

基于 FTIR 离散平稳小波特征提取的 SVM 应用于中药材紫花地丁与同属植物的鉴别研究

黄有强(华东医药股份有限公司, 杭州 310009)

摘要: 目的 探索基于 FTIR 离散平稳小波变换结合支持向量机(support vector machine, SVM)分类法的中药紫花地丁的质量控制新模式。方法 采用衰减全反射傅里叶变换红外光谱法直接快速测定中药紫花地丁与同属植物多花堇菜和戟叶堇菜的 FTIR, 运用基于离散平稳小波变换进行特征向量的提取, 通过分析比较后选取第 4、5 层分解层的特征向量用于支持向量机的训练与验证。结果 通过对不同产地的 90 个样本的验证, 紫花地丁与同属植物多花堇菜和戟叶堇菜的识别率达 100%。结论 基于 FTIR 离散平稳小波变换结合支持向量机分类法的中药紫花地丁与同属植物多花堇菜和戟叶堇菜的分类鉴别方法具有非常好的效果。

关键词: 紫花地丁; 傅里叶变换红外光谱; 离散平稳小波; 支持向量机; 模式识别; 多花堇菜; 戟叶堇菜

中图分类号: R284.1

文献标志码: B

文章编号: 1007-7693(2017)05-0692-05

DOI: 10.13748/j.cnki.issn1007-7693.2017.05.012

引用本文: 黄有强. 基于 FTIR 离散平稳小波特征提取的 SVM 应用于中药材紫花地丁与同属植物的鉴别研究[J]. 中国现代应用药学, 2017, 34(5): 692-696.

Recognition Method Research of *Violae Herba* and Its Sibling Plants Based on FTIR-DSWT and SVM Classification Method

HUANG Youqiang(*Huadong Medicine Co., Ltd., Hangzhou 310009, China*)

ABSTRACT: OBJECTIVE To explore quality control pattern of new traditional Chinese medicine *Violae Herba* based on attenuated total reflection-Fourier transform infrared(FTIR) spectroscopy-discrete stationary wavelet transform(DSWT) combined with support vector machine(SVM). **METHODS** FTIR of *Violae Herba* and its sibling plants *Viola pseudo-monbeigii* Chang and *Viola betonicifolia* J. E. Smith were obtained directly, quickly and accurately by Fourier transform infrared spectrometer with attenuated total reflection accessory. DSWT was used to extrude local region of FTIR of *Violae Herba*, *Viola pseudo-monbeigii* Chang and *Viola betonicifolia* J. E. Smith. After comparison analysis, the d4 and d5 scale were used to extract the feature vectors, which were used to train and test the SVM. **RESULTS** According to 90 testing samples of the different production from different places, the recognition rate of *Violae Herba* and its sibling plants were all 100%. **CONCLUSION** The experimental results show that it is effective to apply DSWT combined with SVM on the basis of FTIR to identify the *Violae Herba* and its sibling plants *Viola pseudo-monbeigii* .Chang and *Viola betonicifolia* J. E. Smith.

KEY WORDS: *Violae Herba*; FTIR; DSWT; SVM; pattern recognition; *Viola pseudo-monbeigii* Chang; *Viola betonicifolia* J. E. Smith

堇菜属植物在我国约有 110 余种, 大多数种类用作园林观赏植物, 同时也因其具有清热解毒及较强的止血、散瘀、消肿作用而作为一中药材使用。该属植物目前仅有紫花地丁已被收入中国药典^[1]。堇菜属植物药用历史悠久, 但品种繁杂难以辨认。药检部门在日常对中药材进行检查的过程中, 常会发现非正品多花堇菜与戟叶堇菜混杂在紫花地丁中。传统中药讲究临床用药的质量, 由此产生了多种真伪、优劣的鉴定手段^[2]。中药传统经验中常用的有看、摸、尝、闻等, 紫

花地丁在药典标准中采用的是显微鉴别法及薄层色谱后确认荧光主斑点的位置来进行确认所含物质差异程度^[1]。紫花地丁、多花堇菜与戟叶堇菜这 3 种植物所含化学成分, 到目前为止仅见紫花地丁有报道^[3-4]。中药质量分析研究涉及多方面、多学科的综合研究课题, 采用现代仪器分析法, 建立既能反映中药材整体固有特征, 又能给出稳定的中药质量评价标准的鉴定方法, 以解决传统中药学与当代世界药学接轨中面临的质量标准问题。因此, 中药的质量控制和评价一直都是中医

