

基于近红外光谱分析的高丹草种子发芽率检测研究

惠云婷¹, 王德成¹, 唐欣², 彭要奇¹, 王红达¹, 张海凤¹, 尤泳^{1*}

1. 中国农业大学工学院, 北京 100083

2. 山东农业工程学院, 山东 济南 250100

摘要 高丹草中粗蛋白质以及碳水化合物的含量丰富, 适合青贮处理。优质的高丹草种子是发展畜牧业十分重要的前提, 发芽率是检验种子质量最常规的指标之一, 播前种子发芽率检测与筛选十分必要。现阶段采用发芽试验法进行种子发芽率的检测, 周期长、成本高。基于此, 提出利用近红外光谱对高丹草种子进行发芽率的快速、无损检测。选择适量的高丹草种子样品, 采集近红外漫反射光谱, 进行一阶导和二阶导预处理以及对比分析 R^2 , R_p^2 , RESEC 和 RMSEP。采用支持向量机(SVM)建模, 使用 MATLAB 中调用的 LIBSVM 软件包来实现 SVM 训练和检测过程, 以检测不同发芽率的高丹草种子。来自不同省份的 100 组高丹草种子先剔除种子内的杂物、破损以及不能满足试验条件的种子后, 用人工气候培养箱进行种子发芽试验, 获得 100 组种子样本的发芽率, 其发芽率分布在 41%~64% 的范围。采用美国 Unity Scientific 2600XT 近红外光谱仪对样本进行光谱扫描。随机分成校正集 70 份和检验集 30 份。分别采用一阶导和二阶导进行了高丹草种子光谱的预处理, 将预处理之后的数据采用支持向量机的方法建模, 并对其参数进行了分析和讨论。结果表明, 近红外光谱预测模型训练集相关系数(R^2)和测试集相关系数(R_p^2)分别为 0.94 和 0.92, 校正均方根误差(RMSEC)、预测均方根误差(RMSEP)分别为 0.21 和 0.25, 两个产地的高丹草种子数据采用一阶导预处理时模型最优。支持向量机的方法建模采用 Rbf 核函数, 当支持向量机惩罚因子 $c=2896.3094$ 和核函数 $g=0.5$ 时, 测试集种子发芽率的检测准确率为 96.6667%(29/30)。该模型预测种子发芽率是可行的, 可以作为初步检测高丹草种子发芽率快速无损检测的手段之一, 能够有效的促进种子生产。

关键词 近红外光谱; 高丹草种子; 发芽率; 导数法; 支持向量机

中图分类号: S339.3+1

文献标识码: A

DOI: 10.3964/j.issn.1000-0593(2022)02-0423-05

引言

近年来, 随着畜牧业的快速发展, 饲草料的需求也急剧增加, 饲草料的短缺已经成为制约畜牧业发展的重要因素之一, 亟待解决。高丹草(sorghum-sudan grass)作为一种新品种, 是由高粱(sorghum bicolor)和苏丹草杂交培育而成, 它结合了二者的优良特性, 富含粗蛋白及碳水化合物, 适合青贮, 且在干旱、寒冷、盐碱地等环境下抵抗能力强^[1]。高丹草的优良特性使其成为许多地区的家畜首选饲料。高丹草的种植改变了一些地区的农业结构, 对于地区发展畜牧业有着突出的贡献。优质的高丹草种子可有效促进畜牧业的发展, 发芽率是衡量种子质量的常规指标之一^[2]。研究表明, 长时间贮藏

高丹草种子会影响种子活力, 发芽率也随之变化。因此高效判断高丹草种子质量可有效促进畜牧业发展。

现阶段实验室检测种子发芽率多采用发芽试验法^[3], 在人工气候培养箱内严格按照种子发芽技术规定的条件(包括湿度、光照、时间等)进行试验, 试验条件要求较高, 一般需要 10~15 d, 周期较长, 且发芽试验对种子具有破坏性, 试验后的种子不能重复使用, 试验投入的人工成本及时间成本较高, 不能满足快速、准确、无损测试的要求, 也不能适应快速发展的畜牧业需求。因此, 为降低成本, 提高检测速度及精度亟需一种无损、快速且准确的发芽率的检测方法^[3-4]。近红外光谱技术的应用广泛, 可以得到有关物质的特征信息^[5], 可以不破坏样品且无需利用化学试剂, 更高效、成本更加低的进行样品分析^[6]。故而, 许多研究人员开始将这一

