

激光与光电子学进展

基于混合机器学习法的太赫兹波
鉴别草种的研究王芳¹, 张春红¹, 赵景峰^{2*}, 哈斯巴特尔², 张玉¹¹中国石油大学(北京)理学院, 北京 102249;²内蒙古自治区草原工作站, 内蒙古 呼和浩特 010020

摘要 利用太赫兹时域光谱技术对黄耆类牧草种子样品进行测试, 得到 5 种常见沙打旺牧草种子在 0.2~1.2 THz 有效频率范围内的太赫兹时域谱, 然后通过快速傅里叶变换得到了各牧草种子样品的吸收系数、折射率等光学参数。研究后发现: 在有效频率范围内, 样品时域谱的峰值强度和延迟时间均不同, 且每条谱线的平均吸收系数和标准差也有明显差异, 各样品的平均折射率也有较大差异。同时, 本文提出了一种将主成分分析(PCA)与随机森林(RF)机器学习算法相结合的优化实验数据的混合模型 PCA-RF, 并基于太赫兹折射率谱, 采用 PCA-RF 模型和 RF 模型对 5 种牧草种子的 200 个数据集进行了统计计算。结果表明: 混合模型 PCA-RF 的平均分类准确率为 91.20%; 与 RF 模型相比, 不管是总的平均分类准确率, 还是每种样品的分类准确率, PCA-RF 模型都优于 RF 模型。研究结果表明, 太赫兹时域光谱技术结合混合机器学习算法的 PCA-RF 模型是一种无损鉴定牧草种子真伪的有效手段, 可用于鉴别同族且差异较小的牧草品种。

关键词 光谱学; 太赫兹时域光谱; 主成分分析; 随机森林; 定性鉴别; 定量分析

中图分类号 O433.4

文献标志码 A

doi: 10.3788/LOP202158.0330001

Identification of a Grass Species Using a Terahertz Wave Based on
Hybrid Machine Learning MethodWang Fang¹, Zhang Chunhong¹, Zhao Jingfeng^{2*}, Sibateer Ha², Zhang Yu¹¹College of Science, China University of Petroleum, Beijing 102249, China;²Inner Mongolia Grassland Station, Huhhot, Inner Mongolia 010020, China

Abstract In this study, the terahertz time-domain spectroscopy (THz-TDS) technology was used to conduct experimental tests on seed samples of astragalus japonica. We obtained the terahertz time-domain spectra of five kinds of *Astragalus adsurgens* Pall. seeds in the effective frequency range of 0.2-1.2 THz, and used the fast Fourier transform analysis to study the optical parameters such as the absorption coefficient and refractive index of each grass-seed sample. It was found that in the effective frequency range, the peak intensity and delay time of the time-domain spectrum of the samples were different, and the average absorption coefficient and standard deviation of each spectrum line were significantly different. In addition, the average refractive index of the samples was significantly different. At the same time, this study proposes a hybrid model of optimized experimental data that combines principal component analysis (PCA) with random forest machine learning (RF). Based on the terahertz refractive index spectrum, 200 datasets of five forage species were statistically calculated, and the calculated results were compared with the calculated

收稿日期: 2020-05-03; 修回日期: 2020-06-12; 录用日期: 2020-06-19

基金项目: 内蒙古自治区科技计划(201602054)

*E-mail: jingfengzhao62@163.com

results of the RF model. The results show that the average classification accuracy of principal component analysis-random forest (PCA-RF) in the mixed model is 91.20%. Compared with the RF model, both total average classification accuracy and the classification accuracy of each sample of the PCA-RF model are better than those of the RF model. The study shows that the PCA-RF model combining THz-TDS with the hybrid machine learning algorithm can be used as an effective method for the nondestructive identification of the authenticity of forage grass seeds. In particular, it can be used for the classification of forage grass varieties of the same family with little difference.

