

基于生物阻抗谱的细胞悬浮液浓度识别方法研究

刘圣龙 杨璐 朱程君 刘凯 韩伟 姚佳烽

A method of identifying cell suspension concentration based on bioimpedance spectroscopy

Liu Sheng-Long Yang Lu Zhu Cheng-Jun Liu Kai Han Wei Yao Jia-Feng

引用信息 Citation: *Acta Physica Sinica*, 71, 078701 (2022) DOI: 10.7498/aps.71.20211837

在线阅读 View online: <https://doi.org/10.7498/aps.71.20211837>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

您可能感兴趣的其他文章

Articles you may be interested in

基于生物阻抗谱的细胞电学特性研究

Electrical characteristics of cells with electrical impedance spectroscopy

物理学报. 2020, 69(16): 163301 <https://doi.org/10.7498/aps.69.20200601>

基于生物阻抗谱的舌体肿瘤组织识别方法

Tongue tumor tissue recognition based on bioelectrical impedance spectroscopy

物理学报. 2021, 70(15): 158704 <https://doi.org/10.7498/aps.70.20210297>

基于生物阻抗谱成像的生物组织检测方法

Biological tissue detection based on electrical impedance spectroscopic tomography

物理学报. 2022, 71(4): 048706 <https://doi.org/10.7498/aps.71.20211600>

基于标准样品回归算法和腔增强光谱的NO₂检测方法

NO₂ gas detection based on standard sample regression algorithm and cavity enhanced spectroscopy

物理学报. 2021, 70(5): 050702 <https://doi.org/10.7498/aps.70.20201322>

红外及太赫兹辐照下细胞膜生物效应的研究进展

Research progress of biological effects of cell membrane under infrared and terahertz irradiation

物理学报. 2021, 70(24): 248707 <https://doi.org/10.7498/aps.70.20212030>

稳态强磁场的细胞生物学效应

Biological effects on cells in strong static magnetic field

物理学报. 2018, 67(14): 148701 <https://doi.org/10.7498/aps.67.20180378>

基于生物阻抗谱的细胞悬浮液浓度识别方法研究*

刘圣龙¹⁾ 杨璐¹⁾ 朱程君²⁾ 刘凯¹⁾ 韩伟³⁾ 姚佳烽^{1)†}

1) (南京航空航天大学机电学院, 南京 210016)

2) (南京医科大学第一附属医院肿瘤科, 南京 210029)

3) (南京大学医学院附属口腔医院, 南京 210008)

(2021年10月4日收到; 2021年12月6日收到修改稿)

基于生物阻抗谱技术提出一种细胞悬浮液浓度自动识别方法, 该方法结合了多元线性回归算法和生物阻抗谱技术, 能够快速准确地识别细胞悬浮液的浓度. 首先, 提出一种细胞位置随机分布策略, 模拟细胞的真实存在状态; 其次, 采用数值仿真的方法生成 2400 组不同浓度的正常、癌变以及混合的细胞模型并计算生物阻抗谱数据; 然后, 利用多元线性回归、支持向量机和梯度提升三种回归算法分别对癌变细胞浓度进行鉴别, 仿真结果表明, 多元线性回归算法为最佳回归模型, 其平均拟合优度和均方误差分别是 0.9997 和 0.0008; 最后, 将多元线性回归算法应用于不同浓度的红细胞悬浮液的识别中, 实验结果显示其平均拟合优度和均方误差分别是 0.9998 和 0.0079, 说明该方法具有较高的细胞悬浮液浓度识别能力.

关键词: 细胞浓度检测, 生物阻抗谱, 多元线性回归, 细胞随机分布策略

PACS: 87.55.dk, 87.19.xj, 07.50.Qx

DOI: 10.7498/aps.71.20211837

1 引言

随着现代医学的不断发展, 医学检测方法已经进入细胞层次, 其中细胞浓度在病情分析^[1]、器官培养^[2,3]、血型血清学试验^[4]以及肿瘤细胞黏附^[5]等方面具有重要作用, 已经引起许多研究者的关注. 目前, 一些成熟的细胞浓度检测方法已经在临床上应用, 这些检测方法各有优点, 但同时存在不足之处. 例如, 免疫细胞化学法 (immunocytochemistry, ICC)^[6]可以准确地检测细胞浓度, 但检测速度慢、操作复杂; 多聚酶链反应法 (polymerase chain reaction, PCR)^[7]操作方便、灵敏度高, 但其应用条件苛刻; 流式细胞术 (flow cytometry, FCM)^[8]可以快速准确地检测细胞浓度, 但需要对被测细胞进行荧光标记. 近些年, 一些新的细胞浓度检测方法被研究学者提出. 2009年 Nilsson 等^[9]提出了一

种基于散斑效应的细胞浓度检测方法, 该方法可以在体外快速地检测细胞浓度, 并且可以观察细胞的分布情况, 但主要应用于红细胞的检测, 并且易受到生理因素的干扰. 2014年 Guo 等^[10]利用光的折射率原理, 设计了一款 TFBG 折射率传感器, 该传感器实现了对细胞浓度的高精度测量, 但可以检测的细胞浓度范围有限, 并且灵敏度较低. 总之, 现有的细胞浓度检测技术都存在一定的局限性, 为了进一步普及细胞浓度在医学检测方面的应用, 临床上急需一种高精度、低成本且适用于连续监测的全新检测方法.

