

# 傅里叶中红外光谱结合稀疏表示分类方法鉴别小麦赤霉病感染等级

梁 琨<sup>1,2</sup>, 张夏夏<sup>1,2</sup>, 丁 静<sup>1,2</sup>, 徐剑宏<sup>3</sup>, 韩东燊<sup>1,2</sup>, 沈明霞<sup>1,2\*</sup>

1. 南京农业大学工学院, 江苏 南京 210031
2. 江苏省现代设施农业技术与装备工程实验室, 江苏 南京 210031
3. 江苏省食品质量安全重点实验室一省部共建国家重点实验室培育基地/  
江苏省农业科学院农产品质量安全与营养研究所, 江苏 南京 210014

**摘要** 旨在探索感染不同等级赤霉病的小麦中主要成分含量变化引起的傅里叶中红外光谱信息响应,并结合模式识别方法实现基于傅里叶变换中红外光谱的小麦赤霉病等级无损检测。以感染不同等级赤霉病小麦为研究对象,在 $4000\sim 400\text{ cm}^{-1}$ 波数范围内采集95个小麦样本的傅里叶中红外光谱数据,利用载荷系数法(XLW)与随机森林算法(RF)分析选取小麦样本傅里叶中红外光谱中的敏感波长,利用稀疏表示分类(SRC)算法建模识别小麦感染赤霉病等级。结果表明:XLW算法和RF算法选择的特征波长作为定性分析模型的输入时模型鉴别准确率与全波段光谱数据作输入时均达90%以上,特征波长提取算法可以有效简化模型并提高效率。RF-SRC模型鉴别效果最好,建模集鉴别准确率达97%,测试集鉴别准确率达96%。小麦感染赤霉病等级的不同会引起小麦中水分、淀粉、纤维素、可溶性氮素、蛋白质、脂肪等物质含量的变化,采用RF算法选择的特征波长均反映了这些物质所对应的傅里叶中红外光谱透射光谱特征的差异,结合SRC模型进行小麦赤霉病等级鉴别可达到最好的鉴别效果。因此,利用傅里叶中红外光谱技术结合模式识别方法对小麦赤霉病等级鉴别是可行的,解释了傅里叶中红外光谱技术检测小麦赤霉病等级的机理。

**关键词** 傅里叶中红外光谱; 小麦; 赤霉病; 稀疏表示分类

**中图分类号:** TP391.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)10-3251-05

## 引 言

赤霉病是世界性的麦类病害之一,多发生在穗期多雨、气候潮湿的地区,具有周期性流行的特点。小麦受赤霉病菌感染后,麦穗枯萎,籽粒腐烂,失去种用和工业价值。同时,染病籽粒会产生多种毒素如脱氧雪腐镰刀菌烯醇(deoxynivalenol, DON),人畜食用这些病粒后,会产生不同程度的中毒反应,引起食品安全问题<sup>[1-3]</sup>。目前用于小麦赤霉病病毒检测的方法主要有:化学生物检测法和基于图像、光谱的无损检测方法等。常用的化学生物检测法操作过程繁琐、重现性差,且所需的检测器件价格昂贵,样品处理过程复杂,操作时需要专门的技术人员,不便推广应用,也不适合大批量样品的检测<sup>[4]</sup>。近年来,光谱技术在农产品无损检测中得到迅速发展。高光谱成像技术因其图谱合一、多波段、高分辨

