

# 茶叶夹杂昆虫异物 THz 光谱检测研究

孙旭东, 刘俊彬

华东交通大学机电与车辆工程学院, 江西 南昌 330013

**摘要** 中国红茶因其悠远的文化底蕴和多种保健功效而广受推崇。红茶经杀青、揉捻、晒干、筛选、干燥制成成品。但复杂的加工工艺也增加了物理污染的风险,在加工过程中难免会夹杂非茶类异物。物理污染属于随机事件,占食品安全投诉事件的19.8%。降低食品污染是未来造成厂家与消费者、进出口贸易双方争议的关键。昆虫等有机异物尚属于X射线不可检测范畴的食品异物,但昆虫异物混入频率较高,容易造成消费投诉,迫切需要开发相应的无损检测方法。太赫兹时域光谱(THz-TDS)技术作为一种新兴的检测技术,在农产品、食品以及药品无损检测方面均具有良好的应用潜力。THz具有良好的低能透射和指纹光谱特性、且无电离辐射损伤,能透过食品基质获取潜藏异物的光谱和图像信息,是农产品、食品无损检测的较佳选择。为了实现茶叶夹杂低密度有机异物的高效检测,本文基于THz光谱技术,探索红茶夹杂昆虫异物的无损检测新方法。在0.2~3.0 THz范围内,采集了红茶基质、昆虫异物以及夹杂昆虫异物红茶的THz光谱。分析了茶叶基质和昆虫异物的THz吸收系数和介电损耗响应特性,从频谱图可以看出茶叶基质与昆虫异物的THz吸收系数和介电损耗存在显著的差异,主要由昆虫异物的蛋白质和脂肪成分引起,这奠定了红茶夹杂昆虫异物的THz光谱检测基础。但茶叶与昆虫异物吸收系数均无明显的特征峰,且在2.0~3.0 THz高频波段内存在较为明显的噪声。采用主成分分析方法分别对吸收系数和介电损耗进行降维处理,通过得分图可以判断出红茶基质与夹杂昆虫异物的红茶之间存在明显的区别,且吸收系数的聚类效果优于介电损耗系数。分别选取0.5~1.0 THz范围的THz吸收系数和介电损耗为输入向量,建立了支持向量机(SVM)和线性判别分析(LDA)判别模型。实验结果表明,基于THz吸收系数的LDA判别模型精度最高,新样品的正确识别率为73.68%。这说明应用THz时域光谱无损检测红茶夹杂的昆虫异物具有可行性,同时THz光谱结合模式识别算法为茶叶夹杂昆虫异物无损检测提供了新方法,也可为其他农产品和食品检测提供参考。

**关键词** 太赫兹; 光谱; 茶叶; 昆虫异物; 判别分析

**中图分类号:** O439 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)09-2723-06

## 引言

茶是一种深受人们喜爱的健康饮品,但传统茶叶加工工艺流程比较复杂。以红茶为例,鲜叶要经过萎凋、揉捻、发酵、烘焙、复焙等一系列工序制作为成品茶。茶叶在采摘和生产过程中极易混入各种低密度有机异物,例如:昆虫、塑料、木屑、铁屑等,这些已成为茶叶的主要污染源,也是消费者投诉的主要食品异物<sup>[1]</sup>。当前,世界各国都在强化进口茶叶的卫生标准,因此研究茶叶夹杂有机异物无损检测技术,具有重要的理论价值和应用前景。

X射线是现行的食品异物检测技术,利用食品基质与异

物的密度差,能有效检出金属异物,并逐步延伸到塑料等高密度异物<sup>[2-3]</sup>。但茶叶的生产加工环境参差不齐,在茶叶加工过程中昆虫等低密度有机异物混入频率较高。根据新闻报道,在2009年—2010年欧盟RASFF连续两年通报我国出口的乌龙茶和普洱茶中存在死昆虫,消费者投诉茶叶中存在昆虫类污染物逐渐增多<sup>[4]</sup>。昆虫异物混入不仅造成感官上的不悦,而且易于引入致病细菌等,造成食品安全事故。昆虫异物污染占食品物理污染投诉事件的32.2%<sup>[1]</sup>,但昆虫异物与食品基质较为接近,X射线很难适用。太赫兹(THz)是介于红外光与微波之间的电磁波,具有低能透射特性,可以作为X射线检测技术的必要补充。THz辐射的能量低,频率为1 THz的光子能量约为4.1 meV,不会造成电离辐射损伤。