Key words spectroscopy; terahertz time-domain spectroscopy; principal component analysis; random forest; qualitative identification; quantitative analysis

OCIS codes 300.6495; 300.1030

1 引言

沙打旺又名直立黄耆,是多年生草本豆科牧草植物,具有药用价值和饲用价值。牧草沙打旺的优劣会直接影响到畜牧业经济效益的高低,因此,沙打旺品种的鉴定是草业科学研究的一项基本内容,对于摸清沙打旺品种、挖掘地方优良品种、进行品种选育等具有重要意义^[1]。

近几年,随着光谱无损检测技术的发展,作为红外光谱辅助手段的太赫兹技术在植物种子及种间差异的鉴别中显示出了其独有的优势。太赫兹的单光子能量很低(约为 4.1 MeV),对生物体无电离损伤^[2],不会对生物组织和细胞造成损害;同时,生物有机分子的分子间弱相互作用和晶体中晶格的低频振动吸收对应太赫兹波段^[3],所以可以通过特征共振和吸收对物质进行指纹谱分析,而这一特性也为植物种子的分类鉴别奠定了理论基础。尽管植物种子是一种混合物,成分较为复杂,很难出现指纹谱,但有些学者将太赫兹技术与机器学习算法相结合,通过定量计算分析也取得了一定进展。例如:2016年,Liu等^[4]利用太赫兹时域光谱成像技术对稻米进行了太赫兹图像的采集,随后利用随机森林(RF)算法对获得的图像进行分类鉴别,得到了比较满意的结果;2016年,龙怡霖等^[5]利用随机森林分类器对有缺损的杂草种子图像进行分类识别,提出了一种实用性更强的模型;2019年,杨玉平等^[6]利用太赫兹光谱技术结合主成分分析法成功地对中草药藏红花和草红花以及天然、人工牛黄进行了比较准确的区分;2019年,周月等^[7]利用主成分分析法结合随机森林算法对转基因油菜种子的太赫兹吸收光谱进行了识别,结果表明,该方法对转基因油菜的分类准确率达到91.00%。以上这些研究的成功为利用太赫兹波技术进行牧草种子的有效鉴定识别奠定了理论及实验基础。

本文在太赫兹实验研究的基础上,提取了 5 种

牧草沙打旺种子的太赫兹光谱,然后进行了牧草种子及种间差异的定性鉴定识别研究。特别地,本文提出了一种将主成分分析(PCA)与随机森林算法相结合的 PCA-RF 算法,然后采用该算法对实验测试得到的 5 种沙打旺牧草种子的折射率数据进行了量化识别计算,并将计算结果与传统的单一随机森林算法的计算结果进行对比分析,从而确定了 PCA-RF 算法的优势。所提算法为快速有效地对牧草种子进行鉴别提供了一种新思路和新方法。

2 实验及结果分析

2.1 实验装置及样品

本文采用的实验装置是钛蓝宝石飞秒锁模激光器,其可产生中心波长为 800 nm、重复频率为 80 MHz、脉宽为 100 fs 的激光脉冲,该激光器的输出功率为 960 mW。采集太赫兹光谱的太赫兹时域光谱(THz-TDS)系统的实验装置如图 1 所示^[8]。所有的测试均在室温(23 °C)条件下进行。

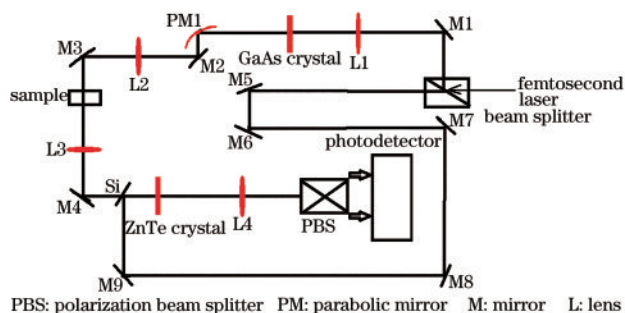


图 1 THz-TDS 系统示意图

Fig. 1 Schematic of THz-TDS system

实验采用的 5 种样品均由内蒙古草原站提供,来自内蒙古各草种基地。5 种样品的名称、年份及产地如表 1 所示。所有牧草样品种子均呈颗粒状,且外形相似。测试时,先对样品进行挑选、粉碎、研磨、烘干处理,得到沙打旺种子的粉末,称取一定质量的样品,将其与聚乙烯粉末充分均匀混合(样品与聚乙烯的质量

表 1 5 种牧草沙打旺种子的相关信息

Table 1 Relevant information of five samples of *Astragalus adsurgens* Pall. seeds

Number	Name	Place of origin	Year
Sample 1	Sha Da Wang 1	Helin	2010
Sample 2	Sha Da Wang 2	Helin	2012
Sample 3	Sha Da Wang 3	Helin	Before 2010
Sample 4	Sha Da Wang 4	Helin	2016
Sample 5	Sha Da Wang 5	Helin	2016

