

# ATR-FTIR 及人工神经网络分类法鉴别薏苡仁与伪品草珠子种仁

吕惠玲<sup>1\*</sup>, 万茶艳<sup>2</sup>, 程存归<sup>2</sup>(1. 金华市中心医院药剂科, 浙江 金华 321000; 2. 浙江师范大学化学与生命科学学院, 浙江 金华 321004)

中图分类号 R284.1 文献标志码 A 文章编号 1001-0408(2015)06-0845-03

DOI 10.6039/j.issn.1001-0408.2015.06.44

**摘要** 目的: 建立快速鉴别及评价薏苡仁及其伪品草珠子种仁的方法。方法: 采用衰减全反射傅里叶变换红外光谱(ATR-FTIR)法测定薏苡仁与草珠子种仁的红外光谱, 选择4个特征吸收峰所对应的吸光度值作为特征向量, 建立人工神经网络(ANN)模型, 对不同薏苡仁与草珠子种仁样本所得红外光谱的特征向量进行分类。结果: 所建立的方法对薏苡仁与草珠子种仁的识别率均为100%。结论: 该方法判别准确度高, 可用于薏苡仁与草珠子种仁的快速鉴别。

**关键词** 薏苡仁; 草珠子种仁; 衰减全反射傅里叶变换红外光谱法; 人工神经网络; 鉴别

## Identification between *Coicis Semen* and *Coix lachrymal-jobi* by ATR-FTIR and ANN Classification

LYU Hui-ling<sup>1</sup>, WAN Cha-yan<sup>2</sup>, CHENG Cun-gui<sup>2</sup>(1. Dept. of Pharmacy, Jinhua Municipal Central Hospital, Zhejiang Jinhua 321000, China; 2. College of Chemistry and Life Science, Zhejiang Normal University, Zhejiang Jinhua 321004, China)

**ABSTRACT** OBJECTIVE: To establish the method for rapid identification and evaluation of *Coicis Semen* and *Coix lachrymal-jobi*. METHODS: The infrared spectrum of *Coicis Semen* and *C. lachrymal-jobi* were identified by ATR-FTIR. Using absorbance value of 4 characteristic absorption peaks as feature vector, artificial neural network (ANN) model was established to classify feature vector of infrared spectrum of *Coicis Semen* and *C. lachrymal-jobi*. RESULTS: Established method was highly accurate, and the correct rates of *Coicis Semen* and *C. lachrymal-jobi* identification were both 100%. CONCLUSIONS: The method has high accuracy, and can be used for rapid identification of *Coicis Semen* and *C. lachrymal-jobi*.

**KEYWORDS** *Coicis Semen*; *Coix lachrymal-jobi*; ATR-FTIR; Artificial neural network; Identification

薏苡仁又名薏仁、苡米、苡仁、土玉米、薏米、起实、薏珠子、草珠珠、回回米、米仁、六谷子, 其性微寒, 味甘、淡, 具有利水消肿、健脾去湿、舒筋除痹、清热排脓等功效, 为常用利水渗湿药<sup>[1]</sup>。同科属植物草珠子种仁的形态与薏苡仁相似, 为常见伪品<sup>[2]</sup>。

傅里叶变换红外光谱(FTIR)法由于能给出复合体系的所有物质的信息, 因而在中药鉴别方面得到了较多的应用<sup>[3-7]</sup>。由于同科属植物种子所含化学成分也较为相近, 采用一维红外光谱分析法较难得到理想的效果<sup>[8]</sup>。本研究采用衰减全反射傅里叶变换红外光谱(ATR-FTIR)法测定了中药材薏苡仁及其伪品草珠子种仁的红外光谱, 并尝试选择特征吸收峰所对应的吸光度构建特征矢量, 然后运用反向传播人工神经网络(ANN)进行分类, 以达到薏苡仁与草珠子种仁鉴别之目的。

## 1 材料

### 1.1 仪器

NEXUS 670 型 ATR-FTIR 仪, 包括 DTGS 检测器、金刚石

ATR、OMNIC E.S.P.5.1 智能操作软件(美国 Nicolet 公司)。

### 1.2 药材

所有样品均由浙江省金华市食品药品检定所中药科提供, 并经过该所李冰岚主任中药师及浙江师范大学生物系陈建华教授鉴定。薏苡仁为禾本科植物薏苡 *Coix lacryma-jobi* L. 的干燥成熟种仁; 草珠子种仁为禾本科植物草珠子 *C. lachrymal-jobi* L. 的干燥成熟种仁。