收稿日期: 2021-01-04, 修订日期: 2021-04-08

基金项目: 国家牧草产业技术体系项目(CARS-34), 国家重点研发计划项目(2016YFD0701705)和农作物种子播前处理后时效性研究及其应用项目(2019GNC106155)资助

作者简介: 惠云婷, 女, 1992年生, 中国农业大学工学院博士后 e-mail: HYT1556521354@163.com

* 通讯作者 e-mail: youyong@cau.edu.cn

技术应用到种子领域。

在农作物的营养检测方面,近红外光谱技术^[7]也展现出了独特优势,如评估高粱内部成份含量^[8]、预测植物叶片含水率^[9]、小麦内在品质的无损检测^[10-11];戴子云等利用近红外光谱技术对结缕草种子的标准发芽率进行分析,并利用定量偏最小二乘法进行数据分析;李毅念等为检测不同老化程度的杂交水稻种子的发芽率,采集数据并进行建模,校正预处理时采用了全波段和标准化+正交信号这种方法;金文玲等设计并搭建了一套投射式近红外光谱检测系统,结合光谱特性建立了主成分分析模型,同时通过偏最小二乘法建立了不同年份水稻种子的活力判别模型,研究结果表明,水稻种子的活力梯度与其近红外吸收光谱的吸收峰值相关^[12]。国内外现有活力检测分析的研究对象主要是针对小麦和水稻,鲜有高丹草种子发芽率分析的报道。本工作针对高丹草种子检测发芽率存在的问题,提出一种利用近红外光谱技术快速检测高丹草种子发芽率的方法。

采集 2 个品种的高丹草种子的近红外光谱,同时进行种子发芽试验;利用导数法对采集的光谱数据进行预处理,并将预处理之后的数据采用支持向量机(SVM)建立高丹草种子发芽率检测模型,以期分析高丹草种子发芽率的光谱特征,提高发芽率检测的速度及精度,从而为基于近红外光谱技术检测高丹草种子发芽率提供依据。

1 实验部分

1.1 材料

高丹草种子 100 份,分别来自内蒙(50 份)、山东(50 份),随机选择相同粒数的种子进行发芽试验,试验地点为中国农业大学牧草机械实验室。

1.2 光谱采集

图 1 所示为美国 Unity Scientific 2600XT 近红外光谱仪,波长范围:680~2 600 nm,分辨率:1 nm;该仪器内含二极管阵列检测器检测固体样品的近红外光谱,检测范围宽;在食品、农业、医药领域得到广泛应用。



图 1 近红外光谱仪

Fig. 1 Near-infrared spectrometer

在测试样品前,剔除杂物、破损粒、霉变粒等坏种粒;称取相同质量 $[(8 \pm 0.1) \text{ g}]$ 的高丹草种子样品置于样品杯中,倒入时,高丹草的种子样品至样品杯的 10 mL 刻度处,并通过旋转视窗设计和漫反射的方式进行扫描。重复扫描 24

次求取平均光谱。

1.3 种子发芽试验

根据农业部《农作物种子检验规程—发芽试验》(GB/T 2930.4—2001)标准,进行高丹草种子发芽试验。将培养皿清洗、烘干,然后在其中平铺两层滤纸,将去离子水沿壁缓慢的滴入培养皿,使滤纸完全浸湿,并将筛选出的 100 粒高丹草种子均匀的放置在浸湿的滤纸上。种子之间保持适当的距离,避免过于紧密或者距离太大影响种子发芽率。每组试验重复 3 次;将培养皿放入培养箱(型号:BD-PRX,南京贝帝实验仪器有限公司),每天观察种子生长情况,当芽长超过种子自身长度 1/2 视为发芽,并记录发芽情况,及时补充水分,且对发芽数进行记录,当实验进行至 12 d 时,统计高丹草种子的总发芽数。

种子发芽率的计算如式(1)所示

$$GR = \frac{n}{N} \times 100 \quad (1)$$

式(1)中,GR 为种子发芽率;N 为供试种子粒数;n 为发芽种子粒数。

1.4 数据分析与处理

采用仪器所带的 UScan™ 软件及美国 Unity 科学公司 SpectraStar 系列第二代仪器的日常定量分析软件,使用 matlabR2020a 软件调用 LIBSVM 软件包。预处理:将采集样品的光谱数据利用导数法进行分析,并将预处理之后的数据采用支持向量机(SVM)建立高丹草种子发芽率检测模型,并且通过该模型对高丹草种子的发芽率进行快速检测与验证。