生物阻抗谱 (bioimpedance spectroscopy, BIS) 技术是一种以多频率、复阻抗为基础的生理参数检测技术^[11]. 该技术利用一对或多对电极向被测物施加微小幅值的正弦电流或电压激励, 通过采集被测物在各个频率分量上的响应信号, 经数字解调后得到被测物完整的响应频谱. 对测量得到的阻抗频

* 国家自然科学基金 (批准号: 62071224) 和精密测试及仪器国家重点实验室开放基金 (批准号: pilab2107) 资助的课题.

† 通信作者. E-mail: jiaf.yao@nuaa.edu.cn

谱进行分析, 可以获取到诸如细胞尺寸大小、种类、存活状态等丰富的生理化学信息^[12].

由于 BIS 采用非侵入式测量方法, 无需使用荧光标记, 并且可以反映细胞微观尺度下的电学特性, 被认为是未来最具潜力的疾病早期诊断手段之一^[13]. 近些年来 BIS 在肿瘤细胞早期诊断^[14]、组织缺血监测^[15]、乳腺癌淋巴水肿探测^[16] 等方面取得了很多有价值的研究成果.

此外, 不同浓度的细胞溶液, 其电学特性有明显差异, BIS 可以从电学特性的差异性中提取有效的电信息, 用来鉴别细胞的浓度. 但是 BIS 只能定性分析不同浓度的细胞溶液, 不能准确对细胞样本进行判断. 因此, 为了定量检测出细胞的浓度, 考虑使用机器学习算法结合生物阻抗谱技术进行细胞浓度的准确检测.

人工智能领域应用的机器学习算法在人脸识别^[17,18]、自然语言处理^[19] 等领域都表现出良好的适应性. 其中各种回归算法也被广泛应用到社会的各个场合. 2016 年武汉大学刘泉声等^[20] 将支持向量机回归算法 (support vector regression, SVR) 应用到地应力场的反演中, 并且反演得到的规律经验证准确可靠. 2017 年 Keprate 等^[21] 利用梯度提升回归算法 (gradient boosting regressor, GBR) 预测小口径管道中裂纹扩展的应力强度因子, 并且具有较高的预测精度和较少的计算时间. 2020 年长安大学王征征等^[22] 运用主成分分析法 (principal component analysis, PCA) 和多元线性回归算法 (multiple linear regression, MLR) 预测出生人口, 极大地提高了出生人口的预测精度且具有一定的理论参考意义.

由于细胞在悬浮液中是随机分布的, 而细胞不同的位置会影响整体的电场分布特性, 这会造成多次测量同一浓度细胞的悬浮液, 其生物阻抗谱数据有较大波动, 进一步地造成数据集样本与对应标签的混乱, 最终大大地影响了分类算法的精度. 通过建立合理的仿真模型, 可以节省大量的时间成本; 而目前急需利用一个模拟细胞随机分布的方法, 建立合理的仿真模型, 使仿真生成与真实细胞存在状态相似的生物阻抗谱数据集, 继而找到适合用于处理真实数据的回归模型.

本文首先在数值仿真时, 利用提出的细胞随机分布策略, 模拟细胞的真实分布, 生成 2400 组不同浓度的正常、癌变以及混合的细胞模型并计算相

应的生物阻抗谱数据; 然后利用 SVR, GBR 和 MLR 三种回归算法对仿真得到的数据集进行训练, 结果表明, MLR 是最优的回归算法; 最后在实验中通过 MLR 对不同浓度的红细胞悬浮液进行训练建模和定量分析, 结果表明, 该方法可以准确地识别细胞悬浮液浓度, 可以为医学检测细胞浓度提供一种全新的方法.

2 仿真模型与方法

2.1 双壳细胞仿真模型

免疫 B 细胞是指在淋巴细胞中的抗体所形成的细胞前体, 具有中和毒素、抗感染以及免疫调理等功能. 免疫 B 细胞是人体内具有代表性的免疫细胞, 所以仿真对象选取人体免疫 B 细胞来验证本文提出的方法^[23]. 如表 1 所列, 人体免疫 B 细胞主要由细胞膜、细胞质、细胞核组成, 细胞核又由核膜和核质组成, 细胞膜的电导率为 5.6×10^{-5} S/m, 相对介电常数为 12.8, 核膜的电导率为 1.11×10^{-2} S/m, 相对介电常数为 106. 细胞悬浮液可等效为电阻和电容元件的组合^[24], 在低频情况下, 电阻成分发挥主要作用, 由于细胞膜的导电性能低, 因此, 可阻止大部分电场线穿过细胞; 而在高频情况下, 电容成分发挥主要作用, 细胞膜的介电常数显著下降, 对电场线的屏蔽效应降低, 但细胞核膜的介电常数较高, 可阻止部分电场线穿过细胞核. 为了研究不同种类、浓度的细胞悬浮液对阻抗谱的影响, 建立了正常 B 细胞和癌变 B 细胞的仿真模型, 如图 1 所示. 仿真所用到的全部参数如表 1 所列, 仿真中设置电极两端的激励电流为 1 mA, 设置激励频率范围为 1 Hz—1 GHz.

