

基于近红外光谱及邻域粗糙集算法的 稻谷贮藏品质无损鉴别

Non-destructive identification of the storage quality of paddy using near
infrared spectroscopy with neighborhood rough set algorithm

杨 东 常 青 苑江浩 曹 阳 赵会义

YANG Dong CHANG Qing YUAN Jiang-hao CAO Yang ZHAO Hui-yi

(国家粮食和物资储备局科学研究院, 北京 100037)

(Academy of National Food and Strategic Reserves Administration, Beijing 100037, China)

摘要:为了能够无损、准确检测出稻谷在贮藏过程中的宜存状况,利用近红外光谱技术结合数据分析方法建立了稻谷贮藏品质的鉴别模型。采集 1 000~1 800 nm 范围内 285 份样品近红外光谱数据,依据实测脂肪酸值将样品宜存状况划分为宜存、轻度不宜存、重度不宜存三类,采用邻域粗糙集(neighborhood rough set, NRS)算法终选出最优的 10 个特征波长结合随机森林(random forest, RF)算法所建立的稻谷贮藏品质鉴别模型性能最优,其校正集与测试集正确识别率分别为 96.31% 和 93.68%,敏感性和特异性参数分布在 0.93~0.99。经分析比较,该模型性能同样优于采用连续投影算法(successive projections algorithm, SPA)和主成分分析(principal component analysis, PCA)结合 RF 算法各自建立的分类模型。结果表明,近红外光谱技术结合 NRS 和 RF 算法用于稻谷贮藏品质的鉴定是可行的,适用于储粮品质安全现场快速筛查。

关键词: 稻谷; 脂肪酸; 无损检测; 邻域粗糙集; 近红外光谱

Abstract: In order to quickly and accurately detect the status of paddy during storage, the near infrared (NIR) technique was used to establish classification models for identifying the storage quality of paddy. The NIR data of 285 samples in the range of 1 000~1 800 nm were collected. Based on the measured fatty acid, the storage state of the sample was divided into three categories (good storage quality, moderate storage quality obviously tending to decline, and poor storage quality). The optimal 10

wavelengths were selected by neighborhood rough set (NRS) algorithm. The best classification model was established by the combination of NRS and random forest (RF) algorithm. The correct classification rate (CCR) of the calibration set and the test set are 96.31% and 93.68%, respectively. The results of sensitivity and specificity are distributed in a range of 0.93 to 0.99. Furthermore, the performance of the model is also superior to the other models established by using successive projections algorithm (SPA) and principal component analysis (PCA) algorithms combined with RF. The result indicated that the fusion of NIR technique and the NRS and RF algorithms is feasible for the identification of paddy storage quality, and which can provide a reference for the development of on-site rapid inspection equipment for grain quality and safety.

Keywords: paddy; fatty acid; non-destructive detection; neighborhood rough set; near infrared spectroscopy

稻谷是中国三大储备原粮之一,其贮藏品质及食用品质备受国民关注^[1-2]。稻谷在贮藏过程中易受仓内温湿度失衡、粮虫侵蚀、霉菌侵染等影响使其品质发生劣变,导致储存稻谷出库时数量及质量损失严重^[3-4]。因此,了解稻谷在贮藏过程中的品质变化趋势,明确稻谷在贮藏过程中的宜存状况,建立适宜的检测方法实现稻谷贮藏品质的有效鉴别,是当前粮食贮藏领域急需攻克的技术难题^[5]。

脂肪酸值是衡量稻谷贮藏品质变化趋势的重要指标之一^[6],又可通过其实测值区间范围确定当前稻谷样品的宜存状况。然而,稻谷中脂肪酸含量的常规检测方法通常以理化试验为主,存在检测周期长、仪器成本较高且损伤样品等问题,不适合现场快速、无损、批量检测的需求^[7]。

基金项目: 国家重大科学仪器设备开发专项(编号: 2014YQ47037704)

