

## 近红外光谱和模式识别的菊花品种无损鉴别

李嘉仪<sup>1</sup>, 余梅<sup>1</sup>, 李脉泉<sup>1</sup>, 郑郁<sup>2\*</sup>, 李跑<sup>1,3\*</sup>

1. 湖南农业大学食品科学技术学院, 湖南长沙 410128
2. 湖南师范大学医学院, 湖南长沙 410013
3. 湖南省农业科学院, 湖南省农产品加工研究所, 湖南长沙 410125

**摘要** 菊花为菊科植物菊的头状花序, 滁菊、贡菊、杭菊和亳菊是常见的几类药用品种菊花。不同品种菊花在外观上具有极大的相似性, 非专业人员仅凭肉眼难以对其进行准确鉴别分析。常规仪器分析法检测成本较高, 分析时间较长, 且需要对样品进行破坏性处理, 影响了产品的二次销售。近红外光谱技术作为近年来快速发展起来的一种绿色、简单、快速的新型检测技术, 在中药鉴别领域取得了很大的进展。基于便携式近红外光谱仪结合化学计量学方法建立了一种菊花品种无损鉴别方法。利用便携式近红外光谱仪采集了滁菊、贡菊、杭菊和亳菊完整以及粉末状两种物理形态样品的光谱, 采用单一以及组合光谱预处理方法消除光谱中存在的干扰, 结合不同模式识别方法(主成分分析法、软独立模式分类法和 Fisher 线性判别分析法)分别构建了不同品种菊花的鉴别模型。结果表明: 由于仪器的限制及样品物理性状的原因, 光谱中存在较为明显的背景、基线漂移以及噪声的干扰, 完整样品由于物理性状的原因, 基线漂移干扰尤为严重; 采用主成分分析法结合光谱预处理方法无法实现不同品种菊花的准确鉴别, 完整样品最佳鉴别正确率仅为 8.33%, 粉末样品最佳鉴别正确率为 52.38%; 通过软独立模式分类法结合预处理方法可以得到较为准确的鉴别结果, 完整样品光谱数据经一阶导数+多元散射校正优化后鉴别正确率为 95%, 粉末状样品数据采用原始数据的鉴别正确率为 92.5%; Fisher 线性判别分析方法结果最佳, 完整样品数据经连续小波变换优化后可以得到 97.5% 的鉴别正确率, 粉末状样品采用原始光谱便可得到 100% 鉴别正确率。以上结果表明, 当采用合适的预处理和建模方法, 完整样品和粉末状样品鉴别结果较为一致, 基于便携式近红外光谱仪结合化学计量学可实现对不同品种菊花的准确无损鉴别分析, 为食药同源产品的无损鉴别分析提供了新途径。

**关键词** 便携式近红外光谱仪; 菊花; 无损鉴别; Fisher 线性判别分析

**中图分类号:** O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)04-1129-05

### 引言

菊花为菊科植物菊的干燥头状花序。滁菊、贡菊、杭菊和亳菊是常见的几类药用菊花品种<sup>[1-2]</sup>。不同品种菊花化学组成如萜类、黄酮、有机酸和生物碱等存在一定差异, 药理作用和功能也有所差别, 但其外观相似度较大, 非专业人员仅通过肉眼难以实现对不同品种菊花的准确鉴别分析。

目前用于菊花品种鉴别的方法主要有感官评价法和仪器分析法。前者通过对外形、气味等进行分析以实现菊花品种的鉴别, 易受主观因素影响, 鉴别正确率较低; 后者是通

过对不同品种菊花的化学组成和含量进行检测以实现鉴别分析。Wang 等<sup>[3]</sup>通过高效液相色谱-二极管阵列-电喷雾串联质谱对 9 个产地菊花的主要化学成分进行了表征, 实现了菊花产地的准确鉴别分析。肖作兵等<sup>[4]</sup>采用气相色谱-质谱/气相色谱-电子鼻对精油特征香气成分的分析实现了不同厂家菊花精油的鉴别。然而此类方法前处理繁杂、检测成本较高、耗时较长, 且对样品具有破坏性, 影响了产品的二次销售。因此, 亟需开发一种不同品种菊花无损鉴别新方法。

近红外光谱的波长介于可见光谱和中红外光谱之间, 有机物中的含氢基团化学键伸缩振动的倍频及合频信息与这一谱区相关<sup>[5]</sup>。然而近红外光谱谱带较宽、吸收峰重叠严重、

