

## 电子鼻检测常见伤口感染细菌实验研究

陆彦邑, 曾琳, 严博文, 黎敏, 何庆华

陆军军医大学大坪医院/国家创伤烧伤复合伤重点实验室, 重庆 400042

**【摘要】**为了进一步探讨电子鼻用于伤口细菌感染快速筛查的可能性,使用自制电子鼻检测鲍曼不动杆菌、大肠杆菌、肺炎克雷伯杆菌、金黄色葡萄球菌、铜绿假单胞菌的巯基乙酸酯(TH)培养液及纯培养液,经过预处理后,使用支持向量机、BP神经网络、逻辑回归3种算法进行分类。结果显示,使用BP算法,总体识别率可达93%以上。对于单个识别率,检测大肠杆菌、肺炎克雷伯杆菌、金黄色葡萄球菌可达98%以上,检测鲍曼不动杆菌、铜绿假单胞菌可达90%左右,表明电子鼻用于伤口感染细菌的快速检测和识别具有一定的可行性。

**【关键词】**伤口感染细菌;电子鼻;分类识别

**【中图分类号】**R318;Q93-331

**【文献标志码】**A

**【文章编号】**1005-202X(2021)10-1268-05

### Experimental study on detection of common bacteria in wound infection by electronic nose

LU Yanyi, ZENG Lin, YAN Bowen, LI Min, HE Qinghua

State Key Laboratory of Trauma, Burns and Combined Injury/Daping Hospital, Army Medical University, Chongqing 400042, China

**Abstract:** To further explore the application of electronic nose in the rapid screening of bacterial infection of wounds, a self-made electronic nose is used to detect Thioglycolate (TH) broth of *Acinetobacter baumannii* (*A. Baumannii*), *Escherichia coli* (*E. Coli*), *Klebsiella pneumoniae* (*K. pneumoniae*), *Staphylococcus aureus* (*S. aureus*), and *Pseudomonas aeruginosa* (*P. aeruginosa*) and sterile TH broth. After preprocessing, the data are classified by support vector machine, BP neural network and logistic regression, separately. The results show that BP neural network can achieve an overall recognition rate over 90%, and specific recognition rates over 98% for *E. Coli*, *K. pneumoniae* and *S. aureus*, and that its recognition rates for *A. Baumannii* and *P. aeruginosa* are around 90%, indicating that electronic nose is feasible for the rapid detection and recognition of bacteria in wound infection.

**Keywords:** bacteria in wound infection; electronic nose; classification and recognition

### 前言

电子鼻是一种利用气体传感器阵列来检测和识别物质气味的检测技术,其核心部件是各种类型的气体传感器。由多个传感器对检测物质的响应构成了传感器阵列对该气味的响应数据,不同类别的物质形成不同类别的数据,采用合适的模式识别方法可对信号进行分类和识别<sup>[1-2]</sup>。

基于传感器的多样化,电子鼻可用于医学诊断<sup>[3]</sup>、环境监测<sup>[4]</sup>、食品检测<sup>[5-6]</sup>等领域,具有广泛的应用价值。

在医学诊断领域,伤口感染检测在临床工作中极其重要,但传统伤口感染检测方法速度较慢,无法迅速确定伤口感染类型,快速有效的检测手段能极大地提高治疗效率。临床中伤口感染细菌种类繁多,常见的有大肠杆菌、金黄色葡萄球菌、铜绿假单胞菌、鲍曼不动杆菌、肺炎克雷伯杆菌等。电子鼻用于伤口感染检测已有研究,涉及到的细菌有大肠杆菌和铜绿假单胞菌<sup>[7-8]</sup>、金黄色葡萄球菌<sup>[7-10]</sup>、粪肠球菌<sup>[11]</sup>、梭菌属和脆弱拟杆菌<sup>[12]</sup>等。国内对电子鼻用于伤口细菌感染检测已有较多的研究<sup>[13-15]</sup>,主要涉及到大肠杆菌、金黄色葡萄球菌、铜绿假单胞菌3种细菌。

本文在现有的研究基础上,使用自制电子鼻检测临床中的5种常见伤口感染细菌(大肠杆菌、金黄色葡萄球菌、铜绿假单胞菌、鲍曼不动杆菌、肺炎克雷伯杆菌)的细菌培养液,并使用模式识别算法进行细菌的分类识别,以期为伤口感染的快速检测提供更多的可能性。