作者简介: 杨东(1987—),男,国家粮食和物资储备局科学研究院助理研究员,博士。E-mail: yd521703@163.com

收稿日期: 2019-06-25

近红外光谱技术因其快速、无损、操作简单、样品无需预处理等特性,已被广泛应用于水果、蔬菜、肉类等农产品品质检测方面^[8-9],也有其应用于稻米品质检测的报道^[10]。而且国家已出台了有关稻谷水分、粗蛋白质等指标近红外分析模型的建立标准^[11]。蒋晓杰等^[12-13]以主成分分析及偏最小二乘回归为核心算法,开展了稻谷中脂肪酸值近红外光谱建模研究工作,并取得一定成效。文韬等^[14]研究了稻谷的霉变状况与脂肪酸值的变化规律,建立了二者的关系模型。

特征波段优选是光谱分析建模的重要环节之一,采用适宜的算法优选出能够表征样品本质特性的波长变量,可以提升建模效率和预测精度。邻域粗糙集(neighborhood rough set, NRS)是针对目标不确定性进行数据约减或分类的算法,已在光谱波段选择中得到了应用,Liu 等^[15]在基于高光谱技术检测大豆品种研究中利用了粗糙集算法进行波段选择,从而实现大豆品种的有效分类;朱启兵等^[16]将邻域粗糙集算法与高光谱散射图像相结合,成功地应用于苹果粉质化程度检测。但多数研究所建模型的复杂度可进一步简化,准确性与稳定性还有待提升。此外,将稻谷的贮藏品质作为分类标准,建立近红外光谱鉴别稻谷宜存状态的定性分析模型的研究还未见报道。

试验拟以贮藏稻谷为分析对象,以脂肪酸值为指标将稻谷样品存储状态划分为三类,引入邻域粗糙集算法进行光谱特征优选,与随机森林算法结合建立稻谷宜存状态的鉴别模型,并对模型进行验证和比较,为稻谷品质安全现场快检技术发展提供方法借鉴。

1 材料与方法

1.1 样本制备

稻谷样品主要来源于黑龙江、吉林、江苏、安徽等中国主要粳稻主产区某些地方粮食储备库,选取当地具有代表性,种植品种较多的库存稻谷样品进行样杆,所有样品入库水分含量均在 13.4%~14.3%,满足国家收购标准,300 余份样品收集后标记密封保存,运到实验室进行后期参数测定。经整理后有效样品 285 份,首先进行光谱数据采集,随后利用 GB/T 15684—2015《谷物碾磨制品 脂肪酸值的测定》方法测定出样品中的脂肪酸值含量。随机选择 190 份样品作为校正集用于模型的构建,剩余样品作为测试集用于模型的验证。

1.2 仪器设备

近红外光谱分析仪:Sup NIR-3000 型,聚光科技(杭州)有限公司;

全温振荡器:HZQ-Q 型,广州永程科学设备有限公司;

电子天平:ML 204 型,梅特勒—托利多仪器有限

公司;

组织捣碎机:8010 BU 型,青岛圣吉仪器系统有限公司;

碾米机:JNM-II 型,吉林省鼎立机械设备有限公司;

试验磨谷机:THU35B 型,北京东孚久恒仪器技术有限公司。

1.3 近红外光谱采集

采集光谱范围 1 000~1 800 nm,波长准确性 ± 0.2 nm,波长重复性 < 0.05 nm,分辨率 (10.9 ± 0.3) nm,每间隔 1 nm 取一次光谱数据,每个样品重复 3 次装样,扫描 32 次取平均得到最终样本光谱数据。

1.4 脂肪酸值测定

按 GB/T 15684—2015《谷物碾磨制品 脂肪酸值的测定》执行,每个样本作 3 次平行试验,取平均值作为最终结果。根据 GB/T 20569—2006《稻谷储存品质判定规则》中相关规定,粳稻谷脂肪酸值(KOH/干基)可将稻谷宜存状况划分为宜存(≤ 25 mg/100 g)、轻度不宜存(≤ 35 mg/100 g)、重度不宜存(> 35 mg/100 g)三类。试验将这三类状态作为模型的输出结果,开展稻谷贮藏品质鉴定方法的研究。