率等优点<sup>[5]</sup>,已被广泛应用于农业、食品等领域。国内外很多学者<sup>[6-7]</sup>将高光谱技术用于小麦赤霉病研究,研究发现高光谱技术可以应用于小麦赤霉病检测,提高了检测效率,为小麦赤霉病快速、无损、智能检测提供了研究方法。与高光谱成像技术相比,傅里叶变换红外光谱由被测对象中某些官能团和极性键振动引起<sup>[8]</sup>,可以有效反映被测对象的组成成分。Shi<sup>[9]</sup>等利用小麦样本 $1700\sim 1100\text{ cm}^{-1}$ 范围内的傅里叶变换红外光谱数据建立PLS模型对小麦中粗蛋白含量进行测定,相关系数达0.9,具有较好的预测能力。Suchowilka<sup>[10]</sup>等对小麦样本的傅里叶变换中红外光谱数据进行主成分分析,实现了对不同小麦品种的鉴别。小麦感染赤霉病等级不同其DON毒素含量也不同,当小麦病粒率达30%以上时,麦粒DON含量大于 $2\text{ mg}\cdot\text{kg}^{-1}$ <sup>[1]</sup>,含量微弱,很难利用光谱直接对小麦中DON毒素含量进行检测。而小麦种子受赤霉菌感染后,会引起穗枯和粒腐,病粒中含有毒素,且

收稿日期: 2018-08-18, 修订日期: 2018-12-27

基金项目: 国家自然科学基金青年基金项目(31401610), 中央高校基本科研业务费专项资金项目(KJQN201557), 江苏省自主创新基金项目(CX(17)1003), 南京农业大学工学院优秀青年人才科技基金项目(YQ201603)资助

作者简介: 梁 琨, 女, 1983年生, 南京农业大学工学院副教授 e-mail: lkbb2006@126.com

\* 通讯联系人 e-mail: mingxia@njau.edu.cn

随赤霉病等级的增加小麦中水分、淀粉、纤维素减少,可溶性氮素、蛋白质、脂肪含量增加<sup>[11]</sup>,因此,本文旨在探索感染不同等级赤霉病的小麦中主要成分含量变化引起的傅里叶中红外光谱信息响应,并结合稀疏表示分类方法进行小麦赤霉病等级鉴别。

## 1 实验部分

### 1.1 材料与仪器

从江苏省农业科学研究院农产品质量安全与营养研究所选取 2017 年江苏地区的小麦,品种为济麦 22。人工分拣除去小麦样本中的石子、秸秆、草籽、土块等杂质后利用 PM8188A 型无损水分测量仪(泰州市维科特仪器仪表有限公司)对小麦进行水分含量检测,对含水量较高的样本进行通风干燥,含水量较低的样本暴露在潮湿空气中,最终控制所有小麦样本水分含量在 12%~13%<sup>[12]</sup>,消除水分对后续实验的影响。每 25 g 小麦作为一个样本,共称取 95 个样本,依次标号,随机选取 72 个样本作为建模集,23 个作为测试集。

(1)磨粉:用 XA-1 型固体样品粉碎机(常州越新仪器制造有限公司)对每个样本进行磨粉至全部颗粒均可通过 20 目筛,将其混合均匀,从中取 0.002 g 小麦粉待用。

(2)压片:将 0.2 g 干燥的纯溴化钾(南京晚晴化玻仪器有限公司)在玛瑙研钵(南京晚晴化玻仪器有限公司)中充分研磨直至溴化钾粉末粘附在研钵壁上,然后加入称取好的小麦粉样本,边研磨边使样本与溴化钾充分混匀,最后将磨好的混合粉末装入 YP-2 型压片机(上海山岳科学仪器有限公司)的模具内压制成片。

(3)扫描:用 Nicolet iS10 傅里叶变换红外光谱仪(美国 Thermo Fisher 公司)对压片进行光谱数据的采集,其光谱范围为 4 000~400  $\text{cm}^{-1}$ ,最高分辨率为 4.0  $\text{cm}^{-1}$ ,扫描次数为 16 次。