**【收稿日期】**2021-06-20

**【基金项目】**国家国际科技合作专项(2014DFA31560);重庆市科技研发基地建设计划(国际科技合作)项目(cstc2013ghz10003)

**【作者简介】**陆彦邑,硕士,助理研究员,主要研究方向:生物医学信号检测与处理,E-mail: 90911xbh@163.com

**【通信作者】**何庆华,博士,副研究员,主要研究方向:生物医学信号检测与处理,E-mail: qinghuahe@126.com

1 实验

1.1 实验装置

采用重庆大学自制电子鼻装置作为电子鼻检测仪器<sup>[16]</sup>,如图1所示。该电子鼻系统分为样品单元、检测单元和控制单元3个部分。传感器阵列包括1个温度传感器、1个湿度传感器、1个气压传感器、1个电压传感器和30个气体传感器(其中传感器GSBT-11已损坏)。采集过程分为基线采集、样本采集、系统清洁3个阶段,采集频率1 Hz。实验时3个阶段设置时长分别为180、180、240 s,共600 s。重复实验,以收集更多的样本。

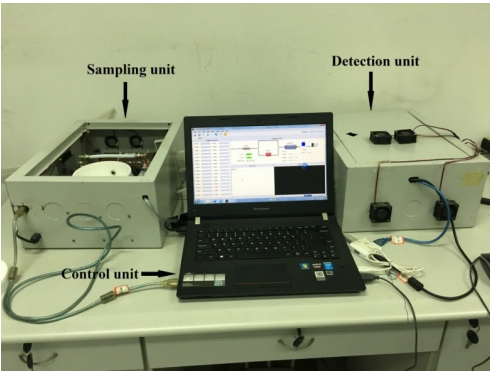


图1 自制电子鼻实验装置  
Fig.1 Experiment device of self-made electronic nose

1.2 实验样品

本实验有6种样品:使用巯基乙酸酯(Thioglycolate, TH)培养液培养的大肠杆菌、金黄色葡萄球菌、铜绿假单胞菌、鲍曼不动杆菌、肺炎克雷伯杆菌的细菌培养液及纯TH培养液,其中大肠杆菌(ATCC25922)、金黄色葡萄球菌(ATCC25923)、铜绿假单胞菌(ATCC27853)为陆军军医大学大坪医院检验科提供的ATCC标准菌株,肺炎克雷伯杆菌为陆军军医大学大坪医院检验科提供的从临床患者伤口分泌物中分离出来的菌株,鲍曼不动杆菌(ATCC19606)为陆军军医大学西南医院烧伤科提供的标准菌株。所有细菌均转种至规格为5 mL的TH培养液中培养16~20 h得到用于检测的细菌培养液。

2 数据处理

基于仪器特性,本研究所用的电子鼻典型传感器响应曲线(传感器MQ135)如图2所示,每一个传感器有一组响应曲线,共600个数据点。

传感器GSBT-11已损坏,温度传感器、湿度传感器、气压传感器及电压传感器响应较为恒定,因此剔除这5个传感器的响应数据,使用剩余29个传感器

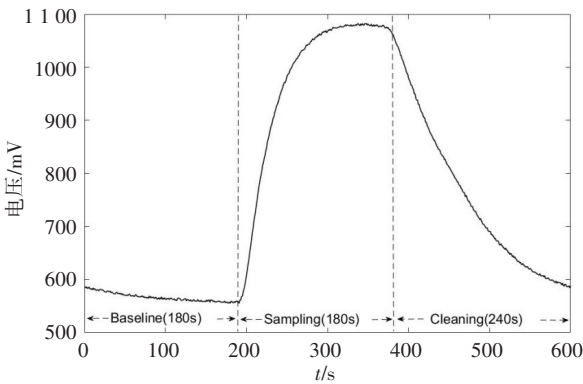


图2 典型的传感器响应曲线  
Fig.2 A typical response curve of a sensor

响应数据进行分析。每一个类别每个传感器均有一组响应数据,一组响应数据有600个数据点,因此原始样本数据维度为 $m \times 600 \times 29$ , $m$ 为收集到的样本组数,600为一组数据的长度,29为传感器个数。本研究中,每类样本各收集到200组数据, $m=1\ 200$ 。由于原始样本数据量较大且存在干扰,因此对样本数据先进行预处理再进行分类识别,数据处理流程见图3。