1.5 数据分析方法

1.5.1 随机森林 随机森林(RF)是一种非参数、非线性的分类和回归算法,利用其集成学习方法的优势,可以快速处理高维数据,能够有效地防止过拟合现象^[17-18]。试验中决策树数量的范围设置为 1~1 000,决策树具体数量与对应分裂变量个数利用 10 折交叉验证方法优选。

1.5.2 邻域粗糙集概念 经典粗糙集(Rough Set, RS)理论将知识理解为对数据的划分,将分类理解为在特定空间上的等价关系,可对不确定或不精确的知识利用特征属性进行描述或约减^[19]。但利用 RS 理论进行连续数据处理时需首先对其进行离散化,会导致原始数据特征属性出现损失,因此引入邻域的概念到 RS 理论中,形成邻域粗糙集模型用于解决集中数值型特征变量离散化的过程^[16,20],具体描述如下:

待处理非空实数样本集合为 $U = \{x_1, x_2, \dots, x_n\}$, C 为条件属性, D 为决策属性,假定 M 为描述 C 的邻域关系,则称 $NDT = (U, M, D)$ 为邻域决策系统。其中,对任意的 $x_i \in U$, x_i 的邻域可定义为 $\delta(x_i) = \{x_j | x_i \in U, \Delta(x_i, x_j) \leq \delta\}$, $\delta \geq 0$ 。式中 $\Delta(x_i, x_j)$ 代表距离度量函数,通常利用 P 范数来表达, $\Delta_p(x_1, x_2) = \left[\sum_{i=1}^N |f(x_1, a_i) - f(x_2, a_i)|^p \right]^{\frac{1}{p}}$ 。另有,针对 $NDT = \langle U, M, D \rangle$ 为邻域决策系统,集合 U 被决策属性 D 划分为 N 个等价类 $(X_1, X_2, \dots, X_N)$, $\forall Q \subseteq M, D$ 关于 Q 邻域的下近似 $(\underline{N_Q D})$ 和上近似 $(\overline{N_Q D})$ 分别表示为:

$$\underline{N_Q D} = \bigcup_{i=1}^N \underline{N_Q X_i}, \quad (1)$$

$$\overline{N_Q D} = \bigcup_{i=1}^N \overline{N_Q X_i}, \quad (2)$$

$$\overline{N_Q X_i} = \{x_i | \delta_Q(x_i) \subseteq X, x_i \in U\}, \quad (3)$$

$$\overline{N_Q X_i} = \{x_i | \delta_Q(x_i) \cap X \neq \Phi, x_i \in U\}, \quad (4)$$

式中:

Q ——邻域属性,归属于整体邻域系统;

D ——整体邻域系统的决策属性;

$N_Q D$ ——决策属性 D 关于 Q 邻域的下近似或上近似;

X_i ——等价类,集合 U 被决策属性 D 划分为 N 个等价类;

x_i ——样本集数据;

$\delta_Q(x_i)$ ——由属性 Q 和度量 $\Delta(x_i)$ 生成的邻域信息粒子。

可变精度(β)是邻域粗糙集模型中重要参数之一,通常 $0.5 < \beta \leq 1.0$ 。 β 与邻域范围(δ)的参数设定直接影响到邻域粗糙集模型的性能及对数据的分析结果,因此 β 与 δ 参数匹配与选择是模型建立的重要环节^[21]。

1.6 模型评价

模型的评价标准主要是通过正确判别率(correct classification rate, CCR)检验,即正确判别的样本个数(N_c)占总样本个数(N_t)的百分比。此外,敏感性和特异性也用于分类模型的评判^[22],通常敏感性和特异性的值越接近于 1 表明模型的性能越好。公式为:

$$Se = \frac{TP}{FN + TP}, \quad (5)$$

$$Sp = \frac{TN}{FP + TN}, \quad (6)$$

式中:

