

电子鼻融合 BP 神经网络鉴别生、醋广西莪术 及姜黄素类成分的含量预测

蓝振威¹, 季德², 王淑美¹, 陆兔林², 孟江^{1*}

(1. 广东药科大学 中药学院 国家中医药管理局中药数字化质量评价技术重点实验室 广东高校(省) 中药质量工程技术研究中心, 广东 广州 510006; 2. 南京中医药大学 药学院, 江苏 南京 210023)

[摘要] 为建立一种快速准确识别生、醋广西莪术并预测其姜黄素类化合物含量的科学评价方法,采用一整套仿生学的识别模式,通过电子鼻获取莪术及其醋制品的数字化气味信号,应用反向传播(back propagation, BP)神经网络算法分析数据,以精确度、敏感性和特异性指标评估判别模型,以相关系数和均方误差评估回归模型。实验结果表明,通过 BP 神经网络算法建立的电子鼻信号判别模型在训练集、校正集和预测集中的 3 个指标均为 100%,明显优于传统的决策树、朴素贝叶斯、支持向量机、K 最邻近和集成分类,能准确地区分生、醋广西莪术;回归模型预测集相关系数和均方误差分别为 0.974 8 和 0.117 5,能很好地预测广西莪术中姜黄素类化合物含量,展示了模仿生物识别模式在中药分析中的优越性。电子鼻气味指纹图谱结合 BP 神经网络算法,快速、便捷、准确地实现了的判别和回归,这提示了可以有更多的仿生学信息获取和识别模式在中药领域中联合使用,为中药质量的快速评价和标准化提供了思路和方法。

[关键词] 广西莪术; 醋莪术; 电子鼻; BP 神经网络; 姜黄素; 含量预测

Identification of curcumin content in raw and vinegar-processed rhizomes of *Curcuma kwangsiensis* based on electronic nose combined with back propagation neural network

LAN Zhen-wei¹, JI De², WANG Shu-mei¹, LU Tu-lin², MENG Jiang^{1*}

(1. Engineering Technology Research Center for Chinese Materia Medica Quality of Universities in Guangdong Province, Key Laboratory of Digital Quality Evaluation of Chinese Materia Medica, State Administration of Traditional Chinese Medicine, School of Traditional Chinese Medicine, Guangdong Pharmaceutical University, Guangzhou 510006, China;
2. School of Pharmacy, Nanjing University of Chinese Medicine, Nanjing 210023, China)

[Abstract] This study aimed to establish a rapid and accurate method for identification of raw and vinegar-processed rhizomes of *Curcuma kwangsiensis*, in order to predict the content of curcumin compounds for scientific evaluation. A complete set of bionics recognition mode was adopted. The digital odor signal of raw and vinegar-processed rhizomes of *Curcuma kwangsiensis* were obtained by e-nose, and analyzed by back propagation (BP) neural network algorithm, with the accuracy, the sensitivity and specificity in discriminant model, correlation coefficient as well as the mean square error in regression model as the evaluation indexes. The experimental results showed that the three indexes of the e-nose signal discrimination model established by the neural network algorithm were 100% in training set, correction set and prediction set, which were obviously better than the traditional decision tree, naive bayes, support vector machine, K nearest neighbor and boost classification, and could accurately differentiate the raw and vinegar products. Correlation

[收稿日期] 2020-03-25

[基金项目] 2016 年度新兴产业重大工程中药标准化项目 (ZYBZH-Y-SC-40); 广东省教育厅广东药科大学创新强校工程项目 (2016KTSCX064, 2018KZDXM040); 广州市科技局项目 (201707010170)

[通信作者] * 孟江, E-mail: jiangmeng666@126.com

[作者简介] 蓝振威, 硕士研究生, E-mail: 1046517519@qq.com

coefficient and mean square error of the regression model in prediction set were 0.974 8 and 0.117 5 respectively, and could well predict curcumin compounds content in *Curcuma kwangsiensis*, and demonstrate the superiority of the simulation biometrics model in the analysis of traditional Chinese medicine. By BP neural network algorithm, e-nose odor fingerprint could quickly, conveniently and accurately realize the discrimination and regression, which suggested that more bionics information acquisition and identification patterns could be combined in the field of traditional Chinese medicine, so as to provide ideas and methods for the rapid evaluation and standardization of the quality of traditional Chinese medicine.

