

可见-近红外多光谱数据对水稻种子成活率的判定

罗龙强¹, 姚辛励¹, 何赛灵^{1, 2*}

1. 浙江大学光电科学与工程学院光及电磁波研究中心, 浙江 杭州 310058
2. 苏州瑞蓝环保科技有限公司, 江苏 常熟 215558

摘要 种子活性受到存储条件的影响很大。收集了真实情况下受到不同存储条件影响的种子, 通过发芽实验验证了其成活率存在差异。再从中选择适量的种子样本, 采集其单颗种子的可见-近红外反射光谱, 运用不同的光谱预处理技术, 结合不同的机器学习建模手段, 以区分不同成活率的种子。比较了不同的光谱预处理方法, 比如标准反射光谱校正、多元散射校正等。从识别准确度的角度, 认为标准反射光谱校正的方法, 能够很大程度上提升不同存活率种子的光谱差异性, 从而经过机器学习判断达到更高的识别准确度。同时比较了支持向量机、K 邻近和距离判别分析等机器监督学习建模方法, 发现利用标准反射光谱校正的方法结合距离判别分析, 能够对种子样本实现完全准确的判定。更进一步, 为了满足实际运用中快速识别的要求, 将高分辨率的光谱数据压缩成为低分辨率多通道带通光谱数据, 这样可以大大降低的光谱数据长度, 节约各种机器学习器在训练和判断中所用的时间。使用简化后的多通道带通光谱数据判定种子存活率, 其识别准确度仍然接近 90%。充分说明了, 利用多通道宽带光谱数据, 并选择合适的机器学习建模方式, 足以满足实际选种产业的一般性需求, 有潜力作为未来粮种成活率快速鉴别的技术手段。还采用了多种带通宽度以简化光谱, 分析比较不同带通宽度对识别精度的影响。总体来说由于带宽增大, 数据量减少, 识别速度更快, 但是识别精度降低。从 10~50 nm 改变光谱带宽, 标准反射校正后的简化光谱的识别精度从 87.50% 下降到 58.75%。在实际运用中, 需要权衡识别速率和预期识别精度, 合理的选择带宽。验证了根据简化后的可见近红外反射光谱, 能够较快速且准确的识别水稻种子存活率, 为以后的基于带通滤波片的快速种子存活率识别奠定了基础。

关键词 水稻种子; 可见近红外光谱; 简化光谱; 光谱预处理; 机器学习; 成活率

中图分类号: O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)01-0221-06

引言

中国是农业大国, 是传统水稻生产大国, 水稻种子的选种育种是重中之重。根据种子的发芽率来选种育种, 是最可靠的选育方法。水稻种子的发芽率受到储藏时间和储藏条件的影响, 储藏时间长、储存条件恶劣, 水稻种子发芽率低^[1]。最直接的方法是采用发芽实验法抽样测量种子发芽率, 该方法检测周期为 14 天, 需要人工水稻培育环境, 比较耗时费力, 且不能针对单个种子进行发芽率判定。其他诸如染色比色法^[2], 电导率测量方法^[3]等, 需要通过物理化学实验的方法, 实现对种子品质的测量。虽然避免了传统的发芽实验周期长等缺点, 但是需要对种子样品进行处理加工, 且不能完

整保存测试种子。

近些年来, 研究人员针对种子中特定官能团的特征吸收, 利用近红外光谱检测种子^[4-6], 采用合适的化学计量法诸如主成分分析等方法, 识别种子的品种, 检测种子纯度; 此外近红外光谱也能结合监督学习的方法, 测量种子生命力。与近红外光谱相对比, 采集可见波段的光谱信息更加容易且准确稳定, 即使不能直接对应检测物中特定物质, 但是结合纹理特征后亦可以准确判定种子品种^[7]。现在商用的种子色选仪, 根据种子可见波段的 RGB 三种颜色的特征, 自动化分离杂交种子的父代与子代, 但却不能对同一类的种子更进一步的分选。针对种子检测的高光谱图像数据采集系统^[8-9], 采集了目标的图谱信息, 数据量更加丰富。但通常需要利用光栅分光型成像光谱仪等设备, 仪器比较昂贵且光谱