Se ——敏感性值;

Sp ——特异性值;

TP ——真正类,即真样本被判别为真样本的数量;

FN ——假正类,即真样本被判别为伪样本的数量;

TN ——真负类,即伪样本被判别为伪样本的数量;

FP ——假负类,即伪样本被判别为真样本的数量;

2 结果与讨论

2.1 脂肪酸值实测数据分析

表 1 为稻谷样本脂肪酸值实测数据基础参数统计结果。其中校正集样本($n=190$)均值和方差分别为 24.15, 10.67 mg/100 g, 测试集样本($n=95$)分别为 22.47, 9.27 mg/100 g, 两个数据集具有可比较的参数统计量,此外,测试集样本脂肪酸含量数据范围(12.98~60.21)恰好被校正集样本数据范围(12.11~67.90)所覆盖,符合基本建模需求,进一步表明样本数据集的划分具有一定的合理性。图 1 为校正集样本脂肪酸值实测数据散点图。由图 1 可知,稻谷样本多处于宜存或轻度不宜存状态,只有少量样本脂肪酸值 > 35 mg/100 g, 处于重度不宜存状态。三类状态样本点数据离散的分布在各自阈值区间内,可作为分类模型的参考依据,但在阈值(25, 35 mg/100 g)边缘处存在易混状态样本,可能会影响模型的性能。

2.2 光谱特性分析

图 2 为采集到稻谷样品经过平滑处理后的光谱数据分布图。由图 2 可知,在测试范围内(1 000~1 800 nm),所有稻谷样品光谱曲线分布趋势基本一致,不同脂肪酸

表 1 校正集与测试集稻谷样品脂肪酸值统计结果

Table 1 Measured values of fatty acid in paddy based on the calibration and test sets

集合	均值/ (10^{-2} mg · g $^{-1}$)	方差/ (10^{-2} mg · g $^{-1}$)	最大值/ (10^{-2} mg · g $^{-1}$)	最小值/ (10^{-2} mg · g $^{-1}$)	范围
校正集	24.15	10.67	67.90	12.11	12.11~67.90
测试集	22.47	9.27	60.21	12.98	12.98~60.21

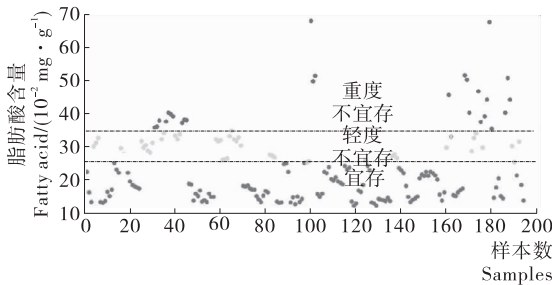


图 1 校正集稻谷样品脂肪酸含量散点分布图

Figure 1 Scatter distribution of fatty acid values of paddy samples in calibration set

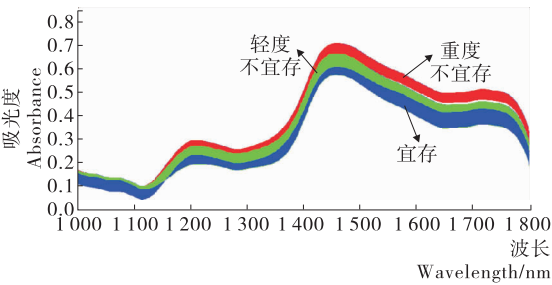


图 2 不同宜存状态稻谷样品的光谱趋势图

Figure 2 Spectra of paddy samples for different storage status

含量样本在吸光度方向上呈现梯度式变化趋势。脂肪酸含量的变化致使其光谱吸收特性存在差异,可能与稻谷在贮藏过程品质变化趋势相关。测试范围内含有多个不同的吸收波段(1 116,1 204,1 450 nm),可能与稻谷样品中包含 OH、CH、NH 等含氢基团的内部组分及其含量的不同相关^[23]。这些特征差异可作为定性分析模型的建立基础。