比为 8:1), 然后用粉末压片机将制得的粉末在 20 MPa 压力下压制成厚度为 1.2 mm 的圆片。样片表面无裂缝, 两平面保持平行(可减少测量时的多重反射)。

2.2 实验结果及分析

通过实验采集 5 种样品在 0.2~1.2 THz 有效频段内的太赫兹时域光谱图, 如图 2(a) 所示。可以看出, 5 种样品的峰值强度和延迟时间略有不同, 峰值强度从高到低分别是样品 1(0.10544)、样

品 4(0.09983)、样品 3(0.09942)、样品 2(0.09827)、样品 5(0.09756), 延迟时间从左到右分别是样品 1(7.010 ps)、样品 3(7.030 ps)、样品 2(7.080 ps)、样品 4(7.150 ps)、样品 5(7.180 ps)。由于样品的实验测试初始条件一致, 而且每个样品的厚度都是 1.2 mm, 质量都是 0.9 g, 且颗粒粒径集中在 107~125 μm 之间, 烘干时间也相同, 所以样品不同的延迟时间和峰值强度说明了样品对太赫兹波的折射与吸收是不同的, 这也说明了利用太赫兹时域光谱技术定性鉴定识别黄耆类牧草种子及种间差异的可行性。但由于 5 种沙打旺种子属于同族牧草系列, 所以其差异性不太大。图 2(b) 是 5 种样品的太赫兹频域图谱, 可以看出, 样品在 0.2 THz 之前以及在 1.2 THz 之后的振幅几乎为 0, 对数据分析几乎没有影响。因此, 本文选择 0.2~1.2 THz 频段的光谱数据进行处理和分析。

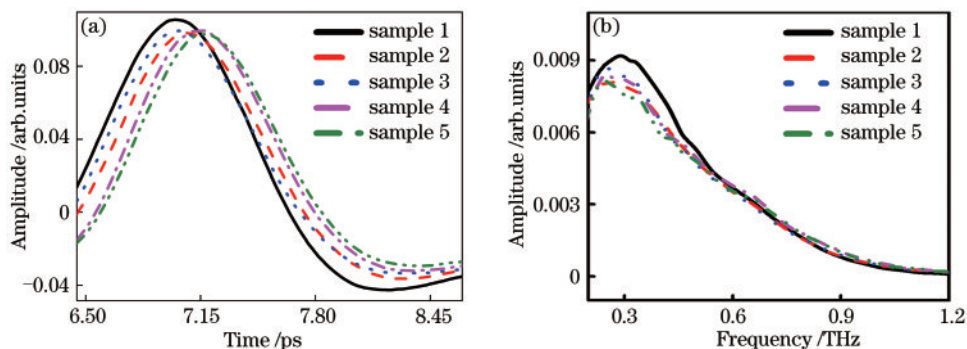


图 2 5 种牧草沙打旺样品的太赫兹时域、频域光谱图。(a) 时域光谱图; (b) 频域光谱图

Fig. 2 Terahertz time and frequency domain spectral waveforms of five *Astragalus adsurgens* Pall. seeds. (a) Terahertz time domain spectral waveforms; (b) terahertz frequency domain spectral waveforms

图 3 是 5 种样品的吸收系数谱图。由图 3 可以看出, 5 种样品的吸收光谱曲线较为类似, 没有明显的吸收峰。这是因为牧草种子是一种混合物, 成分复杂, 各种有机分子之间的特征吸收峰可能会出现重叠, 导致样品的波形较为平滑。具体来说, 在低频区(0.2~0.7 THz), 5 种样品的波形几乎完全重叠, 难以直接区分; 而在高频区(0.7~1.2 THz), 5 种样品的区分度较好。这说明, 随着频率增大, 5 种样品对太赫兹波越来越敏感, 可辨别性越来越大。

本文将 5 种样品在高频区的吸收系数进行了平均和标准差分析, 分析结果如图 4 所示。可以看出, 在 5 种样品中, 样品 1 的平均吸收系数最大(为 2.010), 样品 5 的吸收系数最小(1.523)。因此, 可以根据吸收系数对 5 种样品进行辨别区分。在标准

