以期保证镜头位于样本正 上方,并将样本均匀铺开以避免亚麻籽相互重叠,然后进行光谱图像采集。每次光谱图像采集结束后, 将50粒样本倒入牛皮纸袋,将袋口封紧并摇晃纸袋30秒,使前一次选取的亚麻籽样本与其他亚麻籽均 匀混合,再次倒出全部亚麻籽随机选取50粒亚麻籽作为样本进行光谱采集,同一样本品种采集5张高光 谱图像。

2.2.2. 光谱校正

为了消除摄像头中因暗电流存在及光源强度在不同波段下分布不均而产生的噪声影响,需要对获得的高光谱图像进行黑白校正。打开 SpectraVIEW 点击标准板对反射率为 99%的标准白色校正板进行图像 采集,可获得纯白的标定图像 W,然后点击背景,近红外高光谱仪自动关闭镜头采集黑色的标定图像。黑白帧校对公式为:

$$U = \frac{U_i - B}{W - B} \tag{1}$$

式中, U_i ——原始的高光谱图像, U——校正后的图像。

2.3. 高光谱图像分析

2.3.1. 高光谱数据的提取

使用 ENVI 5.3 软件打开校正后的高光谱图像,如图 3(a)所示。为了消除图像背景信息干扰,从 50 粒亚麻籽中随机选取 5 粒作为 ROI1 (红色),选取背景区域为 ROI2 (蓝色),如图 3(b)所示,用支持向量机 对根据选定的感兴趣区域对图像进行二值化处理,如图 3(c)所示。在得到二值化图像之后将图像矢量化 并进行掩膜处理,如图 3(d)所示。通过应用掩膜功能将掩膜图像应用到高光谱图像上面,使图像中的亚 麻籽从背景中提取出来,生成 50 粒亚麻籽样本的感兴趣区域,如图 3(e)所示。以所有样本作为感兴趣区 域计算样本的平均光谱值,并导出亚麻籽的平均光谱曲线,如图 3(f)所示。为了扩大数据集,本次实验 对每张高光谱图像进行光谱数据提取过程中,将上述过程重复 6 次,每次从光谱图片的 50 粒亚麻籽中随 机选取 5 粒作为感兴趣区域,求出感兴趣区域的平均光谱,共得到 300 组光谱数据作为实验数据。

2.3.2. 光谱数据的预处理

在高光谱图像采集过程中,采集环境、仪器状态等因素都可能引起测量误差,导致采集到的高光谱 图像中存在随机噪声和干扰信息,从而对模型的准确性和鲁棒性带来不利影响。因此,在进一步分析高 光谱图像之前,需要采用合适的方法进行预处理,从而提高光谱数据的信噪比。本研究中采用了 SG 滤 波平滑法、SG-2ndDer 和 SG-MN 三种数据预处理方法。

1) SG 滤波平滑法

Savitzky Golay 滤波平滑算法是由 Savitzky 和 Golag 提出来的,被广泛的用于数据除噪,是一种基于最小二乘原理的多项式平滑算法。SG 滤波器可以有效地保留高光谱图像中的有用信息,同时降低高频噪音,提高数据的信噪比。SG 滤波平滑法的主要原理是对移动窗口内的光谱信号进行指定次数的多项式拟合,替代含有噪声信息的原曲线。

2) 二阶求导(2ndDer)

光谱的一阶和二阶导数常用于光谱基线校正,可以有效地消除基线漂移或其他背景的干扰,提高分

 (a) 高光谱图像
 (b)选取 ROI
 (c)图像二值化

 (d)图像掩膜
 (e)应用掩膜
 (f)亚麻光谱

辨率与灵敏度。但由于光谱中存在部分高频噪声,若处理不当,导数法会放大这些噪声,降低信噪比。

Figure 3. Spectral data extraction 图 3. 光谱数据提取

3) 最大归一化法(MN)

最大值归一化法是机器学习中常见的一种数据预处理方法。其基本原理是找出一组数据集当中的最 小值,每个数据减去最小值并除以最大值与最小值的差值。由于不同特征之间的范围差异可能会较大, 进而影响算法的性能,最大值归一化法可以很好地消除这种影响。

2.4. 建模方法和模型评价标准

2.4.1. 不同的区分方法

*K*近邻算法的基本原理是根据高维空间的距离度量,检索出*K*个最相似近邻,并根据近邻样本所属 的类别标签进行分析,判断待测样本的类别[17]。*K*近邻算法具有支持增量学习的特性,且不同类别间的 子分类器不需要进行先验训练,这使*K*近邻算法具有很强的自适应性[18]。在样本数据充分的情况下,*K* 近邻算法的分类性能最佳,并且少量的噪声对并不会对算法造成干扰,使*K*近邻算法的抗变化性和对噪 声的鲁棒性得到了极大的增强。

当待测样本近邻类别数 K 确定的情况下,可以采用不同的距离度量方法来判断待测样本的类别。但 亚麻籽的内部作用机理较为复杂,在获得亚麻籽的光谱信息后无法准确判断采用哪一种距离度量方法更 合理。因此,本文选用 Matlab 内置函数中的 8 种不同的距离度量方法对 10 个不同品种的亚麻籽进行分 类,其计算方法如下:

城市街区距离:

$$D_{Cityblock}\left(X,Y\right) = \sum_{i=1}^{n} \left|X_{i} - Y_{i}\right|$$

$$\tag{2}$$

切比雪夫距离:

$$D_{Chebychev}(X,Y) = max |X_i - Y_i|$$
(3)