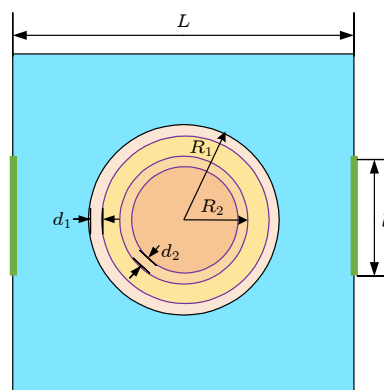


图 1 细胞仿真模型

Fig. 1. Cell simulation model.

表 1 仿真参数汇总^[23]

Table 1. Summary of simulation parameters.

| | 参数 | 正常B细胞 | 癌变B细胞 |
|-------------------------------------------|-------------------|-----------------------|----------------------|
| 电导率 $\sigma/(\text{S}\cdot\text{m}^{-1})$ | 背景溶液 | 0.6 | 0.6 |
| | 细胞膜 | 5.6×10^{-5} | 9.1×10^{-6} |
| | 细胞质 | 1.31 | 0.48 |
| | 核膜 | 1.11×10^{-2} | 4.4×10^{-3} |
| | 核质 | 2.04 | 1.07 |
| 相对介电常数 ϵ | 背景溶液 | 80 | 80 |
| | 细胞膜 | 12.8 | 9.8 |
| | 细胞质 | 60 | 60 |
| | 核膜 | 106 | 60.3 |
| | 核质 | 120 | 120 |
| 几何参数/ μm | 仿真区域 $L \times L$ | 20×20 | 20×20 |
| | 电极长度 l | 4 | 4 |
| | 细胞半径 R_1 | 3.3 | 5.2 |
| | 细胞核半径 R_2 | 2.8 | 4.4 |
| | 细胞膜厚 d_1 | 0.007 | 0.007 |
| | 核膜厚度 d_2 | 0.04 | 0.04 |

2.2 细胞位置随机分布策略

细胞浓度仿真需要在选定区域内进行大量的细胞建模, 并且利用编程将繁杂的建模过程交给计算机自动完成, 缩短仿真周期. 其中仿真区域边长设为 $L_1 = 110 \mu\text{m}$, 然后将仿真区域均匀划分为 10×10 共 100 个网格, 每个小网格的边长为 $l_s = 11 \mu\text{m}$, 该网格可以同时兼容正常和癌变细胞的尺寸, 以满足两种细胞混合仿真的需要. 仿真区域所能容纳的最大细胞数 $n_{\text{max}} = 100$, 细胞浓度用仿真时的细胞数量 n_{sim} 相对最大细胞数量的百分比

O 表示, 即 $O = n_{\text{sim}}/n_{\text{max}} \times 100\%$.

为了模拟真实细胞在悬浮液中的随机分布情况, 本文提出一种位置随机分布策略. 具体如下: MATLAB 具有库函数 `randperm` 和 `reshape`, `randperm` 函数可以生成随机打乱的数字序列, `reshape` 函数可以对矩阵的元素进行重新排列. 因此利用 `randperm` 函数生成了 1—100 自然数的随机数列, 将生成模型的细胞浓度作为判断阈值, 随机数列中小于该阈值的所有位置重新赋值为 1, 剩余位置赋值为 0; 然后通过函数 `reshape` 将该随机数列重塑为一个 10×10 的矩阵, 与仿真区域的网格对应, 矩阵中数值为 1 的位置对应的仿真区域进行细胞建模, 数值为 0 的位置则不进行任何操作. 图 2 展示了 $O_{\text{总}} = 10\%$ 而 $O_{\text{癌}} = 3\%$ 时的某次细胞位置随机分布情况.

2.3 数据集的准备与预处理

机器学习方法可以从人眼无法区分的相似特征数据集中寻找背后的规律, 以便实现更为复杂的回归鉴别工作. 因此, 本文选择机器学习方法来实现不同细胞组浓度的识别, 具体过程如下:

1) 正常细胞组: 此仿真组中, 仿真区域放置的全部为正常细胞, 细胞个数 1 到 100 等间隔划分为 100 组, 每组在位置随机分布策略下仿真 12 次, 共得到 1200 条样本;

2) 癌变细胞组: 此仿真组中, 仿真区域放置的全部为癌变细胞, 细胞个数从 1 到 10 等间隔划分为 10 组, 每组在位置随机分布策略下仿真 20 次, 共得到 200 条样本;