收稿日期: 2021-03-01, 修订日期: 2021-05-31

基金项目: 国家自然科学基金项目(31601551, 32001680), 湖南省自然科学基金青年科学基金项目(2019JJ50240), 湖南省教育厅科学研究项目(19C0933), 中国博士后科学基金面上项目(2019M650187)资助

作者简介: 李嘉仪, 1997 年生, 湖南农业大学食品科学技术学院本科生 e-mail: lijy\_li@yeah.net

\* 通讯作者 e-mail: lixiangzheng@sina.com; lipao@mail.nankai.edu.cn

吸收较弱,信息复杂,需要结合化学计量学方法对其进行解析。光谱预处理及其组合算法被提出用于消除光谱中的基线漂移、背景以及噪声干扰<sup>[6-8]</sup>。此外,大量模式识别方法被提出用于挖掘样品的类别信息。无监督模式识别的主成分分析方法(PCA)<sup>[9]</sup>以及有监督的软独立模式分类法(SIMCA)<sup>[10]</sup>和 Fisher 线性判别分析(FLD)<sup>[11]</sup>是最常用的模式识别方法。近红外光谱因具备检测速度快、重现性好、可实现无损检测等优点,近年来在中药品种识别、产地鉴别以及真伪鉴定等领域得到了广泛应用。Li 等<sup>[11]</sup>利用傅里叶台式近红外光谱仪结合 FLD 方法实现了不同年份陈皮的准确鉴别分析。Han 等<sup>[12]</sup>通过傅里叶近红外光谱仪结合偏最小二乘判别分析算法建立了不同品种菊花粉末的鉴别模型。然而该方法依旧无法实现完整菊花的鉴别分析。傅里叶台式近红外光谱仪分辨率高、光通量大,但是抗震性差,较难做成便携式仪器,且价格昂贵。固定光路阵列检测型便携式近红外光谱仪是最常见的便携式近红外仪器,因其价格低、体积小等优点得到了广泛关注,但不同阵列像元间有串扰,仪器的信噪比较低<sup>[13]</sup>。李庆等<sup>[14]</sup>利用云端一互联便携式近红外技术结合化学计量学方法实现了对西红花真伪鉴别和掺伪品掺伪量预测。然而,菊花不同于其他样品,其物理性状不均匀,且便携式近红外信号中易受到较大背景和噪声的干扰,现阶段尚缺乏菊花品种无损鉴别的研究。

因此,提出了一种基于便携式近红外光谱仪结合化学计量学方法的菊花品种无损鉴别方法。通过便携式近红外光谱仪采集不同品种菊花完整和粉末状样品的光谱信息,同时考察物理性状对光谱的影响,基于预处理方法以及模式识别方法建立菊花品种准确鉴别模型。

## 1 实验部分

### 1.1 材料

从本地药店购买滁菊、贡菊、杭菊和亳菊干样各 30 份(直径约为 2.5 cm),共计 120 份完整样品。同时考察样品物理性状对光谱的影响,对每个品种菊花打粉,各制备 30 份,共计 120 份粉末状样品。

### 1.2 仪器与光谱采集

便携式近红外光谱仪(必达泰克光电科技(上海)有限公

司)用于采集不同品种菊花漫反射光谱信息,范围 11 100~6 200  $\text{cm}^{-1}$ ,共采集 433 个数据点。将完整样品以及装有粉末样品的石英样品杯直接置于光斑的中心处且完全覆盖光斑,进行光谱采集。每份样品重复测定 3 次,取平均值作为该样品的原始光谱。

### 1.3 光谱数据分析

使用 MATLAB R2010b(The Mathworks, Natick, USA)软件对光谱数据进行分析。120 份完整、粉末状样品数据根据 Kennard-Stone 方法按照 7:3 的比例分为 84 个校正集和 36 个预测集。采用去偏移(De-bias)、去趋势(DT)、标准正态变量变换(SNV)、多元散射校正(MSC)、最大最小归一化(Min-Max)、一阶导数(1st)、二阶导数(2nd)和连续小波变换(CWT)等 8 个光谱预处理对数据进行优化,CWT 预处理的小波基为“haar”小波基,小波基的尺度参数为 20。为了消除光谱中的多种干扰,考虑 CWT 以及 1st 等求导算法可以有效消除光谱中的背景以及基线漂移干扰,SNV 和 MSC 方法可以消除菊花固体样品带来的光散射干扰,因此采用 1st-DT,1st-SNV,1st-MSC,CWT-SNV 和 CWT-MSC 等 5 种预处理组合对光谱进行优化。为了实现菊花品种的准确鉴别,采用 PCA、SIMCA 和 FLD 方法分别建立鉴别模型。