相关距离:

$$D_{Correlation}(X,Y) = \frac{C_{ov}(X,Y)}{V_{AR}(X)V_{AR}(Y)}$$
(4)

式中, C_{ov} ——协方差; V_{AR} ——方差。

余弦相似度:

$$D_{Cos}(X,Y) = \frac{\langle X_i, Y_i \rangle}{\|X_i\| * \|Y_i\|}$$
(5)

欧几里得距离:

$$D_{Euclidean}\left(X,Y\right) = \sqrt{\sum_{i=0}^{n} \left(X_i - Y_i\right)^2}$$
(6)

马氏距离:

$$D^{2}(x,\mu) = (x-\mu)' v^{-1}(x-\mu)$$
(7)

式中,定义*X*与母体*G*的距离 $D^2(x,G)$ 为*X*与均值 μ 的距离 $D^2(x,\mu)$,则 $D^2(x,\mu)$ 为马氏距离。 明克夫斯基距离:

$$D_{Minkowski}\left(X,Y\right) = \sqrt[l]{\sum_{i=0}^{n} \left(X_i - Y_i\right)^i}$$
(8)

标准化欧式距离:

$$D_{Seuclidean}\left(X,Y\right) = \sqrt{\sum_{i=1}^{n} \left(\frac{X_i - \overline{X}}{V_{AR}\left(X\right)} - \frac{Y_i - \overline{Y}}{V_{AR}\left(Y\right)}\right)^2}$$
(9)

由于 KNN 算法通过度量待测样本与其近邻间的距离来判断二者间的相似度,如果对特征值进行距离 计算,则取值范围较大的特征值会对结果产生绝对的影响,但较小的特征值却几乎没有任何作用,这导 致数据的数值范围变化容易影响*K*近邻算法的分类结果。因此,需先对样本数据进行归一化后再进行距 离的计算和判断。

2.4.2. 研究技术路线图与论文结构

本文的实验流程图,如图4所示。

3. 结果与分析

3.1. 特征波长的提取与分析

由于高光谱曲线在 870.07~947.47 nm 以及在 1652.31~1709.45 nm 范围内的光谱信息受到噪声的影响 变得非常模糊,因此本次试验实际有效波长范围为 950~1680 nm 之间共 445 个波段的光谱数据。图 5 为 10 个不同样本在 870~1709 nm 近红外光谱图,反映了样本中的 C-H、O-H、N-H、S-H 等含氢基团的特 征信息。由于亚麻籽中含有较高含量的蛋白质和油分等化学物质,化学键不尽相同,故在 1120~1300 nm、 1350~1600 nm、1600~1900 nm 和 1900~2100 nm 有明显的不同吸收峰[19]。由于 10 个不同品种的亚麻籽 在同一波长范围内的反射强度和波长变化曲线高度一致,各品种光谱曲线之间没有明显的分离,双亚 15 号与雁杂 10 号,双油麻 1 号、宁亚 10 号与晋亚 7 号光谱曲线分别有交叉,证明各品种光谱曲线之间相 似度较高,难以靠肉眼观察进行不同品种之间的准确区分。因此,建立分类模型对不同品种的光谱数据 进行分析和品种的区分是必要的。



Figure 4. Experimental flowchart 图 4. 实验流程图



Figure 5. Original spectral curve of flax seed 图 5. 亚麻籽原始光谱曲线

常见的特征提取方法一般被分为波段提取与波点提取两大类,前者基于波段进行分析从而提取出特征波段作为特征波长,而后者则基于波点进行分析从而提取出特征波点作为特征波长[20]。本文采用竞争性自适应重加权算法(CARS) [21]和连续投影算法(SPA) [22]提取特征波点作为特征波长。

CARS 算法通常被应用在机器学习和模式识别中,CARS 算法是由多个学习算子组成,这些算子根据自身的重要性进行加权。用自适应加权采样技术筛选出 PLS 模型中回归系数绝对值较大的波长点,去除权重较小的波长点,并通过交互校验优筛选出均方根误差最小时所对应的波点作为特征波长。

设置蒙特卡洛采样次数为 50,采用 10 折交叉验证法提取特征波长,如图 6 所示。图 6(a)表示 CARS 优选特征波长变量数目随采样次数增大而不断减小,在第 23 次采样前,优选特征波长变量数急剧下降,第 23 次采样后,优选特征波长数目缓慢下降,证明 CARS 优选特征波长数从粗选到精选的一个过程;图 6(b)表示交叉验证均方根误差(RMSECV)随采样次数增加先缓慢减小后逐渐增大,说明在第 11 采样后 CARS 出现了筛选过度的情况,导致剔除了含有有效信息的敏感波长变量,导致模型精确度下降,RMSECV 值逐渐增大;图 6(c)表示在第 11 次采样时 RMSECV 值最小,此时优选出 165 个特征波段变量。



Figure 6. Process of extracting feature wavelengths using CARS algorithm 图 6. CARS 算法提取特征波长过程

SPA 算法是由 Bregman 提出,是一种向前循环特征提取算法,这种算法可以使波段间的共线性最小化。SPA 算法可以对光谱数据进行降维,有效减少光谱中的噪声信息,提升模型的准确性与稳定性。

设置优选最大波长数目为 140, 共优选出 138 个波长变量, 如图 7 所示。图 7(a)表示随变量个数的增加, 均方根误差(RMSECV)整体呈现先急剧下降后缓慢减小的趋势。当变量个数为 138 时, RMSECV 变