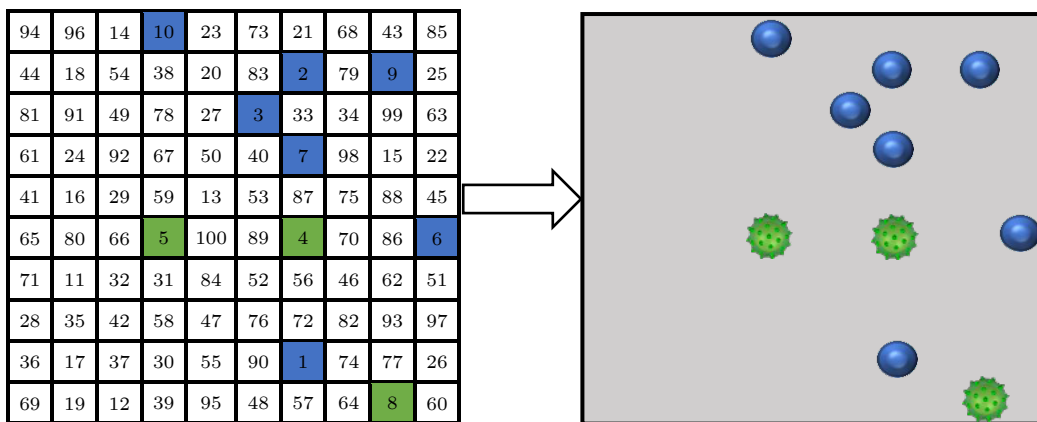


图 2 细胞位置随机分布策略示意图. 蓝色表示正常细胞, 绿色表示癌变细胞

Fig. 2. Schematic diagram of random distribution strategy of cell location. Blue indicates normal cells, and green indicates cancerous cells.

3) 混合细胞组: 此仿真组将两种细胞按不同比例混合, 细胞总数从 10 到 100 等间隔划分为 10 组, 每组内部按癌变细胞个数从 1 到 10 等间隔划分为 10 组, 每个小组内部按位置随机分布策略仿真 10 次, 共得到 1000 条样本。

上述三个仿真组共得到 2400 个样本数据, 首先针对癌变细胞作出浓度划分, 然后作出癌变细胞浓度为 1% 时的 5 次随机位置的 Nyquist 图, 如图 3(a) 所示, 从图中可以得知, 相同浓度的癌变细胞, 其位置的不同, 会严重影响生物阻抗; 最后作出其部分癌变细胞浓度的 Nyquist 图, 如图 3(b) 所示, 从图中可以看出, 其数据集存在比较严重的混杂, 人眼无法识别。另外每个样本数据有多个复阻抗值, 若全部作为特征用于训练, 则会十分耗时, 因此对多个复阻抗值进行系统抽样; 另外复阻抗值的实部和虚部反映待测物的电学特性 (如电导率、相对介电常数) 的比重不同, 所以将复阻抗值的实部和虚部分离, 并分别作为训练特征。

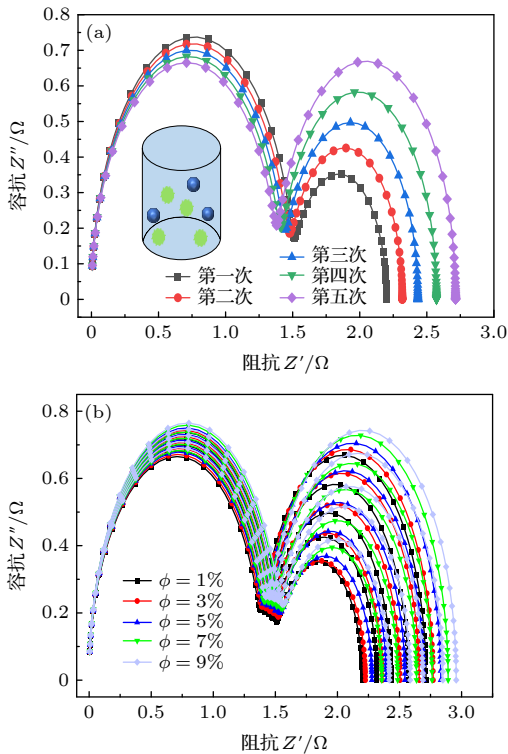


图 3 部分浓度癌变细胞的 Nyquist 图 (a) 癌变细胞浓度为 1% 时的 5 次细胞随机分布的 Nyquist 图; (b) 不同癌变细胞浓度的 Nyquist 图

Fig. 3. Nyquist plot of cancerous cells at partial concentrations: (a) Nyquist plots of randomly distributed cells at 5 times when the concentration of cancerous cells is 1%; (b) Nyquist plots of different cancerous cell concentrations.

2.4 基于机器学习的回归模型

1) 支持向量机回归算法

支持向量机 (support vector machine, SVM) 的基本目的是在给定的特征空间中寻找一个最佳的分离超平面, 使得两类数据正确分离, 并且间隔最大. SVM 还支持使用核函数 $\Phi(x)$ 用来解决回归问题, 即 SVR^[25,26]. SVM 和 SVR 的原理基本相同. 其不同之处在于 SVR 是回归模型, 没有类别, 其基本目的: 让训练集中的每个样本点 (x_i, y_i) , 尽量拟合到一个线性模型. SVR 算法适用了样本容量小、非线性的问题, 并且在引入松弛变量后, 提高了模型的泛化能力。

2) 梯度提升回归算法

GBR 是一种集成学习算法, 主要由损失函数、弱学习器和加法模型三部分组成^[21]. 该算法的根本思想: 通过多个弱学习器依次迭代并且拟合之前模型累加的损失函数的负梯度, 总的损失函数朝着负梯度的方向减少. GBR 通过集成大量表现不好的学习算法, 组成一个比较强大的学习算法, 使其能够同时处理多种类型的数据集、具有较高的预测能力, 但是由于多个弱学习器必须串行处理, 也导致其训练速度过慢。