收稿日期: 2020-07-18, 修订日期: 2020-11-09

基金项目: 国家自然科学基金项目(31960497), 江西省自然科学基金项目(20202BAB205009), 江西省研究生创新资金项目(YC2019-S241)资助

作者简介: 孙旭东, 1978年生, 华东交通大学机电与车辆工程学院副教授 e-mail: sunxudong\_18@163.com



目前 SVM 模型中常用的核函数有：线性核函数，多项式核函数以及高斯径向函数，部分核函数式(2)和式(3)

线性(linear)核函数

$$K(x_i, x_j) = (x_i, x_j) \quad (2)$$

高斯径向基函数(RBF)

$$K(x_i, x_j) = \exp\{-\gamma |x_i - x_j|^2\}, \gamma > 0 \quad (3)$$

LDA 属于一种有监督学习的降维方法，数据集中每个样品类型是有类别输出的，将一个类型之间的方差最小化，不同类型之间方差最大化<sup>[16]</sup>与线性判别分析类似，二次判别分析是另外一种线性判别分析算法，二者拥有类似的算法特征，区别仅在于：当不同分类样本的协方差矩阵相同时，使用线性判别分析；当不同分类样本的协方差矩阵不同时，则应该使用二次判别。

## 2 结果与讨论

### 2.1 昆虫异物的 THz 光谱响应特性

对茶叶与纯异物的吸收系数与介电损耗光谱信号求取平

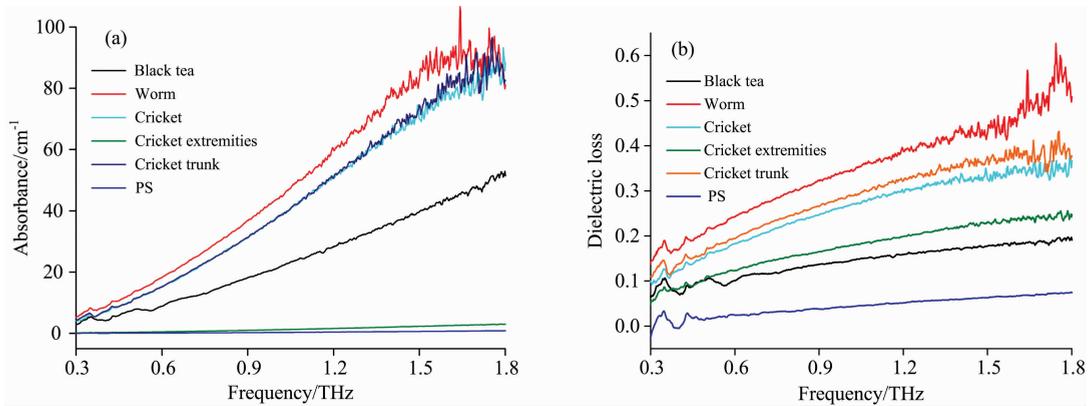


图 3 茶叶与异物的 THz 光谱吸收系数 (a) 和介电损耗 (b)

Fig. 3 THz spectral absorption coefficient (a) and dielectric loss (b) of tea and foreign matters

红茶与异物混合后的 THz 光谱响应特性如图 4 所示。从图 4(a)中可以观察到，不同异物与茶叶混合后的吸收系数与纯红茶的光谱区分明显。由于蟋蟀与蟋蟀躯干在生物成分