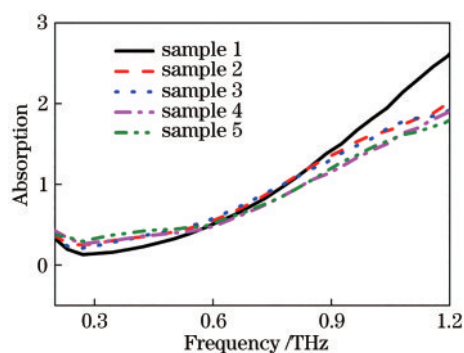


图 3 5 种牧草沙打旺样品的吸收系数图谱

Fig. 3 Absorption coefficient spectra of five *Astragalus adsurgens* Pall. seeds

差分析中可以看到: 样品 1 的误差棒最长, 说明其大部分吸收系数的数值与平均值之间的差异较大, 即

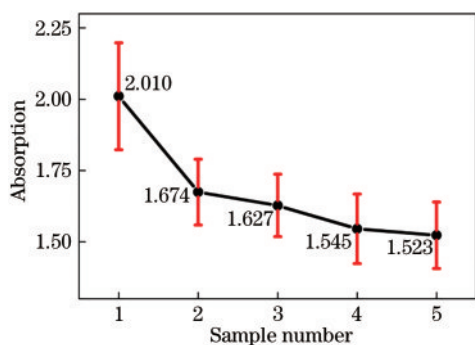


图 4 5 种牧草沙打旺样品的平均吸收系数和标准差

Fig. 4 Average absorption coefficient and standard deviation of five *Astragalus adsurgens* Pall. seeds

该样品对太赫兹波吸收的变化最明显;样品 3 的误差棒最短,说明随频率的升高,其对太赫兹波吸收的变化不明显。

图 5 是 5 种样品的折射率谱图。很明显,5 种样品的折射率具有相同的起伏变化,且整体随着频率增大呈现降低的趋势。样品 1 和样品 4 的折射率相近,它们在有效频段内的平均折射率分别为 1.588 和 1.584;样品 3 和样品 5 的平均折射率相近,分别为 1.557 和 1.562;样品 2 的折射率最小,其平均折射率为 1.543。从吸收系数和折射率谱图对比来看,吸收谱图在高频区的区分度较好,而折射率谱图在整个有效频段内的可识别性较大,所以,为了准确、快速地鉴别差异较小的 5 种样品,本文采用折射率数据进一步与机器学习算法相结合进行研究。

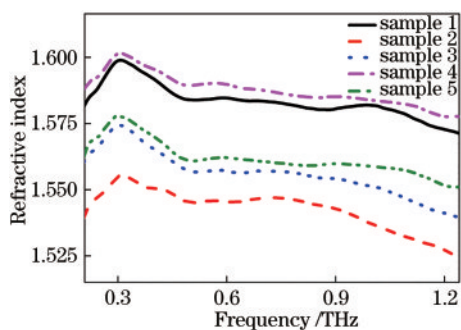


图 5 5 种牧草沙打旺样品的折射率谱图

Fig. 5 Refractive index spectra of five *Astragalus adsurgens* Pall. seeds

3 随机森林模型及 PCA-RF 模型 的比较

3.1 主成分分析算法及随机森林算法

主成分分析是机器学习中一种常用的降维方法,该方法的特征提取是通过将多个变量转化为少

数几个变量实现的^[9]。主成分分析算法通过构造一组正交基,将高维度的数据投影至一个超平面上,将高维数据降至低维,并且使降维后的数据方差尽可能大,这样降维后的数据就保留了原始高维数据中的大部分信息。对一个样本集合 $X = \{x_1, x_2, \dots, x_n\}$ 进行主成分分析的步骤如下:

- 1) 对所有样本进行中心化, $x_i = x_i - \frac{1}{n} \sum_{i=1}^n x_i$;

- 2) 计算样本集合的协方差矩阵, $Cov = \sum_{i=1}^n x_i x_i^T$;

- 3) 对协方差矩阵进行特征值分解;

- 4) 取前 k 个最大特征值,即 k 个目标维数,并计算对应的特征向量 $W = (\omega_1, \omega_2, \dots, \omega_k)$,以 W 为一组基构造新的低维样本集合 X' 。