收稿日期: 2018-11-21, 修订日期: 2019-03-15

基金项目: 国家自然科学基金项目(11621101), 国家重点研发项目(2018YFC1407503), 江苏省重点研发计划项目(BE2015653)资助

作者简介: 罗龙强, 1991 年生, 浙江大学光电科学与工程学院博士研究生 e-mail: luolongqiang@zju.edu.cn

* 通讯联系人 e-mail: sailing@zju.edu.cn

采集时间较长。并且,在进行种子特性判断时,还需要利用图谱的形态特征,因此建模数据变量较多,判定时间很长。

近年来研究人员还关注光谱分析方法上的发展,不同的光谱分析技术被提出以用来分析种子光谱,Loewe^[10]等检测不同种子的近红外光谱,并采用距离判别分析的方法识别种子种类,其识别准确率为 87.8%。Jia 等^[11]采用主成分分析法和偏最小二乘法提取光谱特征,同时利用支持向量机等方法鉴别活种,精度很高。

本实验采集真实条件下由于储存条件而质量不同的种子样本的可见-近红外光谱,较其他研究人员采用的人工老化种子样本而言,种子样本间的差异性并不由人工批量控制,最终实验的结果更有信服力。采用标准反射板光谱数据校正和多元散射校正等不同光谱预处理方法,处理种子样本的原始反射光谱,用之训练了支持向量机、K 邻近和距离判别分析等监督学习方法的模型,比较了不同光谱预处理方法和不同机器学习方法的识别精度,并且分析了这些差异的原因。更进一步,通过计算方法主动压缩光谱分辨率,降低光谱数据长度,计算了简化后的光谱对种子发芽率的判断准确度。说明简化光谱在大大降低光谱数据量的情况下,若是选择合适的机器学习方法,也能保持较高的识别准确度。我们还比较了不同的光谱带宽划分的条件下,机器学习器判定的准确度,证明了光谱划分逐渐粗糙,识别的准确度逐渐降低。这便可以综合考虑判定准确度和光谱数据长度,选择合适的光谱划分带宽,取得符合要求的判定准确度。为基于滤波片的多光谱数据判定种子成活率,提供了理论支持。

1 实验部分

1.1 仪器和参数

光谱仪采用海洋光学 USB2000+, 配备 2 048 个像素的 CCD 感光元件,测量波长范围为 340~1 020 nm。光源用卤素灯 HL2000, 波长范围为 360~2 500 nm, 功率为 10 W。采用 P-TIP 漫反射光纤探头作为光纤接线端,照明角度为 45°, 光谱仪采集角度为 90°, 加载聚光透镜将空间光耦合进光纤。采用 1 000 μm 大芯径光纤,连接光源、探头和光谱仪。

1.2 样本制备

水稻种子由隆平高科提供,出产年份均为 2017 年,水稻品种是梦两优黄莉占(H)。水稻样本源于不同的储藏仓库,由于仓库储存过程中温度、湿度和堆积密度等条件的不同,种子的活性也不一样。区别于高温老化等人工降低种子活力的方法,储藏条件的差异难以得到量化表征。因此在出库前,对 4 个仓库的种子采样进行发芽实验,来确定储藏条件对于种子的影响。测量其发芽率分别为 80%(H1), 77.5%(H2), 71%(H3)和 61.5%(H4)。出仓之后进行低温保存,运送到实验室后,分别取用 40 个样本单个测量光谱数据。

1.3 方法

卤钨灯光源经过 10 min 预热,使出光强度和波形基本稳定。搭建好测量光路,调节漫反射光纤探头,使得出光光斑直径小于水稻粒径。调节积分时间为 20 ms 时,光谱仪数