化不再明显,由于继续增加变量数目会使模型运算量和复杂度进一步增加,因此最终选取 138 个变量作 为最终变量数目。



Figure 7. Process of extracting feature wavelengths using SPA algorithm 图 7. SPA 算法提取特征波长过程

为避免 CARS 和 SPA 单一特征波长提取方法提取出的变量具有共线性,本文再采用 CARS + SPA 和 CARS-SPA 两种组合方法分别进行提取特征波长。其中, CARS + SPA 表示两者分别提取特征波长后取 并集; CARS-SPA 表示用 SPA 对 CARS 提取的特征波长变量二次提取, 消除共线性问题。

3.2. KNN 模型建立与分析

KNN 方法在类别决策过程中,只与少量的相邻样本有关。由于 KNN 方法不依靠判别类域的方式来 判断样本的所属类别,而是通过附近的有限个数的邻近样本来进行判断其类别。本研究对原始数据分别 使用上文所述三种不同的数据预处理方法进行处理,得到的结果作为三个不同的数据集用于后续实验使 用,并在实验过程中将每个数据集都按照 7:3 的比例随机划分为训练集与测试集。将 CARS、SPA、CARS + SPA, CARS-SPA 四种特征提取方法与 KNN 模型结合组建新的分类模型,当 *K* 取值为 1、3、5、7、9 时,本文选用的 8 种距离度量方法在不同模型中的分类结果分别如表 1~12 所示。

K	城市街区 距离	切比雪夫 距离	相关距离	余弦距离	欧几里得 距离	马氏距离	明克夫斯基	标准 欧几里得
1	53.84%	57.14%	21.97%	59.34%	52.74%	84.61%	52.74%	56.04%
3	42.85%	51.64%	28.57%	50.54%	41.75%	72.52%	41.75%	48.35%
5	43.95%	47.25%	27.47%	46.15%	48.35%	70.33%	48.35%	42.85%
7	45.05%	46.15%	28.57%	46.15%	47.25%	64.83%	47.25%	38.46%
9	41.75%	45.05%	29.67%	40.65%	45.05%	62.63%	45.05%	37.36%

Table 1. SG-CARS-KNN (test set) 表 1. SG-CARS-KNN (测试集)

表 2. SG-SPA-KNN (测试集)										
K	城市街区 距离	切比雪夫 距离	相关距离	余弦距离	欧几里得 距离	马氏距离	明克夫斯基	标准 欧几里得		
1	81.31%	84.61%	98.90%	97.80%	82.41%	98.90%	82.41%	78.02%		
3	72.52%	78.02%	98.90%	96.70%	69.23%	98.90%	69.23%	72.52%		
5	70.33%	72.52%	98.90%	93.40%	67.03%	98.90%	67.03%	67.03%		
7	61.53%	69.23%	80.22%	79.12%	56.04%	89.01%	56.04%	63.73%		
9	58.24%	69.23%	67.03%	70.33%	60.44%	85.71%	60.44%	60.43%		

Table 2. SG-SPA-KNN (test set) 表 2 SG-SPA-KNN (测试集)

Table 3. SG-CARS + SPA-KNN (test set) 表 3. SG-CARS + SPA-KNN (测试集)

K	城市街区 距离	切比雪夫 距离	相关距离	余弦距离	欧几里得 距离	马氏距离	明克夫斯基	标准 欧几里得
1	72.52%	84.61%	98.90%	97.80%	82.41%	98.90%	82.41%	78.02%
3	72.52%	78.02%	98.90%	96.70%	69.23%	98.90%	69.23%	72.52%
5	70.33%	72.52%	98.90%	93.40%	67.03%	97.80%	67.03%	65.93%
7	62.63%	69.23%	80.22%	78.02%	56.04%	87.91%	56.04%	63.73%
9	58.24%	69.23%	67.03%	72.52%	60.44%	76.92%	60.44%	62.63%

Table 4. SG-CARS-SPA-KNN (test set) 表 4. SG-CARS-SPA-KNN (测试集)

~4. SG-CARS-SPA-KNN (测试集)	

K	城市街区 距离	切比雪夫 距离	相关距离	余弦距离	欧几里得 距离	马氏距离	明克夫斯基	标准 欧几里得
1	70.33%	82.41%	98.90%	93.40%	71.42%	98.90%	71.42%	68.13%
3	54.94%	59.34%	97.80%	86.81%	56.04%	98.90%	56.04%	63.73%
5	50.54%	51.64%	89.01%	81.31%	50.54%	94.50%	50.54%	57.14%
7	48.35%	45.05%	75.82%	65.93%	47.25%	90.11%	47.25%	52.74%
9	48.35%	50.54%	54.94%	64.83%	45.05%	87.91%	45.05%	51.64%

Table 5. (SG-MN)-CARS-KNN (test set) 表 5. (SG-MN)-CARS-KNN (测试集)

K	城市街区 距离	切比雪夫 距离	相关距离	余弦距离	欧几里得 距离	马氏距离	明克夫斯基	标准 欧几里得
1	57.14%	64.83%	63.73%	73.62%	59.34%	98.90%	59.34%	58.24%
3	47.25%	48.35%	59.34%	64.83%	48.35%	89.01%	48.35%	52.74%
5	39.56%	49.45%	49.45%	58.24%	46.15%	87.91%	46.15%	49.45%
7	39.56%	48.35%	48.35%	51.64%	41.75%	85.71%	41.75%	48.35%
9	40.65%	49.45%	51.64%	54.94%	39.56%	84.61%	39.56%	41.75%