[Key words] *Curcuma kwangsiensis*; vinegar-processed rhizomes of *Curcuma kwangsiensis*; electronic nose; BP neural network; curcumin; content prediction

doi:10.19540/j.cnki.cjmm.20200523.304

莪术是姜科植物蓬莪术 *Curcuma phaeocalis* Val.、广西莪术 *C. kwangsiensis* S. G. Lee et C. F. Liang 或温郁金 *C. wenyujin* Y. H. Chen et C. Ling 的干燥根茎,具有行气破血、消积止痛的功效^[1]。莪术生品行气止痛,破血祛瘀力强,为气中血药,用于癥瘕痞块,瘀血经闭,胸痹心痛,食积胀满^[2]。醋莪术为莪术的醋制品,主入肝经血分,散瘀止痛作用增强^[3]。2015年版《中国药典》一部收录了含有莪术的中成药制剂共38种,其中明确要求使用醋莪术的有16种,因此,莪术及其炮制品的准确区分对临床精准用药和复方的制备至关重要。当前对于莪术和醋莪术的鉴别方法主要有性状鉴别^[4]、显微鉴别、理化鉴别^[5]及分子鉴别^[6],然而性状鉴别存在主观性、模糊性,显微鉴定和成分测定操作繁琐、耗时费力,使得中药材快速准确的质量控制一直难以实现。

气味是中药鉴别的重要特征,大多数中药都具有固定的气味,2015年版《中国药典》对莪术的气味特征描述为“辛,微香”,醋莪术描述为“微有醋香气”,该标准反映了莪术和醋莪术的气味不同,但该标准过于主观和模糊,故需要一种可以量化气味指标的方法,实现通过气味差异对中药进行快速准确的鉴别。“辛香”是莪术非常具有特色的气味指标,姜黄素类成分是其来源之一,该类成分不仅是一种天然色素,同时也是一种具有多种生理活性的药效物质,现代研究表明,姜黄素类化合物具有降血脂,抗肿瘤,抗氧化和抗炎等作用^[7],因此,姜黄素类化合物的含量对莪术饮片及其醋制品的气味和质量评价具有重要意义。

电子鼻是一种仿生学检测仪器,通过模拟生物的嗅觉系统获得气味指纹图谱,实现对气味信号的量化和客观分析,能够数字化、客观化地表征中药气味^[8],在中药的气味识别中已有广泛应用^[9-12];人工神经网络(artificial neural network, ANN)是一种模

仿生物神经网络行为特征进行分布式并行信息处理的数学模型,反向传播(back propagation, BP)神经网络是一种按照误差逆向传播算法训练的多层前馈神经网络,是应用最高广泛的神经网络算法之一,对各种复杂的模式分类和多维函数映射都有优良的能力^[13],已经实现了快速准确的假酒识别^[14]、星系分类^[15]、食品质量评价^[16]、疾病预测^[17]等。

本文以广西莪术为主要来源,基于莪术及其醋制品炮制的气味差异,通过便携式电子鼻检测器获取气味信息,得到样品的气味指纹图谱。再通过BP神经网络模拟生物识别模式对莪术及其醋制品进行鉴别,并对药材中的3种姜黄素类化合物的含量进行预测,实现二者的快速鉴别和姜黄素类化合物含量的快速预测,为莪术的客观气味评价提供参考。

1 材料

便携式电子鼻PEN3(Airsense Analytics, Schererin, 德国)由10个金属氧化物半导体组成传感器阵列,传感器详细信息见表1,配有数据处理软件WinMuster。摇摆式高速粉碎机(DFY-400D),1/10万电子天平(CPA225D,德国Sartorius公司)。

表1 PEN3便携式电子鼻传感器检测范围

Table 1 Detection range of PEN3 electronic nose sensors

No.	传感器	检测范围
1	W1C	芳香有机化合物
2	W5S	灵敏度大,对氮氧化物灵敏
3	W3C	氨水和芳香成分
4	W6S	主要检测氢气
5	W5C	烷烃,芳香化合物和非极性化合物
6	W1S	对甲烷敏感,有机化合物检测范围广
7	W1W	对硫化物敏感
8	W2S	对乙醇和部分芳香化合物敏感
9	W2W	芳香化合物和硫化物
10	W3S	高浓度甲烷和脂肪族化合物

实验用莪术饮片购买于全国不同的饮片厂,经广东药科大学刘基柱教授鉴定为姜科植物广西莪术的干燥根茎,来源信息见表 2。将各

批次的莪术饮片取一半样品,按照《中国药典》醋煮法(通则 0213)煮至透心,取出,稍凉,干燥。

表 2 30 批广西莪术炮制前后饮片信息

Table 2 Information of 30 batches of rhizomes of *Curcuma kwangsiensis* before and after processing

