

电子鼻技术在棉花早期棉铃虫虫害检测中的应用

代雨婷¹, 周 博², 王 俊^{1*}

(1. 浙江大学生物系统工程与食品科学学院, 杭州 310058; 2. 盐城工学院机械工程学院, 盐城 224051)

摘 要: 为了更好地获取棉花虫害信息, 该文使用电子鼻和气质联用技术对受到不同数量棉铃虫早期危害的棉花进行检测。基于气质联用技术获得了棉花挥发物的成分和含量, 基于电子鼻响应曲线提取了稳定值、面积值、平均微分值、小波能量值和多项式拟合曲线参数值 5 种特征值, 筛选出 3 种较优单特征: 稳定值、平均微分值和面积值, 之后基于多特征分别使用多层感知神经网络、径向基函数神经网络和极限学习机 3 种神经网络方法进行分类分析。最后采用支持向量机回归分别基于 3 种较优单特征及多特征对危害棉花的棉铃虫数量进行回归预测。结果表明: 多特征的分类效果优于单特征, 基于多特征“稳定值和平均微分值”和极限学习机分类效果最好, 训练集和测试集的分类正确率均达到 100%。多特征的预测能力优于单特征, 基于多特征“面积值和平均微分值”的回归模型预测效果最佳, 训练集回归模型的决定系数 (R^2) 和均方根误差 (RMSE) 分别为 0.994 0 和 0.086 0, 测试集回归模型的 R^2 和 RMSE 分别为 0.923 0 和 0.370 9, 电子鼻对棉花早期棉铃虫虫害具有较好的区分和预测能力, 电子鼻在棉花早期棉铃虫虫害中的检测具有一定的应用潜力。

关键词: 电子鼻; 神经网络; 预测; 棉花; 棉铃虫; 特征选择

doi: 10.11975/j.issn.1002-6819.2020.03.038

中图分类号: S224.3

文献标志码: A

文章编号: 1002-6819(2020)-03-0313-08

代雨婷, 周 博, 王 俊. 电子鼻技术在棉花早期棉铃虫虫害检测中的应用[J]. 农业工程学报, 2020, 36(3): 313—320.

doi: 10.11975/j.issn.1002-6819.2020.03.038 <http://www.tcsae.org>

Dai Yuting, Zhou Bo, Wang Jun. Application of electronic nose in detection of cotton bollworm infestation at an early stage[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(3): 313—320. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2020.03.038 <http://www.tcsae.org>

0 引 言

棉铃虫是棉花的三大主要害虫之一^[1], 常以幼虫钻蛀蕾铃为害, 引起棉铃腐烂或造成僵瓣, 给棉花生产造成巨大损失。因此, 对其进行检测并及时防治对于提高棉花生产和质量具有重要意义。然而传统方法如声学检测法^[2]、图像识别法^[3]和光谱成像技术^[4]等在复杂的棉田环境中的使用具有局限性。已有大量研究表明, 植物在遭受虫害时挥发物会发生变化^[5-6], 并且遭受虫害的情况不同, 挥发物有所差异^[7-8]。因此, 可以考虑通过植物挥发物的检测来对棉花的棉铃虫虫害情况进行判断, 而电子鼻是一种分析、识别和检测复杂气味和大多数挥发性成分的仪器, 具有快速、无损等优点^[9], 是实现植物挥发物检测的较佳手段之一。

随着电子鼻技术不断发展和完善, 电子鼻在食品^[10-11]、农业^[12-14]、医药^[15-16]、环境^[17-18]等方面有不少应用和研究成果。目前, 国外已经有一些研究者将电子鼻用于棉花虫害的检测, 并且取得了初步性的成果, 但是仍然存在对棉花受虫害情况诊断准确率不高的问题^[19-21]。特征提取是影响电子鼻识别准确率的关键因素之一, 特征提取

是基于一组特征参数进行的, 为了提高电子鼻的虫害识别准确率, 可以对特征参数的提取进行优化^[22-24]。本文针对棉花受到不同数量棉铃虫的早期危害进行研究, 采用电子鼻和气质联用技术对受虫害棉花进行检测, 气质联用技术与电子鼻相结合, 可更加可靠地实现对挥发物的检测。为了提高电子鼻识别虫害的准确率, 分别采用不同种类的单特征参数和多特征参数进行电子鼻的虫害区分能力和预测能力的研究, 探索电子鼻技术在棉花棉铃虫虫害检测中应用的可行性。

1 材料与方法

1.1 材 料

试验材料: 本试验所采用的棉花品种为冀棉 14 号, 棉花种子大小 24 cm (直径) × 19 cm (盆高) 的花盆中, 普通泥土和营养土比例为 1:1^[25], 生长于室外自然条件, 不施农药, 定期浇水, 将液体肥 (海博氏超浓缩深海全营养促长精华, 购于四川海博氏生物科技有限公司) 按 1:1 000 的体积比溶于水, 分别于棉花幼苗期、蕾期和花铃期按需施肥, 待棉花长出棉铃时用于试验, 约为 12 周。

本试验采用的 2 龄棉铃虫由河南省济源白云实业有限公司提供, 试验前, 将棉铃虫置于温度 (26±1) °C、相对湿度 40%~50% 的环境中, 使其尽快恢复正常的生长发育。

1.2 仪器设备

本试验采用的电子鼻为 PEN3 型电子鼻系统 (德国

收稿日期: 2019-09-26 修订日期: 2020-01-30

基金项目: 国家自然科学基金资助项目 (31671583)

作者简介: 代雨婷, 主要从事电子鼻在棉花虫害检测中的应用。

Email: 2994636283@qq.com

*通信作者: 王 俊, 教授, 博士生导师, 主要从事电子鼻电子舌技术开发及其智能检测应用。Email: jwang@zju.edu.cn

Airsense 公司), 该电子鼻的气体传感器阵列由 10 个金属氧化物半导体传感器组成, 不同传感器对不同类型的挥发物灵敏性不同, 具体如表 1。本试验使用的气相色谱-质谱联用仪 (GC-MS) 为安捷伦 7890A 型号的气相色谱和 5975C 型号的质谱。