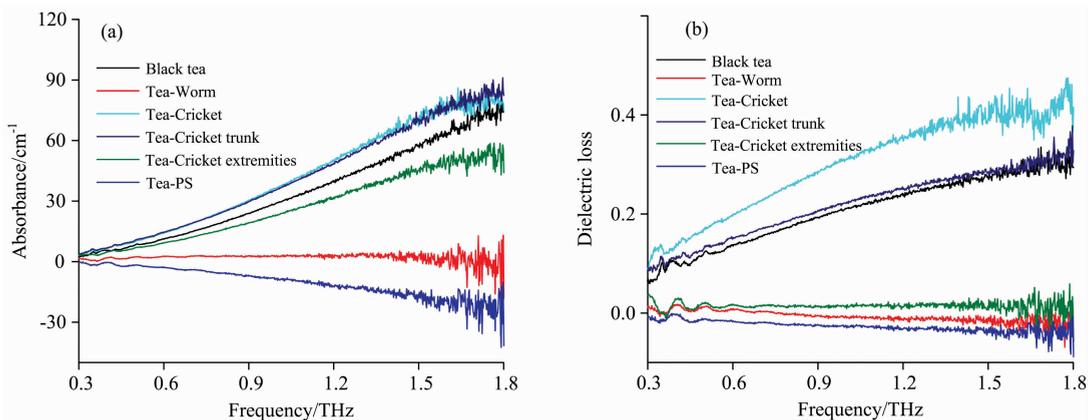


图 4 红茶掺杂异物的吸收系数 (a) 与介电损耗 (b)

Fig. 4 Analysis of response of spectroscopy to insect foreign bodies hidden in tea, absorption coefficient (a) and dielectric loss (b)

均频段光谱，为了有效的减少系统噪声对数据的干扰，本研究通过对光谱进行筛选，选取了 0.3~1.8 THz 波段作为吸收系数与介电损耗的有效数据进行研究。茶叶与异物的吸收系数与介电损耗如图 3 所示。烘干后的蠕虫中粗蛋白和粗脂肪分别占到了 49.5% 和 38.9%<sup>[17]</sup>，而蟋蟀烘干后主要成分为蛋白质，达到了 61%<sup>[18]</sup>。从图 3(a)中可以看出富含蛋白质和脂肪的蠕虫以及蟋蟀对 THz 辐射吸收很高，随着频率的增加，异物的吸收光谱呈现快速上升趋势<sup>[19]</sup>。蟋蟀四肢主要成分为甲壳质，该异物则对 THz 波吸收比较微弱。聚苯乙烯与蟋蟀四肢的吸收系数整体处于水平状态，而且重叠区域较多，很难区分。红茶富含多酚类物质以及碳水化合物等，与其他异物存在较大的差别。

图 3(b)中可以观察到完整的蟋蟀与蟋蟀躯干之间介电损耗存在轻微的分，红茶的光谱与其他异物样品的光谱间隔明显。在 0.3~1.8 THz 之间，无论是吸收系数还是介电损耗，红茶与不同异物的吸收系数存在很明显的区分，这使得 THz 光谱检测技术可以实现茶叶中异物的判别。

上十分接近，两者与茶叶混合后的吸收光谱具有很高的相似性，但与纯茶叶吸收光谱具有一定差异。此外红茶夹杂聚苯乙烯的吸收光谱明显低于纯红茶样品，对 THz 波吸收依然

很弱,与红茶光谱很容易区分。由于红茶与昆虫异物混合物对高频太赫兹辐射的强烈吸收和散射,混合物的吸收光谱在 1.0~1.8 THz 频率范围内振荡。从吸收光谱可以看出各类异物样品与茶叶光谱在 0.5~1.0 THz 波段下区分最大,但对于部分异物种类仍然难以区分。如图 4(b)所示,不同异物的 THz 介电损耗光谱之间存在较大差别。在 0.5~1.0 THz 范围内,光谱整体噪声较小,该波段内红茶与混有蟋蟀躯干异物的样品光谱区分并不明显。其中掺杂蠕虫、完整的蟋蟀与聚苯乙烯的压片光谱和纯红茶谱线间隔明显,掺杂聚苯乙烯的样本介电损耗参数随着频率下降而下降,含有蠕虫样品的频谱趋势较为平缓。混合蟋蟀和蟋蟀躯干异物的红茶压片介电损耗参数在 0.5~1.5 THz 波段呈线性增长趋势,在

1.0~1.8 THz 波段内开始剧烈振荡,故 0.5~1.0 THz 频段被用于后续模型分析使用。

## 2.2 THz 光谱主成分分析

主成分分析法可以对多个变量之间的相关性进行统计分析,是一种减少数据维度的有效方法。对各组数据进行正交变换,降维后的主成分数据之间无线性相关且按得分高低降序排列。通常前  $k$  个主成分累计方差贡献率足够大时,原始数据可以近似的由前  $k$  个主成分代替。运用主成分分析法对含有不同异物的样品进行处理,选取 0.5~1.0 THz 频段范围内的吸收光谱和介电损耗,两个参数主成分得分图如图 5 所示。其中吸收系数与介电损耗前两个主成分在数据集中占比都超过了 95%。