生品	炮制品	购买单位	生品	炮制品	购买单位
CR1	PCR1	玉林本草堂中药饮片厂	CR16	PCR16	广西玉林市和谐药业
CR2	PCR2	玉林本草堂中药饮片厂	CR17	PCR17	广西玉林市和谐药业
CR3	PCR3	康美中药饮片厂	CR18	PCR18	广西玉林市和谐药业
CR4	PCR4	康美中药饮片厂	CR19	PCR19	广西玉林市和谐药业
CR5	PCR5	康美中药饮片厂	CR20	PCR20	广西玉林市和谐药业
CR6	PCR6	江西国康中药饮片厂	CR21	PCR21	广西玉林市和谐药业
CR7	PCR7	康美中药饮片厂	CR22	PCR22	广西玉林市和谐药业
CR8	PCR8	广东和翔制药有限公司	CR23	PCR23	广西玉林市和谐药业
CR9	PCR9	玉林本草堂中药饮片厂	CR24	PCR24	亳州永刚饮片厂
CR10	PCR10	玉林本草堂中药饮片厂	CR25	PCR25	亳州中亚饮片厂
CR11	PCR11	亳州永刚饮片厂	CR26	PCR26	康美中药饮片厂
CR12	PCR12	北京同仁堂饮片有限公司	CR27	PCR27	信远中药科技有限公司
CR13	PCR13	河北金叶子药业	CR28	PCR28	瑞博中药饮片
CR14	PCR14	广西玉林市和谐药业	CR29	PCR29	天诚中药饮片厂
CR15	PCR15	广西玉林市和谐药业	CR30	PCR30	中强中药饮片有限公司

2 方法与结果

2.1 气味指纹图谱获取

2.1.1 检测参数 电子鼻载气为经活性炭清洗装置处理的自然空气,检测室温调整至 25 ℃,静置 30 min,生成稳定的顶空气体后开始测量,采样时间和清洗时间均为 120 s,进样针气流为 300 mL·min⁻¹。

2.1.2 样品制备 取样品粉碎过 4 号筛。称取 20 g 生、醋莪术粉末装入 100 mL 顶空进样瓶中,加薄膜密封,每个样品平行测定 6 次取平均值,获取的气味指纹图谱以 PCR1 为例,见图 1。本文以 10 个传感器的最大响应值为分析指标建立 BP 神经网络判别和回归模型。

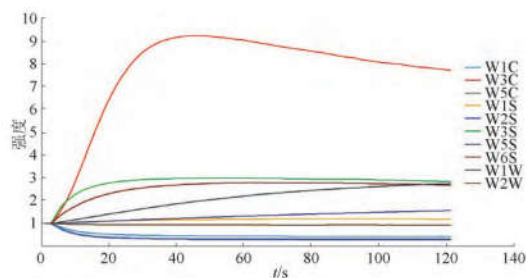


图 1 PCR1 气味指纹图谱

Fig. 1 Odor fingerprint of PCR1

2.2 BP 神经网络算法

与电子鼻相同,神经网络算法的灵感来源于动物的中枢神经系统,神经网络的工作流程见图 2,其中隐藏层是最重要的部分,当数据被输入后,从每个神经元流入下一层对应的神经元中,然后在隐藏层经简单的函数运算并传递,最后到达输出层进行处理。这种仿生的算法只需从提供的数据中建立输入与输出的映射关系,经过有限次迭代计算而获得一个反映实验数据内在规律的数学模型,非常适合解决多维非线性系统问题。BP 神经网络的信息在“神经元”中传播的误差通过反向传播训练多层前馈网络,利用梯度搜索技术使输出值与期望值的误差均方差最小^[18]。更具体地说,误差的输出是按照图 2 方向向前传播,而权值和阈值的调整则是通过相反的传播路径,从而将误差分配给各层所有单元,作为权值的调整依据,并通过误差的梯度下降调整节点之间的阈值^[13]。

2.3 莪术及其炮制品的识别

将电子鼻 10 个感应器的最大响应值数据输入神经网络模型的输入层,将生品莪术和醋制莪术分为两类作为模型的目标。故输入层神经元个数为 10,输出层神经元个数为 2。实验中 60 批样品的电子鼻数据被随机分配 42 批用于模型的训练,其中生