表 1 PEN3 型电子鼻传感器组成及主要特性
Table 1 Sensors equipped in PEN3 electronic nose and their main characteristics

传感器编号 Sensors No.	传感器名称 Sensors name	传感器特性 Sensors characteristics	参考物及检测限 Reference substance and detection limits/ (mg·kg ⁻¹)
S1	W1C	对芳香类化合物灵敏	甲苯, 10
S2	W5S	具有广谱性响应性, 对氮氧化物灵敏	NO ₂ , 1
S3	W3C	对氨和芳香类化合物灵敏	苯, 10
S4	W6S	对氢气灵敏	H ₂ , 100
S5	W5C	对烷烃和芳香类化合物灵敏	丙烷, 1
S6	W1S	对甲烷灵敏	CH ₄ , 100
S7	W1W	对硫化物灵敏	H ₂ S, 1
S8	W2S	对乙醇灵敏	CO, 100
S9	W2W	对芳香类化合物和有机硫化物灵敏	H ₂ S, 1
S10	W3S	对烷烃灵敏	CH ₄ , 100

1.3 试验方法

1.3.1 电子鼻检测方法

为了探究电子鼻对受到不同数量棉铃虫危害的棉花的响应, 试验组安排如下: 对照组 (健康棉花)、“一株棉花+1 条棉铃虫”、“一株棉花+2 条棉铃虫”和“一株棉花+3 条棉铃虫”, 共 4 组, 每组处理设置 20 个重复样品, 共 80 盆棉花用于试验。试验前, 使棉铃虫饥饿 24 h。分别按照试验组安排于 8:00 将相应数量棉铃虫放置于棉花植株上, 之后立即采用保鲜袋 (60 cm×40 cm) 将棉花植株套住, 采用棉线将保鲜袋底部扎住, 保证顶空气体稳定, 棉花受棉铃虫危害 6 h 后进行检测, 棉铃虫低龄幼虫 (2 龄) 和短时间侵染 (6 h) 的设置是为了实现早期检测。电子鼻的检测参数为: 清洗传感器时间 60 s, 速率为 600 mL/min, 检测时间为 90 s, 速率为 400 mL/min。

1.3.2 气相色谱-质谱联用技术检测方法

棉花挥发物的收集: 采用吹扫捕集装置收集挥发物, 采用保鲜袋完全包裹住棉花植株的花盆和土壤, 试验组安排同电子鼻试验, 每组处理设置 3 个重复样品。接虫后将棉花植株放入吹扫捕集装置, 待装置收集 6 h 后, 取下吸附柱进行检测。

挥发物检测: 取下吸附柱后, 使用 0.3 mL 的二氯甲烷溶液分 3 次进行淋洗, 共 0.9 mL, 在淋洗后的溶液加入 10 μL 体积分数为 5×10⁻³ μL/μL 的乙酸辛酯 (溶质) 二氯甲烷 (溶剂) 溶液作为内标物, 取样针取 2 μL 溶液进样检测。

色谱检测条件: 采用 HP-5MS 色谱柱 (30 m×0.25 mm×0.25 μm), 载气为氦气, 流速 2 mL/min。升温程序: 初始温度 50 ℃保持 2 min, 以 5 ℃/min 的速率升温至

180 ℃, 保持 0.1 min, 然后以 10 ℃/min 的速率升温至 250 ℃, 保持 20 min。质谱条件: 离子化方式为电子轰击式电离 (EI), 电子能量为 70 eV, 四极杆温度为 150 ℃, 离子源温度为 230 ℃, 接口温度为 280 ℃, 质量范围为 30~50 amu。

1.4 数据分析

电子鼻数据分析: 从电子鼻传感器曲线中提取 5 种特征值: 稳定值 (85 s 响应值)、面积值、平均微分值、小波能量值和多项式拟合曲线参数值, 采用主成分分析 (principle component analysis, PCA) 对 5 种特征值进行初步分析, 再分别基于多层感知器神经网络^[26] (multi-layer perceptron neural network, MLPNN)、径向基函数神经网络^[27] (radial basis function neural network, RBFNN) 和极限学习机^[28] (extreme learning machine, ELM) 对 5 种特征值的分类效果进行分析, 这 3 种神经网络目前应用广泛且各有特点, MLPNN 具有较好的泛化能力和容错能力, RNFNN 具有较好的收敛性和较快的学习速度, 而 ELM 相比传统的神经网络可在保证学习精度的情况下具有更快的学习速度, 这 3 种神经网络方法均可用于非线性分类, 并且 MLPNN 和 RBFNN 均可直接在 IBM SPSS Statistic 24.0 软件 (美国 SPSS 公司) 上运行, 方便易操作, 因此选取这 3 种神经网络方法用于本试验的结果分析。筛选出稳定值、面积值和平均微分值 3 种较优特征值, 将 3 种特征进行随机组合, 在此基础上再进行分类分析。最后采用支持向量机回归^[29] (support vector machine regression, SVR) 分别基于 3 种较优单特征值和多特征值建立危害棉花的棉铃虫数量的回归预测模型, 选出较优的回归模型。

挥发物的定性和定量分析: 未知化合物的定性主要通过 NIST 质谱库进行检索, 并与发表文献比较, 未知化合物使用内标法进行定量, 定量计算如公式 (1) 和公式 (2) 所示。

$$m_0 = V_0 \rho_0 \tag{1}$$

$$m = \frac{Sm_0}{S_0} \tag{2}$$

式中 m_0 和 m 分别为内标物和未知化合物的质量, μg。 V_0 和 ρ_0 分别为内标物的体积 (L) 和密度 (g/cm³), 均已知, 即 m_0 可知。 S 为未知挥发物经 GC-MS 检测后在 GC-MS 图谱中的峰面积, S_0 为内标物在 GC-MS 图谱中的峰面积。经检索和计算后得出健康组棉花和各虫害组棉花的挥发物的主要成分和含量, 进行比较分析。