作者简介: 黄有强, 男, 主管中药师 Tel: 13957113957 E-mail: HuangYouQiang2008@163.com

药研究的热点和难点^[5-6]。傅里叶变换红外光谱分析法(Fourier transform infrared spectroscopy, FTIR)能给出复合体系几乎所有物质的结构信息,与化学计量学相结合在中药分析及其他领域得到了较为广泛的应用^[7-9]。支持向量机(support vector machine, SVM)是十大经典机器学习算法之一,它是一种监督式学习的方法,属于一般化线性分类器,也被称为最大边缘区分类器,其特点是能够同时达到经验误差最小化和几何边缘区最大化。因此, SVM 常应用于分类及回归分析^[10-11]。本研究旨在借助金刚石衰减全反射附件,采用 FTIR 中的粉末直接测定法对紫花地丁与同属植物多花堇菜和戟叶堇菜进行红外光谱分析的基础上,进行离散平稳小波变换(discrete stationary wavelet transform, DSWT)后提取特征向量,最后再进行 SVM 的分类鉴别,探究中药质量控制的新模式。

1 DSWT 算法基本原理

DSWT 也称平移不变小波变换,它是一种非正交的小波变换。从矩阵的观点来看, DSWT 可视为矩阵乘法。

$$G=Sg \tag{1}$$

式中, g 是一个 $1 \times N$ 的输入矢量; S 是一个 $(L+1)N \times N$ 矩阵, 其中 S_i 是一个 $N \times N$ 的方阵, S_i 的各列是单个矢量 S_i 的周期平移的结果, S_i 通常是 DSWT 在第 i 个尺度下的基, 而 S_{L+1} 是最粗尺度下的尺度函数。

由于 DSWT 具有冗余性, 因而具有多种逆变换, Guo 等^[12-13]给出了如下逆变换形式:

$$S^{-1} = \left[\frac{1}{2} S_1, \frac{1}{2^2} S_2, \dots, \frac{1}{2^L} S_L, \frac{1}{2^{L+1}} S_{L+1} \right] \tag{2}$$

式中, 常量因子矢量 $(1/2, 1/2^2, \dots, 1/2^L, 1/2^{L+1})$ 是随着尺度变得越来越粗糙时所需增加的 DSWT 的冗余度的偏移量。直接计算式(1)需要的计算复杂度为 $O[(L+1)N^2]$, 但是由于存在快速算法, 因而总共所需的计算复杂度最多为 $O(N \log N)$ 。

DSWT 的最大特点是冗余性和平移不变性, 能对连续小波变换给出一个更为近似的估计。信号经 DSWT 时不对信号本身进行下抽样处理, 而是在每 2 个滤波器系数间插入零值来实现滤波器的延展, 因此, 变换后的逼近信号和细节信号长度都与原信号长度相同, 可以有效避免由于信号下抽样但小波基不具有平移不变性而造成重构信号产生 Gibbs 震荡的问题, 因而更适合于处理相关

性问题。

2 材料与方法

2.1 仪器与试剂

NEXUS 670 型傅里叶变换红外光谱仪(美国尼高力仪器公司), DTGS 检测器, OMNI E.S.P.5.1 智能操作软件, 金刚石 ATR 附件, 扫描光谱范围设置为 $4\,000 \sim 650\text{ cm}^{-1}$, 分辨率 2 cm^{-1} , 扫描累加次数 32 次。

实验所用试剂样本信息见表 1。

表 1 试验样本信息

Tab. 1 The information of test samples

样本名称	产地	训练样本数	检验样本数
紫花地丁 <i>Viola yedoensis</i> Makino	浙江金华	20	10
	贵州贵阳	20	10
	山东青岛	20	10
多花堇菜 <i>Viola pseudo-monbeigii</i> Chang	浙江金华	20	10
	贵州贵阳	20	10
	山东青岛	20	10
戟叶堇菜 <i>Viola betonicifolia</i> J. E. Smith	浙江金华	20	10
	贵州贵阳	20	10
	山东青岛	20	10