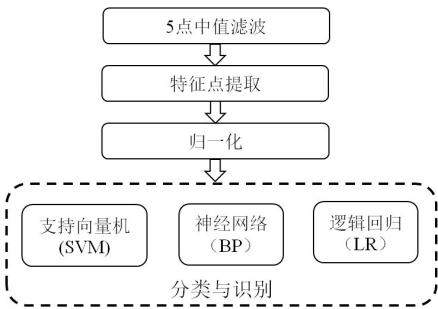


图3 数据处理流程图  
Fig.3 Flowchart of data processing

2.1 预处理

2.1.1 滤波 实验数据存在异常干扰,对于每个传感器的每一组响应数据,按传感器依次采用5点中值滤波来减小干扰。

2.1.2 特征点提取 滤波后,对于每个传感器的响应数据 $v = (v_1, v_2, \dots, v_{600})^T$ ,特征点 $x_{\text{feature}}$ 提取方法如下:

$$x_{\text{feature}} = v_{\max} - v_{\text{baseline}} \tag{1}$$

其中, $v_{\max}$ 为响应曲线中样本采集阶段的最大值, $v_{\text{baseline}}$ 为响应曲线中基线采集阶段其中一段的平均值,定义如下:

$$v_{\max} = \max_{181 \leq j \leq 360} (v_j) \tag{2}$$

$$v_{\text{baseline}} = \frac{1}{50} \sum_{j=131}^{180} (v_j) \tag{3}$$

对每个样本的一组数据,有多少个传感器就有

多少个特征点,此时一组完整的传感器相应特征点可表示如下:

$$x = (x_1, x_2, \cdots, x_n)$$

(4)

其中, $n$ 为特征点总数。

**2.1.3 归一化** 特征点提取后的样本数据,依然按组进行归一化,方法如下:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

(5)

2.2 分类与识别

经过以上处理的数据,去除了干扰,提取了特征点,可进行样本的分类和识别。使用逻辑回归(Logistic Regression, LR)、BP(Back propagation)神经网络、支持向量机(Support Vector Machine, SVM)3种算法对6种样品进行分类识别。

**2.2.1 LR** LR是一种根据预测函数 $h_{\theta}(x)$ 来实现二分类的分类算法,通过样本训练得到预测函数,后使用预测函数对未知样本进行预测和归类。 $h_{\theta}(x)$ 定义如下:

$$h_{\theta}(x) = g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}}$$

(6)

$x$ 为样本特征向量, $\theta$ 为各个特征的参数。当 $h_{\theta}(x) \geq 0.5$ 时,样本属于正类(第一类); $h_{\theta}(x) < 0.5$ 时,样本属于反类(第二类)。

模型的损失函数为:

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log(h_{\theta}(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)}))] + \frac{\lambda}{2m} \sum_{j=1}^n \theta_j^2$$

(7)

其中, $m$ 为训练样本总数, $n$ 为特征总数。为避免过拟合,引入正则化项, $\lambda$ 为正则化系数。损失函数用于评估预测的准确性,损失函数值越小,代表预测越准确。可使用梯度下降算法对 $J(\theta)$ 求偏导数,寻找最优的 $\theta$ 使得 $J(\theta)$ 最小,此时的预测函数 $h_{\theta}(x)$ 即为最优预测函数。

本研究属于多分类问题,可以使用一对多(one vs rest)的方式对每个类别训练一个二元分类器:对类别①,类别①为正类,其余为反类,搭建二分类器 $h_{\theta}^1(x)$ ;对类别②,类别②为正类,其余为反类,搭建二分类器 $h_{\theta}^2(x)$ 。对需要预测的新样本,依次使用以上二分类器预测输出,输出值最大的那组预测结果即为新样本的所属类别。

**2.2.2 BP神经网络** BP网络即反向传播网络,是目前应用非常广泛的神经网络模型之一。神经网络典型结构如图4所示,分为输入层、隐含层、输出层。BP网络的训练过程可分为前向传播和反向传播两个过程。前向传播根据输入的样本特征向量

$(x_1, x_2, \cdots, x_n)$ ,通过各隐含层的权重值 $w$ 和偏置项 $b$ 计算输出值 $(y_1, y_2, \cdots, y_c)$ ,以及输出值与实际值之间的误差项 $\delta$ ,其中, $c$ 为总类别数。若误差值在给定范围内,则网络训练完毕。若误差项不在给定的范围内,则进行反向传播,通过误差项 $\delta$ 回传给各级隐含层,重新计算各级的权重值 $w$ 和偏置项 $b$ ,数次迭代直到误差项在给定范围内,此时网络训练完毕。