随机森林算法是以大量决策树的结果汇总来提高模型精度的,该算法避免了过拟合问题,非常适合折射率、吸收系数等非线性数据的建模,并在数据缺失或不平衡的情况下仍能保持稳定^[10]。随机森林是一种有监督的模式识别算法^[11],它通常包含 ID3 (Iterative Dichotomiser 3)、C4.5 (Classification 4.5)、CART (Classification and Regression Tree) 三种算法^[12]。本文采用 ID3 算法进行计算。ID3 算法将基于香农信息论的信息增益 (information gain) 作为划分准则。某个属性的信息增益越大,使用该属性进行划分后的子集就越纯。

设 D 为训练的样本集合, d 为样本的某个属性,共有 V 个可能的取值 (d_1, d_2, \dots, d_V) 。 D^v 为属性 d 中所有取值为 d_v 的样本集合。信息增益的定义由 (1) 式给出, $\text{Gain}(D, d)$ 表示样品集合 D 中属性为 d 的信息增益。 $\text{Ent}(D)$ 表示样本集合 D 的信息熵 (information entropy), 定义由 (2) 式给出。其中, p_i 表示第 i 个样本所占的比例, $i = 1, 2, \dots, n$ 。

$$\text{Gain}(D, d) = \text{Ent}(D) - \sum_{v=1}^V \frac{|D^v|}{D} \text{Ent}(D^v), \quad (1)$$

$$\text{Ent}(D) = - \sum_{i=1}^n p_i \log_2 p_i. \quad (2)$$

由以上可知,主成分分析算法对数据具有良好的处理能力,可以在对大量数据进行降维处理减少数据量的同时保留数据中的主要信息,且可以去除无用噪声。但是,对于一组数据而言,该算法本身无法给出任何有用的信息,而随机森林分类器不仅可以对数据进行有效分类,还可以有效防止数据过拟合。为了体现主成分分析和随机森林分类器结合后的优势,本文分别建立了随机森林模型与主成分分

析-随机森林模型来对 5 种样品进行分类鉴别,从而确定一种更加准确的用于牧草种子的分类方式。

3.2 随机森林模型及计算结果分析

为了建立样品的数据集,本文首先选取了 5 种牧草沙打旺样品在有效频段(0.2~1.2 THz)内的折射率数据。为了缩短模型的学习时间,提高预测精确度,本文同时选取了 5 种样品的振幅系数,建立 5 种样品的数据集(共 200 个),每种样品各 40 组数据。以训练集 170 组、测试集 30 组的建模方法建立

随机森林模型,并以测试集分类准确率(test accuracy)来评价模型的好坏(即判断测试集与训练集所建立模型的吻合程度),分类准确率越高,吻合性越好,模型精度越高^[13]。表 2 为随机森林模型对 5 种牧草沙打旺样品的 10 次平均分类准确率为 85.00%;在 10 次训练过程中,样品 5 的分类准确率最高,为 90.90%,样品 1 的分类准确率最低,为 81.60%;整体的分类结果较低,结果不太理想。

表 2 随机森林模型对 5 种样品的分类结果

Table 2 Classification results of RF model

No.	Classification accuracy of all kinds of samples/%					Classification accuracy of five samples/%	
	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5		
1	75	100	86	67	86	83.30	
2	80	89	86	80	100	86.70	
3	91	60	100	86	67	83.30	
4	60	100	100	67	86	83.30	
5	80	67	67	100	100	83.30	
6	75	100	89	80	100	86.70	
7	100	63	100	100	80	86.70	
8	80	100	78	71	100	83.30	
9	75	85	75	88	100	86.70	
10	100	83	80	80	90	86.70	
Average classification accuracy/%		81.60	85.00	86.10	81.90	90.90	85.00

3.3 数据处理及建立 PCA-RF 模型

为了提高随机森林模型的精度,考虑到折射率图谱中的数据点较多,具有一定的噪声干扰,所以需要原始数据进行处理,提取特征量后再进行建模分析。先从 5 种牧草样品的折射率数据出发,对数据进行主成分分析,然后选取贡献率最高的三个主成分(几乎能代表所有折射率光谱数据信息),再结合有效频段内的振幅系数,以随机森林算法模型为基础,建立 PCA-RF 模型。对折射率数据进行主成分分析后,得到了前三个特征分量的贡献率及累计贡献率,如表 3 所示。其中主成分 1 的贡献率超过 95.00%,为 95.79%,前三个主成分的累计贡献率达到了 99.68%,说明前