据较为合适。依次将水稻种子置于光斑处,移动光纤探头,观察光谱仪光强幅值最大处记录水稻种子的反射光谱数据。光谱仪采样时间为 10 ms,平滑度为 5 个像素点,光谱平均 10 次以消除光谱仪随机噪声。光谱仪软件采用海洋光学 SpectraSuite, MATLAB2017b 软件用于数据分析和机器学习。

2 结果与讨论

2.1 可见近红外光谱分析

光谱仪采样范围为 340~1 020 nm,但在光谱仪的测量光谱的极限附近,光谱抖动很大,所以截取光谱范围为 400~1 000 nm。通过标准反射板的光谱数据进行校正,可以消除背景光和光源的影响,其校正公式为

$$R = \frac{R - R_b}{R_w - R_b}$$

其中 R 为测量到的光谱数据, R_b 为背景光谱数据, R_w 为白板光源光谱数据。在试验中,背景光谱数据为光谱仪暗噪声,软件已经自动去除。在试验中发现,白板反射光谱数据并不稳定,所以每测量 20 个种子光谱,测量一次标准反射白板的反射光谱,用来校正这 20 个种子由于光源不稳定导致的光谱数据漂移。多元散射校正(MSC)也是一种常用的对原始光谱进行预处理的方法,通过多元散射校正可以有效的减少因为漫反射的影响,消除样本间的基线平移和漂移现象,增强光谱特异性。

在本次试验中的预处理手段有三个,分别是:(1)标准反射板的光谱数据校正,通过式(1)对原始光谱进行预处理;(2)采用多元散射校正的方法对光谱进行预处理;(3)联用标准反射板的光谱数据校正和多元散射校正两种方法,对光谱预处理。在后文中通过机器学习器的识别准确度,来衡量这三种预处理方法对种子识别判定的影响。

通过标准反射板的光谱数据对光谱处理之后,可以得到不同类别存活率种子的反射系数,如图 1 所示。不同存活率的种子组别之间的差异性通过校正后可以观察到较大的差异,但是组间的差异并不是线性,且差异甚至要小于反射系数的组内差异。这意味着必须通过合适的机器学习的算法才能提取组间差异而忽略组内差异,从而实现很好的对于种子发芽率的判断精度。

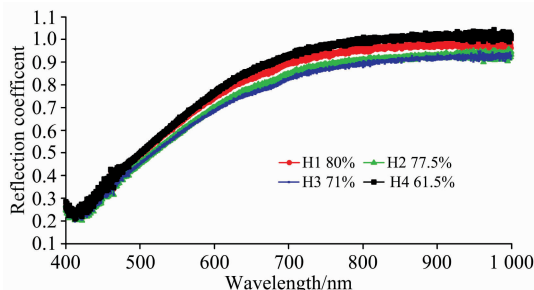


图 1 通过标准反射板对光谱进行校正

Fig. 1 Spectral correction with standard reflector

如图 2 所示为原始光谱经过多元散射校正之后,并且为

了凸显光谱之间的差异性,将每组反射光谱与全体平均光谱平均值求差,得到的差值光谱曲线。可以发现在 550 nm 波长处,各组差值光谱之间的差异性较大,其次在 680 和 750 ~850 nm 处,也能明显的观察到不同组别的种子光谱有较大的区别。种子光谱的差异,体现了种子内部生物活性酶含量和种类差异,而酶活性能很大程度上反映种子活性。这充分说明了因储存差异造成的不同存活率种子其光谱有明显的差异,而且这些差异在不同波长下的显著程度不同。运用合适的机器学习的方法,可以根据光谱对种子存活率进行较为精确的判断。

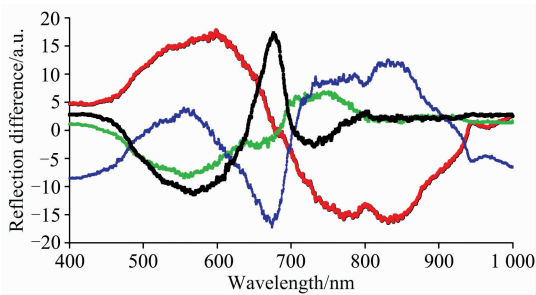


图 2 多元散射校正后各种存活率种子差值光谱
Fig. 2 Differential spectra of seeds with various germination rates after MSC pretreatment