(4)赤霉病等级划分与 DON 毒素含量测定:以粮食卫生标准 GB 2715—2005 中规定的小麦及全麦粉中 DON 限量标准 1  $\text{mg} \cdot \text{kg}^{-1}$  为分界线, DON 含量超过 1  $\text{mg} \cdot \text{kg}^{-1}$  的样本为感染赤霉病样本,不超过 1  $\text{mg} \cdot \text{kg}^{-1}$  的为健康样本。DON 含量的检测方法参考标准 SN/T3137—2012 中规定的液相色谱-质谱法,提取液由乙腈和水按 84:16 的比例配置,离心时转速为 2 500  $\text{r} \cdot \text{min}^{-1}$ ,活化时使用 3 mL 提取液,提取的样品过柱时的流速为 1  $\text{mL} \cdot \text{min}^{-1}$ 。利用型号为 3500QTRAP 色谱仪-液相色谱质谱联用仪(ABSCIEX 公司)进行毒素含量测定,流动相 A 为 5  $\text{mmol} \cdot \text{L}^{-1}$  的醋酸铵水,按照时间 0, 3, 7, 13, 13.1 和 16 min 时,流动相 A 和流动相 B 分别按照 85:15, 30:70, 20:80, 10:90, 85:15, 85:15 的浓度梯度进行操作,流速为 0.6  $\text{mL} \cdot \text{min}^{-1}$ ,进样量为 5  $\mu\text{L}$ 。毒素检出限为 20  $\mu\text{g} \cdot \text{kg}^{-1}$ 。

### 1.2 方法

#### 1.2.1 数据采集与校正

(1)对小麦样本扫描前需要对纯 KBr 压片进行扫描,作为背景谱图。(2)将制备好的小麦样本压片装入样品池,扫描样本谱图。(3)利用 OMNIC 8.0 软件对采集到的光谱图进

行大气背景扣除和基线校正。将每个样本各个波数处的透射率作为该样本的光谱数据,在 MATLAB 2009a 软件(美国 MathWorks 公司)中做后续处理。

#### 1.2.2 数据预处理方法

由于小麦样本压片表面会残留一些细小粉末,在傅里叶红外透射光谱测量过程中存在光程误差,影响样本的光谱信息,本文采用多元散射校正法(multiplicative scatter correction, MSC)对采集的光谱数据进行预处理。

#### 1.2.3 特征波长提取方法

实验采集的光谱在 4 000~400  $\text{cm}^{-1}$  波数范围内共有 3 734 个数据点,其包含的信息量大,但并不是所有的光谱信息都对小麦赤霉病等级鉴别有价值。且不同样本的透射率光谱曲线大致相似,只在某些波段范围或波长处存在较大差异,因此可对最能代表样品信息的特征波长进行提取,减少后期建模的输入参量,提高数据处理效率。本文分别采用载荷系数法和随机森林进行特征波长提取。

载荷系数法(X-loading weights, XLW)是基于 PLS 建模结果产生的,在每个隐含变量(latent variable, LV)下均可得到各个波长点所对应的载荷系数。载荷系数的大小说明了该波长对所建模型预测性能影响的大小<sup>[13]</sup>,因此可根据各波长所对应的载荷系数的大小来提取特征波长。

利用随机森林(random forest, RF)可计算波长变量重要性的特点,采用 Bootstrap 重采样方法生成多个训练集,后利用 CART 算法构建决策树,随机选取若干特征作为节点的候选特征,根据基尼指数最小原则度量候选特征的重要性<sup>[14]</sup>,最后根据随机森林的特征重要性评分对所有特征进行排序并选取大于设定阈值的特征。特征 F 的重要性评分  $Sc(i)$  表示如下

$$Sc(i) = \frac{1}{m} \sum_{j=1}^m \text{Gini}(f_{ij}) \quad (1)$$

其中  $m$  为训练集的个数,  $f_{ij}$  为第  $j$  个训练集中第  $i$  维特征,  $i = 1, 2, \dots, 3 734$ ,  $\text{Gini}(f_{ij})$  为特征  $f_{ij}$  的基尼指数。