2 结果与分析

2.1 受不同数量棉铃虫危害的棉花挥发物变化

经气质联用技术检测发现, 健康棉花和受虫害棉花的挥发物共有约 40 种, 包括萜烯类 16 种, 质量占比 4.58%~14.21%, 芳香族化合物 16 种, 质量占比 85.27%~94.61%, 烷烃、酸类、酯类和醇类 7 种, 质量占比 0.4%~2.6%。表 2 为主要的代表性挥发物, 采用最小显著差数法 (least significant difference, LSD) 进行多重比较, 挥

发物含量平均值后的字母代表不同的显著性，字母相同，则为差异不显著 ($P>0.05$) [25]。

由表 2 可得，对照组棉花和受虫害棉花组的挥发物在成分和含量上均存在差异，特别是 1,4-二甲苯和 2-甲基萘等芳香族化合物以及 α -蒎烯、 β -石竹烯和花侧柏烯等萜烯类化合物存在显著差异，而受不同数量棉铃虫危害的棉花的挥发物在成分组成上差异较小，但在含量上有一定差异，受虫害棉花释放的挥发物总量普遍显著高于对照组健康棉花释放的挥发物总量。并且由表中可以

看出，有些化合物仅在健康棉花中出现或仅在受虫害棉花中出现，如 2-甲基萘和棕榈酸在受虫害的棉花中是新产生的， β -石竹烯在受虫害的棉花中不再产生，推测棉花在受虫害后合成了一些新的化合物用于启动防御机制，以及棉花受虫害后某些化合物减少了合成和释放。总的来看，健康棉花和受虫害棉花的挥发物存在差别，受不同数量棉铃虫危害的棉花挥发物也有一定差异，而电子鼻作为一种基于挥发物检测的技术，在棉花的棉铃虫害检测上具有一定潜力。

表 2 基于气质联用技术的健康棉花和受虫害棉花的主要挥发物
Table 2 Main identified VOCs emitted by undamaged cotton plants and cotton plants infested with cotton bollworms of different amounts detected by GC-MS μg

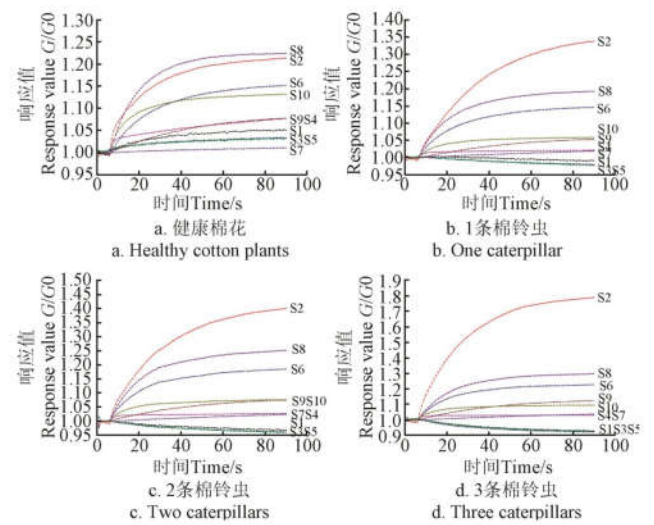
化合物 Volatile compounds	对照组 Control	棉铃虫数量 Caterpillars number		
		1	2	3
α -蒎烯 α -Pinene	3.69±2.65b	9.97±0.43a	2.24±0.82b	4.05±5.23b
1,4-二乙基苯 Benzene, 1,4-diethyl-	10.81±10.80b	80.18±61.38a	177.45±85.03a	128.45±52.27a
4-乙基苯甲醛 Benzaldehyde, 4-ethyl-	3.02±1.73a	5.78±4.46a	4.82±5.03a	3.61±0.79a
萘 Naphthalene	2.01±0.61a	1.78±1.32a	1.31±0.74a	1.16±0.54a
十二烷 Dodecane	1.00±0.37a	0.70±0.05a	0.66±0.23a	1.12±0.15a
对乙基苯乙酮 Ethanone, 1-(4-ethylphenyl)-	49.14±32.88a	25.38±10.10a	24.46±19.94a	23.78±12.03a
2-甲基萘 Naphthalene, 2-methyl-	0c	0.37±0.04b	0.32±0.11b	1.25±0.07a
β -雪松烯 β -Cedrene	1.55±1.23a	0.80±0.44a	0.95±0.48a	0.87±0.66a
顺-罗汉柏烯 cis-Thujopsene	44.84±30.45a	19.89±10.65a	30.14±11.95a	21.82±17.92a
雪松醇 Cedrol	0.74±0.17a	0.48±0.23a	0.40±0.27a	0.65±0.39a
β -石竹烯 β -Caryophyllene	0.66±0.53a	0b	0b	0b
花侧柏烯 Cuparene	1.08±0.54a	0.57±0.07b	0.64±0.09a	1.50±0.73a
棕榈酸 Palmitic acid	0b	3.67±2.48a	1.14±1.35a	1.05±1.23a
总计	87.91±30.10b	144.58±59.29ab	225.19±39.77a	197.10±42.22a

注：平均值后的相同字母表示差异不显著 ($P>0.05$)。
Note: Means followed by the same letter are not significantly different ($P>0.05$).