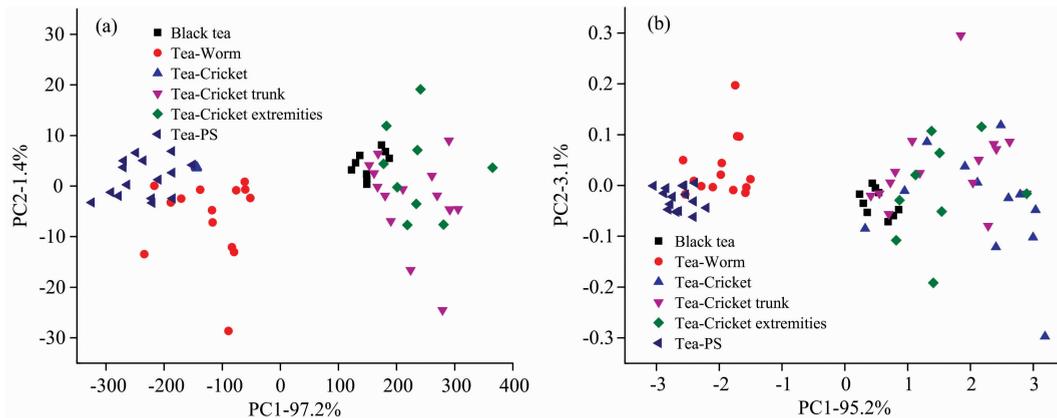


图 5 红茶与异物样品光谱主成分分析图

(a): 吸收系数; (b): 介电损耗

Fig. 5 Principal component analysis scores plot of black tea and foreign body samples

(a): Absorption coefficient; (b): Dielectric loss

从图 5(a)吸收系数主成分分析可以看出,六种茶叶与异物样品并没有形成良好的聚类效果,其中纯红茶与掺杂蟋蟀躯干的红茶之间难以区分。但是混合 PS、蠕虫、蟋蟀四肢异物的样品与红茶没有相互交错,这与各类样品的吸收光谱特征相符合。在图 5(b)介电损耗主成分分析中可以观察到纯红茶与掺杂蟋蟀躯干的异物样品之间没有明显的区分,而且含有蟋蟀与蟋蟀躯干异物的红茶主成分之间重叠度高,很难区分。对比两项参数的主成分分析,可以发现吸收系数的聚类效果比介电损耗较好<sup>[10]</sup>。为了进一步快速准确识别茶叶中混入的异物种类,采用模式识别方法建立茶叶中异物判别模型。

## 2.3 茶叶夹杂异物定性判别模型

由于红茶与混合各类异物的吸收光谱以及介电损耗很难直接区分它们之间的类型,在实际检测过程中,光谱变量与特性存在的线性关系并不十分明显,因此采用 SVM 和 LDA 两种模型对六类样品进行分类判别。通过观察光谱信号特征,挑选 0.5~1.0 THz 波段信号作为吸收光谱和介电损耗的特征数据。将六种数据随机分配,其中建模集与预测集的比例为 3:1。首先使用 SVM 模型对样品进行分类,其中核函数采用线性(Linear)核函数与径向基(RBF)核函数。惩罚参数  $c$  和核函数参数对 SVM 分类模型运行结果会产生很大

的影响。参数  $c$  过大会降低模型分类效果并且会过度拟合,对分类精度造成影响;反之参数  $c$  太小,则会降低误差经验数值大小。为了进一步优化 SVM 模型中惩罚函数( $c$ )和核函数系数( $g$ )内部参数,采用网络搜索和交叉验证获取最优的参数组合。根据吸收系数与介电损耗两项参数建立的 SVM 定性分析模型结果如表 1 所示,在无预处理状态下,吸收光谱下的 SVM 与线性核函数对茶叶中异物种类识别精度较好,达到了 68.42%。

运用 LDA 模型分类之前,首先需要对原始的吸收光谱和介电损耗进行主成分分析,运用 PCA 降低数据维度并从原始光谱中获取特征光谱。吸收系数谱中前三个主成分的方差贡献率分别为 97.2%, 1.4% 和 0.5%。前三个特征向量累计方差达到了 99.1%,有效的减少了光谱数量,并最大程度保持了原始数据的光谱特征。将三个主成分输入到 LDA 模型中进行分类判别,模型获取了比较好的预测结果,如表 1 所示。LDA 模型中二次判别分析对茶叶中混入的异物类型识别率最高,达到了 73.68%。