## 2 方法

在采集数据前, 按仪器测试要求把金刚石 ATR 附件水平放置于 FTIR 仪的样品仓中, 分别取种仁样品适量在粉碎机中粉碎后, 混合均匀后再从中取少量置于玛瑙研钵上磨成细小均匀的粉末, 过 200 目筛, 测定样品的 FTIR。为了降低测定误差, 图谱基线采用自动校正法进行校正。光谱范围为 4 000~650  $\text{cm}^{-1}$ , 分辨率为 2  $\text{cm}^{-1}$ , 扫描累加次数为 32 次。从样品的 FTIR 图谱中选择 4 个特征吸收峰对应的吸光度值作为特征向量。在进行 ANN 识别时, 以薏苡仁和草珠子种仁的 FTIR 作为样本, 薏苡仁和草珠子种仁的训练数及预测样本数各选择

究近况[J]. 药学进展, 2009, 33(12):542.

[6] 杜薇, 邹巧根, 孙莉莉, 等. 色谱及其联用技术在药物杂质

\* 中药师。研究方向: 中药质量分析。E-mail: 416078926@qq.com

分析中的应用[J]. 海峡药学, 2013, 25(5):1.

[7] 张立峰, 刘灿仿. 离子对 RP-HPLC 法测定利塞膦酸钠原料药及其制剂的含量[J]. 中国药房, 2013, 24(41):3 912.

(收稿日期: 2014-07-25 修回日期: 2014-11-28)

(编辑: 陈宏)

100个。样本经过FTIR测定后,以所选择的特征向量值作为特征信息进行ANN训练与预测。样品的FTIR图谱见图1。

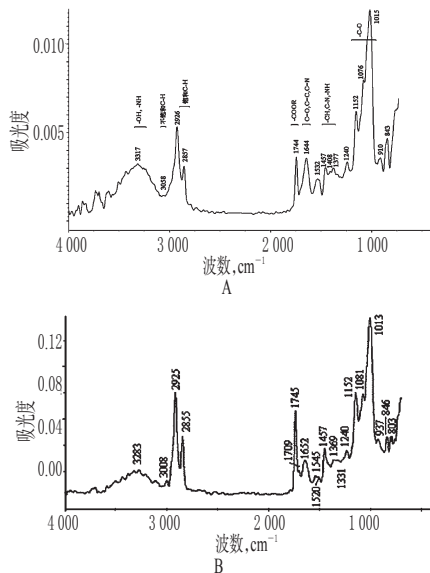


图1 样品的FTIR图谱

A.薏苡仁;B.草珠子种仁

Fig 1 FTIR of samples

A.Coicis Semen;B. *C. lachrymal-jobi*

### 3 结果与分析

#### 3.1 FTIR图谱分析

由图1可以看出,薏苡仁和草珠子种仁均在 $3\ 300\ \text{cm}^{-1}$ 左右及 $1\ 000\sim 1\ 200\ \text{cm}^{-1}$ 之间出现了碳水化合物的强吸收峰,说明两者均含有较多的淀粉等碳水化合物。薏苡仁在 $3\ 058\ \text{cm}^{-1}$ 出现了吸收,说明含有较多的末端烯烃类物质;而草珠子种仁则在 $3\ 008\ \text{cm}^{-1}$ 出现了吸收,说明含有较多的非末端烯烃类物质;在 $1\ 745\ \text{cm}^{-1}$ 左右两者均有吸收,说明两者均含有酯类物质,但相对含量则有所不同。这表明薏苡仁、草珠子种仁的FTIR存在着一定的差异性。

#### 3.2 特征向量的选择

在进行ANN分类时,需用样品的特征向量值作为ANN的输入值。考虑到本研究的样品为种仁,除去所含淀粉的—OH伸缩振动及C—O键的伸缩振动外,本研究选择酯羰基伸缩振动吸收峰 $1\ 745\ \text{cm}^{-1}$ 、C=C伸缩振动吸收峰 $1\ 652\ \text{cm}^{-1}$ 、C—H弯曲振动吸收峰之一的 $1\ 457\ \text{cm}^{-1}$ 和C—N键的伸缩振动吸收峰 $1\ 331\ \text{cm}^{-1}$ 所对应的吸光度值构建特征向量。因此,输入层共有4个特征向量值。

#### 3.3 误差反向传播(BP)算法

本研究采用ANN分类中的前馈型神经网络(BPNN)模型<sup>[9-10]</sup>及随机输入的方式对薏苡仁及草珠子种仁进行分类。

对于BPNN,当各节点均采用Sigmoid型函数时,1个隐含层就足以解决任意判决分类问题。ANN计算流程见图2。

第一阶段如图2(a)所示,在选择网络模型和学习规则后,根据已知的输入和输出学习数据(也称作目标输出数据),通过网络的学习训练,固定神经网络节点的权值和节点阈值,然后比较ANN输出的数据与目标输出数据间的误差,不断微调网络的权值和阈值,直到学习样本的输出与目标输出数据的