2.2 传感器响应曲线分析

图 1 为不同数量棉铃虫危害棉花的电子鼻传感器响应曲线，图 1a~图 1d 分别为健康棉花以及棉花受 1 条、2 条或 3 条棉铃虫危害的电子鼻传感器响应曲线，横轴为检测时间 0~90 s，纵轴为电子鼻的响应值。从图 1 中可以看出，各棉花样品的电子鼻传感器响应曲线存在差异性。首先，健康棉花与受虫害棉花对于电子鼻传感器的响应不同，对于健康棉花（图 1a），传感器 S2、S6 和 S8 响应差异较小，且 S1、S3 和 S5 响应值变化较小，而对于受虫害棉花（图 1b—图 1d），传感器 S2 的响应值明显高于 S6 和 S8，且 S1、S3 和 S5 传感器响应值变化较大。由表 1 可知，S1、S3 和 S5 对芳香族化合物灵敏，根据气质联用仪分析结果，健康棉花和受虫害棉花的芳香族类化合物确实存在差异，主要是 1,4-二甲苯和 2-甲基萘等芳香族化合物存在差异，这也说明了电子鼻结果的可靠性。其次，受不同数量棉铃虫危害的棉花的电子鼻传感器响应也存在差异，随着对棉花危害的棉铃虫的数量的增加，传感器 S2 的响应值也逐渐增大，这与 S2 传感器的广谱响应性有关，根据气质联用检测结果，棉花受害后挥发物释放总量增加，并且棉铃虫数量增加后，棉花释放总量有所增加（参照总量平均值），这说明电子鼻检测结果和气质联用结果基本保持一致。传感器 S1、S3 和 S5 的响应值变化逐渐增加，说明不同数量棉铃虫危

害棉花的芳香族类化合物也有所差异，这与上述气质联用仪检测的结果一致。由以上各样品的电子鼻传感器的响应曲线差异分析可以表明，电子鼻在棉铃虫危害棉花的检测上具有可行性。



注：G 和 G_0 分别为传感器检测样品气体和洁净空气时的电阻值，S1~S10 分别为电子鼻的 10 个传感器。
Note: G and G_0 are the resistance values when sensors detect sample gas and clean air, respectively. S1 to S10 are the ten sensors of the electronic nose, respectively.

图 1 不同数量棉铃虫危害棉花的电子鼻传感器响应曲线
Fig.1 Response curves of electronic nose sensors to cotton plants infested with different number of caterpillars

2.3 主成分分析

本文从电子鼻的传感器曲线中提取了 5 种特征值,采用主成分分析方法对 5 种特征值的区分效果进行初步分析,图 2a~图 2e 分别为基于稳定值、平均微分值、面积值、小波能量值和多项式拟合曲线参数值的分析结果,图中不同的棉花受害组以相应的棉铃虫害虫数进行标示。5 种特征值图 2a~图 2e 第一主成分和第二主成分之和分别为 83.01%、97.18%、88.58%、88.38%和 77.09%,前 4 种特征值的第一主成分和第二主成分之和都较大,而多项式拟合参数值的两主成分之和较小,说明前 4 种特征值可以用来代表电子鼻信号对棉花受不同数量虫害的区分情况,而多项式拟合曲线参数值不能很好的代表电子鼻信息。从图 2 中可以看出,基于稳定值,面

积值和小波能量值的区分效果比较相似,健康棉花组和受虫害棉花组可以明显区分,但各棉花受不同数量棉铃虫害虫组存在重叠。说明这 3 种特征值有利于表达健康棉花组和受虫害棉花组的信息差异性,但是不足以表达各棉花受不同数量棉铃虫害虫组的信息差异性。基于平均微分值,健康棉花组和各棉花受害组都可以明显区分,说明平均微分值能充分体现各试验组之间的信息差异性。而基于多项式拟合曲线参数值,各试验组均有重叠,说明多项式拟合曲线参数值对各试验组的差异性的表达能力较差,不适用于后续分析。因此,初步分析认为平均微分值、稳定值、面积值和小波能量值适用于后续分析,而多项式拟合曲线参数值对电子鼻信息的代表性较差。

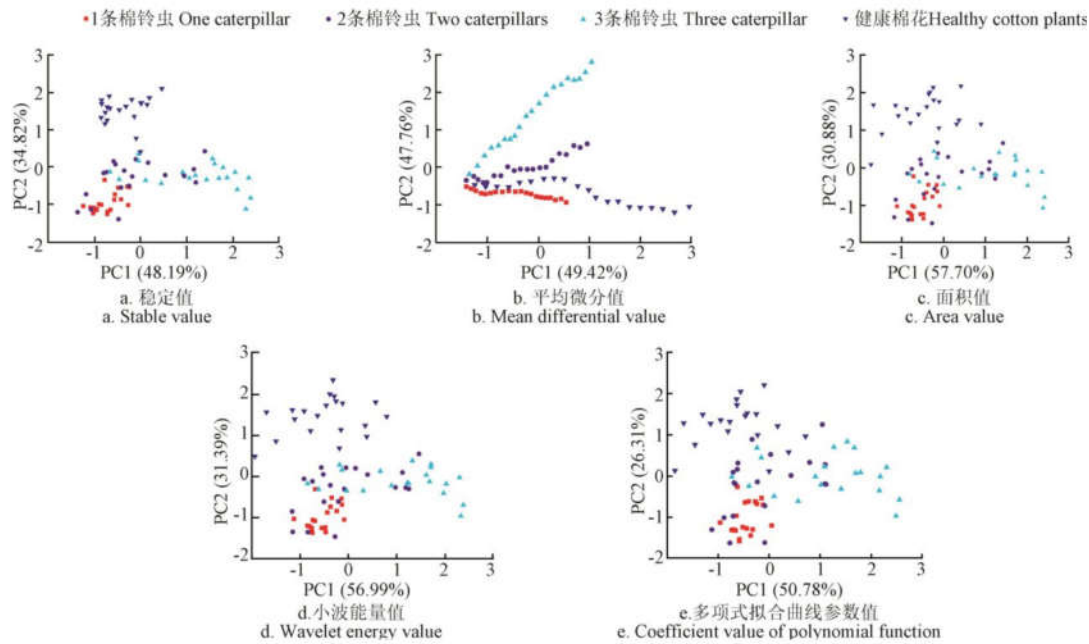


图 2 基于 5 种特征值的棉花受不同数量棉铃虫害的主成分分析结果图
Fig.2 Discrimination results of PCA on cotton plants infested by different number of caterpillars based on five kinds of feature parameters