## 2 结果与讨论

### 2.1 不同品种菊花原始光谱特征

图 1 为不同品种菊花完整和粉末状样品的原始光谱图。样品的光谱存在相同的变化趋势,完整及粉末状样品的光谱均在 8 500  $\text{cm}^{-1}$  处出现较为明显的吸收峰,可能是与 C—H 第三泛频带振动有关,在 7 100  $\text{cm}^{-1}$  出现的吸收峰与 C—H 第二泛频带振动相关。受便携式仪器及样品物理性状的影响,光谱存在明显的噪声干扰。一方面便携式仪器不同阵列像元间有串扰,仪器的信噪比要差一些,另一方面,菊花样品中水分含量较少,近红外吸收强度较弱,导致噪声非常明显。图 1(a)中完整样品光谱中基线漂移极其严重;由于减少了物理性状的干扰,图 1(b)中粉末状样品光谱基线漂移干扰有所缓解。然而,无论是完整样品还是粉末状样品,均无法从原始光谱中发现不同品种菊花的差异信息。

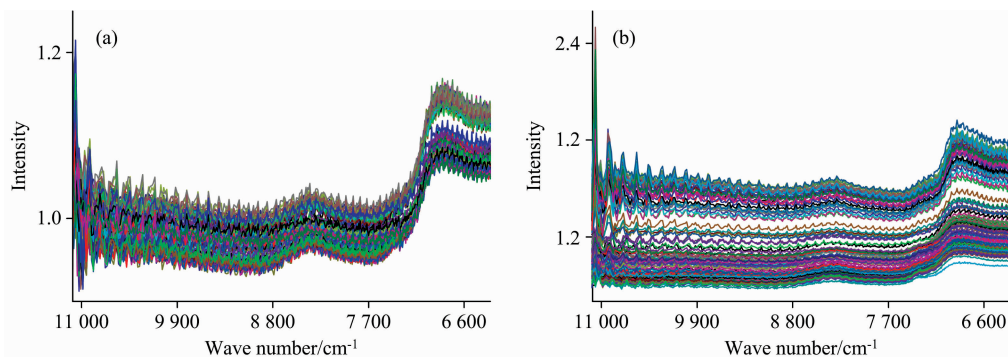


图 1 完整样品(a)和粉末状样品(b)原始光谱图

Fig. 1 Original spectra of the complete (a) and powdered samples (b)

2.2 基于 PCA 的菊花品种鉴别分析

为了实现不同品种菊花的鉴别分析，采用 PCA 方法对光谱数据进行处理。图 2 为 4 个品种菊花完整以及粉末状样品原始光谱的 PCA 结果，红色、蓝色、紫色和绿色图标分别代表滁菊、贡菊、杭菊和亳菊。图中不同形状的实心图标代表校正集，空心图标代表预测集。由于第一主成分(PC1)与第二主成分(PC2)的累计方差贡献率之和大于 90%，因此选用 PC1 及 PC2 绘制 PCA 图。由图 2(a)可知，不同品种完整菊花样品的置信椭圆呈现重叠交织状，校正集和预测集鉴别

正确率均为 8.33%。由图 2(b)可知，由于减少了物理性状的干扰，粉末状样品的鉴别结果优于完整样品的结果，贡菊与杭菊得到了准确鉴别，而亳菊和滁菊的置信椭圆呈现重叠交织状态，校正集鉴别正确率为 52.38%，验证集鉴别正确率为 61.11%。此外，无论是完整样品还是粉末状样品，结合预处理的 PCA 鉴别结果均低于结合原始光谱 PCA 结果。仅利用完整及粉末状样品的原始光谱及优化后的光谱无法实现对不同品种菊花的准确鉴别分析。