#### 1.2.4 定性建模方法

稀疏表示分类方法(sparse representation based classification, SRC)是近年来出现的一种新的分类算法。分别用全波段和特征波长提取后的光谱数据构成数据词典矩阵,用该数据词典矩阵近似线性的表示每一个测试样本,使用 CVX 3.0 凸优化工具箱求得稀疏表示系数,进而重构样本,最后分别计算重构样本与测试样本之间的残差平方和,将测试样本确定为残差平方和较小的类<sup>[15]</sup>。

测试样本  $Y$  的第  $i$  类重构样本  $\hat{Y}_i$

$$\hat{Y}_i = \mathbf{C} \hat{\mathbf{V}}_i \quad (i = 1, 2) \quad (2)$$

其中  $\mathbf{C}$  表示数据词典矩阵,  $\hat{\mathbf{V}}_i$  为稀疏表示系数。

## 2 结果与讨论

### 2.1 样本中红外透射光谱分析

感染不同等级赤霉病的小麦样本在 4 000~400  $\text{cm}^{-1}$  波数范围内的光谱透射曲线如图 1 所示。观察图 1(a)可知,感染不同等级赤霉病的小麦样本的透射率光谱曲线相似,在

3 600~3 000, 3 000~2 800, 1 700~1 600, 1 570~1 550 和 1 000  $\text{cm}^{-1}$  附近有明显的光谱吸收峰, 各吸收峰所对应的振动模式如下: 3 600~3 000  $\text{cm}^{-1}$  范围内为分子内羟基(—OH)伸缩振动谱带, 此处的吸收峰强度越大, 表明样本中所含—OH数量越多, 分子内—OH主要来自淀粉、纤维素、半纤维素、多糖、单糖<sup>[16]</sup>及水分<sup>[9]</sup>。3 000~2 800  $\text{cm}^{-1}$  为淀粉中 C—H 伸缩振动引起的吸收峰。1 700~1 600  $\text{cm}^{-1}$  为蛋白质中肽基 C=O 伸缩振动引起的吸收峰与脂肪族中 C=C 和 C=N 伸缩振动引起的特征峰的重叠部分, 1 570~1 550  $\text{cm}^{-1}$  吸收峰主要由蛋白质中 N—H 弯曲振动和 C—N 伸缩振动以及脂肪族硝基化合物  $\text{NO}_2$  反对称伸缩振动引起, 其中 N—H 弯曲振动峰强度反映了小麦中蛋白质的含量<sup>[17]</sup>。1 000  $\text{cm}^{-1}$  附近为碳水化合物中多糖类的 C—O 伸缩振动吸收峰。1 000~650  $\text{cm}^{-1}$  区域内的谱带多而弱, 吸收峰较密集, 复杂多变而且不容易辨识<sup>[18]</sup>。由上述分析可知, 感染不同等级赤霉病的小麦中纤维素、多糖、水分、淀粉、蛋白质、脂肪等吸收峰强度和尖锐程度不同。采用多元散射校正法对光谱进行预处理以消除压片表面粉末引起的散射效应。图 1(b)为预处理后的光谱图。经过散射校正后的光谱数据增强了与成分含量和组织结构相关的光谱吸收信息。

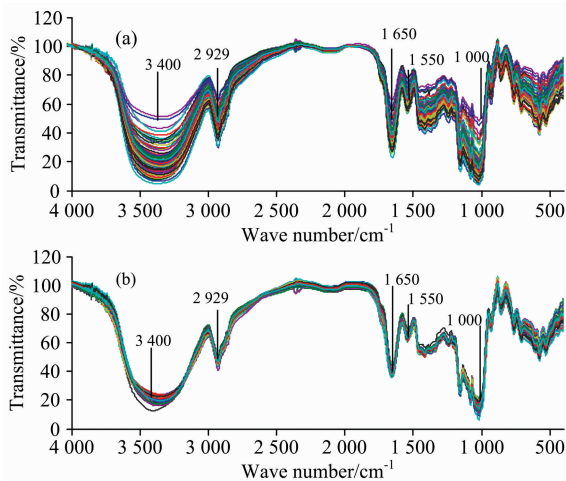


图 1 小麦样本透射率光谱曲线图

Fig. 1 Spectral curve of transmittance of wheat samples

- (a): Spectral curve of transmittance of wheat samples;  
(b): Transmission spectrum curve after pretreatment