所有试剂均为 10 月采收, 鲜采的样本经洗净, 除去杂质, 晒干, 不经任何化学处理。试验选取采自于浙江金华、贵州贵阳和山东青岛的紫花地丁及其同属植物多花堇菜、戟叶堇菜作为分析对象, 为了保证样本的代表性和试验的客观性, 在每个地区每种植物分别选取了 30 个样本, 每种植物的样本总数为 90 个, 共计 270 个样本。所有样本均经过浙江师范大学化学与生命科学学院植物学学科陈建华教授鉴定。考虑到药用中药紫花地丁为全株, 故实验过程中也取每份样品的全株作为分析对象, 干燥样品于玛瑙研钵中研磨, 至样品为均匀细小粉末状。将粉末状样品密封, 待测。

2.2 方法

在采集数据前, 按照仪器测试要求把金刚石 ATR 附件置于傅里叶变换红外光谱仪的样品仓中, 且调整到平整水平位置。分别取各样品按所给定的测试条件直接测定样品的 ATR-FTIR, 为了减小测定的误差, 每个样品测定前均对背景进行扫描, 得到的图谱基线采用自动校正法进行校正。

2.3 数据处理

根据各种样本的 FTIR 数据, 以 $4\,000 \sim 650\text{ cm}^{-1}$ 内的红外光谱吸光度值为指标, 以 3 种堇菜属植

物为对象,制作成 Excel 表格,保存为 xls 格式。考虑到样品主要吸收峰出现在 $2\,000\sim 650\text{ cm}^{-1}$,因此进行 DSWT 处理时选取 $2\,000\sim 650\text{ cm}^{-1}$ 内的红外光谱数据。把所有数据导入 Matlab 7.0 软件进行 DSWT,从而得到特征向量值;SVM 算法和建模,采用 Matlab 7.0 软件进行样本的回归估计,以紫花地丁、多花堇菜和戟叶堇菜为实验样本,实验中各产地各种植物选择 30 个,其中训练样本数选择 180 个,另外 90 个作为检验样本。

3 结果与分析

3.1 红外光谱分析

以采自于浙江金华的紫花地丁、多花堇菜和戟叶堇菜为例,三者的典型 FTIR 见图 1。红外光谱初步解析如下:3 种堇菜属植物的峰位非常一致,在 $3\,300\text{ cm}^{-1}$ 有一很强的吸收峰,是 N-H 和 O-H 的伸缩振动吸收,主要由萜类、碳水化合物如纤维素、甾醇类化合物及蛋白质等含有羟基的化合物产生;吸收带位于 $3\,000\text{ cm}^{-1}$ 右侧的是样品中所含物质的饱和 C-H 键的伸缩振动吸收,主要是由纤维素和脂肪等分子中的饱和 C-H 键所引起^[14];在 $1\,735\text{ cm}^{-1}$ 有较强的尖锐吸收峰,为酯羰基的伸缩振动吸收峰,三者比较而言,紫花地丁的吸收峰是最弱的,多花堇菜和戟叶堇菜的吸收峰较强,可能紫花地丁中酯类化合物含量相对较少所造成。在 $1\,100\text{ cm}^{-1}$ 附近的指纹区,包括了 C-O 键的伸缩振动,堇菜属植物内所含有的碳水化合物(如纤维素)主要形成了这些吸收带。C-H 弯曲振动吸收位于 $1\,500\text{ cm}^{-1}$ 到 $1\,200\text{ cm}^{-1}$ 。这些结构信息为鉴别这 3 种植物提供了一定的依据。

对于 3 种植物的区别,仅仅从 FTIR 图的吸收峰形来判断显然难以区别,因此本研究采用小波变换进行 FTIR 吸收信息的放大处理。考虑到高波数段的 FTIR 吸收主要是 OH 或 NH 键的伸缩振动吸收,和 $3\,000\text{ cm}^{-1}$ 左右的 C-H 键伸缩振动吸收,从吸收情况来看来不是很特征,三者差别最大的应该是包括指纹区以内的官能团吸收区域 $650\sim 2\,000\text{ cm}^{-1}$,所以本研究选择 $650\sim 2\,000\text{ cm}^{-1}$ 进行离散小波变换分析。