K	城市街区 距离	切比雪夫 距离	相关距离	余弦距离	欧几里得 距离	马氏距离	明克夫斯基	标准 欧几里得
1	97.80%	98.90%	98.90%	98.90%	98.90%	98.90%	98.90%	86.81%
3	93.40%	98.90%	98.90%	96.70%	97.80%	98.90%	97.80%	78.02%
5	90.11%	93.40%	96.70%	93.40%	94.50%	97.80%	94.50%	69.23%
7	85.71%	73.62%	81.31%	81.31%	78.02%	91.20%	78.02%	63.73%
9	80.22%	64.83%	64.83%	68.13%	65.93%	85.71%	65.93%	62.63%

Table 6. (SG-MN)-SPA-KNN (test set) 表 6. (SG-MN)-SPA-KNN (测试集)

Table 7. (SG-MN)-CARS + SPA-KNN (test set) 表 7. (SG-MN)-CARS + SPA-KNN (测试集)

Κ	城市街区 距离	切比雪夫 距离	相关距离	余弦距离	欧几里得 距离	马氏距离	明克夫斯基	标准 欧几里得
1	97.80%	98.90%	98.90%	98.90%	98.90%	98.90%	98.90%	81.31%
3	93.40%	98.90%	98.90%	96.70%	96.70%	98.90%	96.70%	73.62%
5	84.61%	93.40%	96.70%	92.30%	93.40%	95.60%	93.40%	62.63%
7	78.02%	73.62%	81.31%	81.31%	80.22%	94.50%	80.22%	64.83%
9	73.62%	64.83%	64.83%	73.62%	70.33%	89.01%	70.33%	64.83%

Table 8. (SG-MN)-CARS-SPA-KNN (test set) 表 8. (SG-MN)-CARS-SPA-KNN (测试集)

K	城市街区 距离	切比雪夫 距离	相关距离	余弦距离	欧几里得 距离	马氏距离	明克夫斯基	标准 欧几里得
1	79.12%	92.30%	97.80%	80.22%	90.11%	98.90%	90.11%	72.52%
3	68.13%	96.70%	91.20%	72.52%	85.71%	98.90%	85.71%	61.53%
5	69.23%	91.20%	83.51%	61.53%	82.41%	94.50%	82.41%	58.24%
7	64.83%	72.52%	69.23%	54.94%	74.72%	89.01%	74.72%	56.04%
9	61.53%	63.73%	68.13%	59.34%	71.42%	87.91%	71.42%	52.74%

Table 9. (SG-2ndDer)-CARS-KNN (test set) 表 9. (SG-2ndDer)-CARS-KNN (测试集)

K	城市街区 距离	切比雪夫 距离	相关距离	余弦距离	欧几里得 距离	马氏距离	明克夫斯基	标准 欧几里得
1	97.78%	83.33%	100.00%	100.00%	92.22%	94.44%	92.22%	100.00%
3	94.44%	75.56%	94.44%	94.44%	84.44%%	72.22%	84.44%	98.89%%
5	93.33%	65.56%	86.67%	85.56%	76.67%	74.44%	76.67%	95.56%
7	85.56%	60.00%	72.22%	74.44%	64.44%	71.11%	64.44%	93.33%
9	80.00%	56.67%	67.78%	67.78%	61.11%	68.89%	61.11%	91.11%

K	城市街区 距离	切比雪夫 距离	相关距离	余弦距离	欧几里得 距离	马氏距离	明克夫斯基	标准 欧几里得
1	100.00%	96.67%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
3	94.44%	87.78%	94.44%	94.44%	91.11%	93.33%	91.11%	94.44%
5	90.00%	83.33%	80.00%	81.11%	90.00%	82.22%	90.00%	88.89%
7	75.56%	72.22%	57.78%	57.78%	78.89%	75.56%	78.89%	80.00%
9	60.00%	61.11%	41.11%	43.33%	63.33%	54.44%	63.33%	70.00%

Table 10. (SG-2ndDer)-SPA-KNN (test set) 表 10. (SG-2ndDer)-SPA-KNN (测试集)

Table 11. (SG-2ndDer)-CARS + SPA-KNN (test set) 表 11. (SG-2ndDer)-CARS + SPA-KNN (测试集)

K	城市街区 距离	切比雪夫 距离	相关距离	余弦距离	欧几里得 距离	马氏距离	明克夫斯基	标准 欧几里得
1	100.00%	96.67%	100.00%	100.00%	100.00%	87.78%	100.00%	100.00%
3	94.44%	87.78%	94.44%	94.44%	91.11%	58.89%	91.11%	95.56%
5	90.00%	83.33%	80.00%	81.11%	90.00%	57.78%	90.00%	93.33%
7	75.56%	72.22%	57.78%	57.78%	78.89%	50.00%	78.89%	91.11%
9	60.00%	61.11%	42.22%	43.33%	63.33%	47.78%	63.33%	91.11%

Table 12. (SG-2ndDer)-CARS-SPA-KNN (test set) 表 12. (SG-2ndDer)-CARS-SPA-KNN (测试集)

K	城市街区 距离	切比雪夫 距离	相关距离	余弦距离	欧几里得 距离	马氏距离	明克夫斯基	标准 欧几里得
1	96.67%	87.78%	94.44%	94.44%	90.00%	100.00%	90.00%	100.00%
3	87.78%	80.00%	88.89%	88.89%	81.11%	94.44%	81.11%	94.44%
5	82.22%	71.11%	80.00%	80.00%	73.33%	85.56%	73.33%	84.44%
7	77.78%	62.22%	75.56%	75.56%	70.00%	73.33%	70.00%	71.11%
9	64.44%	55.56%	53.33%	52.22%	58.89%	61.11%	58.89%	56.67%