## 2.2 特征波长提取

### 2.2.1 基于 XLW 的特征波长提取

用预处理后的光谱数据建立偏最小二乘判别模型, 当模型的最佳主成分数为 4 时交叉验证的准确率达 87.52%, 前四个主成分的载荷分布如图 2 所示。将每个主成分载荷系数最大处所对应的特征波长作为该主成分选取的特征波长。从图 2 可以看出选择的特征波长所对应的波数分别为 1 000, 2 855, 3 416 和 3 676  $\text{cm}^{-1}$ 。1 000  $\text{cm}^{-1}$  处于碳水化合物中多糖类的 C—O 伸缩振动吸收峰附近, 可反映小麦样本中多糖类如纤维素、半纤维素等含量的差异; 2 855  $\text{cm}^{-1}$  处于淀粉中 C—H 伸缩振动引起的吸收峰, 反映了小麦样本中淀粉

含量的差异; 3 416 和 3 676  $\text{cm}^{-1}$  均处于分子内羟基(—OH)伸缩振动谱带, 反映了感染不同等级赤霉病小麦样本中淀粉、纤维素、半纤维素、多糖、单糖及水分含量的差异。

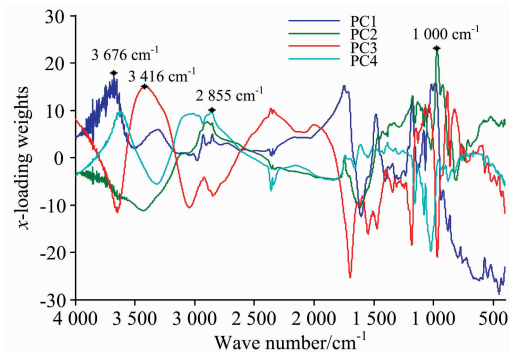


图 2 不同波数处的载荷系数

Fig. 2 The load coefficient at different wavenumber

### 2.2.2 基于 RF 的特征波长提取

将预处理后的所有样本的光谱数据作为随机森林算法输入, 求得各特征的重要性得分如图 3 所示。在阈值为 0.0028 的情况下选择 20 个特征波长。分别对应的波数为 1 029, 1 030, 1 031, 1 592, 1 594, 1 600, 1 613, 1 614, 1 615, 1 620, 1 621, 3 388, 3 397, 3 398, 3 419, 3 420, 3 425, 3 433, 3 438 和 3 472  $\text{cm}^{-1}$ 。所选的特征波长对应的波数主要集中在 1 029~1 031, 1 600~1 621 和 3 388~3 472  $\text{cm}^{-1}$  三个区域, 其中 1 029~1 031  $\text{cm}^{-1}$  处于碳水化合物中多糖类的 C—O 伸缩振动吸收峰附近, 1 600~1 621  $\text{cm}^{-1}$  处于蛋白质中肽基 C=O 伸缩振动引起的吸收峰与脂肪族中 C=C 和 C=N 伸缩振动引起的特征峰的重叠部分, 与小麦样本中脂肪酸含量有关; 3 388~3 472  $\text{cm}^{-1}$  处于分子内羟基(—OH)伸缩振动谱带, 反映了小麦样本中淀粉、纤维素、半纤维素、多糖、单糖及水分含量的差异。