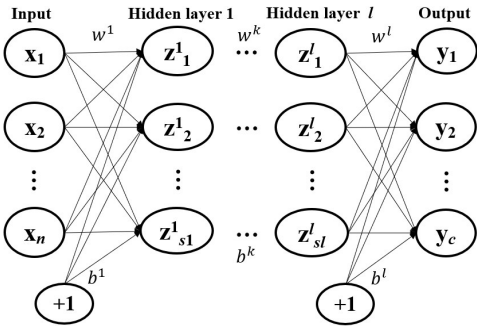


图4 BP神经网络结构  
Fig.4 Structure of BP neural network

设隐含层共有 $l$ 层,各隐含层节点数为 $(s1, s2, \cdots, sl)$ , $z^k, h^k, f(z^k)$ 分别为第 $k$ 层的节点值、输出值、激活函数, $k = (1, 2, \cdots, l)$ ,则各隐含层的节点值和输出值计算如下:

$$z_i^{(k+1)} = \sum_{j=1}^{s_k} w_{ij}^k h_j^k + b_i^k$$

(8)

$$h_i^{k+1} = f(z_i^{k+1})$$

(9)

其中, $z_i^{(k+1)}$ 为第 $(k+1)$ 层第 $i$ 个节点的值, $w_{ij}^k$ 为第 $k$ 层第 $i$ 个节点上第 $j$ 个输入的权重值, $b_i^k$ 为第 $k$ 层第 $i$ 个节点的偏置项值。给定各级网络初始权重值和偏置项值,根据以上公式即可算出各级隐含层和输出层的节点值和输出值。

设 $\hat{y} = (\hat{y}_1, \hat{y}_2, \cdots, \hat{y}_c)$ 为真实值,当输出值 $y = (y_1, y_2, \cdots, y_c)$ 与真实值的误差不在给定范围内时,通过将误差项回传给各级隐含层来更新各级权重值和偏置项,不断迭代直到误差项在给定范围内,此过程可称为网络参数的训练。

网络训练方法有多种,常见的训练方法有梯度下降法、共轭梯度法、拟牛顿算法、列文伯格-马夸尔特法(Levenberg-Marquardt, LM)法等。网络训练完毕,将待测试的样本输入网络,即可预测出其类别。

**2.2.3 SVM** SVM本身是一个二分类器,其目的是找到一个最优超平面,使两类数据点正确地分在超平面的两侧。如图5所示,样本线性可分时,可以用一条直线(在高维空间则是一个决策曲面)把两类样本分开。图5中虚线上的点称为支持向量,当两类样本的支持向量到该直线(决策曲面)的间隔 $d$ 最大时,该直线(决策曲



面)即为最优超平面。决策曲面定义如下:

$$\boldsymbol{w}^T \boldsymbol{x} + b = 0$$

(10)

其中,  $\boldsymbol{x}$  为样本特征向量;  $\boldsymbol{w}$  为可调权值;  $b$  为偏置, 代表决策面相对于原点的偏移。求解决策曲面的过程即是寻找最优超平面的过程。经过计算可得间隔  $d$ :

$$d = \frac{1}{\|\boldsymbol{w}\|}$$

(11)

因此, 最大化  $d$  等价于最小化权值向量  $\boldsymbol{w}$  的欧几里得范数  $\|\boldsymbol{w}\|$ 。最终计算可得:

$$\boldsymbol{w} = \sum_{i=1}^n \alpha_i y_i \boldsymbol{x}_i$$

(12)

其中,  $\alpha_i$  是拉格朗日乘子;  $\boldsymbol{x}_i$  表示特征向量  $\boldsymbol{x}$  中某一个特征值,  $i = (1, 2, \cdots, n)$ ,  $n$  表示特征向量维度;  $y_i$  表示相应类的标识,  $y_i$  为 +1 时代表第一类(正类),  $y_i$  为 -1 时代表第二类(反类)。式(12)作为 SVM 的支持向量和最优超平面分类器。多数情况下, 样本是线性不可分的, 此时可以通过核函数将样本映射到高维空间, 变为线性可分样本再进行分类。常见的核函数有线性核、多项式核、径向基函数核(Radial Basis Function, RBF)以及 sigmoid 核。