3.2 FTIR-DSWT 分析

DSWT 的逆变换有多种形式,本研究采用 Lang 等^[15]给出的逆变换形式。DSWT 通过 Matlab 7.0 软件来实现,在进行离散小波分解时,应该根

据信号的光谱特性选择适当的小波基函数和分解层数。在小波多分辨率分解过程中,根据 IR 信号的特性,并比较不同分辨率下信号分解的效果,确定合适的小波基及小波尺度。其标准是突出原始光谱中的若干个特征峰,并选取平滑性好的小波基。常用的小波基有 Mexican hat, Meyer, Morlet, Daubechies, Coiflet 及 Symlets 等,以小波基形状与待分析信号形状是否更接近及衰减信号是否更快作为选择的前提,经比较分析,选取了 Daubechies 小波作为“分析小波”。对 3 种样本的 FTIR 进行 DSWT(在 6 尺度下的小波系数),结果见图 2。

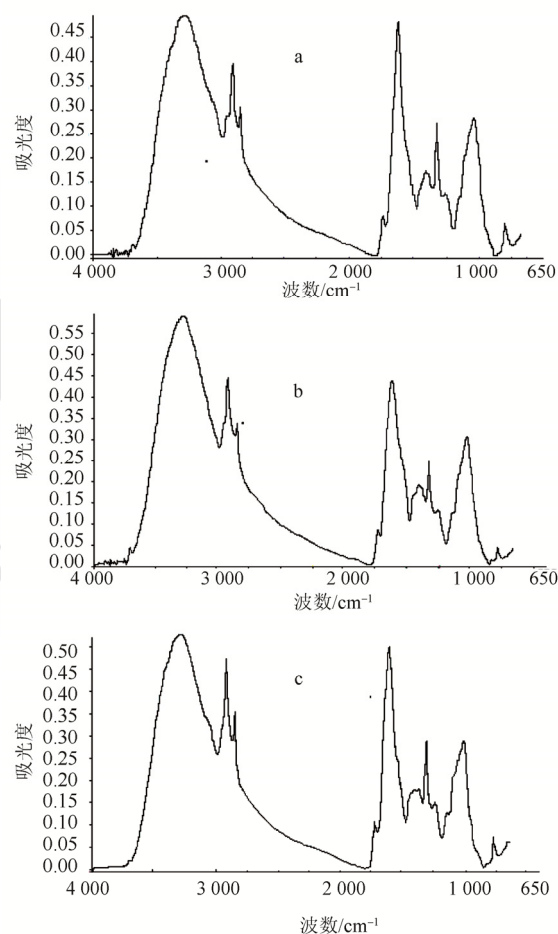


图 1 3 种样品的 FTIR

a-紫花地丁; b-多花堇菜; c-戟叶堇菜。

Fig. 1 FTIR of three kinds of samples

a-*Viola yedoensis* Makino; b-*Viola pseudo-monbeigii* Chang; c-*Viola betonicifolia* J. E. Smith.

3.3 特征提取

由图 2 可见,当 DSWT 的分解尺度比较小的时候,细节信号对光谱变化比较敏感,对原始光

谱中各特征吸收峰反应过于强烈,而尺度太大时又过于平缓,均不利于特征的提取。基于此,本研究选择 d4 和 d5 这 2 个尺度的 DSWT 细节信号作为特征变量提取空间。特征变量的定义为 DSWT 信号的 d4 和 d5 这 2 个尺度下共 4 个区域的光谱能量,即小波系数平方和,DSWT 细节信号特征区间划分结果见图 3。