2 结果与讨论

2.1 样本的发芽率和近红外光谱

通过种子发芽试验得到的 100 份样品的发芽率分布于 41%~64%之间(表 1),平均值为 52.86%;将试验的 100 份种子根据发芽率进行分类,发芽率 41%~52%属于第一类(类标签:0),52%~100%为第二类(类标签:1);按照 7:3 的比例划分训练集和测试集,可得训练集样本 70 份,测试集 30 份。

表 1 样本发芽率

Table 1 Germination rate of sample

发芽率/%	样本数
40~45	18
45~50	19
50~55	29
55~60	18
60~65	16

近红外漫反射光谱如图 2,从图中可以看出,不同种类的高丹草种子样品的光谱图像相似,样品的漫反射吸光度值范围是 0.19~0.94。光谱图像有重叠,在 1 110, 1 200, 1 300, 1 460, 1 900, 2 000, 2 100, 2 200 和 2 500 nm 附近有明显的吸收峰、吸收谷存在。

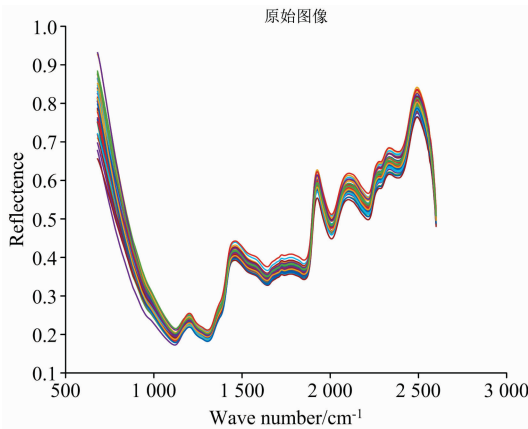


图 2 高丹草种子的原始光谱
Fig. 2 The original NIR spectra of sorghum-sudan grass seeds

2.2 不同光谱预处理方法对高丹草种子建模效果的影响

采集样品光谱信息的目的是采集该样品所包含的化学组成信息，然而一些其他无关信息也有可能被采集，因此对采集到的光谱进行预处理以去除电噪声、样品背景和杂散光无关信息的干扰，降低模型的复杂程度，从而提高模型的稳定^[13]。在多种预处理方法^[14]中，使用导数光谱导数法来消除由于背景色或其他因素引起的光谱基线漂移或位移，以区分重叠的波峰和波谷，提高分辨率，降低系统所带来的误差以及基线校正。在本研究中，由于采集的高丹草种子光谱段波长样点多且分辨率较高，采用直接差分法求导，以此减少误差。

根据光谱特性，对 9 个近红外的光谱波段 1 000~1 200, 1 200~1 300, 1 300~1 400, 1 400~1 800, 1 800~2 000, 2 000~2 100, 2 100~2 200, 2 200~2 300, 2 300~2 600 和全波段 1 000~2 600 进行光谱预处理，对 9 个近红外波段建模，比较分析 R^2 , R_p^2 , SEC 和 SEP 所得参数值(如表 2)。由前人的研究可知，采用 R^2 与 RMSEP 作用判别预处理模型优劣^[2]。一个好的模型结果其相关系数(R)应该较大，而模型的 RMSEC 应该较小并且模型的 RMSEP 也应该较小，且已知 R 的值越接近 1、RMSEC 和 RMSEP 的值越小且二者之间的差值越小，说明所建立的模型比较稳定且较好。通过表格可知，一阶导为最优的预处理方法，预处理之后不限定特定波段，采用该预处理方法后得到的 R^2 和 R_p^2 的值分别为 0.94 和 0.92，RMSEC 和 RMSEP 的值分别为 0.21 和 0.25，

表 2 高丹草种子的光谱预处理结果

Table 2 Spectral pretreatment results of the sorghum-sudan grass seed

光谱预处理方法	训练集		测试集	
	R^2	RMSEC	R_p^2	RMSEP
无	0.90	0.29	0.87	0.36
1-Der	0.94	0.21	0.92	0.25
2-Der	0.81	0.43	0.79	0.48