从参数分析上看,吸收系数比介电损耗建模精度更高,这一结论与两项参数主成分分析结果相符合,吸收光谱的特征数据更有利于判断茶叶中的异物类型。而对比两个模型,可以发现 LDA 分类模型比 SVM 模型对茶叶和混入异物的

类型识别精度更高。由于茶叶中掺杂的异物类型多,但数据类型模式比较简单。SVM 分类器适合用于将非线性可分的问题映射到更高维的空间,转化为线性可分问题,寻找不同类别间差别最大的向量。SVM 作为一种有监督学习方法,在小样本数据和非线性问题上具有更好的分类和预测。由于该

特征数据非线性相关并不明显,且数据结构简单,所以 SVM 分类效果一般。但是对茶叶中掺杂异物种类的判别, LDA 更适用于线性分类问题,分类精度更好,为茶叶中混入的异物识别提供一种可行性分析方法。

表 1 茶叶与异物分类模型结果

Table 1 Tea and foreign bodies classification model results

Classification method	kernel function	Absorption coefficient		Dielectric loss	
		Accuracy in the calibration set	Accuracy in the prediction set	Accuracy in the calibration set	Accuracy in the prediction set
SVM	RBF kernel	86.37%	65.78%	44.17%	47.37%
	Linear kernel	87.04%	68.42%	58.14%	63.16%
LDA	Quadratic	88.37%	73.68%	83.72%	71.05%
	Linear	76.74%	63.16%	72.09%	68.42%

### 3 结 论

应用 THz 时域光谱技术,在 0.5~1 THz 范围内进行了红茶夹杂昆虫异物的实验研究,实验结果证明了红茶夹杂昆虫异物 THz 光谱检测的可行性。

(1)红茶的 THz 吸收系数、介电损耗系数与昆虫异物均有较为明显的区别,但吸收系数略优于介电损耗系数,其 LDA 定性判别模型的正确识别率达到了 73.68%,建议将吸收系数作为茶叶夹杂昆虫异物 THz 检测的主要特征参数。

(2)外骨骼类昆虫异物和蠕虫类异物与茶叶基质在蛋白质等化学组分上存在较为明显的区别,同时 THz 光谱具有较好的指纹光谱特性,适用于化学性状的表型,但对于物理性状的表型则 THz 图像更具优势,THz 光谱和图像的高效融合能充分发挥光谱和图像的特点,也是茶叶夹杂有机异物检测的重要发展方向。

(3)THz 技术为茶叶以及其他食品基质夹杂有机异物检测提供了一种新的方法,但 THz 光谱数据处理目前仍沿袭一些传统的化学计量学方法,在后续研究中开发面向 THz 光谱的新算法,充分挖掘其有效信息,也是重要的方向。

### References

- [1] Edwards M. Woodhead Publishing Limited, 2004.
- [2] Einarsdóttir H, Emerson M J, Clemmensen L H, et al. Food Control, 2016, 67(1): 39.
- [3] Mery D, Lillo I, Loebel H, et al. Journal of Food Engineering, 2011, 105(3): 485.
- [4] TAO Xue-ming, ZHENG Yu-yan, HONG Deng-hua, et al(陶学明, 郑玉艳, 洪登华, 等). Anhui Agricultural Science Bulletin(安徽农学通报), 2014, (11): 8.
- [5] Jiang Y Y, Ge H Y, Zhang Y, et al. Food Chemistry, 2020, 307(3): 125533.
- [6] Wang Q, Hameed S, Xie L J, et al. Journal of Food Measurement and Characterization, 2020, 14(5): 2453.
- [7] Lee Y K, Chol S W, Han S T, et al. Journal of Food Protection, 2012, 75(1): 179.
- [8] Ok G, Choi S W, Park K H, et al. Sensors, 2013, 13(1): 71.
- [9] Ok G, Park K H, Lim M C, et al. Journal of Food Engineering, 2018, 221(3): 124.
- [10] Shin H J, Choi S W, Ok G. Food Chemistry, 2018, 245(4): 282.
- [11] Jordens C, Koch M. Optical Engineering, 2008, 47(3): 037003.
- [12] Wang C, Zhou R, Huang Y, et al. Food Control, 2019, 97(3): 100.
- [13] Jiang Y Y, Ge H Y, Zhang Y. Optik, 2019, 181(3): 1130.
- [14] Kim K W, Kim K S, Kim H, et al. Optics Express, 2012, 20(9): 9476.
- [15] Qin B, Li Z, Chen T, et al. Optik, 2017, 142(8): 576.
- [16] WANG Jing-rong, ZHANG Zhuo-yong, YANG Yu-ping, et al. Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2017, 37(5): 1606.
- [17] Kulma M, Kourimska L, Homolkova D, et al. Journal of Food Composition and Analysis, 2020, 92(1): 103570.
- [18] Stone A K, Tanaka T, Nickerson M T. Journal of Food Science and Technology, 2019, 56(7): 3355.
- [19] Li M L, Dai G B, Chang T Y, et al. Applied Sciences, 2017, 7(2): 172.