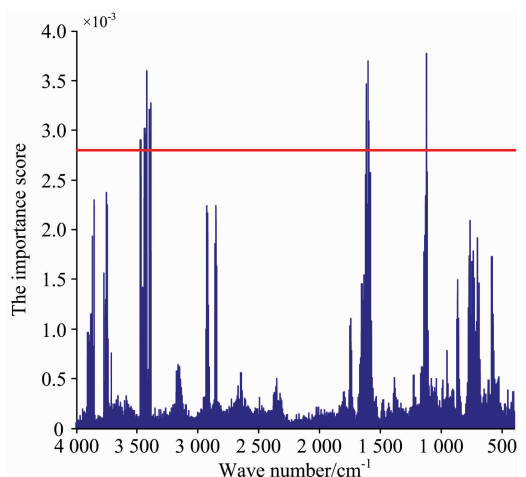


图 3 波数的重要性评分

Fig. 3 The importance score of each wavenumber

XLW 和 RF 算法选择的特征波长在小麦样本平均光谱上的位置见图 4。两种算法均选择了 1 000 和 3 400  $\text{cm}^{-1}$  附

近的波长,对应成分反映小麦样本中淀粉、纤维素、半纤维素、多糖、单糖及水分含量信息。而 XLW 算法选择的位于  $2\ 855\ \text{cm}^{-1}$  (淀粉中 C—H 伸缩振动吸收峰)和  $3\ 676\ \text{cm}^{-1}$  (纤维素、半纤维素、多糖、单糖及水分分子内—OH 伸缩振动谱带)反映的物质信息与  $1\ 000$  和  $3\ 400\ \text{cm}^{-1}$  反映的物质信息一致;RF 算法选择的  $1\ 600\ \text{cm}^{-1}$  附近的 6 个特征波数还反映了小麦样本中蛋白质和脂肪含量差异,因此,与 XLW 算法相比,RF 算法选择的特征波长多包含了蛋白质和脂肪含量信息,这为后续建立模型准确性的差异分析提供了理论依据。

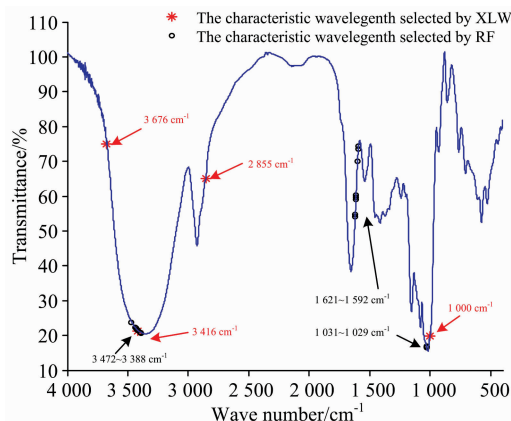


图 4 XLW 和 RF 算法选择的特征波长

Fig. 4 The characteristic wavelength selected by the XLW and RF algorithms

### 2.3 基于 SRC 的小麦赤霉病等级鉴别模型

分别将建模集中 72 个样本的全波段光谱数据及经特征提取后的光谱数据作为 SRC 模型的数据词典矩阵,利用 CVX3.0 凸优化工具箱求解稀疏表示系数,进而重构样本。根据重构样本与测试样本间的残差平方和对测试样本进识别。为说明该模型的有效性对建模集样本进行十折交叉得到

表 1 基于 SRC 的小麦赤霉病等级鉴别结果

Table 1 Identification results of wheat scab grade based on SRC

模型类别	模型输入变量数	建模集鉴别准确率/%	测试集鉴别准确率/%	模型处理时间/s
基于全波段的 SRC 鉴别模型	3 734	96	91	63.105 140
基于 XLW 的 SRC 鉴别模型	4	90	91	36.186 938
基于 RF 的 SRC 鉴别模型	20	97	96	32.242 960