通过表 1~12 可以看出,当 K 选取 1、3、5 时,亚麻籽分类模型正确识别率大于 K 取 7、9 的结果。这 是由于当 K 值选取过大时,在特征空间中覆盖的局部区域过大,分类器的拟合性能会被降低,因此当 K 值为 1、3、5 时,模型能够在拟合和泛化性能间获得较好的均衡。在 SG-CARS-KNN 分类模型中,基于 马氏距离公式的分类表现果最好,准确率达到 84.61%;在 SG-SPA-KNN 分类模型中,基于相关距离公 式和基于马氏距离公式的分类效果最好,准确率最高达到 98.9%;在 SG-CARS + SPA-KNN 分类模型中, 基于相关距离公式和基于马氏距离公式的分类效果最好,准确率最高达到 98.9%;在 SG-CARS-SPA-KNN 分类模型中,基于相关距离公式和基于马氏距离公式的分类效果最好,准确率最高达到 98.9%;在 (SG-MN)-CARS-KNN 分类模型中,基于马氏距离的分类效果最好,准确率最高达到 98.9%;在 (SG-MN)-SPA-KNN 分类模型中,基于相关距离、余弦相似度、欧几里得距离、马氏距离、切比雪夫距 离、明克夫斯距离的分类效果最好,分类准确率最高达到 98.9%;在(SG-MN)-CARS + SPA-KNN 分类模型中,基于相关距离、余弦相似度、欧几里得距离、马氏距离、切比雪夫距离、明克夫斯距离的分类效果最好,分类准确率最高达到 98.9%;在(SG-MN)-CARS-SPA-KNN 分类模型中,基于马氏距离公式的分类效果最好,最高达到 98.9%;在(SG-2ndDer)-CARS-SPA-KNN 分类模型中,基于相关距离、余弦相似度、标准欧式化距离公式下的分类效果最好,准确率最高达到 100%;在(SG-2ndDer)-SPA-KNN 分类模型中,基于城市街区距离、相关距离、余弦相似度、欧几里得距离、马氏距离、明克夫斯距离、标准欧式化距离公式的分类效果最好,准确率最高达到 100%;在(SG-2ndDer)-CARS + SPA-KNN 分类模型中,基于城市街区距离、相关距离、余弦相似度、欧几里得距离、明克夫斯距离、标准欧式化距离公式的分类效果最好,准确率最高达到 100%;在(SG-2ndDer)-CARS + SPA-KNN 分类模型中,基于城市街区距离、相关距离、余弦相似度、欧几里得距离、明克夫斯距离、标准欧式化距离公式的分类效果最好,准确率最高达到 100%。对比原始数据经同一种预处理方法后用不同的单个特征提取方法提取特征变量进行分类的结果可以发现,基于 SPA 特征提取方法的分类模型准确率高于基于CARS 特征提取方法的分类模型准确率。对比数据经同一种预处理方法后用不同的组合特征提取方法提取特征变量进行分类的结果可以发现,基于 CARS + SPA 特征提取方法的分类模型准确率高于基于CARS-SPA 特征提取方法的分类模型准确率。

3.3. 模型泛化性测试

为进一步验证分类模型的泛化性以及模型是否出现过拟合现象,本文依照 2.3.1 节所述方式采集 10 个不同品种亚麻籽粒样本 200 组高光谱数据,并使用相同数据预处理方法对其进行处理后作为验证集数 据,验证集在模型中的分类效果如表 13~24 所示。

K	城市街区 距离	切比雪夫 距离	相关距离	余弦距离	欧几里得 距离	马氏距离	明克夫斯基	标准 欧几里得
1	58.94%	53.64%	50.99%	50.99%	58.27%	81.45%	58.27%	66.88%
3	50.99%	46.35%	49.01%	48.34%	50.33%	78.14%	50.33%	54.30%
5	50.33%	42.38%	46.35%	44.37%	45.69%	74.17%	45.69%	52.31%
7	52.31%	48.34%	47.02%	45.03%	49.67%	69.54%	49.67%	48.34%
9	53.64%	43.70%	48.34%	45.03%	48.34%	68.87%	48.34%	48.34%

Table 13. SG-CARS-KNN (validation set) 表 13. SG-CARS-KNN (验证集)

Table 14. SG-SPA-KNN (validation set) 表 14. SG-SPA-KNN (验证集)

Κ	城市街区 距离	切比雪夫 距离	相关距离	余弦距离	欧几里得 距离	马氏距离	明克夫斯基	标准 欧几里得
1	88.74%	90.07%	99.34%	98.67%	90.07%	99.34%	90.07%	88.74%
3	80.13%	85.43%	99.34%	95.36%	83.44%	99.34%	83.44%	85.43%
5	75.49%	78.80%	99.34%	94.04%	75.50%	99.34%	75.50%	79.47%
7	70.86%	78.80%	99.34%	92.05%	74.17%	99.34%	74.17%	78.14%
9	72.18%	76.82%	99.34%	90.73%	70.86%	99.34%	70.86%	74.17%