3) 多元线性回归算法

MLR 是研究某一因变量与两个或两个以上自变量之间的关系, 是一种成熟的定量分析方法^[27]. 因此多元线性拟合回归方程可表示为

$$y_i = \sum_{j=1}^n w_j x_{i,j} + b, \quad (1)$$

其中 $y_i (i = 1, 2, \dots, m)$ 为第 i 个样本的标签; $x_{i,j} (j = 1, 2, \dots, n)$ 为第 i 个样本的第 j 个特征; w_j 为对应特征的权重; b 为偏置量. 通常为了简化表达式, 将偏置量 b 归纳到向量 w 中, 即有表达式:

$$y_i = \sum_{j=1}^{n+1} w_j x_{i,j}. \quad (2)$$

则回归模型可表示为

$$y = xw. \quad (3)$$

利用最小二乘法求得最优参数 w , 即确定了回归模型. MLR 原理简单、容易理解, 建模速度很快, 能够得到明确的数学解, 有利于决策分析; 但是不适合对具有很强相关性的数据特征集建模。

2.5 基于回归模型的仿真结果

利用 2.4 节介绍三种回归模型分别用于 2.3 节仿真数据集的回归建模, 并且选择一个最佳的回归算法用于下文实验数据的建模处理. 事先对各个模型进行调参, 使它们在最佳的状态进行比较, 并引入均方误差 (mean-square error, MSE) 和拟合优度 (R^2) 作为评价指标. 将数据集按照 7:3 的比例划分成训练集和测试集, 用于上述三种回归模型的训练, 得到的结果如图 4 所示. 由图 4 可知, 无论在训练集还是在测试集中结果均为 $R^2_{\text{MLR}} > R^2_{\text{GBR}} > R^2_{\text{SVR}}$, $\text{MSE}_{\text{MLR}} < \text{MSE}_{\text{GBR}} < \text{MSE}_{\text{SVR}}$; 说明 MLR 是最佳的回归模型. 为了更形象地比较三种回归模型的优劣, 特地采用系统抽样的方法选取 80 组数据, 并比较三种回归模型的预测值与真实值的接近程度, 结果如图 5 所示, 可以直观地看出 MLR 的预测值更接近真实值.

由图 4 可知, 三种回归模型的 R^2 值十分接近, 为了避免偶然误差的影响, 又通过五折交叉验证求取平均值的方法, 再次验证阻抗谱数据集更适合哪种回归模型. 由表 2 可知: 五折交叉验证的结果与之前调参后的结果基本一致, 其中 MLR 的结果十

分稳定, 平均拟合优度 R^2 和均方误差 MSE 分别达到 0.9997 和 0.0008; 而 SVR 和 GBR 的结果有轻微波动, 但其平均拟合优度 R^2 和均方误差 MSE 也分别达到 0.9974, 0.0302 和 0.9988, 0.0150. 本文选取的三种回归算法表现都较为优异, SVR 适合非线性程度比较高的回归问题, GBR 擅长于解决数据规律不明显的回归问题, 可归纳总结大量样本的潜在信息, 而 MLR 更适合解决线性回归问题. 在细胞浓度检测这一问题上, MLR 表现较好, 说明本文所要解决的回归问题更接近于线性回归. 因此将 MLR 用于下文的实验数据处理.

3 实验与验证

3.1 实验设备

本文分别用不同浓度的红细胞悬浮液作为实验对象. 图 6(a) 给出了实验设备原理图: 一台 PC 机、一台阻抗分析仪 (IM3570)、一个屏蔽装置和一个传感器. 图 6(b) 是传感器的结构图, 由一对平行的测量电极嵌入亚克力容器侧面组成. 测量电极长度 $d_c = 10 \text{ mm}$, 高度 $h_c = 20 \text{ mm}$, 极板间宽

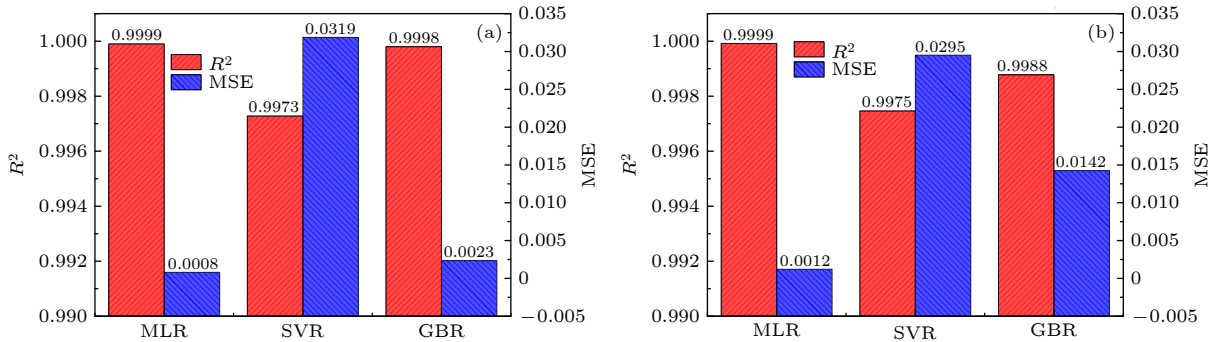


图 4 三种回归模型对仿真数据集的训练结果 (a) 训练集的训练结果; (b) 测试集的预测结果

Fig. 4. The training results of the three regression models on the simulation data set: (a) Training results of the training set; (b) predicted results on the test set.