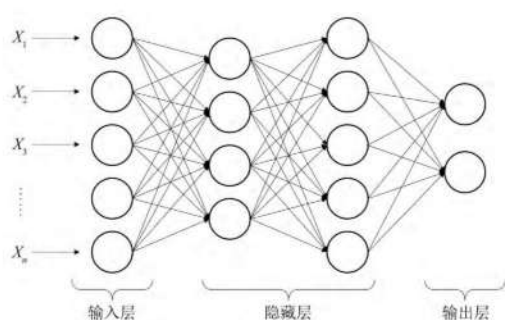


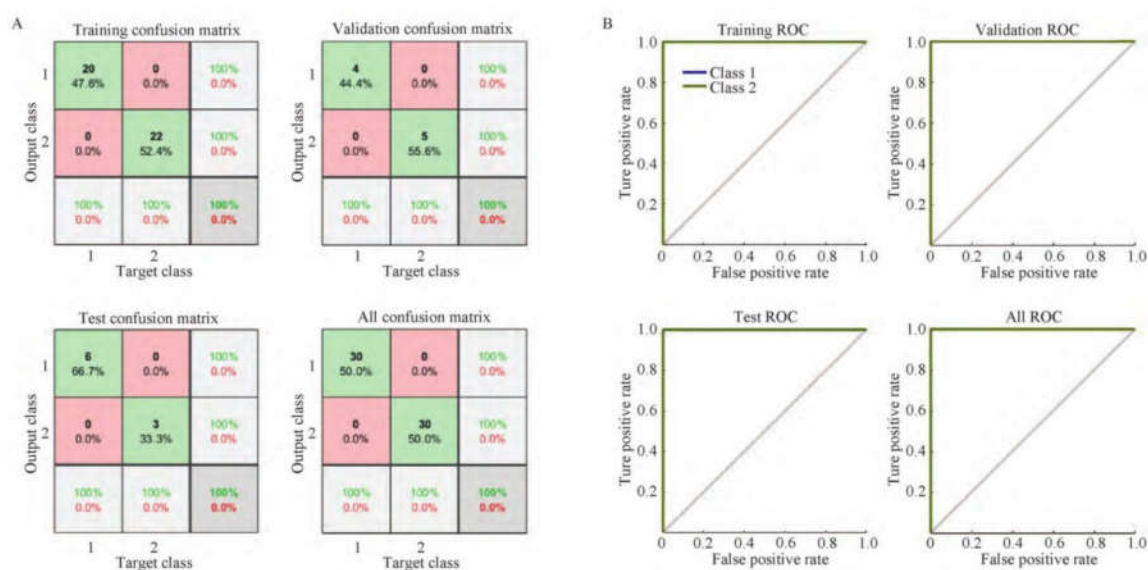
图2 神经网络工作流程

Fig. 2 Neural network workflow

品 20 批, 醋制品 22 批; 9 批样品的电子鼻数据用于模型学习的校正, 含有生品 4 批, 醋制品 5 批; 测试集含有 6 批生品和 3 批醋制品, 共 9 批。神经网络算法经 Matlab 2019b 中的 Deep Learning Toolbox 13.0 工具包实现, 经试凑法确定隐藏的神经元数量为 12, 采用学习率自适应的梯度下降 BP 算法对训练

集进行学习, 迭代次数设置为 1 000, 使用交叉熵计算代价, 确认检查值默认为 6, 即当交叉熵误差连续增加 6 次时停止训练, 保留误差最小的模型。经过 36 次迭代后确定最优模型的学习率为 5.0287×10^{-4} 。

建立的模型在训练集、校正集、预测集和全集中的正确率和 ROC 曲线下面积见图 3, 测试集的表现更能说明模型的性能, 其中图 3A 中 1 代表生品, 2 代表醋制品, 横坐标代表目标分类, 纵坐标代表经过建立的模型对输入层运算得到的输出类, 同类相交的点代表正确判断的样本数, 异类相交的点表示错判的样本数。可见测试集中模型输出的分类结果和真实类别是完全一致的, 判断准确率均为 100%; 图 3B 为模型的 ROC 曲线, 其中 Class1 代表生品, Class2 代表醋品, 横坐标表示虚报概率, 纵坐标表示击中概率, 测试集中两类样品的曲线下面积均为 1。上述结果说明利用神经网络算法建立的判别模型可以实现对莪术饮片生品及其醋制品的精准鉴别。



A. 混淆矩阵; B. ROC 曲线。

图3 BP神经网络判别模型训练集,校正集,预测集和全集

Fig. 3 Results of training set, correction set, prediction set and universal set of BP neural network discriminant model

2.4 BP-ANN,DTM,NBM,SVM,KNN,Boost 模型比较

为了证明模仿人和动物中枢神经系统建立的BP人工神经网络模型(BP-ANN)能更好地对仿生的电子鼻信息进行学习和决策, 选择比较了决策树(decision tree, DTM), 朴素贝叶斯(naive bayes model, NBM), 支持向量机(support vector machine,