2.3 特征波长优选

利用 NRS 算法约减光谱变量,参阅文献[21]可知,通常邻域尺寸取值以 0.05 为步长,范围标准化到[0,1]区间, β 取值以 0.05 为步长在[0.5,1]区间变化。结合文献[16]的参数取值结果,试验中 β 设定为 0.75 和 0.85, δ 设定为 0.05,0.10,0.15,0.20, β 与 δ 分别进行组合完成邻域粗糙集优选算法参数的设置(分别设定为组合 1 和组合 2),经过多次校正实现特征波长的优选,结果如表 2 所示。

组合 1 针对不同邻域尺寸分别优选出 16,13,12,10 个波长变量,分布状况如图 3(a)所示,优选出的特征波段基本都分布在主要吸收峰附近。进一步分析可知,

不同邻域尺寸优选出的特征波长存在一定的共性,图 3(a)中标识出的波长变量均被共同优选出,因此针对组合 1 的优选结果进一步凝练出 10 个具有共性的特征波长变量(1 057,1 116,1 204,1 282,1 298,1 354,1 450,1 592,1 651,1 762 nm)用于分类模型的建立。

组合 2 针对不同邻域尺寸分别优选出 12,10,8,6 个波长变量,分布状况如图 3(b)所示,组合 2 共同优选出的特征波长数量减少,与组合 1 优选结果的吸收波段位置存在差异,但同样分布在主要吸收峰附件。针对组合 2 的优选结果凝练出 8 个特征波长变量(1 086,1 148,1 234,1 276,1 400,1 576,1 682,1 754 nm)用于分类模型的建立。优选出的特征波长变量能否表征稻谷的宜存状态还需进一步通过建立的分类模型来验证。

2.4 分类模型建立

基于 NRS 算法,利用组合 1 和组合 2 各自优选出的特征波长变量与 RF 相结合,分别建立 NRS-RF-1 和 NRS-RF-2 分类模型用于稻谷贮藏品质的鉴定,分类结果如表 3 所示。模型 NRS-RF-1 和 NRS-RF-2 在校正集的分类准确率相互接近(CCR均>94%),均达到较理想分

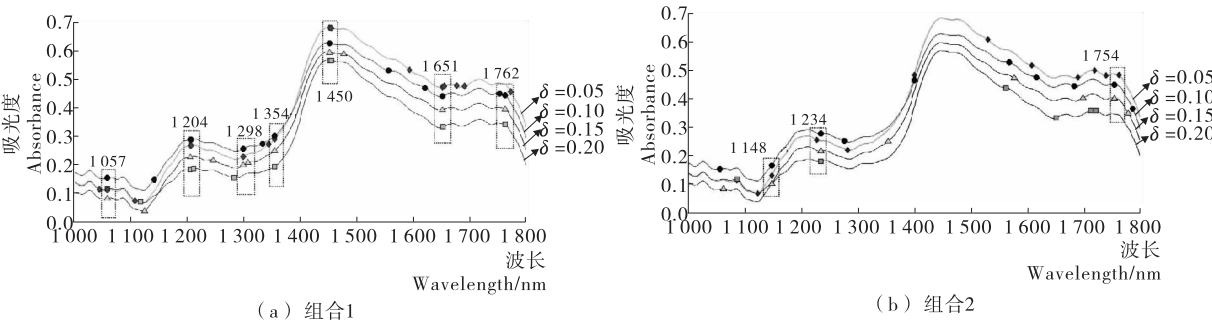


图 3 采用 NRS 算法优选特征波长结果

Figure 3 Selection of optimal wavelengths by NRS-combination 1(a) and combination 2(b)

表 2 邻域粗糙集算法优选特征波长统计结果[†]