2.4 受不同数量棉铃虫害的棉花的分类分析

2.4.1 基于单特征值的分类模型构建

由以上主成分分析结果可以看出,棉花受不同数量虫害组仍不能明显区分。为了对棉花受不同数量棉铃虫害进行更好的区分,分别基于 3 种神经网络方法和 5

种特征值进行分类分析。分析时,将 4 组棉花样本(共 80 个样本)分为训练集和测试集,其中每组 14 个样本作为训练集,6 个样本作为测试集,因此 56 个训练集样本和 24 个测试集样本用于 4 分类模型的构建和测试,表 3 为分类结果。

表 3 基于 3 种神经网络和 5 种单特征值的棉花受不同数量棉铃虫害的分类结果

Table 3 Classification results of cotton plants infested by different number of caterpillars based on three neural network methods and five kinds of feature parameters						
神经网络 Neural networks	数据集 Data set	分类正确率 Classification accuracy rate/%				
		稳定值 Stable value	平均微分值 Mean differential value	面积值 Area value	小波能量值 Wavelet energy value	多项式拟合曲线参数值 Coefficient value of polynomial function
MLPNN	训练集	92.74	100.00	93.84	93.18	93.68
	测试集	85.50	100.00	85.48	88.44	85.20
RBFNN	训练集	85.46	83.68	83.92	82.88	82.88
	测试集	68.68	68.44	68.50	69.08	69.08
ELM	训练集	100.00	100.00	98.92	97.50	100.00
	测试集	91.67	95.00	82.50	78.33	62.50

由表 3 可以看出,从分类方法上来看,在 3 种神经网络方法中,MLPNN 方法下基于各特征值下的训练集和测试集的分类正确率总体最高,训练集的分类正确率均在 90%以上,测试集的分类正确率均在 85%以上,说明 MLPNN 方法比较适用于本试验数据集的分类分析。而基于 RBFNN 方法各特征值下的训练集和测试集的分类正确率总体最低,且各特征值下的测试集的分类正确率均低于 70%,因此认为 RBFNN 方法不适用于本试验的棉花受不同数量棉铃虫危害的分类分析。从各特征值上来看,基于稳定值和平均微分值的平均分类正确率相对高,而基于多项式拟合曲线参数值的平均分类正确率总体最差,基于面积值的平均分类正确率与基于小波能量值的平均分类正确率较为相当,基于面积值的平均分类正确率稍高于基于小波能量值的分类正确率,这与主成分分析的结果基本一致,也进一步说明稳定值、平均微分值和面积值能够更好地代表电子鼻检测棉花虫害的信息,可以作为今后电子鼻实际应用中采用特征值种类的参考。总体来看,在 MLPNN 方法下基于各特征值的分类效果总体最好,稳定值、平均微分值和面积值在 5 种特征值中分类效果相对较好。

2.4.2 基于多特征值的分类模型优化

根据单特征值的分类结果来看,稳定值、平均微分值和面积值的总体分类效果在 5 种特征值中相对较好,但是仍然需要进行优化。为了更好的表达棉花受虫害的挥发物的电子鼻信号,考虑综合采用稳定值、平均微分值和面积值 3 种特征值用于分类分析,即将稳定值、平均微分值和面积值 3 个特征值进行两两组合和 3 项组合,再进行分类分析。在组合各类特征值前探究了传感器阵列的交叉敏感性对多特征值分类效果的影响,因此进行了传感器阵列优化的预试验,传感器阵列优化方法采用了因子载荷分析和方差分析,结果发现传感器阵列优化前后,基于多特征值的分类

效果没有较大提升或者反而下降,原因可能为传感器阵列优化后,有用信息可能也被丢失,造成分类效果提升少或下降。因此本试验中可忽略不同传感器交叉敏感性的影响,考虑直接组合不同类的特征值,不进行传感器阵列的优化,简化试验分析过程。根据表 3 中的结果,由于基于 RBFNN 方法的分类表现较差,因此仅将 MLPNN 方法和 ELM 方法用于下一步的分类分析,构建分类模型时采用的训练集和测试集的设置同采用单特征值时进行分类分析时的设置相同。表 4 为基于 2 种神经网络方法和 3 种特征值组合对棉花受不同数量棉铃虫危害的分类结果。

由表 4 可以看出,总体来看,与单特征值相比,基于 2 种神经网络方法的多特征值的分类正确率均有较大提高。2 种神经网络方法中,基于 MLPNN 方法各多特征值的分类表现仍然优于基于 ELM 方法的分类表现,训练集的分类正确率均在 99%以上,测试集的分类正确率均在 90%以上。基于 ELM 方法,与单特征值相比,各多特征值的测试集的分类正确率有较大提高,均在 85%以上。总体来看,分类正确率较高的分别为基于 ELM 方法及“稳定值+平均微分值”、基于 MLPNN 方法及“稳定值+平均微分值+面积值”和基于 MLPNN 方法及“面积值+平均微分值”。因此,可以认为基于多特征值可以提高电子鼻对棉花受不同数量棉铃虫危害的分类效果。虽然将各单特征值组合后,数据集变大,但是电子鼻数据集本就是小样本数据,组合后的数据集采用神经网络方法进行分析仍然能较快的得到分析结果,而且由于组合后的特征综合了电子鼻传感器响应曲线的各类信息,加大了各棉花受不同数量棉铃虫危害组的信息差异性,因此显著提高了电子鼻对棉花受不同数量棉铃虫危害组的分类正确率。由以上分析表明,多特征值更适合用于表达棉花虫害挥发物的电子鼻信号,更适合用于棉花受不同数量棉铃虫危害的分类分析。

表 4 基于 3 种神经网络和多特征值的棉花受不同数量棉铃虫危害的分类结果

Table 4 Classification results of cotton plants infested by different number of caterpillars based on three neural network methods and multiple-features parameters

神经网络 Neural networks	数据集 Data set	分类正确率 Classification accuracy rate/%			
		面积值+平均微分值 Area value and mean differential value	稳定值+平均微分值 Stable value and mean differential value	面积值+稳定值 Area value and stable value	稳定值+平均微分值+面积值 Stable value, mean differential value and area value
MLPNN	训练集	99.60	99.14	99.30	100.00
	测试集	96.02	90.70	92.54	97.74
ELM	训练集	100.00	100.00	100.00	100.00
	测试集	86.67	100.00	90.83	90.83