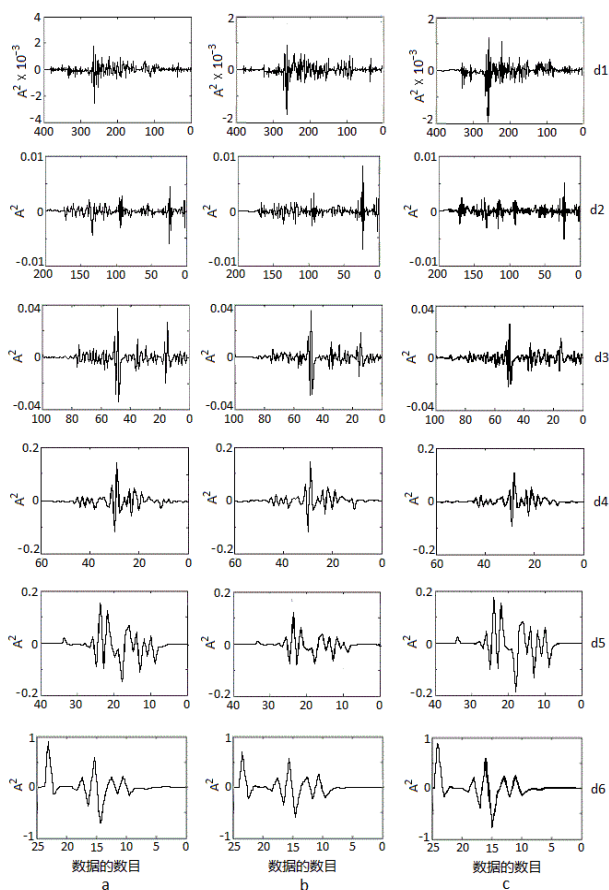


图 2 DSWT 结果

a-紫花地丁; b-多花堇菜; c-戟叶堇菜。

Fig. 2 The result of DSWT of samples

a-*Viola yedoensis* Makino; b-*Viola pseudo-monbeigii* Chang; c-*Viola betonicifolia* J. E. Smith.

3.4 模式识别

SVM 是 Burges 等^[16]根据统计学理论中结构风险最小化原则提出的一种新的机器学习算法,以训练误差作为优化问题的约束条件,以置信范围值最小化作为优化目标,即采用结构风险最小化准则的学习方法,使模型具有良好的泛化能力。

SVM 采用函数来将非线性分类问题转化为高维空间中的线性问题,其常用的核函数有多项式函数、径向基函数和 Sigmoid 函数。本研究过程中

对 3 种核函数均进行了学习分类,结果发现径向基函数作为核函数时,识别率最高。

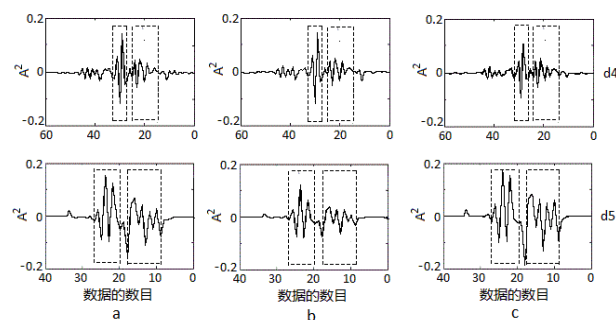


图 3 DSWT 细节信号特征区间划分示意图

a-紫花地丁; b-多花堇菜; c-戟叶堇菜。

Fig. 3 Division of feature region in the DSWT domain

a-*Viola yedoensis* Makino; b-*Viola pseudo-monbeigii* Chang; c-*Viola betonicifolia* J. E. Smith.

3.5 研究实例

把采自于浙江金华、贵州贵阳和山东青岛的紫花地丁及其同属植物多花堇菜、戟叶堇菜作为分析对象,在 ATR-FTIR 测定的基础上进行 DSWT,继续进行特征向量的提取,每个样本得到 4 个特征值,经过 SVM 的学习训练和识别。由于用的是二分类器,最终输出的结果为各个样本是或者不是紫花地丁的种类归属。模式识别过程见图 4。

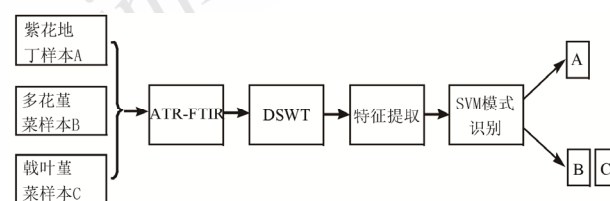


图 4 模式识别过程

Fig. 4 The process of pattern recognition

为了对紫花地丁与其同属植物多花堇菜、戟叶堇菜的分类试验中从 FTIR 经过离散平稳小波提取到特征向量进行有效性验证,本研究先选择各产地的 180 个样本即各产地植物选取 20 个样本进行学习训练。然后选择 90 个样本作为验证样本,即各产地各种植物选取 10 个样本。当采用径向基函数作为核函数时, SVM 模式识别率为 100%。