Table 2 Selection results of optimal wavelengths by NRS

可变精度(β)	邻域尺寸(δ)	波长数量	优选波长/nm
0.75	0.05	16	1 043, 1 057 , 1 106, 1 204 , 1 206 , 1 298 , 1 343, 1 354 , 1 450, 1 454 , 1 592, 1 651 , 1 654, 1 676, 1 690, 1 771
	0.10	13	1 057 , 1 140, 1 204 , 1 206 , 1 298 , 1 332, 1 354 , 1 450 , 1 555, 1 620, 1 650, 1 752, 1 762
	0.15	12	1 057 , 1 123, 1 204 , 1 245, 1 298 , 1 306, 1 354 , 1 450 , 1 458, 1 476, 1 651 , 1 762,
	0.20	10	1 057 , 1 116, 1 204 , 1 210, 1 282, 1 354, 1 450 , 1 454 , 1 651 , 1 762,
0.85	0.05	12	1 086 , 1 122, 1 148 , 1 227, 1 282, 1 400 , 1 530, 1 608, 1 689 , 1 719, 1 739, 1 762
	0.10	10	1 055, 1 148 , 1 234 , 1 276, 1 400 , 1 567, 1 616, 1 682, 1 754 , 1 786
	0.15	8	1 062, 1 148 , 1 234 , 1 354, 1 576, 1 698 , 1 754 , 1 778
	0.20	6	1 086 , 1 234 , 1 562, 1 650, 1 712, 1 721

[†] 加粗数据表示不同 β 和 δ 组合共同优选出的特征波长。

类结果,其中,对重度不宜存状态的稻谷样品的判别均只误判了1个。进一步分析可知,由于轻度不宜存属于过度存储状态,致使处在宜存和重度不宜存状态的稻谷样品易被误分为该状态(脂肪酸含量处于阈值边缘的样本)。针对测试集样本,两个模型的分类能力均呈现出下降趋势,NRS-RF-1的CCR为93.68%,错误判别了6个样本数据,而NRS-RF-2模型的分类能力下降明显,CCR为86.31%,13个样本被错误划分为其他类。综合考虑可知,NRS-RF-1模型的分类能力与稳定性要优于NRS-RF-2模型,组合1优选出的10个特征波长变量更能表征稻谷贮藏品质的本质特性。

2.5 与其他模型比较分析

为了验证NRS-RF-1模型的分类能力,进一步采用常规的主成分分析(PCA)和连续投影算法(SPA)分别对原始光谱数据进行降维处理,沿用RF算法分别建立PCA-RF和SPA-RF分类模型用于稻谷贮藏品质的鉴别,结果如表4所示。PCA-RF模型校正集与测试集分类准确率均<90%,分别为88.94%,86.31%,其中校正集与测

试集分别有21,13个稻谷样本的宜存状态被误判到其他类属。模型分类能力明显不及NRS-RF-1模型。SPA-RF模型的分类能力略优于PCA-RF模型,校正集CCR达到94.73%,仅10个样本状态判别错误,测试集CCR下降到89.47%,表明SPA-RF模型的分类稳定性还需进一步提升。

选取NRS-RF-1和SPA-RF模型通过敏感性和特异性指标的计算结果进一步比较模型性能。由表5可知,NRS-RF-1和SPA-RF模型校正集敏感性和特异性指标分布在0.96~0.99和0.92~0.99,没有明显差异,分类结果较为理想。测试集中,NRS-RF-1模型评价指标分布在0.93~0.98,具有较理想的模型稳定性,而SPA-RF模型的分类准确性和稳定性均出现下降趋势(0.88~0.98)。由此可知,SPA-RF模型针对稻谷样品贮藏品质的整体分类性能同样不及NRS-RF-1模型的鉴别结果,利用NRS算法优选的10个光谱特征波长结合RF算法建立的NRS-RF-1模型用于稻谷样本贮藏品质的鉴定是可行的。

表3 基于NRS-RF模型的稻谷宜存状态分类结果[†]