两个值及其差值均较小，说明该方法预处理越优。通过一阶导(1-Der)和二阶导(2-Der)预处理后获得的高丹草种子预处理后光谱图，如图 3 和图 4。一阶导光谱的特征性展现的更加凸显，吸收峰和吸收谷更加尖锐，光谱的离散程度显著减小。

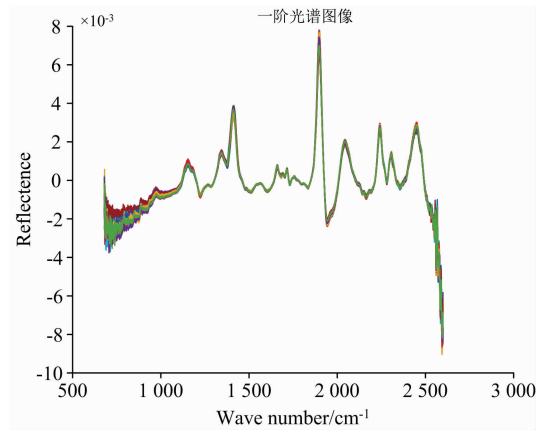


图 3 一阶导预处理后的光谱曲线
Fig. 3 NIR spectrum curves after 1-Der pre-processing

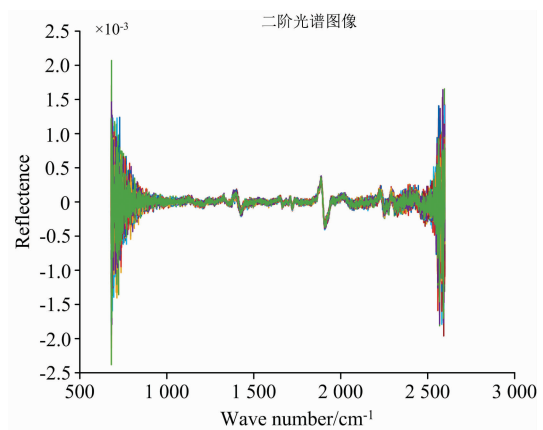


图 4 二阶导预处理后的光谱曲线
Fig. 4 NIR Spectrum curves after 2-Der pre-processing

2.3 基于导数法-支持向量机的高丹草种子发芽率检测模型

支持向量机(support vector machines, SVM) 在解决小样本问题时有其独特之处^[15]，其在优化问题时，将训练误差作为约束条件，相较于一些传统的分析样本的方法来说，具有不可替代的优势性。在对高丹草种子发芽率检测时，该方法十分适用。标记所有通过导数方法预处理的样本，在训练 SVM 核函数及其参数时，随机选取标记好的高丹草种子样本总量的十分之七，剩下的数据样本用作测试集。在本研究中，使用 MATLAB 中调用的 LIBSVM 软件包来实现 SVM 训练和预测过程。

通过使用 3 折和 10 折交叉验证方法进行对比，首先选择核函数类型为 Rbf，采用网格寻优的方法寻找支持向量机惩罚因子 c 和核函数 g 的最优值。参数 g 定义了单个训练样本的影响大小，值越小影响越大，参数 c 在误分类样本和分界面简单性之间进行权衡。在试验时选择随机抽样，使参数

c 和 g 在 $[2^{-5}, 2^{15}]$ 的范围内进行网格优化, 并确定最佳参数, 如图 5(3D 视图)所示和图 6(等高线图)所示。准确率是指被对分的样本数除以所有的样本数, 错分率与正确率相反, 描述被错分的比例; 用准确率和错分率来评价分类模型。结果表明, 3 折和 10 折交叉验证对比, 折线对比区别较小, 检验准确率相同, 3 折检验效率较高, 所以使用 3 折交叉验证。参数 c 和 g 最优分别为 2 896.309 4 和 0.5, 并将这两