交叉建模集识别准确率。具体结果见表 1。

分析表 1 中的鉴别结果可知,基于全波段和基于各特征提取算法建立的 SRC 鉴别模型均达到 90% 以上,特征提取算法建立的 SRC 鉴别模型的处理速度均约为全波段建立的 SRC 鉴别模型的 1/2。表明由 XLW 和 RF 两种特征提取算法选择的特征波长反映的有效信息与全波段光谱数据基本相同,进行特征提取可提高 SRC 模型的鉴别效率。比较两种鉴别模型,RF-SRC 模型的测试集鉴别准确率最高,其原因可能为采用 RF 算法选择的特征波长不仅反映小麦中淀粉、纤维素、半纤维素、多糖、单糖、水分的吸收信息,同时也涵盖了蛋白质、脂肪等物质的吸收信息,与小麦受赤霉病感染后主要成分含量变化情况更吻合。比较 XLW-SRC 模型,RF-SRC 模型除反映样本中淀粉、纤维素、半纤维素、多糖、单糖及水分含量信息的光谱数据外,反映蛋白质、脂肪含量信息的光谱数据也对模型鉴别准确率有显著影响,该数据的缺少会导致鉴别准确率下降。

## 3 结论

(1) 本文探索感染不同等级赤霉病的小麦中主要成分含量变化引起的傅里叶中红外光谱信息响应,并结合模式识别方法实现基于傅里叶变换中红外光谱的小麦赤霉病等级无损检测。随小麦赤霉病等级的增加,小麦中水分、淀粉、纤维素减少,可溶性氮素、蛋白质、脂肪含量增加,这些物质含量的差异会造成傅里叶中红外光谱透射光谱特征的不同。运用 XLW 算法和 RF 算法进行特征波长提取后,简化模型并提高效率。对比分析两种特征提取算法,RF 算法提取的特征波长反映了小麦样本中淀粉、纤维素、半纤维素、多糖、单糖、水分、蛋白质、脂肪含量的差异,较 XLW 算法提取的特征反映蛋白质和脂肪含量的差异有效信息。

(2) 分别将全波段光谱数据以及两种特征提取方法提取的光谱数据作为输入变量,建立 SRC 模型。其中经 RF 算法选择的特征波长作模型输入时, SRC 模型的鉴别准确率最高,建模集鉴别准确率为 97%,测试集鉴别准确率为 96%。结果表明,RF 算法选择的特征波长不仅反映小麦中淀粉、纤维素、半纤维素、多糖、单糖、水分的吸收信息,同时也涵盖了蛋白质、脂肪等物质的吸收信息,与小麦受赤霉病感染后主要成分含量变化情况更吻合。反映蛋白质和脂肪含量信息的光谱数据对 SRC 模型的鉴别准确率影响显著。本文验证了基于傅里叶中红外光谱技术结合模式识别方法进行小麦赤霉病等级鉴别的可行性,解释了傅里叶中红外光谱技术检测小麦赤霉病等级的原理。

## References

- [1] Dahl B, Wilson W W. *Agricultural Systems*, 2018, 162: 145.
- [2] Jaillais B, Roumet P, Pinson-Gadais L. *Food Control*, 2015, 54: 250.
- [3] Dweba C C, Figlan S, Shimelis H A. *Crop Protection*, 2017, 91(2017): 114.
- [4] Barbedo J G A, Tibola C S, Fernandes J M C. *Biosystems Engineering*, 2015, 131: 65.