表 3 主成分特征向量

Table 3 Eigenvector of principle component

Component	Eigenvector		
	Eigenvalue	Variance contribution rate/%	Cumulative variance contribution rate/%
1	84.29	95.79	95.79
2	3.214	3.650	99.44
3	0.2125	0.2400	99.68

三个主成分可以代表原始光谱 99.68% 的信息。所以,本文以表 3 中的前三个主成分数据以及 0.2~1.2 THz 频段内的振幅系数作为数据源,建立每种样品的数据集。同时以 170 组训练集和 30 组测试集建立 PCA-RF 模型。最终模型对 5 种样品的分类准确率如表 4 所示。

表 4 的计算结果表明,PCA-RF 模型对 5 种牧草样品的 10 次平均分类准确率为 91.20%,并且模型对样品 4 的分类准确率最高(为 97.00%),对样品 3 的分类准确率最低(为 88.30%),对其他三类样品的分类准确率也分别达到了 94.20%、92.10% 和 91.40%,均超过了 90.00%。与随机森林模型相比,样品 4 的平均准确率提升得最大,由 81.90% 提升到了 97.00%,提高了 15.1 个百分点,样品 5 的准确率提升得较小,为 0.5 个百分点。总体来说,不管是 10 次总的分类结果,还是 10 次各样品的分类准确率,PCA-RF 模型都优于随机森林模型,说明 PCA-RF 模型在黄耆类牧草种子的分类方面比随机森林模型更适合。因为主成分分析不仅较好地保存了有效信息,还极大地剔除了无效数据,减少了数据的维数,从而极大地提高了模型的精度。

表 4 PCA-RF 模型的分类结果
Table 4 Classification results of PCA-RF model

No.	Classification accuracy of all kinds of samples/%					Classification accuracy of five samples/%	
	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5		
1	80	100	100	90	100	93.30	
2	100	83	89	80	100	90.00	
3	100	83	75	100	100	93.30	
4	100	80	83	100	75	90.00	
5	100	100	80	100	71	90.00	
6	86	100	83	100	100	93.30	
7	86	75	100	100	100	90.00	
8	90	100	100	100	100	96.70	
9	100	100	87	100	88	90.00	
10	100	100	86	100	80	93.30	
Average classification accuracy/%		94.20	92.10	88.30	97.00	91.40	91.20

4 结 论

本文利用太赫兹时域光谱技术对黄耆类牧草种子进行了测试,并进行了种间差异的定性分析。分析结果表明:5种样品的时域光谱,无论是峰值强度还是延迟时间均存在差异,并且黄耆类牧草种子对太赫兹波的敏感性随着频率的增大而增强,样品种间的太赫兹光学特征在高频区域的差异比较明显。这说明利用太赫兹时域光谱技术定性鉴别黄耆类牧草种子及种间差异是可行的。

为了提高测试结果的精确度,剔除无效数据,本文提出了PCA-RF计算模型,采用该模型对光谱实验数据进行了计算,并将计算结果与传统随机森林模型的计算结果进行了对比。对比结果显示:PCA-RF模型对5种样品的平均分类准确率达到91.20%,与RF模型85.00%的分类准确度相比有了较大提升,且PCA-RF模型对每种样品10次测试的分类准确率也都高于RF模型。

以上结果说明太赫兹时域光谱技术结合机器学习算法的PCA-RF模型是一种很好的提高鉴别识别精度的混合方法。这种混合分析方法不仅为牧草品种的鉴别提供了一种快速、有效的办法,同时也为建立牧草种质资源数据库奠定了基础。

参 考 文 献

[1] Li X L, Shao A J, Zhan X R, et al. Study on microscopic identification of *Astragalus complanatus* and *A. adsurgens* seeds[J]. China Journal of Chinese Materia Medica, 2015, 40(7): 1271-1273.
李晓琳, 邵爱娟, 展晓日, 等. 沙苑子及其伪品直立

黄芩的显微鉴别研究[J]. 中国中药杂志, 2015, 40(7): 1271-1273.

[2] Liu X Q, Yao J L, Huang F, et al. Study on detection of penicillin drugs based on terahertz time-domain spectroscopy[J]. Acta Optica Sinica, 2020, 40(6): 0630001.