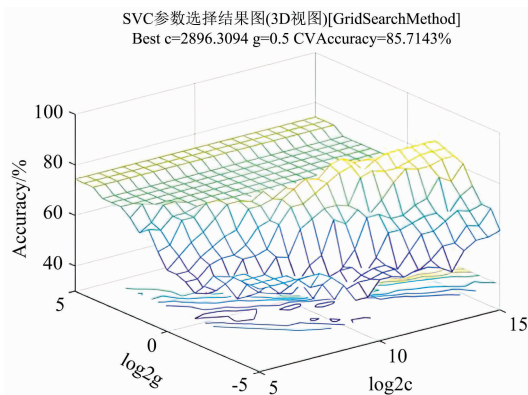


图 5 网格搜索最佳 c, g 3D 视图

Fig. 5 Grid search for best c, g 3D view

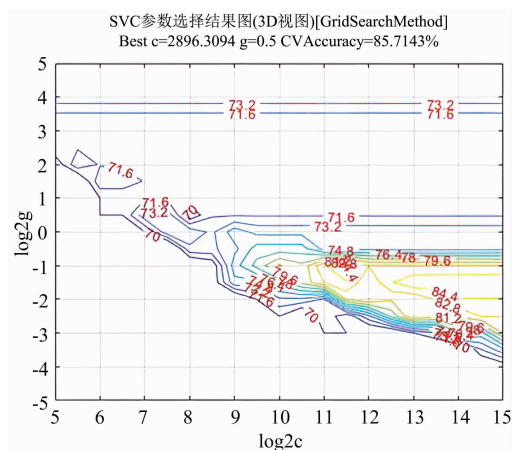


图 6 网格搜索最佳 c, g 等高线视图

Fig. 6 Contour map of grid search for best c, g

个最优参数训练支持向量机。3 折交叉验证下的最佳检测准确率为 96.666 7% (29/30), 如图 7 所示分类情况检测的性能已经可以满足一定的要求。结合预处理和 3 折交叉验证法, 建立导数法-支持向量机高丹草种子检测模型。

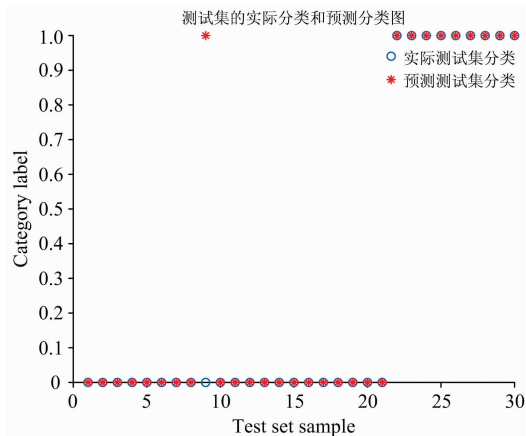


图 7 测试集的实际检测和预测检测图

Fig. 7 The actual classified result and prediction result of testing samples

3 结 论

以高丹草种子为研究对象, 选取 1 000~2 600 nm 波段近红外光谱结合支持向量机(SVM)建立高丹草种子发芽率检测的数学模型, 实现了高丹草种子发芽率的快速检测。通过采用一阶导和二阶导的预处理, 得到近红外光谱预测数学模型 R_c^2 和 R_g^2 的值分别为 0.94 和 0.92, RMSEC 和 RMSEP 的值分别为 0.21 和 0.25。经过对比, 一阶导为最优的预处理方法; 将预处理之后的数据采用支持向量机的方法建模, 同时对模型的参数和核函数 Rbf 进行了详细的分析和讨论, 得到最佳参数 $c=2 896.309 4$ 和 $g=0.5$, 测试集的分类准确率为 96.666 7% (29/30)。研究表明, 利用近红外光谱对高丹草种子发芽率进行建模分析检测是可行的, 且该方法有其特有的优点, 能够避免周期较长的发芽试验, 更高效、成本低。本试验采用的高丹草种子样本来源于两个种植地点, 在以后的研究中, 需要用更多个种植地点、不同品种的样品来完善模型, 使检测结果更加准确可靠。

References

- [1] Peng J, Kim M, Sung K. Agriculture, 2020, 10(4): 137.
- [2] WU Jing-zhu, DONG Wen-fei, LIU Qian, et al(吴静珠, 董文菲, 刘倩, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2015, 31(s2): 272.
- [3] Zhou S, Sun L J, Xing W, et al. Infrared Physics & Technology, 2020, 108.
- [4] Lin J, Yu L, Li W, et al. Applied Spectroscopy, 2018, 72(1): 611.
- [5] SUN Tong, WU Yi-qing, LI Xiao-zhen, et al(孙通, 吴宜青, 李晓珍, 等). Acta Optica Sinica(光学学报), 2015, 35(6): 350.
- [6] Zhu D, Wang K, Zhang D, et al. Sensor Letters, 2011, 9(3): 1144.
- [7] GAO Tong, WU Jing-zhu, MAO Wen-hua, et al(高彤, 吴静珠, 毛文华, 等). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2019, 50(S1): 399.
- [8] Peiris K H S, Bean S R, Jagadish S V K. Cereal Chemistry, 2020, 97(5): 1066.