经过两种不同预处理方法校正之后,分别得到的是反射系数和反射光谱的数据。经过标准反射板校正的结果,在整个光谱段都能体现均一的差异性,但是这也意味着光谱数据容易受到漂移等因素的影响。而多元散射校正的预处理手段,利用整体的光谱平均值,减少了干扰的影响,能体现种子在不同波段光谱的差异性。根据两种预处理方法得到的数据,很难直观的评价两种预处理手段的优劣,需要进一步用两种预处理方法得到的数据,对机器学习器加以训练,观察其判定准确度从而比较预处理方法。机器学习针对的就是光谱的差异特征,训练学习器,从而实现水稻种子的种类判别。所以通过预处理,消除背景噪声并保留光谱的差异部分,以增强种类间的光谱差异,一定程度上能够提升机器学习的判断精度。

2.2 全光谱机器学习准确度分析

上述的光谱数据波长范围在 400~1 000 nm 之间,每组光谱为 1 813 个数据点。将每种类型的种子及其光谱样本平均分成 5 份,从 5 份中依次选择 1 份作为测试集,其余 4 份作为训练集,共进行 5 次样本的选择。这就是所谓的 5 折交叉法取样,每个样本都有一次机会作为测试集,丰富了建模的有效性。所以每次取样过程中,针对每种类型的种子,得到训练集样本容量为 32,测试集样本容量为 8。

分别采用支持向量机(SVM)、k-邻近(KNN)和距离判别分析三种监督学习的方法,利用训练集的数据对学习器进行训练,再输入测试集样本进行识别并观察判定结果,得到经过五折交叉得到结果的混淆矩阵。表 1 是利用多元散射校正的光谱和距离判别学习器计算出的混淆矩阵。以第一行的数据为例,第一行实际为 H1 的种子共 40 颗粒,其中 29 颗被正确的归类为 H1,其余的有 8 颗被误判为 H2,3 颗被误

判为 H4。对角线上的元素是每个种类判断正确的个数,数值越大说明判断的准确的情况越多。

表 1 利用多元散射校正光谱的距离判别学习器识别结果混淆矩阵

Table 1 The confusion matrix of distance discriminant learner recognition result using MSC pre-treated spectrum

| Actual classification | Predicted classification | | | |
|-----------------------|--------------------------|----|----|----|
| | H1 | H2 | H3 | H4 |
| H1 | 29 | 8 | 0 | 3 |
| H2 | 9 | 27 | 2 | 2 |
| H3 | 0 | 0 | 37 | 3 |
| H4 | 3 | 1 | 4 | 32 |

可以通过混淆矩阵对角线上的元素和与混淆矩阵的总元素和之比,计算模型的识别准确率。我们将不同光谱预处理的方法和机器学习模型进行组合,比较它们的识别准确率,并记录到表 2 中。

表 2 不同预处理方式和机器学习器的识别精度

Table 2 Recognition accuracy of different preprocessing methods and machine learners

| Machine learners | Spectra preprocessing methods | | | |
|---------------------|-------------------------------|---------------------|--------|------------|
| | Raw Spectra/% | SDC ^a /% | MSC /% | SDC+MSC /% |
| SVM | 41.88 | 65.63 | 76.25 | 41.25 |
| KNN | 31.25 | 30.63 | 38.75 | 41.25 |
| Discr. ^b | 61.88 | 100 | 78.13 | 100 |