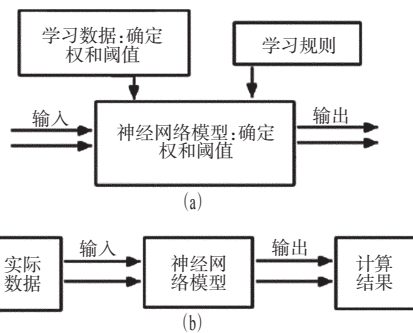


图2 ANN计算流程

Fig 2 The artificial neural network calculation flow chart

误差在允许范围内为止。第二阶段如图2(b)所示,根据经第一阶段固定的模型,输入实际数据后,输出计算结果:首先进行数据预处理,把两种同属植物种子依次对应输出层的2个节点,然后对FTIR小波特征值的9个特征向量分别归一化为0~1之间的数据。

总的来说,分类算法分两个步骤进行,即BP网络的训练和利用训练后的BP网络进行不同种类的分类。BP算法流程图见图3。

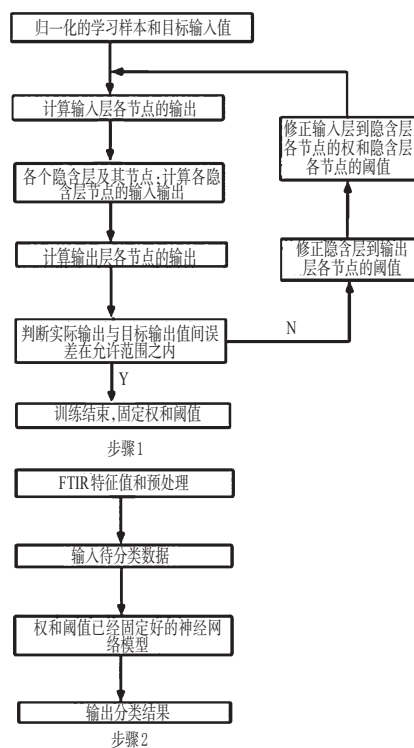


图3 BP算法流程

Fig 3 BP algorithm flow process chart

本研究的计算过程分为两个步骤。步骤1是网络训练过程,对BP网络的各层进行设计,具体如下,输入层:输入归一化的4个特征向量;输出层:由于是用于分类,输出层的每一节点对应1种样品,本研究共2个;中间隐层:对于BP网络,所取的隐层节点数少时,容错性就差。但是,隐层节点数过多易引起学习时间过长,误差也不一定最佳,因此需要多次试验后选取一个最佳的隐层节点数。

步骤2是利用步骤1所得到的网络模型,对归一化的4个特征向量进行计算。激活函数采用 Sigmoid 函数,通过学习不断修正权值和阈值,使对应输入样本 $p$ 的最小二乘误差达到最小。计算最小二乘误差函数的公式为:

$$E_p = 0.5 \times \sum (t_{pj} - O_{pj})^2 \dots \dots \dots (I)$$

其中, $t_{pj}$ 为样本 $p$ 在输出层节点 $j$ 的目标输出值,即样品种类; $O_{pj}$ 为样本 $p$ 在输出层节点 $j$ 的实际输出值。计算实际输出值是从输入层到输出层的方向进行,而误差和权值的调整方向是从输出层到输入层。

(1)计算各节点 $j$ 的输出值 $O_{pj}$ (当节点 $j$ 是输入层的节点,则其输出值=输入值)的公式为:

$$O_{pj} = f(\text{net}_{pj}) = 1 / (1 + e^{-\sum w_{ij} O_{pi} - \theta_j}) \dots \dots \dots (II)$$

其中, $w_{ij}$ 是连接节点 $i$ 和节点 $j$ 的权, $\theta_j$ 是节点 $j$ 的阈值,阈值可认为是从输出恒为1的节点到其他节点间连接的权值,与 $w_{ij}$ 的学习调整过程一样。权值 $w_{ij}$ 的学习调整方法如下。

(2)当 $j$ 是输出层节点时,通过下式修正连接隐层节点 $i$ 到输出层节点 $j$ 的权值 $w_{ij}$ :

$$\Delta w_{ij}(n+1) = \eta \delta_{pj} O_{pi} + \alpha \Delta w_{ij}(n) \dots \dots \dots (III)$$