表 2 仿真数据的三种回归算法五折验证结果

Table 2. Validation results of three regression algorithms for simulation data.

| | MLR | | SVR | | GBR | |
|-----|--------|--------|--------|--------|--------|--------|
| | R^2 | MSE | R^2 | MSE | R^2 | MSE |
| 1 | 0.9999 | 0.0006 | 0.0006 | 0.0257 | 0.9994 | 0.0066 |
| 2 | 0.9995 | 0.0009 | 0.0009 | 0.0313 | 0.9985 | 0.0171 |
| 3 | 0.9998 | 0.0006 | 0.0006 | 0.0336 | 0.9986 | 0.0179 |
| 4 | 0.9999 | 0.0012 | 0.0012 | 0.0267 | 0.9986 | 0.0156 |
| 5 | 0.9996 | 0.0008 | 0.0008 | 0.0335 | 0.9987 | 0.0176 |
| 平均数 | 0.9997 | 0.0008 | 0.0008 | 0.0302 | 0.9988 | 0.0150 |

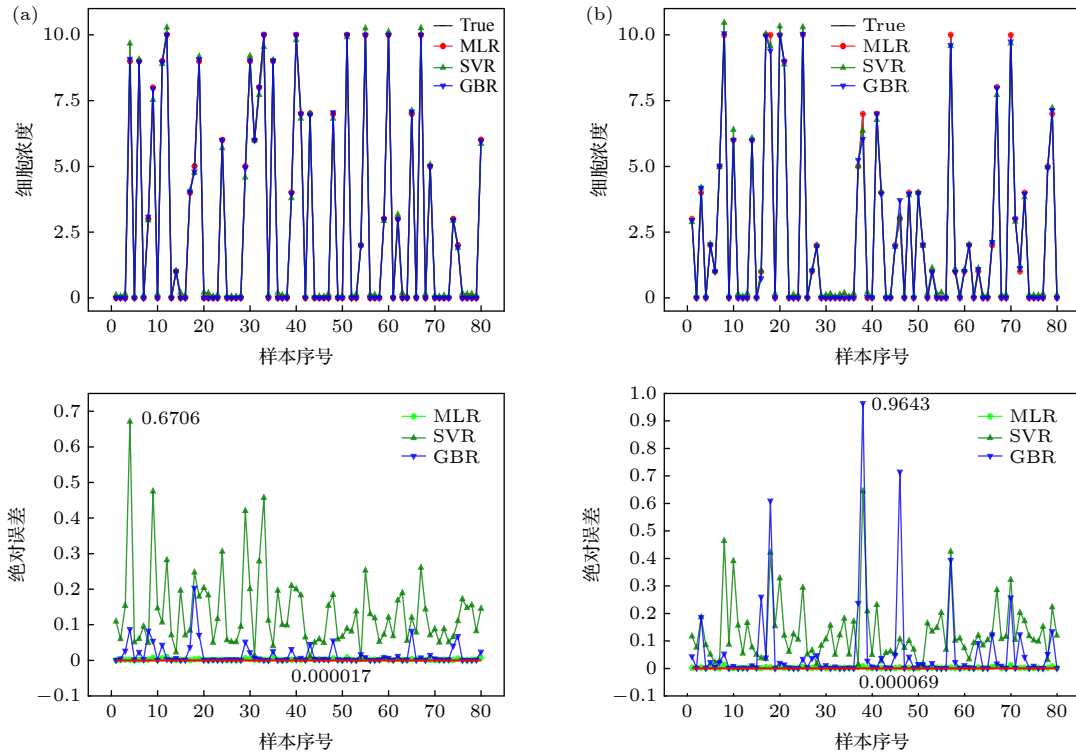


图 5 三种回归模型对部分数据预测值与真实值的差异 (a) 训练集预测值与真实值的差异; (b) 测试集预测值与真实值的差异
 Fig. 5. The difference between the predicted value and the true value of partial data by three regression models: (a) The difference between the predicted value of the training set and the true value; (b) the difference between the predicted value of the test set and the true value.

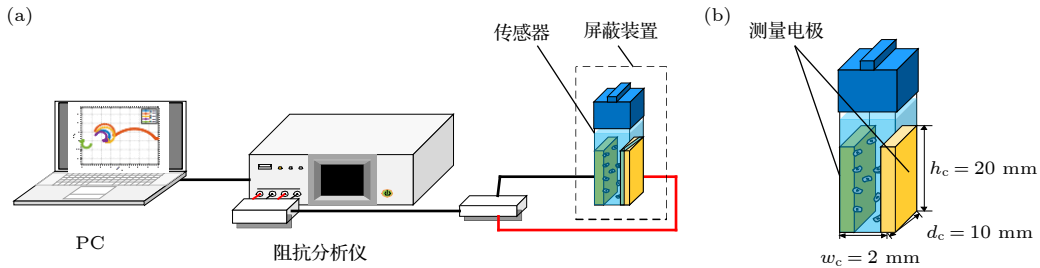


图 6 生物阻抗谱检测实验设备 (a) 实验设备原理图; (b) 传感器结构图

Fig. 6. Bioelectrical impedance spectroscopy instrument: (a) Schematic diagram of experimental equipment; (b) sensor structure diagram.