2.5 危害棉花的棉龄虫数量的预测模型构建

2.5.1 基于单特征值构建危害棉花的棉铃虫数量的预测模型

根据电子鼻对棉花受不同数量棉铃虫危害的分类结果,证明电子鼻在识别健康棉花和受虫害棉花以及识别不同棉铃虫数量具有可行性。为了对电子鼻在预测危害棉花的棉铃虫数量的可行性进行研究,采用支持向量机回归方法建立电子鼻信号和棉铃虫数量的回归模型。自

变量为电子鼻结果特征值,因变量为棉铃虫数量。共有 80 个棉花样本,每组 14 个样本作为训练集,6 个样本作为测试集。分别采用电子鼻检测结果的稳定值、平均微分值和面积值作为自变量建立回归模型,表 5 为基于各特征值建立的回归模型的训练集和预测集的预测效果。

使用决定系数 R^2 和均方根误差 RMSE 评价回归模型的优劣,由表 5 可以看出,基于单特征值的训练集的回归模型预测效果均较好, R^2 分别为 0.982 3, 0.995 7 和

0.905 7, 且 RMSE 较小, 但是基于各特征值的测试集的回归模型预测效果有限, R^2 分别为 0.674 2, 0.723 6 和 0.781 4, 且 RMSE 较大, 训练集和测试集的回归模型预测效果差异较大, 说明基于单特征值建立的回归预测模型容易出现过拟合问题, 回归模型效果预测效果不佳。总体上, 基于单特征值建立的回归模型预测效果排序为: 面积值>平均微分值>稳定值, 推测原因可能为面积值和平均微分值都表达的是电子传感器整条曲线的信息, 较为全面, 且不易受到极端值的影响, 而稳定值采用的是曲线上的某一点的值, 易受到极端值的影响, 会导致回归模型的预测效果较差。因此认为基于电子鼻检测结果的单特征值对危害棉花的棉铃虫数量进行预测不太理想, 仍然需要对回归模型进行优化。

表 5 基于支持向量机回归和单特征值建立的回归模型预测棉铃虫数量结果

Table 5 Prediction of the amounts of cotton bollworm based on SVR and single-features.				
数据集 Data set	评价指标 Evaluation index	稳定值 Stable value	平均微分值 Mean differential value	面积值 Area value
训练集 Training set	R^2	0.982 3	0.995 7	0.905 7
	RMSE	0.156 2	0.078 7	0.347 0
测试集 Training set	R^2	0.674 2	0.723 6	0.781 4
	RMSE	0.771 4	0.685 1	0.520 3

2.5.2 基于多特征值对危害棉花的棉铃虫数量的预测模型的优化

由于采用电子鼻检测结果的单特征值对棉铃虫数量预测的效果并不理想, 尝试采用电子鼻检测结果的多特

征值对棉铃虫数量进行预测。采用支持向量机回归建立回归模型时, 分别将稳定值、平均微分值和面积值进行两两组合和三项组合作为自变量, 棉铃虫数量作为因变量, 训练集和测试集的分布与采用单特征值建立回归模型时相同, 仍然采用 R^2 和 RMSE 作为回归模型的评价指标, 分析模型预测效果。表 6 为基于多特征值和支持向量机回归建立的回归模型的训练集和测试集的预测效果。

由表 6 可以看出, 基于多特征值建立的测试集的回归模型的预测效果相比各单特征值下建立的测试集的回归模型的预测效果明显提高, R^2 均在 0.8 以上, 且 RMSE 较小, 说明采用多特征值可以提高棉铃虫数量预测的效果。由表 6 可知基于“面积值+平均微分值”建立的回归模型预测效果最佳, 训练集的回归模型的 R^2 和 RMSE 分别为 0.994 0、0.086 0, 测试集的回归模型的 R^2 和 RMSE 分别为 0.923 0 和 0.370 9, 基于其他组合特征值建立的回归模型预测效果均不如“面积值+平均微分值”, 比如基于“稳定值+平均微分值+面积值”建立的回归模型的预测效果反而不如“面积值+平均微分值”, 可能的原因是稳定值容易将极端值带入数据集中, 在加入“面积值+平均微分值”的组合特征后, 破坏了原本回归模型的稳定性, 降低了原本回归模型的预测效果。而由表 5 可知, 基于单特征值建立的回归模型平均预测效果排序为: 面积值>平均微分值>稳定值, 根据表 6 中各特征组合的回归模型预测效果比较, 推测单特征值的预测效果越好, 用于特征值组合后对组合后特征的预测效果提升越大, 单特征值的预测效果越差, 用于特征值组合后, 对组合后特征预测效果提升越小或者反而会造成预测效果下降。

表 6 基于支持向量机回归和多特征值建立的回归模型预测棉铃虫数量结果
Table 6 Prediction of the amounts of cotton bollworm based on SVR and multiple-features parameters.

数据集 Data set	评价指标 Evaluation index	稳定值+平均微分值 Stable value and mean differential	面积值+平均微分值 Area value and mean differential value	面积值+稳定值 Area value and stable value	稳定值+平均微分值+面积值 Stable value, mean differential value and area value
训练集 Training set	R^2	0.952 8	0.994 0	0.970 5	0.994 5
	RMSE	0.249 4	0.086 0	0.192 6	0.007 0
测试集 Test set	R^2	0.830 6	0.923 0	0.871 9	0.879 3
	RMSE	0.453 4	0.370 9	0.407 9	0.403 3