其中, $\eta$ 为学习速率, $\alpha$ 为动量项, $\delta_{pj}$ 是输出层节点 $j$ 的误差信号, $\delta_{pj}$ 的计算公式为:

$$\delta_{pj} = (t_{pj} - O_{pj}) O_{pj} (1 - O_{pj}) \dots \dots \dots (IV)$$

(3)当 $j$ 不是输出层节点时,也通过上式修正连接隐层节点 $i$ 到 $j$ 的权值 $w_{ij}$ ,但 $\delta_{pj}$ 的计算变为:

$$\delta_{pj} = O_{pj} (1 - O_{pj}) \sum \delta_{pk} w_{kj} \dots \dots \dots (V)$$

其中, $\delta_{pk}$ 是输入来自节点 $j$ 的输出值的节点 $k$ 的误差信息, $w_{kj}$ 是连接节点 $j$ 到 $k$ 的权值。

### 3.4 网络确定与应用结果

经过试验,本研究确定的BP网络结构为4个输入层、6个隐层和2个输出层,误差为0.05, $\alpha$ 为0.8, $\eta$ 为0.02。

薏苡仁和草珠子种仁样本的训练集和测试集各选取200个,训练数及预测样本数各选100个。样本经过FTIR测定后,所选择的特征向量值作为特征信息进行ANN训练与预测。本研究采用的BP网络结构中共有4个输入层节点,依次归一化为0~1之间的4个特征向量,输出层节点1类为薏苡仁,2类为草珠子种仁。训练后的网络模型对薏苡仁与草珠子种仁样品的FTIR经过小波变换后的特征向量数据的验证结果均为100%,表明ANN分类法对薏苡仁与草珠子种仁的判别准确率非常高。

## 4 结论

采用ATR-FTIR法可以避免压片法或溶剂溶解后用液膜法所产生的一些人为因素影响。对于所得到的每个样品的

FTIR图谱,可以选择比较有特征的吸收峰所对应的吸光度值作为特征向量值。本研究中每个样品均选择样品所含物质官能团的酯羰基伸缩振动吸收峰 $1745 \text{ cm}^{-1}$ 、C=C伸缩振动吸收峰 $1652 \text{ cm}^{-1}$ 、C—H弯曲振动吸收峰之一的 $1457 \text{ cm}^{-1}$ 和C—N键的伸缩振动吸收峰 $1331 \text{ cm}^{-1}$ 所对应的吸光度值构建特征向量,作为ANN的输入值。通过ANN的训练,再用训练出来的网络模型对薏苡仁与草珠子种仁所得FTIR小波提取的特征向量进行分类。由本研究结果来看,其判别准确率很高。此方法可用于薏苡仁与草珠子种仁的鉴别。

## 参考文献

- [1] 国家药典委员会. 中华人民共和国药典:一部[S]. 2010年版. 北京:中国医药科技出版社, 2010: 353.
- [2] 马翠荣, 赵建波, 唐培英. 茜草、薏苡仁及其伪品的鉴别[J]. 中国药业, 2007, 16(14): 57.
- [3] 金文英, 程路遥. 连续小波变换 HATR-FTIR 分析应用于中药枸杞子与同属非正品鉴别的研究[J]. 浙江师范大学学报:自然科学版, 2008, 31(3): 312.
- [4] 洪庆红, 成则丰, 李群力. FTIR 聚类分析法应用于菟丝子真伪的鉴别[J]. 光谱学与光谱分析, 2008, 28(8): 1 803.
- [5] 黄美爱, 陈晓龙, 程存归. 傅里叶变换红外光谱-排序法鉴别土茯苓真伪的研究[J]. 医药导报, 2008, 27(7): 775.
- [6] 孙跃宗, 程存归, 李冰岚, 等. 麦冬及其非正品的傅里叶变换红外光谱鉴别[J]. 医药导报, 2004, 23(11): 813.
- [7] 李冰岚, 成则丰, 寿文虹, 等. 傅里叶变换红外光谱法直接测定并结合主成分分析鉴定白芍及其伪品的研究[J]. 医药导报, 2007, 26(1): 71.
- [8] Cheng CG, Liu J, Zhang CJ, 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]. *Applied Spectroscopy Reviews*, 2010, 45 (2) : 148.
- [9] Cheng CG, Yu P, Wu CS, et al. Recognition of MIR data of semen armeniacae amarum and semen persicae using discrete wavelet transformation and BP-artificial neural network[J]. *Journal of Spectroscopy*, 2012, 27(4): 253.
- [10] 程存归, 田玉梅, 张长江. 基于FTIR-CWT及ANN分类法的中药材菟丝子与同科属金灯藤子的识别方法研究[J]. 化学学报, 2008, 66(7): 793.

(收稿日期:2014-07-12 修回日期:2014-12-12)

(编辑:孙冰)