度 $w_c = 2 \text{ mm}$, 单次测量可容纳的细胞悬浮液体积 $V \approx 0.4 \text{ mL}$. 将传感器置于屏蔽装置中进行测量, 隔离周围环境的电磁干扰, 所述的屏蔽装置是由铁质容器构成并且需要接地, 而且引入平均相对测量误差 s 作为评价屏蔽性能的指标, 其计算公式 (4) 式所示, 使用屏蔽装置后, 连续两次测量阻抗谱的平均相对测量误差 s 为 0.1%, 不使用屏蔽装置的平均相对测量误差 s 为 0.22%, s 显著降低. 传感器通过屏蔽线与阻抗分析仪进行连接; 阻抗分析仪捕获到探头发出的信号后, 将测量数据传输给 PC 端进行处理. 频率测量范围为 $f = 1 \text{ kHz}$ —

5 MHz, 测量点数量为 190 个. 然后选取一定数量的频点上的阻抗信息用于训练.

$$s = \frac{1}{n} \sum_{i=1}^n \left| \frac{m_1^i - m_2^i}{m_1^i + m_2^i} \right| \times 100\%, \quad (4)$$

其中 n 表示测量点的数量; m_1^i, m_2^i 分别表示连续两次所测阻抗谱的第 i 个频率点的阻抗幅值.

3.2 实验过程

制备了不同浓度的红细胞悬浮液, 其制备过程如下: 小鼠在全麻条件下, 取血液约 1 mL, 置于抗

凝管中, 向血液中加入等体积的生理盐水混匀, 然后在温度为 4 °C、转速为 1500 r/m 的条件下离心 5 min, 结束后, 弃去上清液中间白色细胞层, 再用生理盐水重悬红细胞; 重复 5 次, 以去除血液中的纤维蛋白原以及其他成分, 防止细胞凝集; 在室温情况下, 用等渗浓度的蔗糖溶液 (0.308 mol/L) 稀释重悬红细胞. 利用自动细胞计数仪测定红细胞浓度 (5×10^8 个/mL), 然后等渗的蔗糖溶液将红细胞稀释成一系列体积浓度为 10%, 15%, 20%, 25%, 30%, 40%, 50% 的红细胞悬浮液. 另外取适量的蔗糖溶液作为对照组, 记为体积浓度为 0% 的红细胞悬浮液. 红细胞实验共设置 8 组浓度类别, 每组浓度类别利用阻抗分析仪连续测量 20 次, 共获得 160 组阻抗谱数据, 实验时快速测量每种浓度红细胞悬浮液的阻抗谱数据, 并且在测量前重悬传感器, 使细胞达到随机分布的状态.

3.3 实验结果与讨论

通过对实验得到的数据进行正则化处理, 并利用 MLR 回归算法进行训练和调参, 得到了数据集的预测值与真实值, 对比结果如图 7 所示, 可以直观地看到, MLR 的预测值与真实值非常接近. 进一步地统计了数据集每种浓度的平均绝对误差, 如表 3 所列, 可以准确地得到, 红细胞悬浮液每种浓度下的平均绝对误差在 0.0132 附近波动. 最后数据集在利用五折验证, 其结果如表 4 所列, 可以得到 MLR 的平均拟合优度和均方误差分别为 0.9998 和 0.0079.

以上结果充分说明本文提出的细胞悬浮液浓度自动识别方法, 可以准确快速地计算细胞悬浮液的浓度. 而且实验的总体结果与仿真结果基本一致, 说明在数值仿真时, 利用细胞位置随机分布策略生成的生物阻抗谱数据集, 挑选出的最优回归模型是完全可以用于处理实验数据, 也说明了利用细

胞随机分布策略可以模拟出细胞的真实分布情况. 但是实验结果与仿真结果相比, 稍有差异, 原因是实验数据有许多偶然因素 (红细胞每个时刻的活跃程度不同, 每个浓度下红细胞的大小、发育程度不完全相同, 测量数据、配置溶液以及移液过程中不可避免的人为误差等因素), 使其数据有明显波动, 从而导致结果变化.

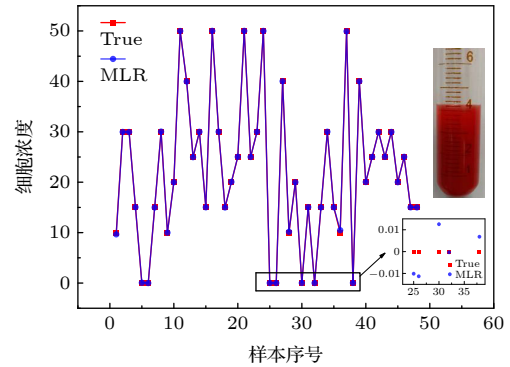


图 7 MLR 对数据集预测值与真实值的差异, 内插图是 MLR 预测值与真实值的误差放大图像

Fig. 7. The difference between the predicted value and the true value of the data set by MLR. The inset is an enlarged image of the error between the MLR predicted value and the true value.