注: a: 标准反射板校正; b: 距离判别分析

Note: a: Standard reflector correlation; b: Distance discrimination analysis

首先纵向比较不同的机器学习模型,其中最优秀的为距离判别分析法,对于相同方法处理过的光谱,其识别准确率均高于其他两种方式。基于支持向量机的机器学习模型对于经过标准反射光谱校正和多元散射光谱校正之后的光谱数据,识别准确度较高,针对原始光谱和标准反射+多元散射光谱校正的预处理方法,识别准确度较低。K 邻近的机器学习方法对于光谱分类并不适用,其识别准确度非常低。所以说明距离判别分析是一种十分有效的针对不同存活率的种子的光谱分析方法。

横向对比不同的光谱预处理的方法,可见标准反射光谱校正和多元散射光谱校正处理的光谱,较原始光谱而言,能够提高机器学习器的识别精度。因为多元散射校正预处理方法,利用全体光谱的平均值作为基准,求出每个光谱与其系数偏差,并消除系数偏差。所以经过处理光谱之间的偶然差异性减少,但却保留了本征的差异性,而本征的差异性能够反映种子的种类。所以多元散射的预处理的方法,能够提高机器学习的判断准确度。标准反射光谱校正的方法虽然需较多的测量数据,但是更符合实际情况。通过标准反射光谱校正的预处理手段,可以消除由于波形漂移带来的偏差,提

升学习器的识别准确度。

但是观察标准反射十多元散射光谱校正方法,相对于标准反射光谱或者多元散射光谱校正,并没有提高机器学习的准确度,而且在支持向量机模型的机器学习器下,其识别准确度反而不及单一预处理方式。这是因为光谱数据经过两次处理,不仅消除了测试中的偶然误差和个体间的差异,同时也可能消除表征种类差异性的光谱特征。这样的光谱训练出来的学习器,判断准确度不一定能够提升。说明在光谱预处理的方式方法要结合具体的机器学习模型,不能过多的预处理,降低种类间的特征光谱差异。

2.3 采用多种模型分析简化光谱

上述的光谱数据保留了 400~1 000 nm 之间的光谱数据,每组光谱数据为 1 340 个数据点。这样的数据量对于机器学习建模和识别所需要的时间都很长,为了满足实际运用快速检测的需求。我们将波长按照 10 nm 的带宽进行平均,每条光谱简化得到 45 个数据点。每个中心波长的数据点为带宽内光谱值的平均值。

采取与分析原始光谱相同的方法:将简化后的光谱按照 5 折交叉法取样,输入不同监督学习器中,分析其识别准确度。将不同模型下和不同预处理方法进行对比,总结识别准确度于表 3。

由于简化光谱的数据长度较原始光谱短,所以包含的信息量较少,包含光谱种间差异性的特征也少。所以用简化光谱来训练机器学习模型,其最终的准确度较原始光谱有所下降。但是观察表 3,针对标准反射光谱校正的数据,训练出来的基于距离判别的机器学习器,仍然可以取得 87.5% 的识别准确度,这个在实际选种行业中是一个可观的准确度。充分说明了,简化后的光谱仍然可以较为准确的识别不同存活率分类的水稻种子,为以后基于可见波段的带通滤波片的种子多光谱鉴别系统,提供了理论支持。进一步说明了识别准确度与光谱的分辨率相关性较小,而与采用的预处理手段和

机器学习模型相关性很大。

表 3 不同预处理方式和机器学习器对 10 nm 带通宽度的简化光谱的识别精度

Table 3 Recognition accuracy of different preprocessing methods and machine learners for simplified spectral recognition with a Bandwidth of 10

| Machine learners | Simplified spectra preprocessing methods | | | |
|------------------|--|--------|--------|------------|
| | Simplified spectra/% | SDC /% | MSC /% | SDC+MSC /% |
| SVM | 68.13 | 27.50 | 67.50 | 55.00 |
| KNN | 31.87 | 28.13 | 38.75 | 27.50 |
| Discr. | 67.50 | 87.50 | 63.12 | 89.38 |