刘晓庆, 姚嘉丽, 黄凡, 等. 基于太赫兹时域光谱的青霉素类药物检测研究[J]. 光学学报, 2020, 40(6): 0630001.

[3] Zhang W T, Li Y W, Zhan P P, et al. Recognition of transgenic soybean oil based on terahertz timedomain spectroscopy and PCA-SVM[J]. Infrared and Laser Engineering, 2017, 46(11): 1125004.

张文涛, 李跃文, 占平平, 等. 基于太赫兹时域光谱技术与PCA-SVM的转基因大豆油鉴别研究[J]. 红外与激光工程, 2017, 46(11): 1125004.

[4] Liu W, Liu C H, Hu X H, et al. Application of terahertz spectroscopy imaging for discrimination of transgenic rice seeds with chemometrics[J]. Food Chemistry, 2016, 210: 415-421.

[5] Long Y L, Cai C. Random forest- based damaged weed seeds recognition[J]. Computer Applications and Software, 2016, 33(8): 185-189.

龙怡霖, 蔡骋. 基于随机森林的缺损杂草种子识别[J]. 计算机应用与软件, 2016, 33(8): 185-189.

[6] Yang Y P, Zhang C, Liu H S, et al. Identification of two types of safflower and bezoar by terahertz spectroscopy[J]. Spectroscopy and Spectral Analysis, 2019, 39(1): 45-49.

杨玉平, 张成, 刘海顺, 等. 两类红花和牛黄的太赫兹光谱法真伪鉴别分析[J]. 光谱学与光谱分析, 2019, 39(1): 45-49.

- [7] Zhou Y, Sun J, Yang S G, et al. Terahertz spectral analysis and recognition based on machine learning [J]. *Radio Engineering*, 2019, 49(12): 1031-1036.
周月, 孙霁, 杨四刚, 等. 基于机器学习的太赫兹光谱分析与识别[J]. *无线电工程*, 2019, 49(12): 1031-1036.
- [8] Bao R M, Zhao K, Tian L, et al. Analysis of THz time-domain spectroscopy in crude oil ultrasound treatment [J]. *Modern Scientific Instruments*, 2013(2): 126-129.
宝日玛, 赵昆, 田璐, 等. 原油超声处理的太赫兹时域光谱分析[J]. *现代科学仪器*, 2013(2): 126-129.
- [9] Liu J X, Du B, Deng Y Q, et al. Terahertz-spectral identification of organic compounds based on differential PCA-SVM method [J]. *Chinese Journal of Lasers*, 2019, 46(6): 0614039.
刘俊秀, 杜彬, 邓玉强, 等. 基于差分-主成分分析-支持向量机的有机化合物太赫兹吸收光谱识别方法[J]. *中国激光*, 2019, 46(6): 0614039.
- [10] Li X H. Using "random forest" for classification and regression [J]. *Chinese Journal of Applied Entomology*, 2013, 50(4): 1190-1197.
李欣海. 随机森林模型在分类与回归分析中的应用[J]. *应用昆虫学报*, 2013, 50(4): 1190-1197.
- [11] Xu H D, Lin L L, Li Z, et al. Nephrite origin identification based on Raman spectroscopy and pattern recognition algorithms [J]. *Acta Optica Sinica*, 2019, 39(3): 0330001.
徐荟迪, 林露璐, 李征, 等. 基于拉曼光谱和模式识别算法的软玉产地鉴别[J]. *光学学报*, 2019, 39(3): 0330001.
- [12] Yuan L S, Lou M Y, Liu Y Q, et al. Palm vein classification based on deep neural network and random forest [J]. *Laser & Optoelectronics Progress*, 2019, 56(10): 101010.
袁丽莎, 娄梦莹, 刘娅琴, 等. 结合深度神经网络和随机森林的手掌静脉分类[J]. *激光与光电子学进展*, 2019, 56(10): 101010.
- [13] Wang B, Wang Q H, Xiao Z, et al. Discrimination of origin of eggs using visible-near-infrared spectroscopy and random forest [J]. *Science and Technology of Food Industry*, 2017, 38(24): 243-247.
王彬, 王巧华, 肖壮, 等. 基于可见-近红外光谱及随机森林的鸡蛋产地溯源[J]. *食品工业科技*, 2017, 38(24): 243-247.