K	城市街区 距离	切比雪夫 距离	相关距离	余弦距离	欧几里得 距离	马氏距离	明克夫斯基	标准 欧几里得
1	59.60%	69.54%	64.90%	52.98%	66.22%	98.01%	66.22%	52.98%
3	55.63%	64.24%	54.30%	50.33%	58.94%	97.35%	58.94%	54.97%
5	54.98%	59.60%	47.02%	47.68%	56.29%	96.02%	56.29%	54.30%
7	51.66%	58.94%	50.33%	45.70%	58.28%	93.38%	58.28%	47.68%
9	48.34%	58.94%	52.31%	46.35%	50.33%	94.04%	50.33%	44.37%

Table 15. SG-CARS + SPA-KNN (validation set) 表 15. SG-CARS + SPA-KNN (验证集)

Table 16. SG-CARS-SPA-KNN (validation set) 表 16. SG-CARS-SPA-KNN (验证集)

K	城市街区 距离	切比雪夫 距离	相关距离	余弦距离	欧几里得 距离	马氏距离	明克夫斯基	标准 欧几里得
1	96.02%	97.35%	99.34%	96.69%	96.69%	99.34%	96.69%	86.09%
3	92.05%	96.02%	98.01%	94.04%	96.02%	99.34%	96.02%	80.13%
5	92.71%	95.36%	98.01%	94.04%	94.04%	99.34%	94.04%	74.17%
7	88.07%	92.71%	94.04%	91.40%	92.05%	97.35%	92.05%	67.55%
9	89.40%	91.39%	93.38%	91.40%	91.40%	95.36%	91.39%	67.55%

Table 17. (SG-MN)-CARS-KNN (validation set) 表 17. (SG-MN)-CARS-KNN (验证集)

K	城市街区 距离	切比雪夫 距离	相关距离	余弦距离	欧几里得 距离	马氏距离	明克夫斯基	标准 欧几里得
1	95.36%	98.67%	99.34%	98.01%	98.01%	99.34%	98.01%	86.09%
3	92.71%	97.35%	99.34%	96.02%	96.69%	99.34%	96.69%	78.80%
5	93.38%	96.69%	98.67%	95.36%	96.02%	99.34%	96.02%	71.52%
7	89.40%	94.70%	97.35%	94.70%	94.04%	98.01%	94.04%	67.55%
9	89.40%	90.06%	96.69%	92.71%	92.71%	98.01%	92.71%	65.56%

Table 18. (SG-MN)-SPA-KNN (validation set) 表 18. (SG-MN)-SPA-KNN (验证集)

K	城市街区 距离	切比雪夫 距离	相关距离	余弦距离	欧几里得 距离	马氏距离	明克夫斯基	标准 欧几里得
1	88.74%	90.06%	99.34%	99.34%	89.40%	99.34%	89.40%	91.39%
3	83.44%	88.07%	99.34%	96.69%	84.10%	99.34%	84.10%	80.79%
5	76.15%	78.80%	99.34%	94.04%	76.82%	99.34%	76.82%	78.80%
7	73.51%	78.14%	99.34%	94.04%	75.50%	99.34%	75.50%	75.50%
9	72.85%	74.84%	99.34%	90.73%	72.85%	99.34%	72.85%	74.83%

K	城市街区 距离	切比雪夫 距离	相关距离	余弦距离	欧几里得 距离	马氏距离	明克夫斯基	标准 欧几里得
1	82.78%	85.43%	98.67%	93.38%	83.44%	99.34%	83.44%	80.13%
3	72.84%	74.17%	98.67%	88.07%	72.18%	99.34%	72.18%	72.18%
5	62.91%	68.87%	98.01%	85.43%	60.26%	99.34%	60.26%	68.21%
7	56.95%	62.91%	98.01%	82.12%	58.94%	99.34%	58.94%	68.87%
9	57.97%	58.28%	94.04%	73.51%	58.94%	96.69%	58.94%	65.56%

Table 19. (SG-MN)-CARS + SPA-KNN (validation set) 表 19. (SG-MN)-CARS + SPA-KNN (验证集)

Table 20. (SG-MN)-CARS-SPA-KN (validation set) 表 20. (SG-MN)-CARS-SPA-KN (验证集)

K	城市街区 距离	切比雪夫 距离	相关距离	余弦距离	欧几里得 距离	马氏距离	明克夫斯基	标准 欧几里得
1	86.09%	98.01%	94.70%	78.14%	94.70%	99.34%	94.70%	78.14%
3	76.15%	95.36%	92.71%	79.47%	89.40%	99.34%	89.40%	68.87%
5	72.18%	95.36%	91.39%	70.86%	87.41%	99.34%	87.41%	65.56%
7	68.21%	97.35%	94.04%	68.21%	85.43%	99.34%	85.43%	59.60%
9	70.19%	95.36%	89.40%	66.22%	82.12%	99.34%	82.12%	59.60%

Table 21. (SG-2ndDer)-CARS-KNN (validation set) 表 21. (SG-2ndDer)-CARS-KNN (验证集)

Κ	城市街区 距离	切比雪夫 距离	相关距离	余弦距离	欧几里得 距离	马氏距离	明克夫斯基	标准 欧几里得
1	97.78%	83.33%	100.00%	100.00%	92.22%	94.44%	92.22%	100.00%
3	94.44%	75.56%	94.44%	94.44%	84.44%	72.22%	84.44%	98.89%
5	93.33%	65.56%	86.67%	85.56%	76.67%	74.44%	76.67%	95.56%
7	85.56%	60.00%	72.22%	74.44%	64.44%	71.11%	64.44%	93.33%
9	80.00%	56.67%	67.78%	67.78%	61.11%	68.89%	61.11%	91.11%