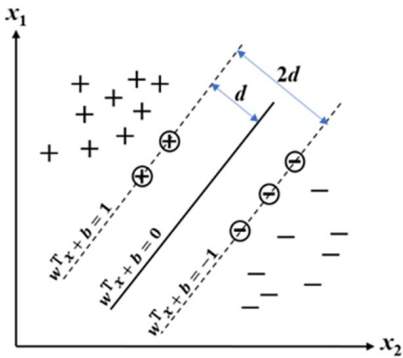


图5 支持向量机原理图  
Fig.5 Schematic diagram of support vector machine

当使用 SVM 用于多分类时, 本文采用一对一的方式进行分类。其实现方式是用 SVM 对任意两类设计一个二分类器, 从而得到  $c(c-1)/2$  个二分类器,  $c$  为总类别数。对于一个需要分类的样本, 用所有二分类器对其进行预测, 得票最多的类别即为该样本所属类别。

3 结果与讨论

3.1 样品检测结果

在同一实验条件下, 使用电子鼻检测 6 种样品, 6 种样品的典型电子鼻雷达图谱如图 6 所示, 可看出 6 种样品各有其图谱特征。

3.2 数据处理结果

6 类样品每类共收集到 200 组样本数据, 经过滤波、特征值提取和归一化的预处理后得到了  $200 \times 29$  的特征值矩阵。

对于每类的 200 组样本数据, 随机选取其中 100 组为训练集, 剩余 100 组为测试集, 分别使用 SVM、BP、LR 来进行分类识别。经过多次实验可分别得到 3 种算法的最优参数: BP 选择一层隐含层, 隐含层节点数为输入层的 1.5 倍, 隐含层和输出层的激活函数分别为 tan-sigmoid 和 purelin, 网络训练方法为 LM 算法; SVM 算法的核函数类型为 RBF 核; 为避免拟合, LR 引入正则化参数, 参数最优值为 1。由于训练样本为随机选取, 因此每一次运行识别率略有不同, 运行 10 次取平均可得平均识别率, 见表 1。

3 种算法中, 总体识别率 BP 最高(93.18%), SVM 次之(86.52%), LR 最低(72.78%)。对于单个类别的识别率, BP 依然最高。从单个类别识别率可看出, BP 和 SVM 对大肠杆菌、肺炎克雷伯杆菌和金黄色葡萄球菌具有良好的识别效果, 能达到 98% 以上。鲍

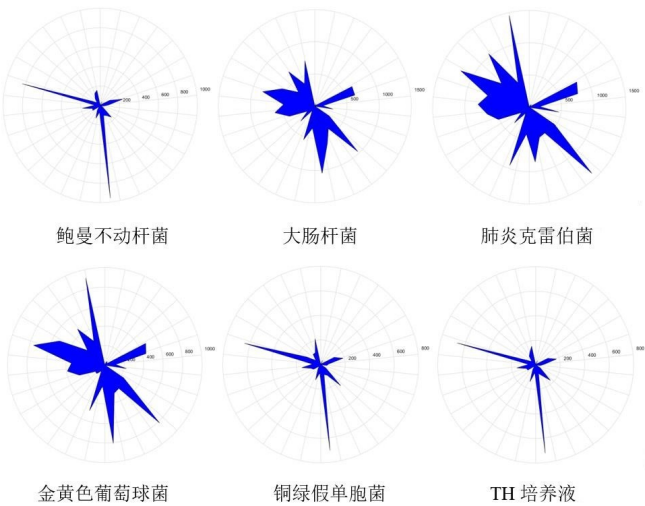


图6 6种样品电子鼻雷达图谱  
Fig.6 Radar patterns of 6 types of samples detected by electric nose

表1 3种算法平均识别率(%)

Tab.1 Average recognition rates of 3 algorithms (%)

检测项目	BP神经网络	支持向量机	逻辑回归
鲍曼不动杆菌	90.10	75.60	64.50
大肠杆菌	99.30	100.00	100.00
肺炎克雷伯杆菌	99.00	98.10	86.90
金黄色葡萄球菌	99.90	99.70	89.50
铜绿假单胞菌	88.30	77.60	34.10
纯TH培养液	82.50	68.10	61.70
总体识别率	93.18	86.52	72.78

曼不动杆菌和铜绿假单胞菌识别率不如以上3种细菌,但使用BP算法也能达到88%以上。以上结果表明,对于常见伤口感染细菌的培养液,使用BP算法识别率和可分性更高。

本研究为电子鼻用于快速筛查常见伤口感染细菌类型提供了一定的实验基础和可行性,但也有其局限性,需改进和继续研究的地方有以下两个方面。一是传感器数量较多,可能存在冗余,需要筛选和优化传感器组合以减小传感器部分体积和复杂度。二是本研究只对细菌培养液进行了检测,实际的临床伤口细菌感染是复杂且多样的,需进一步开展临床样本研究,探索更多的可能性。