2.4 不同带宽下的简化光谱对识别准确度的影响

在整个光谱范围不变的情况下,我们研究了不同的通带宽度下简化光谱在学习器作用下的识别精度。随着通带宽度不断的增加,由 10 nm 增加到 50 nm,通带的个数逐渐的减少,也就意味着输入的变量更少。根据机趋学习的一般性原理,输入学习器的变量越少,就越难从变量中找到与因变量相关的自变量条件。我们设计的带通宽度分别为 10, 15, 20, 30, 40 和 50 nm,采用距离判别分析学习器在不同的带通宽度和不同的光谱校正预处理下的识别精度,其结果表 4 所示。

由表可知,带宽增大其判断精度均降低。偶然的几个异常升高值,可以理解为简化光谱划分的时候,能够反映种间差异性的光谱成分被分到两个不同的通带,更加凸显了种间的差异性。通过以上的数据,可以得到结论:光谱简化的过程中,不能过大的设计光谱通带,这样会失去存活率相关的光谱特征,使得机器学习识别器准确度降低。而常见的 10 nm 的通带,是通常的带通滤波片能够达到的最小带宽,用来做多光谱鉴别水稻种子活力,是比较合适的。

表 4 不同带通宽度下简化光谱识别精度

Table 4 Simplified spectral recognition accuracy with different band-pass widths

| preprocessing methods | Different bandwidths | | | | | |
|-----------------------|----------------------|-------|-------|-------|-------|-------|
| | 10 nm | 15 nm | 20 nm | 30 nm | 40 nm | 50 nm |
| Simplified spectra/% | 67.50 | 66.25 | 72.50 | 64.38 | 65.63 | 60.62 |
| SDC/% | 87.50 | 85.00 | 78.75 | 68.13 | 58.13 | 58.75 |
| MSC/% | 63.12 | 66.25 | 71.25 | 63.75 | 62.50 | 63.12 |
| SDC+MSC/% | 89.38 | 83.75 | 77.50 | 76.25 | 59.38 | 57.50 |

3 结 论

采集由于实际储存条件不同造成存活率不同的种子在可见近红外波段的反射光谱,通过标准反射光谱校正、多元散射校正和两者联用等光谱预处理方法,结合支持向量机、k 邻近和距离判别分析等机器学习模型,来根据光谱预测种子的存活率。通过比较分析,我们认为基于距离判别分析的机器学习器,是最适合光谱分类分析的技术手段。而标准

反射光谱校正后的光谱数据,能够对四种不同成活率的水稻种子,实现 100% 的判别精度。基于可见近红外反射光谱,找到适合真实种子存活率的最佳光谱预处理手段和最优机器学习判断方法。同时我们用 10 nm 的带宽简化了原始光谱,降低样本的光谱数据长度,利用简化后的光谱对种子存活率进行判定,仍然能够取得 87.50% 的识别准确度,这样的高准确度在实际选种产业有巨大的运用潜力。同时我们设计了不同带宽,用来简化光谱。在实验过程中发现,随着简化程度加深、光谱带宽增大,其判定效果降低。所以找到较小的带