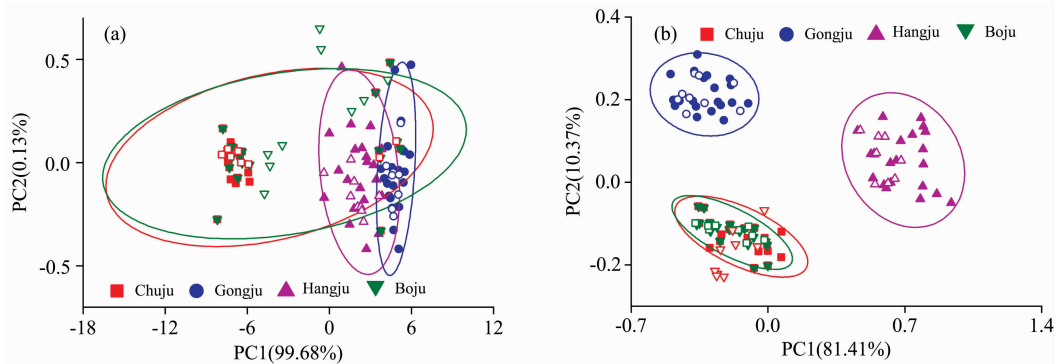


图 2 基于完整样品的 (a) 及粉末状样品 (b) 原始光谱的 PCA 图

Fig. 2 PCA results of the original spectra of complete (a) and powdered sample (b)

2.3 基于 SIMCA 的菊花品种鉴别分析

PCA 方法仅利用方差信息实现数据的分类，鉴别能力较弱；而 SIMCA 方法利用先验信息为每一类样品建立独立的 PCA 模型，从而实现对未知样品的鉴别。表 1 为采用 SIMCA 结合预处理的鉴别结果。与 PCA 模型结果相比，SIMCA 模型的鉴别正确率得到了明显提高，完整样品原始光谱的鉴别正确率为 70%，采用预处理方法优化光谱后，采用 1st-MSC 组合预处理方法可得到最佳鉴别结果，鉴别正确率为 95%。粉末状样品原始光谱数据的鉴别正确率为 92.5%，采用预处理方法优化数据后，其鉴别正确率有所下降，可能原因是预处理方法在扣除干扰信息同时，也扣除了隐藏在光谱中的菊花品种差异信息。

2.4 基于 FLD 的菊花品种鉴别分析

采用 FLD 模式识别方法对光谱数据进行处理。表 1 为 FLD 结合预处理方法后的鉴别结果。FLD 模型鉴别结果优于 PCA 和 SIMCA 模型结果。采用完整样品原始光谱数据便可获得 87.5% 的鉴别正确率。采用预处理方法优化光谱数据后，采用 CWT 预处理可获得最佳的鉴别结果，鉴别正确率为 97.5%。在 FLD 方法的帮助下，粉末状样品采用原始光谱数据便可实现 100% 的鉴别分析，采用 De-bias, DT 和 1st 预处理同样可以获得 100% 鉴别正确率。图 3 为不同品种菊花的 FLD 结果。结果表明：与 PCA 和 SIMCA 模型结果相比，FLD 方法获得的菊花品种鉴别结果最佳，且当采用合适的预处理和建模方法，完整和粉末状样品鉴别结果较为一致，表明该方法可以实现菊花品种的无损鉴别分析。

表 1 不同预处理方法结合 SIMCA 和 FLD 方法的鉴别正确率

Table 1 Identification accuracies obtained by SIMCA/FLD and pretreatment methods

Dataset	Pretreatment	SIMCA/%					FLD/%				
		Chu	Gong	Hang	Bo	Whole	Chu	Gong	Hang	Bo	Whole
Complete	Original	100	50	60	70	70	60	100	100	90	87.5
	De-bias	100	0	70	50	55	70	100	90	80	85
	DT	90	10	60	60	55	70	100	90	90	87.5
	SNV	90	10	60	70	57.5	80	100	80	90	87.5
	Min-Max	80	60	60	100	75	60	100	90	90	85
	MSC	90	10	60	40	50	70	100	90	90	87.5
	1st	100	80	70	60	77.5	80	100	100	100	95
	2nd	100	0	60	40	50	60	100	100	90	87.5
	CWT	100	70	70	50	72.5	90	100	100	100	97.5
	1st-DT	100	80	70	50	75	70	100	100	100	92.5