Table 22. (SG-2ndDer)-SPA-KNN (validation set) 表 22. (SG-2ndDer)-SPA-KNN (验证集)

K	城市街区 距离	切比雪夫 距离	相关距离	余弦距离	欧几里得 距离	马氏距离	明克夫斯基	标准 欧几里得
1	100.00%	96.67%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
3	94.44%	87.78%	94.44%	94.44%	91.11%	93.33%	91.11%	94.44%
5	90.00%	83.33%	80.00%	81.11%	90.00%	82.22%	90.00%	88.89%
7	75.56%	72.22%	57.78%	57.78%	78.89%	75.56%	78.89%	80.00%
9	60.00%	61.11%	41.11%	43.33%	63.33%	54.44%	63.33%	70.00%

K	城市街区 距离	切比雪夫 距离	相关距离	余弦距离	欧几里得 距离	马氏距离	明克夫斯基	标准 欧几里得
1	96.67%	87.78%	94.44%	94.44%	90.00%	100.00%	90.00%	100.00%
3	87.78%	80.00%	88.89%	88.89%	81.11%	94.44%	81.11%	94.44%
5	82.22%	71.11%	80.00%	80.00%	73.33%	85.56%	73.33%	84.44%
7	77.78%	62.22%	75.56%	75.56%	70.00%	73.33%	70.00%	71.11%
9	64.44%	55.56%	53.33%	52.22%	58.89%	61.11%	58.89%	56.67%

 Table 23. (SG-2ndDer)-CARS + SPA-KNN (validation set)

 表 23. (SG-2ndDer)-CARS + SPA-KNN (验证集)

 Table 24. (SG-2ndDer)-CARS-SPA-KNN (validation set)

 表 24. (SG-2ndDer)-CARS-SPA-KNN (验证集)

K	城市街区 距离	切比雪夫 距离	相关距离	余弦距离	欧几里得 距离	马氏距离	明克夫斯基	标准 欧几里得
1	100.00%	96.67%	100.00%	100.00%	100.00%	87.78%	100.00%	100.00%
3	94.44%	87.78%	94.44%	94.44%	91.11%	58.89%	91.11%	95.56%
5	90.00%	83.33%	80.00%	81.11%	90.00%	57.78%	90.00%	93.33%
7	75.56%	72.22%	57.78%	57.78%	78.89%	50.00%	78.89%	91.11%
9	60.00%	61.11%	42.22%	43.33%	63.33%	47.78%	63.33%	91.11%

通过上表可以看出验证集数据在 KNN 分类模型中依旧能取得较好的分类结果。对比测试集与验证集 在模型中的分类准确率可以看到,模型分类准确率没有明显变化,证明基于提取特征光谱的 KNN 亚麻籽 粒分类模型具有良好的泛化性和稳定性,并且未出现过拟合。

4. 总结

本文首先探讨了建立基于高光谱成像技术的 KNN 亚麻籽品种鉴别模型的可行性;其次构建了基于高 光谱成像技术的 KNN 亚麻籽粒快速鉴别分类模型,分析了图像采集、数据预处理、特征提取和 K 近邻 算法的工作原理和特点;然后进行了特征提取方法的组合,有效解决了单个特征提取方法容易造成有效 信息丢失的问题;最后,经过对比模型的分类结果,得出最佳的分类鉴别模型,并对模型泛化性与拟合 性进行了验证,实现了不同品种亚麻籽粒的快速、准确、可靠的分类鉴别。

1) 采用 SG 滤波平滑、SG + MN、SG + 2ndDer 预处理方法对采集的光谱数据进行平滑降噪处理并分 别进行最大归一化和求二阶导数处理,确定 951.04~1680.7 nm 范围内的光谱波段信号为有效光谱。

2) 采用 CARS 和 SPA 单一特征提取法以及 CRAS + SPA 和 CARS-SPA 两种组合筛选法对分别经过 SG 滤波平滑、SG-MN、SG-2ndDer 三种不同方法预处理后的数据进行降维。

3) 基于上文所述 4 种特征提取方法建立的亚麻籽分类模型拥有良好的分类表现,其中(SG-2ndDer)-CARS-KNN、(SG-2ndDer)-SPA-KNN、(SG-2ndDer)-CARS + SPA-KNN、(SG-2ndDer)-CARS-SPA-KNN 这 4 种 分类模型对亚麻籽的分类准确率最高可以达到 100%。

结果表明,基于高光谱成像技术的 KNN 亚麻籽粒品种快速鉴别模型是有效的、稳定的、可靠的。生产上可以利用近红外高光谱成像技术在 870~1709 nm 波段下的吸收值结合 KNN 机器学习方法来设计检测机器来快速、准确地鉴别亚麻籽的品种,该研究为实现不同品种亚麻籽的快速、准确鉴别提供了新的

思路和方法。同时,为了避免单个特征提取方法容易造成有效信息丢失的问题,应根据模型复杂度和实际需求,组合不同的特征提取方法;为了提高模型整体准确率,可以采用不同的距离度量方法进行品种分类,对比分类结果筛选出最合理的距离度量方法。

基金项目

中国国家自然科学基金(编号: 32360437); 甘肃省高等教育创新基金(编号: 2021A-056); 甘肃省高等 教育机构产业支持计划(编号: 2021CYZC-57)。