Table 3 Classification results of paddy storage status by NRS-RF models

模型	类别	测试集				校正集			
		宜存 (n=87)	轻度 不宜存 (n=75)	重度 不宜存 (n=28)	准确率/%	宜存 (n=87)	轻度 不宜存 (n=75)	重度 不宜存 (n=28)	准确率/%
NRS-RF-1	宜存	84	3	0		45	3	0	
	轻度不宜存	2	72	1	96.31	1	27	1	93.68
	重度不宜存	0	1	27		0	1	17	
NRS-RF-2	宜存	82	5	0		41	7	0	
	轻度不宜存	3	70	2	94.21	3	25	1	86.31
	重度不宜存	0	1	27		0	2	16	

† 加粗数据表示分类正确的样本数量及分类准确率。

表4 基于PCA-RF和SPA-RF模型的稻谷贮藏品质分类结果[†]

Table 4 Classification results of paddy storage status by PCA-RF and SPA-RF models

模型	类别	测试集				校正集			
		宜存 (n=87)	轻度 不宜存 (n=75)	重度 不宜存 (n=28)	准确率/%	宜存 (n=87)	轻度 不宜存 (n=75)	重度 不宜存 (n=28)	准确率/%
PCA-RF	宜存	79	8	0		41	7	0	
	轻度不宜存	5	67	3	88.94	3	24	2	86.31
	重度不宜存	0	5	23		0	1	17	
SPA-RF	宜存	83	4	0		43	5	0	
	轻度不宜存	1	71	3	94.73	2	26	1	89.47
	重度不宜存	0	2	26		0	2	16	

† 加粗数据表示分类正确的样本数量及分类准确率。

表 5 基于 NRS-RF-1 和 SPA-RF 模型的敏感性和特异性指标统计结果[†]

Table 5 The results of sensitivity and specificity based on NRS-RF-1 and SPA-RF models

模型	类别	校正集		测试集	
		敏感性	特异性	敏感性	特异性
NRS-RF-1	宜存	0.96	0.98	0.93	0.97
	轻度不宜存	0.96	0.96	0.93	0.93
	重度不宜存	0.96	0.99	0.94	0.98
SPA-RF	宜存	0.95	0.99	0.89	0.95
	轻度不宜存	0.94	0.94	0.89	0.89
	重度不宜存	0.92	0.98	0.88	0.98

[†] 加粗数据表示敏感性和特异性指标的分布范围。

3 结论

采集 1 000~1 800 nm 的稻谷样品近红外光谱数据,利用参数调整后的邻域粗糙集(NRS)算法优选出 10 个能表征稻谷样品脂肪酸值变化趋势及其宜存状态的特征波长变量,并结合随机森林(RF)算法构建了分类模型 NRS-RF-1,用于稻谷贮藏品质的鉴定。该模型校正集与测试集 CCR 均>93%,敏感性和特异性分布在 0.93~0.99,具有可观的分类准确性与稳定性。经分析比较,NRS-RF-1 模型性能同样优于试验中所建立的 SPA-RF 和 PCA-RF 两类模型。结果表明,NRS 算法优选出的特征变量能够代表稻谷样品的本质特性,构建的 NRS-RF-1 模型用于稻谷贮藏品质的鉴定是可行的,后期可进一步融合色泽、品尝评分值、水分等多贮藏品质指标细化稻谷宜存状态等级,改进现有模型算法,建立更加适宜的分类模型,为储粮品质安全现场快检设备研制及多光谱在线检测系统研发提供技术支持。

参考文献

[1] ORINA I, MANLEY M, WILLIAMS P J. Non-destructive techniques for the detection of fungal infection in cereal grains[J]. Food Research International, 2017, 100(1): 74-86.

[2] 曹阳, 魏雷, 赵会义, 等. 我国绿色储粮技术现状与展望[J]. 粮油食品科技, 2015, 23(S1): 11-14.

[3] 唐芳, 欧阳毅, 祁智慧. 基于扫描电镜观察研究真菌孢子检测对稻谷霉变判定[J]. 中国粮油学报, 2018, 33(4): 122-126.