SVM), K最近邻(K-nearest neighbor, KNN)和集成分类(boost)这5种在机器学习和分类中广泛应用的算法进行模型验证。采用精确度(accuracy, AC)、敏感性(sensitivity, SE)、特异性(specificity, SP)和ROC曲线下面积(area under curve, AUC)4个指标来评估模型的识别能力。其中3个评价指标

的计算公式如下。

$$AC = TP + TN / TP + FP + FN + TN \quad (1)$$

$$SE = TP / TP + FN \quad (2)$$

$$SP = TN / FP + TN \quad (3)$$

式中 TP 为真阳性样本数, TN 为真阴性样本数, FP 为假阳性样本数, FN 为真阴性样本数。AC 表示样本预测真实结果的百分比, SE 表示识别真实阳性样本的可能性。SP 表示识别真正不同于类的样本的可能性。本文将生品莪术作为阳性样本, 醋品莪术作为阴性样本。

为了在 5 种算法中寻找各自最优的超参数, 以保证该模型对本实验莪术电子鼻数据发挥最优的分类效果, 采用贝叶斯优化器, 采集函数为每秒期望改善, 迭代运算 100 次寻找最佳的模型, 选取测试集中表现最优的参数组合作为模型开始前预设的参数。优化模型得到 DTM 分类器参数为: 最大分裂数为 6, 分裂准则为最大程度减少偏差; NBM 分类器的核类型为 box; SVM 分类器的参数为: 数据经过标准化后, 采用线性核函数, 约束等级为 200, 核尺度为 160; KNN 分类器参数为: 邻点个数为 1, 距离权重为反距离平方, 并标准化数据; boost 分类器的参数为: 采用 bag 集成方法, 学习类型为 DTM, 最大分裂数为 8, 15 个弱学习器, 学习率为 0.007, 采样的预测变量数为 2。以上模型均采用 5 折交叉验证进行模型校正。

5 种模型与 BP-ANN 的评价指标见表 3, 在精确度, 敏感性, 特异性和 ROC 曲线下面积 AUC 4 个指标上, 5 种用于比较的模型没有可以达到 BP 神经网络模型完全准确判断的优异性能。这是因为电子鼻每一维的数据与传统的液相或者气相等手段获得的数据不同, 它包含了多个成分改变引起的共同变化, 以 W1C 和 W5C 为例, 两者同样对芳香衍生物敏感, 这类化合物经过炮制往往也产生相同的含量变化, 但是 W1C 对芳香衍生物的专属性较强, 这使得 W1C 感应器的响应值所包含的信息比较单一, 能准确反映芳香衍生物的变化信息; 而 W5C 不仅敏感于芳香衍生物, 还对烷烃以及非极性化合物敏感, 这是一个非常泛化的敏感范围, 代表着 W5C 感应器的响应值不仅包含了样品在炮制过程中芳香衍生物的变化, 同时也包含了烷烃和非极性化合物的变化(如蛋白质高温水解产生的硫化物), 这样的不同信息叠加在单一维度的数据中, 使数据具有混沌性, 一般

的模型难以进行准确学习。

表 3 6 种算法性能比较

算法	AC	SE	SP	AUC	%
ANN	100	100	100	100	
DTM	91.7	90.3	93.1	92.0	
NBM	95.0	90.9	100	99.0	
SVM	98.3	96.8	100	100	
KNN	95.0	96.6	93.5	95.0	
Boost	93.3	90.6	96.4	97.0	

BP 神经网络算法模仿人的神经元建立多个函数单元, 很好地模拟了这种混沌的分析模式^[19], 每一个数据在隐藏层的传导类似电信号在神经间的跳跃, 而设定的校正集数据则为这种跳跃施加阻力, 引导数据向阻力最小的方向传导, 最终通过对电子鼻数据非线性的分析与学习, 建立了高精度的判别模型。