3 结 论

本试验采用电子鼻和气质联用技术对受不同数量棉铃虫早期危害的棉花进行检测, 对电子鼻传感器与棉花挥发物的响应关系进行了阐述, 对如何提高电子鼻对棉花虫害的识别能力进行了研究, 分别从电子鼻传感器响应曲线中提取 5 种特征值: 稳定值、面积值、平均微分值、小波能量值和多项式曲线拟合参数值。分别基于多层感知器神经网络、径向基函数神经网络和极限学习机 3 种神经网络对 5 种特征值的分类正确率进行了考察, 优选出稳定值、面积值和平均微分值。分别对稳定值、面积值和平均微分值 3 种单特征值和多特征值进行分类和预测能力的研究, 结果发现:

1) 多特征值的分类能力普遍优于单特征值的分类能力, 基于多特征值“稳定值+平均微分值”和极限学习机

(ELM) 对棉花受不同数量棉铃虫危害的分类能力最佳, 训练集和测试集的分类正确率均可达到 100%。

2) 基于多特征值建立的回归模型的预测效果普遍优于单特征值建立的回归模型预测效果, 基于多特征值“面积值+平均微分值”和支持向量机建立的回归模型对危害棉花的棉铃虫数量预测效果最佳, 训练集回归模型的 R^2 和 RMSE 分别为 0.994 0、0.086 0, 测试集回归模型的 R^2 和 RMSE 分别为 0.923 0 和 0.370 9。

结果表明, 电子鼻在棉花的早期棉铃虫虫害检测上具有可行性, 并且通过特征值优化和特征值组合可以提高电子鼻对受棉铃虫虫害棉花的识别准确率和预测效果, 为电子鼻在棉花虫害检测中的应用提供了思路。

[参 考 文 献]

[1] 曾娟, 陆宴辉, 简良, 等. 中国病虫草害调查诊断与决策支持系统[M]. 北京: 中国农业出版社, 2017: 1—232.

- [2] Mankin R W, Hagstrum, D W, Smith M T, et al. Perspective and promise: A century of insect acoustic detection and monitoring[J]. American Entomologist, 2011, 57(1): 30—44.
- [3] 翟治芬, 徐哲, 周新群, 等. 基于朴素贝叶斯分类器的棉花盲椿象危害等级识别[J]. 农业工程学报, 2015, 31(1): 204—211.
Zhai Zhifen, Xu Zhe, Zhou Xinqun, et al. Recognition of hazard grade for cotton blind stinkbug based on Naive Bayesian classifier[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2015, 31(1): 204—211. (in Chinese with English abstract)
- [4] 张军国, 韩欢庆, 胡春鹤, 等. 基于无人机多光谱图像的云南松虫害区域识别方法[J]. 农业机械学报, 2018, 49(5): 249—255.
Zhang Junguo, Han Huanqing, Hu Chunhe, et al. Identification method of pinus yunnanensis pest area based on UAV multispectral images[J]. Transaction of the Chinese Society for Agricultural Machinery, 2018, 49(5): 249—255. (in Chinese with English abstract)
- [5] Anastasaki E, Balayannis G, Papanikolaou N E, et al. Oviposition induced volatiles in tomato plants[J]. Phytochemistry Letters, 2015, 13: 262—266.
- [6] Erb M, Veyrat N, Robert C A M, et al. Indole is an essential herbivore-induced volatile priming signal in maize[J]. Nature Communications, 2015, 6: 1—10.
- [7] Silva D B, Weldegergis B T, Van Loon J J A, et al. Qualitative and quantitative differences in herbivore-Induced plant volatile blends from tomato plants Infested by either tuta absoluta or bemisia tabaci[J]. Journal of Chemical Ecology, 2017, 43(1): 53—65.
- [8] Clavijo McCormick A, Boeckler G A, Köllner T G, et al. The timing of herbivore-induced volatile emission in black poplar (*Populus nigra*) and the influence of herbivore age and identity affect the value of individual volatiles as cues for herbivore enemies[J]. BMC Plant Biology, 2014, 14: 1—13.
- [9] Hu W W, Wan L T, Jian Y Y, et al. Electronic noses: From advanced materials to sensors aided with data processing[J]. Advanced Materials Technologies, 2019, 4: 1—38.
- [10] Ramirez H L, Soriano A, Gomez S, et al. Evaluation of the food sniffer electronic nose for assessing the shelf life of fresh pork meat compared to physicochemical measurements of meat quality[J]. European Food Research and Technology, 2018, 244(6): 1047—1055.
- [11] Kodogiannis V. Application of an electronic nose coupled with fuzzy-wavelet network for the detection of meat spoilage[J]. Food and Bioprocess Technology, 2017, 10(4): 730—749.
- [12] 程绍明, 王俊, 王永维, 等. 基于电子鼻技术的番茄苗早疫病病害快速检测研究[J]. 科技通报, 2013, 29(7): 68—77.
Chen Shaoming, Wang Jun, Wang Yongwei, et al. Research on distinguishing tomato seedling infected with early blight disease by electronic nose[J]. Bulletin of Science and Technology, 2013, 29(7): 68—77. (in Chinese with English abstract)
- [13] Xu S, Zhou Z Y, Li K L, et al. Recognition of the duration and prediction of insect prevalence of stored rough rice infested by the red flour beetle (*Tribolium castaneum* Herbst) using an electronic nose[J]. Sensors, 2017, 17(4): 1—9.
- [14] 林振华, 姜水, 张红梅, 等. 基于金属氧化物传感器阵列的小麦霉变程度检测[J]. 传感技术学报, 2018, 31(7): 1017—1023.
Lin Zhenhua, Jiang Shui, Zhang Hongmei, et al. Detection of moldy wheat using MOS sensor array in electronic nose[J]. Chinese journal of Sensor and Actuators, 2018, 31(7): 1017—1023. (in Chinese with English abstract)
- [15] Schuermans V N E, Li Z Y, Jongen A, et al. Pilot study: Detection of gastric cancer from exhaled air analyzed with an electronic nose in Chinese patients[J]. Surgical Innovation, 2018, 25(5): 429—434.
- [16] Altomare D F, Porcelli F, Picciariello A, et al. The use of the PEN3 e-nose in the screening of colorectal cancer and polyps[J]. Techniques in Coloproctology, 2016, 20(6): 405—409.
- [17] Pace C, Khalaf W, Latino M, et al. E-nose development for safety monitoring applications in refinery environment[J]. Procedia Engineering, 2012, 47: 1267—1270.
- [18] Blanco-Rodríguez A, Camara V F, Campo F, et al. Development of an electronic nose to characterize odours emitted from different stages in a wastewater treatment plant[J]. Water Research, 2018, 134: 92—100.
- [19] Henderson W G, Khalilian A, Han Y J, et al. Detecting stink bugs/damage in cotton utilizing a portable electronic nose[J]. Computers and Electronics in Agriculture, 2010, 70(1): 157—162.
- [20] Degenhardt D C, Greene J K, Khalilian A. Temporal dynamics and electronic nose detection of stink bug-induced volatile emissions from cotton bolls[J]. Psyche, 2012, 2012: 1—9.
- [21] Lampson B D, Han Y J, Khalilian A, et al. Development of a portable electronic nose for detection of pests and plant damage[J]. Computers and Electronics in Agriculture, 2014, 108: 87—94.
- [22] Yan J, Guo X Z, Duan S K, et al. Electronic nose feature extraction methods: A review[J]. Sensors, 2015, 15(11): 27804—27831.
- [23] Sun Y B, Wang J, Cheng S M. Discrimination among tea plants either with different invasive severities or different invasive times using MOS electronic nose combined with a new feature extraction method[J]. Computers and Electronics in Agriculture, 2017, 143: 293—301.
- [24] Yan J, Tian F C, He Q H, et al. Feature extraction from sensor data for detection of wound pathogen based on electronic nose[J]. Sensors and Materials, 2012, 24(2): 57—73.
- [25] Magalhães D M, Borges M, Laumann R A, et al. Influence of multiple- and single-species infestations on herbivore-induced cotton volatiles and anthonomus grandis behaviour[J]. Journal of Pest Science, 2018, 91(3): 1019—1032.
- [26] Rao M R N, Sridhar G R, Madhu K, et al. A clinical decision support system using multi-layer perceptron neural network