# THz Spectroscopy Detection of Insect Foreign Body Hidden in Tea Products

SUN Xu-dong, LIU Jun-bin

School of Mechatronics and Vehicle Engineering, East China Jiaotong University, Nanchang 330013, China

**Abstract** Chinese black tea is widely respected for its long-term cultural heritage and multiple health benefits. Black tea uses fresh leaves of tea trees as raw materials and is finished by finishing, kneading, drying, screening, and drying. These traditional techniques create a mellow, varied, and rich taste of the famous tea. However, the complex processing technology also increases the risk of physical pollution, and non-tea and foreign tea matter will inevitably be mixed during the processing. Physical contamination is a random event, accounting for 19.8% of food safety complaints. Physical pollution can only be reduced, and it is difficult to eliminate it. It is the key to future disputes between manufacturers and consumers and import and export trade. X-rays use the density difference between the food matrix and foreign objects to detect foreign metal objects effectively and extend to high-density plastics. Organic foreign bodies such as insects are still food foreign bodies that X-rays cannot detect. Insect foreign bodies are mixed with a high frequency, which causes sensory discomfort and easily introduces pathogenic bacteria. There is an urgent need to develop corresponding non-destructive testing methods. As an emerging detection technology, Terahertz time-domain spectroscopy (THz-TDS) technology has good application potential in non-destructive testing of agricultural products, food and medicines. THz has good low-energy transmission and fingerprint spectrum characteristics and has no ionizing radiation damage. It can obtain the spectrum and image information of hidden foreign objects through the food matrix. It is a better choice for non-destructive testing of agricultural products and food. In order to realize the high-efficiency detection of low-density organic foreign matter in tea, this paper explores a new method of the non-destructive detection of black tea-infested insect foreign matter based on THz spectroscopy. In the range of 0.2~3.0 THz, the THz spectra of black tea matrix, insect foreign matter, and black tea mixed with foreign insect matter were collected. The THz absorption coefficient and dielectric loss response characteristics of the tea matrix and insect foreign bodies are analyzed. From the spectrogram, it can be seen that there are significant differences between the THz absorption coefficient and dielectric loss of the tea matrix and the foreign insect bodies, mainly caused by the protein and fat of the foreign insect bodies. The ingredients are caused, laying the foundation for the THz spectrum detection of black tea mixed with insects and foreign bodies. However, the absorption coefficients of tea and insects have no obvious characteristic peaks, and there is more obvious noise in the high-frequency band of 2.0~3.0 THz. The principal component analysis method is used to reduce the dimensionality of the absorption coefficient and the dielectric loss. The score map shows that there is a clear difference between the black tea matrix and the black tea with insect foreign bodies, and the clustering effect of the absorption coefficient is better than that of the dielectric loss factor. The THz absorption coefficient and dielectric loss in the range of 0.5~1.0 THz were selected as input vectors, and support vector machine (SVM) and linear discriminant analysis (LDA) discriminant models were established. The experimental results show that the LDA discrimination model based on the THz absorption coefficient has the highest accuracy, and the correct recognition rate of new samples is 73.68%. It shows that applying THz time-domain spectroscopy for non-destructive detection of black tea inclusions of foreign insect bodies is feasible. At the same time, THz spectrum combined with pattern recognition algorithm provides a new method for non-destructive detection of tea inclusions of foreign insect bodies and can also provide a reference for other agricultural products and food detection.

**Keywords** Terahertz; Spectra; Tea; Foreign body of insects; Discriminant analysis

(Received Jul. 18, 2020; accepted Nov. 9, 2020)