2.5 姜黄素类成分含量预测

2.5.1 姜黄素类成分含量 实验室前期工作通过 HPLC 测量了炮制前后莪术饮片中的姜黄素类化合物含量, 结果见表 4。

2.5.2 特征提取与数据融合 姜黄素类化合物是典型的芳香衍生物, 莪术中主要含有双去甲氧基姜黄素, 去甲氧基姜黄素和姜黄素, 在其炮制过程里, 3 种化合物会转化和流失, 这些信息叠加地包含在对芳香衍生物敏感的电子鼻传感器中。考虑到电子鼻信号表征的是炮制过程中整体的变化, 将姜黄素类化合物总量与 10 个电子鼻特征值建立皮尔逊相关系数, 电子鼻传感器 W1C, W3C, W5C, W1S, W2S, W3S, W5S, W6S, W1W, W2W 的相关系数分别为 0.396 0, 0.304 6, 0.299 3, -0.219 9, -0.225 2, -0.109 1, -0.240 5, -0.141 6, -0.177 7, -0.310 1。可见两者之间的相关性并不强, 说明每个感应器受到莪术炮制过程中多种内在成分变化的影响, 进一步验证了气味信号的复杂性。根据所针对的化合物结构特点和表 1 中感应器的检测范围, 提取出相关性较强, 即包含了较多姜黄素类化合物的信息的 5 个电子鼻检测器特征数据: W1C, W3C, W5C, W2W 和 W2S。提取出的非同源信息具有不同的权重, 对数据进行归一化处理, 得到神经网络模型的输入层数据。

表4 莪术饮片及其炮制品中姜黄素类化合物含量测定

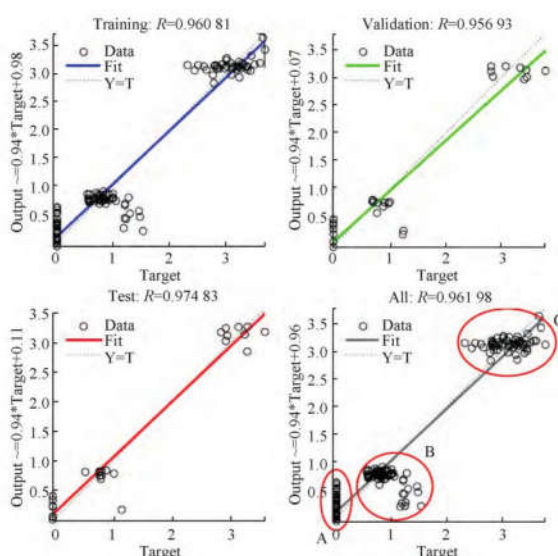
No.	双去甲氧基姜黄素	去甲氧基姜黄素	姜黄素	No.	双去甲氧基姜黄素	去甲氧基姜黄素	姜黄素
SP1	0.605 7	5.313 9	2.028 2	CSP1	0.561 7	3.378 5	1.201 6
SP2	0.574 1	2.987 7	0.000 0	CSP2	0.548 5	2.830 0	0.000 0
SP3	1.002 9	3.275 2	0.000 0	CSP3	0.925 8	3.456 5	0.000 0
SP4	0.824 5	2.449 0	0.000 0	CSP4	0.542 1	3.240 8	0.000 0
SP5	0.833 8	3.016 8	0.000 0	CSP5	0.685 9	2.805 2	0.000 0
SP6	1.028 6	3.570 4	0.000 0	CSP6	0.900 6	3.273 8	0.000 0
SP7	0.981 6	3.437 2	0.000 0	CSP7	0.889 9	3.402 5	0.000 0
SP8	0.786 7	2.935 8	0.000 0	CSP8	0.776 3	3.073 6	0.000 0
SP9	0.670 4	3.376 7	1.459 6	CSP9	0.571 1	2.887 0	1.207 6
SP10	0.573 2	2.774 7	1.211 6	CSP10	0.558 3	2.855 8	1.185 2
SP11	0.867 1	3.444 2	0.000 0	CSP11	0.812 1	3.300 7	0.000 0
SP12	0.792 6	3.387 5	0.000 0	CSP12	0.759 5	3.105 2	0.000 0
SP13	0.990 1	3.517 8	0.000 0	CSP13	0.994 3	3.702 3	0.000 0
SP14	0.828 3	3.094 4	0.000 0	CSP14	0.727 2	2.851 6	0.000 0
SP15	0.601 5	2.658 5	0.000 0	CSP15	0.577 7	2.510 8	0.000 0
SP16	0.712 2	2.811 0	1.239 8	CSP16	0.636 7	2.322 2	0.000 0
SP17	1.005 8	3.780 7	0.000 0	CSP17	0.781 5	3.450 7	0.000 0
SP18	0.918 6	3.331 5	0.000 0	CSP18	0.796 6	3.245 7	0.000 0
SP19	0.812 2	3.271 3	0.000 0	CSP19	0.823 3	3.124 1	0.000 0
SP20	0.774 7	2.789 9	1.273 1	CSP20	0.705 6	2.844 8	1.160 2
SP21	0.845 9	3.223 5	0.000 0	CSP21	0.721 5	3.180 8	0.000 0
SP22	0.991 0	3.610 3	0.000 0	CSP22	0.743 8	2.938 4	0.000 0
SP23	0.766 4	2.613 8	1.229 9	CSP23	0.693 5	2.812 5	1.248 5
SP24	0.819 5	3.228 6	0.000 0	CSP24	0.748 5	3.204 3	0.000 0
SP25	0.751 1	2.820 6	0.000 0	CSP25	0.699 3	3.012 4	0.000 0
SP26	0.868 7	3.244 6	0.000 0	CSP26	0.758 1	2.805 1	0.000 0
SP28	1.047 9	3.120 4	1.295 5	CSP28	0.783 2	2.910 7	0.000 0
SP29	0.986 5	3.314 7	0.000 0	CSP29	0.827 2	3.662 8	0.000 0
SP30	0.930 6	3.126 9	1.475 9	CSP30	0.875 3	3.019 3	1.529 7