- to predict quality of life in diabetes[J]. *Diabetes & Metabolic Syndrome*, 2010, 4(1): 57—59.
- [27] Yin Y, Zhao Y Z. A feature selection strategy of E-nose data based on PCA coupled with Wilks Λ -statistic for discrimination of vinegar samples[J]. *Journal of Food Measurement and Characterization*, 2019, 13(3): 2406—2416.
- [28] Du D D, Xu M, Wang J, et al. Tracing internal quality and aroma of a red-fleshed kiwifruit during ripening by means of GC-MS and E-nose[J]. *Rsc Advances*, 2019, 9(37): 21164—21174.
- [29] Sun Y B, Wang J, Sun L, et al. Evaluation of E-nose data analyses for discrimination of tea plants with different damage types[J]. *Journal of Plant Diseases and Protection*, 2019, 126(1): 29—38.

Application of electronic nose in detection of cotton bollworm infestation at an early stage

Dai Yuting¹, Zhou Bo², Wang Jun^{1*}

(1. College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China;

2. Department of Mechanical Engineering, Yancheng Institute of Technology, Yancheng 224051, China)

Abstract: Cotton bollworm is one of the main pests of cotton. Cotton is under threat of yield loss and poor quality because of the cotton bollworm. However, cotton bollworms tend to hide in the cotton plants so that there are limitations for conventional detection methods, such as acoustic signal method, image recognition method and spectral imaging technology. A lot of researches have shown that volatile organic compounds (VOCs) released by plants will change when they are attacked by pests. So it is possible to get the cotton bollworm damage information by detecting the volatiles. Currently, gas chromatograph-mass spectrometer (GC-MS) can accurately detect the composition and content of volatile matter. However, this method has some disadvantages in practical application, such as time-consuming, high cost and inconvenience. The electronic nose is composed of sensor array, which is an instrument to analyze, identify and detect most of the volatiles. In this study, electronic nose was used to detect the cotton plants infested with cotton bollworm of different amounts at an early stage. The volatile organic compounds (VOCs) in cotton were analyzed by GC-MS. The plant height of cotton used in the study was 50-70 cm, and the boll period was about 12 weeks. Cotton bollworms used in the study were at second-instar. The VOCs emitted by the undamaged and damaged cotton plants detected by GC-MS were different, which indicated that electronic nose had potential in the application of cotton bollworm detection. The curve of electronic nose sensor was obtained for cotton plants infected by different numbers of cotton bollworm. Then five kinds of feature parameters were extracted from the curves of electronic nose sensors: stable value, area value, mean differential value, wavelet energy value and the coefficients of the fitted quadratic polynomial function. Feature parameters were selected based on three kinds of neural network methods: multilayer perceptron neural network (MLPNN), radial basis neural network (RBFNN) and extreme learning machine (ELM). Then stable value, area value and mean differential value were selected because of their better classification performance among the five kinds of feature parameters. Multiple-features were combinations of single-features. The classification analysis was carried out based on multiple-features and three kinds of neural network methods. And support vector machine regression (SVR) models were established based on single-features and multiple-features, respectively. The results showed that the classification performance of multiple-features was better than that of single-features. The classification performance was best based on “stable value and mean differential value” features and ELM. The classification accuracy of training set and test set based on “stable value and mean differential value” features were both 100%. The regression models based on multiple-features were better than that based on single-features. And the regression model was the best based on “area value and mean differential value” features. The coefficient of determination (R^2) and root mean square error (RMSE) of the regression model based on the training set of “area value and mean differential value” were 0.994 0 and 0.086 0. The R^2 and RMSE of the regression model based on the test set of “area value and mean differential value” were 0.923 0 and 0.370 9. The results show that feature election and multiple-features can improve the classification performance of the electronic nose for infested cotton plants. It can be concluded that electronic nose has strong potential for the application of detection of cotton plants infested with cotton bollworm at an early stage.

Keywords: electronic nose; neural network; prediction; cotton plants; cotton bollworm; feature election