宽并采用距离判别分析学习器作为判定方法,能够快速有效的区分不同存活率的水稻种子。为以后的基于带通滤波片的快速种子识别技术奠定了基础。

References

- [1] XU Hui-bin, WEI Yi-dong, LIAN Ling, et al(许慧滨,魏毅东,连玲,等). *Molecular Plant Breeding(分子植物育种)*, 2013, 11(5): 552.
- [2] ZHONG Shu-mei, GUO Xiao-liang, GUO Jie, et al(钟淑梅,郭晓亮,郭杰,等). *Modern Chinese Medicine(中国现代中药)*, 2017, 19(12): 1732.
- [3] LIU Chun-xiang, SUI Ming, HU Pan, et al(刘春香,隋铭,胡畔,等). *Seed(种子)*, 2013, 32(7): 18.
- [4] Montes J M, Paul C, Kusterer B, et al. *Journal of Near Infrared Spectroscopy*, 2006, 14(6): 387.
- [5] Olesen M H, Shetty N, Gislum R, et al. *Journal of Near Infrared Spectroscopy*, 2011, 19(3): 171.
- [6] Agelet L E, Ellis D D, Duvick S, et al. *Journal of Cereal Science*, 2012, 55(2): 160.
- [7] DENG Xiao-qin, ZHU Qi-bing, HUANG Min(邓小琴,朱启兵,黄敏). *Laser & Optoelectronics Progress(激光与光电子学进展)*, 2015, 52(2): 128.
- [8] XU Si, ZHAO Guang-wu, DENG Fei, et al(许思,赵光武,邓飞,等). *Seed(种子)*, 2016, 35(4): 34.
- [9] Santosh S, Matej K, Uros Z, et al. *Sensor and Actuators B: Chemical*, 2016, 237, 1027.
- [10] Veronica L, Rafael M N, Juan G, et al. *Food Control*, 2017, 73: 634.
- [11] Jia S, Yang L, An D, et al. *Journal of Cereal Science*, 2016, 69(Supplement C): 145.

Study on the Method of Determining the Survival Rate of Rice Seeds Based on Visible-Near Infrared Multispectral Data

LUO Long-qiang¹, YAO Xin-li¹, HE Sai-ling^{1, 2*}

1. Centre for Optical and Electromagnetic Research, College of Optical Science and Engineering, Zhejiang University, Hangzhou 310058, China
2. Suzhou Reliatek Environmental Technology Co., Ltd., Changshu 215558, China

Abstract The seed vigor was greatly affected by the storage condition. This work collected seeds that were affected by different storage conditions under real circumstance, and verified their germination rate difference by germination experiment. After that, we took some samples from them and measured the visible-near-infrared reflection spectra of each single grain. By adapting different spectral preprocessing techniques, combined with several machine learning modeling methods, we tried to distinguish the seeds according to their germination rates. In our experiments, some spectral pretreatment methods were compared, such as standard reflection spectrum correction. It also compared supervised machine learning modeling method such as support vector machine, K-near neighbor and discriminant analysis. From the perspective of recognition accuracy, we believed that standard reflection spectrum correction method can greatly improve the spectral difference of seeds with different survival rates, and thus achieve higher recognition accuracy through machine learning. At the same time, we compared the supervised machine learning modeling methods such as support vector, k near neighbor and distance discriminant analysis, and find that the standard reflection spectrum correction method combined with distance discriminant analysis can achieve hundred-percent accuracy of predicting different seeds category. Furthermore, in order to meet the requirements of rapid identification in practical applications, in the experiment we compressed high-resolution spectral data into low-resolution multi-channel band-pass spectral data, which can greatly reduce the spectral data length and save the time spent in training and classifying of various machine learner. The prediction accuracy of models trained by those simplified spectra data is still close to 90%. It fully demonstrates that the use of multi-channel broadband spectral data combining with the selection of appropriate machine learning modeling methods is sufficient to meet the general needs of the actual seed selection industry, and it is a potential technique for rapid identification of rice grain survival rates in the future. The experiment also used a variety of bandpass widths to simplify the spectra, and analyze and compare the effects of different bandpass widths on the recognition accuracy. In general, due to the increase in bandwidth, the length of data is reduced, and the recognition speed is faster, but the recognition accuracy is decreased. In the experiment, we changed the spectral bandwidth from 10 to 50 nm, and the recognition accuracy of the simplified spectrum after standard reflection correction decreased from 87.50% to 58.75%. In practical use, it is necessary to balance the recognition rate and the expected recog-

niton accuracy, and select a reasonable bandwidth. This study verified that the simplified near-infrared reflectance spectroscopy can quickly and accurately identify the survival rate of rice seeds, which lays a foundation for rapid seed survival rate recognition technique based on bandpass filters in the future.

Keywords Rice seeds; Visible-near-infrared spectra; Simplified spectra; Spectra pretreatment; Machine learning; Germination rate

(Received Nov. 21, 2018; accepted Mar. 15, 2019)

* Corresponding author