续表 1

	1st-SNV	100	90	100	60	87.5	90	100	100	70	90
	1st-MSC	100	90	100	90	95	80	90	100	90	90
	CWT-MSC	100	90	90	90	92.5	70	90	100	90	87.5
	CWT-SNV	90	90	100	90	92.5	80	90	100	80	87.5
Powdered	Original	70	100	100	100	92.5	100	100	100	100	100
	De-bias	70	80	100	100	87.5	100	100	100	100	100
	DT	70	70	100	80	80	100	100	100	100	100
	SNV	70	90	20	70	62.5	100	100	100	90	97.5
	Min-Max	60	80	80	90	77.5	100	100	100	90	97.5
	MSC	70	90	20	70	62.5	100	100	100	90	97.5
	1st	60	100	100	100	90	100	100	100	100	100
	2nd	70	30	60	50	52.5	80	80	70	50	70
	CWT	60	90	90	90	82.5	90	100	100	100	97.5
	1st-DT	50	90	90	100	82.5	100	100	100	90	97.5
	1st-SNV	70	50	20	70	52.5	70	80	80	80	77.5
	1st-MSC	70	60	20	80	57.5	70	80	80	70	75
	CWT-MSC	80	60	20	60	55	70	80	80	90	80
	CWT-SNV	70	60	10	60	50	60	80	80	80	75

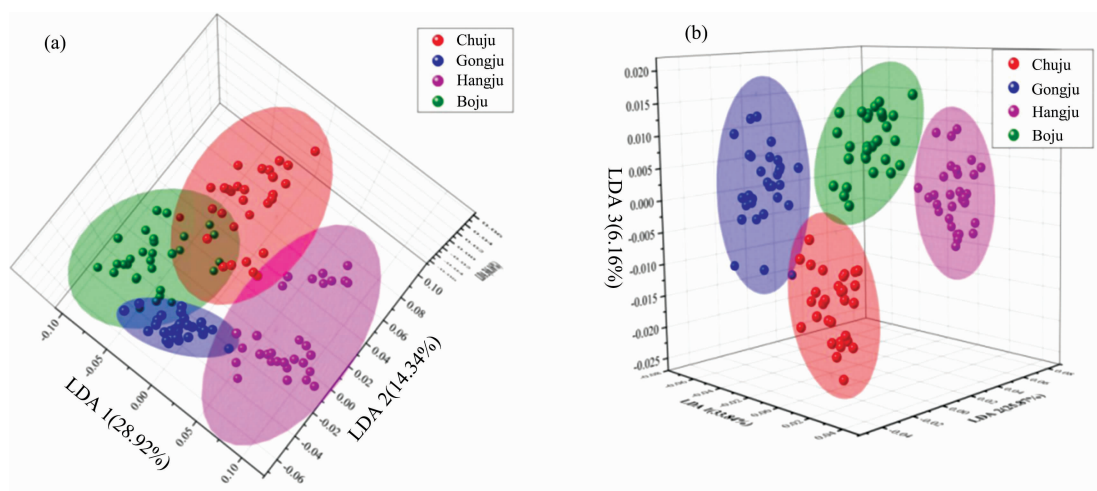


图 3 完整样品 CWT-FLD 结果 (a), 粉末状样品原始光谱 FLD 结果 (b)

Fig. 3 CWT-FLD result of the complete samples (a), FLD result of original spectra of the powdered samples (b)

### 3 结 论

基于便携式近红外光谱仪结合化学计量学方法建立了一种菊花品种无损鉴别方法。结果表明：与 PCA 和 SIMCA 模型结果相比，采用 FLD 方法可以达到更好的鉴别结果，且当

采用合适的预处理和建模方法，完整样品和粉末状样品 FLD 模型鉴别结果较为一致，完整样品光谱数据经 CWT 预处理后的鉴别正确率为 97.5%，粉末状样品采用原始光谱即可获得 100% 鉴别正确率。以上结果表明，便携式近红外光谱仪结合化学计量学方法可以实现菊花品种的无损鉴别分析，为食药同源产品的无损鉴别分析提供了新途径。

### References

- [1] Li Y F, Yang P Y, Luo Y H, et al. Food Chemistry, 2019, 286: 8.
- [2] Chen S, Liu J, Dong G Q, et al. Food Chemistry, 2021, 344: 128733.
- [3] Wang S, Hao L J, Zhu J J, et al. Food Analytical Methods, 2015, 8(1): 40.
- [4] XIAO Zuo-bing, FANG Bin-bin, NIU Yun-wei, et al(肖作兵, 范彬彬, 牛云蔚, 等). Journal of Chinese Institute of Food Science and Technology(中国食品学报), 2017, 17(12): 287.
- [5] CHU Xiao-li, SHI Yun-ying, CHEN Pu, et al(褚小立, 史云颖, 陈 瀑, 等). Journal of Instrumental Analysis(分析测试学报), 2019,

38(5): 603.