2.5.3 姜黄素类成分含量预测 提取出的传感阵列并不是姜黄素类化合物的专属性响应器,其中还包含了类似烷烃,非极性化合物,氨气和硫化物等难以精准定量的物质的影响,和姜黄素类含量的关系也非常复杂,难以通过线性关系进行解释和预测,因此本文继续采用BP神经网络来建立电子鼻的响应数据与姜黄素类化合物的映射关系,在提取的相关电子鼻响应曲线基础上,运用非线性的神经网络回归来解决数据中所包含的高度叠加的样品信息,预测姜黄素类化合物的含量。在这一步,本实验将60批样品随机分配42批为训练样本,9批为校正样本,9批为测试样本,筛选的5个电子鼻特征数据作为神经网络的输入,以校正集的均方误差评价模型训练的性能,通过多次试验测试最佳参数为隐藏神经元个数设置为16个,最优化算法为Levenberg-Marquardt算法,确认检查值默认为6,训练迭代次数3868

上限为1000。

BP神经网络训练过程中的输出数据和相应的期望数据的回归关系见图4,相关系数代表了预测值与实际值之间的拟合程度,从图4中可以看出,样品主要聚集成了不同的3部分,分别对应了莪术生品及其醋制品中的双去甲氧基姜黄素,去甲氧基姜黄素和姜黄素3种姜黄素类化合物的含量,神经网络模型在训练集中的均方误差(MSE_c)为0.1498,相关系数(R_c)为0.9608;校正集的均方误差(MSE_{cv})为0.1488,相关系数(R_{cv})为0.9570;测试集的均方误差(MSE_p)为0.1175,相关系数(R_p)为0.9748,模型表现出较好的预测性,可以用于莪术生品及其醋制品中姜黄素类化合物的含量预测。

2.5.4 模型稳定性分析 为了保证建立的模型的稳定性和准确性,在模型超参数不变的前提下,重复



A. 姜黄素; B. 双去甲氧基姜黄素; C. 去甲氧基姜黄素。

图4 神经网络输出值与目标值的线性关系

Fig. 4 The linear relationship between output value and the target value of neural network

6次随机划分训练集、校正集和测试集的并重复训练10次,训练后模型在测试集中相关系数和均方误差的平均值见表5,结果稳定。

表5 不同测试集的测试结果平均值

Table 5 Average of test results for different test sets

测试集	平均均方误差	平均相关系数
1	0.129 2	0.960 2
2	0.141 9	0.958 7
3	0.167 4	0.967 1
4	0.119 5	0.971 4
5	0.195 6	0.970 7
6	0.125 1	0.978 4

3 讨论

中药的准确鉴别一直以来都是中药鉴定的核心任务之一,也是确保中药质量必不可少的工作。气味是药材非常独特的性状之一,特定的气味往往和特定的结构和成分的含量息息相关,2015年版《中国药典》中对莪术和醋莪术的气味感官描述为:“气香或微香”“微有醋香气”,具有主观性、经验性,模糊性,缺乏客观的评价指标。常用的色谱技术 TLC, HPLC, GC 样品处理过程复杂,分析成本高,耗时,且多针对单一成分或几类成分的定性、定量研究。电子鼻采集中药气味指纹图谱方法重复性良好,能

将传统鉴别经验转化为客观数据,量化不同中药之间的气味差异,且测定时样品无需预处理,快速简单实现样品“气味”的整体性分析。

电子鼻信号本身易受环境影响,并且信号中包含了多种因素的叠加,想要解释其与单一化合物的线性关系非常困难。BP神经网络算法具有非线性和混沌性的特点^[20],优于DTM, NBM, SVM, KNN和Boost 5种算法,可以较好地多种信息叠加的响应值中综合分析多种化合物的影响,因此,本文采用一套仿生的识别分析模式,即电子鼻技术和神经网络算法,对生、醋莪术进行鉴别和含量预测。通过电子鼻技术获取气味信号,将传统药材主观的气味特征转变为客观的信号特征,数据经归一化处理后作为输入神经元通过BP神经网络进行处理;以姜黄素类化合物的气味特点为轴心,提取出对姜黄素类化合物总含量变化相关的电子鼻传感器信息,对莪术饮片中具有芳香烃结构的3种姜黄素类化合物,通过BP神经网络算法进行回归,对莪术生饮片及其醋制品中的姜黄素类化合物进行准确的预测,为莪术及其炮制品的快速准确鉴别和质量评价提供了方向和依据。