4 结 论

本研究使用自制电子鼻检测大肠杆菌、金黄色葡萄球菌、铜绿假单胞菌、鲍曼不动杆菌、肺炎克雷伯杆菌细菌培养液和纯TH培养液共6种样品,得到具有各自图谱特征电子鼻雷达图谱。对6类样本数据进行预处理和提取特征后,使用SVM、BP、LR 3种算法进行了分类识别。结果表明,BP算法识别效果最好,SVM次之,LR最低,提示BP算法对大肠杆菌、肺炎克雷伯杆菌、金黄色葡萄球菌、鲍曼不动杆菌、铜绿假单胞菌的细菌培养液具有良好的可分性,为电子鼻用于伤口细菌感染类型的快速筛查进一步提供了可能性。

【参考文献】

[1] GARDNER J W, BARTLETT P N. A brief history of electronic noses [J]. *Sensor Actuat B-Chem*, 1994, 18-19(1): 210-211.

[2] HINES E L, LIOBET E, GARDNER J W. Electronic noses: a review of signal processing techniques[J]. *IEE P-Circ Dev Syst*, 1999, 146(6): 297-310.

[3] WILSON A D, BAIETTO M. Advances in electronic-nose technologies developed for biomedical applications[J]. *Sensors*, 2011, 11(1): 1105-1176.

[4] ROMAIN A C, NICOLAS J. Long term stability of metal oxide-based gas sensors for E-nose environmental applications: an overview[J]. *Sensor Actuat B-Chem*, 2010, 146(2): 502-506.

[5] KOREL F, BALABAN M O. Microbial and sensory assessment of milk with an electronic nose[J]. *J Food Sci*, 2002, 67(2): 758-764.

[6] WANG D F, WANG X C, LIU T A, et al. Prediction of total viable counts on chilled pork using an electronic nose combined with support vector machine[J]. *Meat Sci*, 2012, 90(2): 373-377.

[7] ABDULLAH A A, YUSUF N, AMMAR Z, et al. Bacteria classification using electronic nose for diabetic wound monitoring[J]. *Appl Mech Mater*, 2013, 339: 167-172.

[8] YUSUF N, ZAKARIA A, OMAR M I, et al. *In-vitro* diagnosis of single and poly microbial species targeted for diabetic foot infection using E-nose technology[J]. *BMC Bioinformatics*, 2015, 16(1): 158.

[9] DUTTA R, MORGAN D, BAKER N, et al. Identification of staphylococcus aureus infections in hospital environment: electronic nose based approach[J]. *Sensor Actuat B-Chem*, 2005, 109(2): 355-362.

[10] DUTTA R, DUTTA R. "Maximum probability rule" based classification of MRSA infections in hospital environment: using electronic nose [J]. *Sensor Actuat B-Chem*, 2006, 120(1): 156-165.

[11] CAREY J R, SUSLICK K S, HULKOWER K I, et al. Rapid identification of bacteria with a disposable colorimetric sensing array [J]. *J Am Chem Soc*, 2011, 133(19): 7571-7576.

[12] PAVLOU A, TURNER A P, MAGAN N. Recognition of anaerobic bacterial isolates *in vitro* using electronic nose technology[J]. *Lett Appl Microbiol*, 2002, 35(5): 366-369.

[13] 贾鹏飞. 面向伤口感染检测的电子鼻智能数据处理算法研究[D]. 重庆: 重庆大学, 2014.

JIA P F. Research of intelligent data processing algorithms for detection of wound infection based on electronic nose [D]. Chongqing: Chongqing University, 2014.

[14] 吕博. 用于伤口感染细菌检测的医用电子鼻研究[D]. 重庆: 重庆大学, 2016.

LÜ B. Study of medical electronic nose for the detection of bacteria in wound infection[D]. Chongqing: Chongqing University, 2016.

[15] 胥勤涛, 田逢春, 闫嘉, 等. 结合气体浓缩的电子鼻伤口病原菌快速检测[J]. *传感技术学报*, 2009, 22(3): 303-306.

XU X T, TIAN F C, YAN J, et al. Rapid detection of wound pathogen by enose with a gas condensation unit[J]. *Chinese Journal of Sensors and Actuators*, 2009, 22(3): 303-306.

[16] SUN H, TIAN F C, LIANG Z F, et al. Sensor array optimization of electronic nose for detection of bacteria in wound infection [J]. *IEEE Trans Ind Electron*, 2017, 64(9): 7350-7358.

(编辑:薛泽玲)