- [9] ZHU Long-jie, ZHANG Hua, ZHUANG Ya-dong, et al(朱龙杰, 张 华, 庄亚东, 等). Tobacco Science & Technology(烟草科技), 2017, 50(9): 55.
- [10] Femenias A, Gatius F, Ramos A J, et al. Food Chemistry, 2020, 2(341): 128206.
- [11] Amigo J M, Alvarez A D O, Engelsen M M, et al. Food Chemistry, 2016, 208: 318.
- [12] JIN Wen-ling, CAO Nai-liang, ZHU Ming-dong, et al(金文玲, 曹乃亮, 朱明东, 等). Chinese Optics(中国光学), 2020, 13(5): 1032.
- [13] ZHANG Ya-kun, LUO Bin, SONG Peng, et al(张亚坤, 罗 斌, 宋 鹏, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2018, 34(18): 195.
- [14] Zhang J, Dai L M, Cheng F. Molecules, 2019, 24(1): 149.
- [15] WANG Xiang-yu, ZHU Chen-guang, FU Ze-tian, et al(王翔宇, 朱晨光, 傅泽田, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2019, 39(6): 1864.

Detection of Sorghum-Sudan Grass Seed Germination Rate Based on Near Infrared Spectroscopy

HUI Yun-ting¹, WANG De-cheng¹, TANG Xin², PENG Yao-qi¹, WANG Hong-da¹, ZHANG Hai-feng¹, YOU Yong^{1*}

1. College of Engineering, China Agricultural University, Beijing 100083, China

2. Shandong Agriculture and Engineering University, Ji'nan 250100, China

Abstract Sorghum-Sudan Grass is rich in crude protein and carbohydrate, suitable for silage treatment. High-quality seeds are a prerequisite for animal husbandry development, and germination rate is one of the most conventional indicators to test the seed quality. Therefore, testing and screening the germination rate of the seeds prior to sowing is essential. The germination test method is currently used to detect seed germination rate, which has a long cycle and high cost. In this study, a rapid and non-destructive method based on NIR was proposed to detect the germination rate of sorghum-sudangrass seeds. The near-infrared diffuse reflectance spectra of the seed samples were collected with 1-Der and 2-Der processing. Moreover, comparative analysis of the parameter values obtained for R_c^2 , R_p^2 , RESEC and RMSEP was also performed. The support vector machine (SVM) was used for modeling, and the LIBSVM software package in Matlab was used to realize the SVM training and detection process to detect the seeds of sorghum-sudangrass seed with different germination rates. Using the Unity scientific 2600 XT Near-infrared spectrometer, 100 groups of sorghum-sudangrass seeds from different provinces were selected as samples. Before the experiment, the broken seeds and seeds that did not germinate were removed, and the germination test was carried out in the incubator. The germination rate of 100 samples was obtained, and the germination rate ranged from 41% to 64%. The seed samples were spectroscopically scanned and were randomly divided into calibration set (70 samples) and test set (30 samples). In this paper, the 1-Der and 2-Der method was used to preprocess the spectrum of sorghum-sudangrass seeds. SVM modeled the preprocessed data, and its parameters were analyzed. The results showed that the correlation coefficients of the training set (R_c^2) and test set (R_p^2) were 0.94 and 0.92 respectively, and the root mean square error of correction (RMSEC) and root mean square error of prediction (RMSEP) were 0.21 and 0.25 respectively, which reflected that the model was the best when the 1-Der was used to preprocess the seed data. When $c=2896.3094$, $g=0.5$, the detection accuracy of the test set was 96.6667% (29/30) by using Rbf core functions of SVM modeling. These results suggested that the model was feasible to predict the seed germination rate, and could be used as one of the rapid and non-destructive detection methods for the preliminary detection of seed germination rate of sorghum-sudangrass could effectively promote the seed production.

Keywords Near infrared spectroscopy; Sorghum-sudan grass seed; Germination rate; Derivative method; Support vector machine

(Received Jan. 4, 2021; accepted Apr. 8, 2021)

* Corresponding author