本实验的另一个特色是将仿生的信息获取模式与仿生的信息学习与判别模式联合使用,将具有混沌性的数据与算法结合到准确的化合物含量中,实现了对样品更准确的判别和分析,为中药的整体多重作用分析提供了新的思路。仿生学仪器和算法的研究开发不仅可以继承动物对模糊信息精准分析的优势,同时数字化的信号将主观的特征转换为客观指标,机器学习则大大加速了训练过程,使得快速、简便、精确的分析成为可能。

[参考文献]

- [1] 国家药典委员会. 中华人民共和国药典. 一部[S]. 北京: 中国医药科技出版社, 2015: 208.
- [2] 林丽娅, 陈素红, 吕圭源. 莪术行气破血、消积止痛的相关药理研究[J]. 国际中医中药杂志, 2008, 30(2): 114.
- [3] 甘彦雄, 赵思蕾, 罗妮妮, 等. 蓬莪术醋制前后“行气破血”功效变化对比研究[J]. 现代中药研究与实践, 2015, 29(2): 29.
- [4] 崔国静, 林迎春, 徐亚. 莪术的鉴别与炮制[J]. 首都医药, 2010, 17(19): 50.
- [5] 沈艳, 李羿, 刘婷婷, 等. 3种基源莪术的显微及理化鉴定[J]. 化学研究与应用, 2014, 26(2): 267.
- [6] 叶永浩, 郝虹, 蔡宇忆, 等. 广西莪术的栽培繁殖及分子鉴定研究进展[J]. 中药材, 2015, 38(10): 2220.

- [7] 赵承光, 梁广, 邵丽丽, 等. 姜黄素类化合物抗炎和细胞保护作用的构效关系研究进展[J]. 中草药, 2008, 39(4): 619.
- [8] 王光耀, 盛良, 王兴华. 应用电子鼻研究中医药[J]. 医学信息, 2013(18): 243.
- [9] 拱健婷, 王佳宇, 李莉, 等. 基于电子鼻气味指纹图谱与XGBoost算法鉴别姜黄属中药[J]. 中国中药杂志, 2019, 44(24): 5375.
- [10] 拱健婷, 赵丽莹, 徐东, 等. 电子鼻联合支持向量机建立苦杏仁不同炮制品及走油程度快速判别模型[J]. 中国中药杂志, 2020, 45(10): 2389.
- [11] 费程浩, 戴辉, 苏杭, 等. 电子鼻技术的研究进展及其在中药行业中的应用[J]. 世界中医药, 2019, 14(2): 257.
- [12] 黄得栋, 何微微, 晋玲, 等. 基于电子鼻技术区分不同产地的南五味子[J]. 中国实验方剂学杂志, 2017, 23(23): 22.
- [13] 成华义. GA-PSO-BP神经网络在大气污染物浓度预测中的应用研究[D]. 武汉: 华中科技大学, 2014.
- [14] 石志标, 左春桢, 杨长明. 仿生鼻系统的试验研究[J]. 农业机械学报, 2005, 36(2): 70.
- [15] 戴加明. 基于深度卷积神经网络的星系形态分类研究[D]. 北京: 中国科学院大学(中国科学院国家空间科学中心), 2018.
- [16] YUSUF S, ISMAIL H. Quality evaluation of chickpeas using an artificial neural network integrated computer vision system[J]. Int J Food Sci Technol, 2011, 46(1): 194.
- [17] AMMA N G B. Cardiovascular disease prediction system using genetic algorithm and neural network[C]. Dindigul: International Conference on Computing, IEEE, 2012.
- [18] HECHT N. Theory of the backpropagation neural network[C]. Washington: International 1989 Joint Conference on Neural Networks, IEEE, 2002.
- [19] WANG X Y, MENG J, TAN G L, et al. Research on the relation of EEG signal chaos characteristics with high-level intelligence activity of human brain[J]. Nonlinear Biomed Phys, 2010, 4(1): 2.
- [20] ZHANG J, WANG K. Research about the cooperation fuzzy chaos neural network[C]. Dalian: World Congress on Intelligent Control & Automation, IEEE, 2006.

[责任编辑 孔晶晶]