4 结语

由于 ATR-FTIR 测试只需要中药样本进行粉碎即可测得样本的 FTIR 图谱,而且能得到中药材所含所有物质的结构信息,比溶剂提取后浸膏加

KBr 压片法测定样品的 FTIR 进行鉴别或样品粉末加 KBr 压片法即所谓的无损 FTIR 鉴别等质量控制有一定的优势。在 FTIR 进行离散平稳小波特征区间的能量积分基础上,采用 SVM 对紫花地丁与其同属植物多花堇菜、戟叶堇菜进行分类,结果与传统中药鉴定和植物分类结果完全一致,结果表明了此种分类鉴别方法具有非常好的效果。此类方法具有准确和客观性,可用以补充和完善现有的中药质量控制手段和方法。

REFERENCES

- [1] 中国药典. 一部[S]. 2015: 337-338.
- [2] CHEN K L, HUANG L F, LIU Y M. Development history of methodology of Chinese medicines' authentication [J]. China J Chin Mater Med(中国中药杂志), 2014, 39(7): 1203-1208.
- [3] ZHOU H Y, QIN M J, HONG J L, et al. Chemical constituents of *Viola yedoensis* [J]. Chin J Nat Med(中国天然药物), 2009, 7(4): 290-292.
- [4] ZHAO Z Z, CHEN H B, XIAO P G. Inheritance and innovation of traditional Chinese medicinal authentication [J]. China J Chin Mater Med(中国中药杂志), 2015, 40(17): 3385-3390.
- [5] SUN Y F, DU L L, ZHOU L, et al. Study on antibacterial active components from *Viola yedoensis* [J]. China J Chin Mater Med(中国中药杂志), 2011, 36(19): 2666-2671.
- [6] CHEN S L, GUO B L, ZHANG G J, et al. Advances of studies on new technology and method for identifying traditional Chinese medicinal materials [J]. China J Chin Mater Med(中国中药杂志), 2012, 37(8): 1043-1054.
- [7] CHENG C G, TIAN Y M, ZHANG C J. Research of recognition method between Semen Cuscutae and its sibling plant Japanese dodder seed based on FTIR-CWT and ANN classification method [J]. Acta Chim Sin(化学学报), 2008, 66(7): 793-798.
- [8] CHENG C G, LIU J, ZHANG C J, et al. An overview of infrared spectroscopy based on continuous wavelet transform combined with machine learning algorithms: application to Chinese medicines, plant classification and cancer diagnosis [J]. Appl Spectrosc Rev, 2010, 45(2): 148-164.
- [9] CHENG C G, LIU J, WANG H, et al. Infrared spectroscopic studies of Chinese medicines [J]. Appl Spectrosc Rev, 2010, 45(3): 165-178.
- [10] TAO M L, GU W T, HOU K H, et al. Exploration of the "internal and external combined" new TCM quality control mode based on support vector machines [J]. Chin Pharm(中国药房), 2016, 27(1): 118-121.
- [11] CHENG C G, XIONG W, JIN W Y. Recognition of *Equisetum avense* L. and *Hippochaete remosissima* (Desf.) Boerner. based on Fourier transform infrared spectrometry combined with chemometrics [J]. Chin J Anal Chem(分析化学), 2009, 37(5): 676-680.
- [12] ZHANG C J, FU M Y, ZHANG Q H, et al. Infrared image de-noising based on discrete stationary wavelet transform [J]. Opt Tech(光学技术), 2003, 29(2): 250-256.
- [13] LANG M, GUO H, EDEGEND J, et al. Nonlinear processing of a shift-invariant DWT for noise reduction [C]. Spies Symposium on Oe/aerospace Sensing & Dual Use Photonics, 2001, 2491(1): 640-651.
- [14] CHENG C G, LIU J, CAO W Q, et al. Classification of two species of *Bidens* based on ftir-discrete stationary wavelet transform extraction combined with probability neural network [J]. Vib Spectrosc, 2010, 54(1): 50-55.
- [15] LANG M, GUO H, EDEGEND J E, et al. Nonlinear processing of a shift-invariant discrete wavelet transform (DWT) for noise reduction [C]. Proceedings of SPIE, 1995, 2491(1): 640-651.
- [16] BURGESS C J C. A tutorial on support vector machines for pattern recognition [J]. Data Min Know Dis, 1998, 2(2): 121-167.

收稿日期: 2016-10-07

(本文责编: 李艳芳)