[4] BIANCOLILLO A, FIRMANI P, BUCCI R, et al. Determination of insect infestation on stored rice by near infrared (NIR) spectroscopy[J]. Microchemical Journal, 2019, 145: 252-258.

[5] 吴子丹, 赵会义, 曹阳, 等. 粮食储藏生态系统的仿真技术应用研究进展[J]. 粮油食品科技, 2014, 22(1): 1-6.

[6] 范维燕, 林家永, 邢邯, 等. 稻谷脂肪酸值近红外光谱快速

测定技术研究[J]. 食品科学, 2009, 30(24): 347-350.

[7] 俞法明, 陆艳婷, 严文潮, 等. 用近红外光谱技术快速测定籼稻品种的蛋白质含量[J]. 中国粮油学报, 2009, 24(5): 134-138.

[8] 田卫新, 何丹丹, 杨东, 等. 一种基于高光谱图像的熟牛肉 TVB-N 含量预测方法[J]. 食品与机械, 2016, 32(12): 70-74.

[9] YANG Dong, LU An-xiang, WANG Ji-hua, et al. Detection of the freshness state of cooked beef during storage using hyperspectral imaging[J]. Applied Spectroscopy, 2017, 71(10): 2 286-2 301.

[10] 徐彦, 李忠海, 付湘晋, 等. 近红外光谱技术在稻米品质快速检测中的应用[J]. 食品与机械, 2011, 27(1): 158-161.

[11] 国家粮食局. GB/T 24895—2010 粮油检验 近红外分析定标模型验证和网络管理与维护通用规则[S]. 北京: 中国标准出版社, 2010: 1-9.

[12] 蒋晓杰, 周旭, 杨志成, 等. 漫反射红外光谱法结合 PLS 测定稻谷脂肪酸值研究[J]. 中国粮油学报, 2019, 34(3): 117-120.

[13] JAMES V A, WITTENBERG A, LI Hui, et al. High throughput phenotyping of Camelina sativa seeds for crude protein, total oil, and fatty acids profile by near infrared spectroscopy[J]. Industrial Crops and Products, 2019, 137: 501-507.

[14] 文韬, 洪添胜, 李立君, 等. 霉变稻谷脂肪酸含量的光谱检测模型构建与优化分析[J]. 农业工程学报, 2016(1): 193-199.

[15] LIU Yao, XIE Hong, CHEN Yue-hua, et al. Neighborhood mutual information and its application on hyperspectral band selection for classification[J]. Chemometrics and Intelligent Laboratory Systems, 2016, 157: 140-151.

[16] 朱启兵, 黄敏, 赵桂林. 基于邻域粗糙集和高光谱散射图像的苹果粉质化检测[J]. 农业机械学报, 2011, 42(10): 154-157.

[17] BREIMAN L. Bagging predictors[J]. Machine Learning, 1996, 24(2): 123-140.

[18] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.

[19] DONG Ying, XIANG Bing-ren, GENG Ying, et al. Rough set based wavelength selection in near-infrared spectral analysis[J]. Chemometrics and Intelligent Laboratory Systems, 2013, 126: 21-29.

[20] 刘瑶, 李梓楠, 吴涛, 等. 基于高光谱图像和邻域粗糙集理论的大豆品种识别算法及其综合性能评估[J]. 大豆科学, 2018, 37(4): 110-119.

[21] 胡清华, 于达仁, 谢宗霞. 基于邻域粒化和粗糙逼近的数值属性约简[J]. 软件学报, 2008, 19(3): 640-649.

[22] PARIKH R, MATHAI A, PARIKH S, et al. Understanding and using sensitivity, specificity and predictive values[J]. Indian Journal of Ophthalmology, 2008, 56(1): 45-50.

[23] 沈飞, 张斌, 刘潇, 等. 基于光谱和气味特征信息的霉变大米无损检测[J]. 中国粮油学报, 2018, 33(4): 127-132.