- [ 5 ] DU Ying-ying, CHEN Xiao-he, LIANG Kun, et al(杜莹莹, 陈小河, 梁 琨, 等). Science and Technology of Food Industry (食品工业科技), 2016, 37(17): 54.
- [ 6 ] CHEN Shu-xi, XIE Chuan-qi, WANG Qiao-nan, et al(程术希, 谢传奇, 王巧男, 等). Spectroscopy and Spectral Analysis (光谱学与光谱分析), 2014, 34(5): 1362.
- [ 7 ] LIANG Kun, DU Ying-ying, LU Wei, et al(梁 琨, 杜莹莹, 卢 伟, 等). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2016, 47(2): 309.
- [ 8 ] Yu S, Sheng L, Zhang C, et al. Spectrochimica Acta Part A Molecular & Biomolecular Spectroscopy, 2018, 198: 88.
- [ 9 ] Shi H, Yu P. Food Control, 2017, 82.
- [10] Suchowilska E, Kandler W, Wiwart M, et al. International Agrophysics, 2012, 26(2): 207.
- [11] Williams P J, Geladi P, Britz T J, et al. Journal of Cereal Science, 2012, 55: 272.
- [12] Mahesh S, Jayas D S, Paliwal J et al. Sensing and Instrumentation for Food Quality and Safety, 2011, 5: 1.
- [13] Wang Y, Gao Y, Yu X, et al. Food Analytical Methods, 2016, 9(1): 131.
- [14] Mu K X, Feng Y Z, Chen W, et al. Chemometrics and Intelligent Laboratory Systems, 2018, 179: 46.
- [15] Zhang S, Wu X, You Z, et al. Computers and Electronics in Agriculture, 2017, 134: 135.
- [16] HE Chun-xia, FU Lei-ming, XIONG Jing, et al(何春霞, 傅雷鸣, 熊 静, 等). Journal of Nanjing Agricultural University(南京农业大学学报), 2016, 39(2): 325.
- [17] Amir R M, Anjum F M, Khan M I, et al. Journal of Food Science & Technology, 2013, 50(5): 1018.
- [18] Singh V K, Devi A, Pathania S, et al. Biocatalysis and Agricultural Biotechnology, 2017 (9): 58.

## Discrimination of Wheat Scab Infection Level by Fourier Mid-Infrared Technology Combined with Sparse Representation Based Classification Method

LIANG Kun<sup>1, 2</sup>, ZHANG Xia-xia<sup>1, 2</sup>, DING Jing<sup>1, 2</sup>, XU Jian-hong<sup>3</sup>, HAN Dong-shen<sup>1, 2</sup>, SHEN Ming-xia<sup>1, 2\*</sup>

1. College of Engineering, Nanjing Agricultural University, Nanjing 210031, China

2. Jiangsu Province Engineering Lab for Modern Facility Agriculture Technology and Equipment, Nanjing 210031, China

3. Jiangsu Key Laboratory for Food Quality and Safety-State Key Laboratory Cultivation Base Ministry of Science and Technology/Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China

**Abstract** This paper aims to explore the response of Fourier transform mid-infrared (FT-MIR) spectra to the changes of the main components in wheat scab with infected different grades and to realize a non-destructive detection of grades of wheat scab based on FT-MIR spectroscopy combined with Sparse Representation based Classification algorithms. The FT-MIR spectra of 95 wheat samples infected with different grades of wheat scab samples were collected in  $4\ 000\sim 400\ \text{cm}^{-1}$ . The sensitive wavelengths in the FT-MIR spectra of wheat samples were selected by X-loading Weights and Random Forest algorithms, and Sparse Representation based Classification algorithms were used to build models to predict grades of wheat scab. The results showed that the characteristic wavelengths selected by XLW algorithm and RF algorithm achieved an accuracy of more than 90% for each qualitative analysis model, thus, the characteristic wavelength extraction algorithms could effectively simplify the model and improve efficiency. RF-SRC model had the best results, because the accuracy of the modeling set was 97% and the accuracy of the test data set was 96%. Being infected different grade wheat scab could cause the change of the content of water, starch, cellulose, soluble nitrogen, protein and fat in wheat samples. The characteristic wavelength selected by the RF algorithm could reflect the difference of the spectral characteristics of the FT-MIR spectra of these materials, so the grades discrimination of wheat scab by the RF-SRC model can achieve the best effect. Therefore, it is feasible to distinguish the grades of FHB in Wheat by using FT-MIR spectroscopy and pattern recognition method. This paper explained the mechanism of measuring the grades of FHB in Wheat by FT-MIR.

**Keywords** Fourier transform mid-infrared spectra; Wheat; Fusarium head blight; Sparse representation based classification

\* Corresponding author

(Received Aug. 18, 2018; accepted Dec. 27, 2018)