参考文献

- Mazza, G. and Biliaderis, C.G. (1989) Functional Properties of Flax Seed Mucilage. *Journal of Food Science*, 54, 1302-1305. <u>https://doi.org/10.1111/j.1365-2621.1989.tb05978.x</u>
- [2] Dorrell, D.G. (1970) Distribution of Fatty Acids within the Seed of Flax. *Canadian Journal of Plant Science*, **50**, 71-75. <u>https://doi.org/10.4141/cjps70-011</u>
- [3] Goyal, A., Sharma, V., Upadhyay, N., Gill, S. and Sihag, M. (2014) Flax and Flaxseed Oil: An Ancient Medicine & Modern Functional Food. *Journal of Food Science and Technology*, 51, 1633-1653. <u>https://doi.org/10.1007/s13197-013-1247-9</u>
- [4] Kaur, P., Waghmare, R., Kumar, V., Rasane, P., Kaur, S. and Gat, Y. (2018) Recent Advances in Utilization of Flaxseed as Potential Source for Value Addition. OCL, 25, Article No. A304. <u>https://doi.org/10.1051/ocl/2018018</u>
- [5] Zhao, L., Wang, L.M., Zhao, W., Dang, Z., Zhang, J.P. and Dang, Z. H, (2016) The Isolation of Crude Fat Acid Components in a RIL Population of Oil Flax and Their Correlation Analysis. *The Crop Journal*, **32**, 33-37.
- [6] Bhatty, R.S. and Cherdkiatgumchai, P. (1990) Compositional Analysis of Laboratory-Prepared and Commercial Samples of Linseed Meal and of Hull Isolated from Flax. *Journal of the American Oil Chemists' Society*, 67, 79-84. <u>https://doi.org/10.1007/bf02540631</u>.
- [7] 王利华, 霍贵成. ω-3 不饱和脂肪酸的生物学作用[J]. 东北农业大学学报, 2001, 32(1): 100-104.
- [8] Suri, K., Singh, B., Kaur, A., Yadav, M.P. and Singh, N. (2020) Influence of Microwave Roasting on Chemical Composition, Oxidative Stability and Fatty Acid Composition of Flaxseed (*Linum usitatissimum* L.) Oil. Food Chemistry, 326, Article ID: 126974. <u>https://doi.org/10.1016/j.foodchem.2020.126974</u>
- [9] 张春利. 小麦抗穗发芽分子标记的发掘与验证[D]: [博士学位论文]. 北京: 中国农业科学院, 2008.
- [10] 周全,朱大洲,王成,等.成像光谱技术在农产品/食品检测中的研究进展[J].食品科学,2010,31(23):423-428.
- [11] Zhu, S., Zhang, J., Chao, M., Xu, X., Song, P., Zhang, J., et al. (2019) A Rapid and Highly Efficient Method for the Identification of Soybean Seed Varieties: Hyperspectral Images Combined with Transfer Learning. *Molecules*, 25, Article 152. <u>https://doi.org/10.3390/molecules25010152</u>
- [12] Moola, W.S., Bijker, W., Belgiu, M. and Li, M. (2021) Vegetable Mapping Using Fuzzy Classification of Dynamic Time Warping Distances from Time Series of Sentinel-1A Images. *International Journal of Applied Earth Observation and Geoinformation*, **102**, Article ID: 102405. <u>https://doi.org/10.1016/j.jag.2021.102405</u>
- [13] 曹宇佳, 孙嘉迎, 孙微, 等. 基于珠海一号高光谱数据的农作物精细化分类研究[J]. 测绘与空间地理信息, 2023, 46(z1): 101-103, 106.
- [14] 袁建清, 仇逊超, 贾银江, 等. 高光谱成像的水稻冠层穗颈瘟早期识别[J]. 西南大学学报(自然科学版), 2023, 45(6): 57-65.
- [15] Zhang, J., Feng, X., Liu, X. and He, Y. (2018) Identification of Hybrid Okra Seeds Based on Near-Infrared Hyperspectral Imaging Technology. *Applied Sciences*, 8, Article 1793. <u>https://doi.org/10.3390/app8101793</u>
- [16] 孙婷. 高粱品种的高光谱分类及多品种混合识别研究[D]: [硕士学位论文]. 自贡: 四川轻化工大学, 2021.
- [17] Chang, C. and Lin, C. (2011) LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2, 1-27. <u>https://doi.org/10.1145/1961189.1961199</u>
- [18] 马惠玲, 王若琳, 蔡骋, 等. 基于高光谱成像的苹果品种快速鉴别[J]. 农业机械学报, 2017, 48(4): 305-312.
- [19] 叶佳丽, 江海霞, 郭栋良, 等. 亚麻籽蛋白质、亚麻酸和木酚素含量近红外预测模型建立与应用[J]. 中国油料作物学报, 2021, 43(2): 353-360.
- [20] 胡建. 基于 FT-IR 的鳜鱼鲜味强度检测方法研究[D]: [硕士学位论文]. 武汉: 华中农业大学, 2023.

- [21] Li, H., Liang, Y., Xu, Q. and Cao, D. (2009) Key Wavelengths Screening Using Competitive Adaptive Reweighted Sampling Method for Multivariate Calibration. *Analytica Chimica Acta*, 648, 77-84. <u>https://doi.org/10.1016/j.aca.2009.06.046</u>.
- [22] Araújo, M.C.U., Saldanha, T.C.B., Galvão, R.K.H., Yoneyama, T., Chame, H.C. and Visani, V. (2001) The Successive Projections Algorithm for Variable Selection in Spectroscopic Multicomponent Analysis. *Chemometrics and Intelli*gent Laboratory Systems, 57, 65-73. <u>https://doi.org/10.1016/s0169-7439(01)00119-8</u>