- [6] Han X, Huang Z X, Chen X D, et al. *Fuel*, 2017, 207: 146.
- [7] Han X, Tan X, Huang Z X, et al. *Analytical Methods*, 2017, 9(24): 3720.
- [8] Bian X H, Wang K Y, Tan E X, et al. *Chemometrics and Intelligent Laboratory Systems*, 2020, 197(2): 103916.
- [9] Li P, Du G R, Cai W S, Shao X G. *Journal of Pharmaceutical and Biomedical Analysis*, 2012, 70: 288.
- [10] Pomerantsev A L, Rodionova O Y. *Journal of Chemometrics*, 2020, 34(8): e3250
- [11] Li P, Zhang X X, Li S K, et al. *Sensors*, 2020, 20(6): 1586.
- [12] Chen C W, Yan H, Han B X. *Revista Brasileira De Farmacognosia*, 2014, 24(1): 33.
- [13] Baca-Bocanegra B, Hernández-Hierro J M, Nogales-Bueno J, et al. *Talanta*, 2019, 192: 353.
- [14] LI Qing, YAN Xiao-jian, ZHAO Kui, et al(李庆, 闫晓剑, 赵魁, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2020, 40(10): 3029.

## Nondestructive Identification of Different Chrysanthemum Varieties Based on Near-Infrared Spectroscopy and Pattern Recognition Methods

LI Jia-yi<sup>1</sup>, YU Mei<sup>1</sup>, LI Mai-quan<sup>1</sup>, ZHENG Yu<sup>2\*</sup>, LI Pao<sup>1,3\*</sup>

1. College of Food Science and Technology, Hunan Agricultural University, Changsha 410128, China

2. School of Medicine, Hunan Normal University, Changsha 410013, China

3. Hunan Agricultural Product Processing Institute, Hunan Academy of Agricultural Sciences, Changsha 410125, China

**Abstract** Chrysanthemum is derived from the capitulum of Chrysanthemum. Chuju, Gongju, Hangju and Boju are common medicinal chrysanthemums. Different chrysanthemum varieties have great similarities in appearance, and it is difficult for laypeople to identify them accurately only by naked eyes. The conventional instrumental analysis method has the disadvantages of high detection cost, long analysis time, and destructive treatment of samples, which affects the secondary sales of the products. As a green, simple and rapid detection technology, near-infrared spectroscopy has made great progress in traditional Chinese medicine identification. This study established a nondestructive identification method of different Chrysanthemum varieties based on portable near-infrared spectrometer and chemometric methods. The spectra of complete and powder samples of Chuju, Gongju, Hangju and Boju were collected by grating portable near-infrared spectrometer. The single and combined spectral pretreatment methods were used to eliminate the interferences in the spectra. The identification models of different Chrysanthemum varieties were constructed by combining principal component analysis, soft independent modeling of class analogy and Fisher linear discriminant analysis methods. The results show that: due to the restrictions of the current measure instruments and the difference of sample particle size and distribution, there are obvious interferences of background, baseline drift and noise in the spectra. The baseline drift interference is particularly serious for the analysis of the complete samples. The principal component analysis combined with spectral pretreatment methods could not identify different varieties of chrysanthemum. The best identification accuracy of complete samples was only 8.33%, and that of powder samples was 52.38%. The soft independent modeling of class analogy can obtain more accurate identification results with preprocessing methods. The identification accuracy of complete sample data is 95% with first derivative+multiple scattering correction, while the identification accuracy of powder sample data is 92.5% with the original data. The results of Fisher linear discriminant analysis are the best. When the complete sample spectra were optimized by continuous wavelet transform, the identification accuracy was 97.5%. When the original spectra of powder samples were used, the identification accuracy could reach 100%. The above results show that the complete and powder samples' identification results are consistent when the appropriate pretreatment and modeling methods are used. Based on the grating portable near-infrared spectrometer combined with chemometrics methods, the accurate identification of different Chrysanthemum varieties can be realized, which provides a new way for the nondestructive identification of food and drug homologous products.

**Keywords** Portable near infrared spectrometer; Chrysanthemum; Nondestructive identification; Fisher linear discriminant analysis

(Received Mar. 1, 2021; accepted May 31, 2021)

\* Corresponding authors