4 结论

本文提出了一种细胞悬浮液浓度自动识别方法, 根据临床经验提出了细胞位置随机分布策略并建立了细胞模型, 通过数值仿真的方法研究了该方法的可行性, 并且进行了实验验证, 得到以下结论:

- 1) 采用了细胞位置随机分布策略, 该策略可以模拟细胞真实分布情况;
- 2) 所提出的细胞悬浮液浓度自动识别方法具有较好的仿真结果. 仿真表明, MLR 是用于细胞浓度识别的最佳回归模型, 通过 MLR 对阻抗谱数据集进行回归处理, 五折交叉验证的平均拟合优度

表 3 每种红细胞悬浮液浓度下的平均绝对误差

Table 3. The average absolute error of each red blood cell suspension concentration.

| 浓度/% | 0 | 10 | 15 | 20 | 25 | 30 | 40 | 50 |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 平均绝对误差 | 0.0104 | 0.0124 | 0.0113 | 0.0134 | 0.0129 | 0.0165 | 0.0158 | 0.0161 |

表 4 MLR 回归算法的五折交叉验证结果

Table 4. Cross validation results of MLR regression algorithm.

| | 1 | 2 | 3 | 4 | 5 | 平均数 |
|-------|--------|--------|--------|--------|--------|--------|
| R^2 | 1.0000 | 0.9998 | 0.9999 | 0.9996 | 0.9999 | 0.9998 |
| MSE | 0.0015 | 0.0021 | 0.0152 | 0.0035 | 0.0173 | 0.0079 |

和均方误差分别达到了 0.9997 和 0.0008;

3) 所提出的细胞悬浮液浓度自动识别方法在实验中表现优异, 能够识别不同浓度的红细胞悬浮液, 通过 MLR 对抗谱数据集进行回归处理, 五折交叉验证的平均拟合优度和均方误差分别达到了 0.9998 和 0.0079, 每种浓度样本的平均绝对误差在 0.0132 附近波动;

4) 所提出的细胞悬浮液浓度自动识别方法可应用于细胞浓度的检测中, 具有检测速度快、操作简便、检测精度高的优点, 并为细胞浓度的检测方法提供了一种新的思路。

本文提供的细胞浓度检测方法通用性较强, 不仅可用于检测细胞浓度, 还能够推广到其他悬浮液浓度的检测中, 例如可检测血栓浓度以实现心脑血管疾病的术后判断和药效评估。

参考文献

- [1] Fu X F, Xiong X B 2017 *Lab. Med. Clin.* **14** 274 (in Chinese) [付小芬, 熊小波 2017 *检验医学与临床* **14** 274]
- [2] Zhang Y, Fang W M, Chen Z F 2005 *Jiangsu. Med. J* **10** 762 (in Chinese) [张瑛, 范卫民, 陈哲峰 2005 *江苏医药* **10** 762]
- [3] Ren L R, He L, Liu T, Wang H Y 2020 *Med. J. West Chin.* **32** 1132 (in Chinese) [任丽蓉, 何蕾, 刘涛, 王浩宇 2020 *西部医学* **32** 1132]
- [4] Han S S, Guo Z H, Zhu Z Y 2015 *Chin. J. Blood Transfusion* **28** 1463 (in Chinese) [韩莎莎, 郭忠慧, 朱自严 2015 *中国输血杂志* **28** 1463]
- [5] Fu Y, Wang Y, Bai F, Wu J, Sun R, Ding Z R, Dong C 2015 *J. Med. Biomech.* **30** 392 (in Chinese) [傅怡, 王洋, 白帆, 吴洁, 孙仁, 丁祖蓉, 董澄 2015 *医用生物力学* **30** 392]
- [6] Kong Y, Jiang D 2006 *Chin. Cancer* **15** 172 (in Chinese) [孔雁, 姜达 2006 *中国肿瘤* **15** 172]
- [7] Feng Y Q 2009 *M. S. Thesis* (Changsha: Central South University) (in Chinese) [冯也倩 2009 硕士学位论文 (长沙: 中南大学)]
- [8] Yang R, Zou M Q 2004 *J. Instrumental. Anal.* **06** 124 (in Chinese) [杨蕊, 邹明强 2004 *分析测试学报* **06** 124]
- [9] Nilsson G E, Zhai H, Chan H P, Farahmand S, Maibach H I 2009 *Skin Res. Technol.* **15** 6
- [10] Guo T, Liu F, Liu Y, Chen N K, Guan B O, Albert J 2014 *Biosens. Bioelectron.* **55** 452
- [11] Yang Y X, Qiao Y 2013 *Chin. J. Sci. Instrum.* **34** 908 (in Chinese) [杨宇祥, 乔洋 2013 *仪器仪表学报* **34** 908]
- [12] Yao J F, Wan J F, Yang L, Liu K, Chen B, Wu H T 2020 *Acta Phys. Sin.* **69** 163301 (in Chinese) [姚佳峰, 万建芬, 杨璐, 刘凯, 陈柏, 吴洪涛 2020 *物理学报* **69** 163301]
- [13] Yao J F, Jiang Z P, Xu Z F, Liu X Y, Chen B, Wu H T 2018 *Biol. Chem. Eng.* **4** 4 (in Chinese) [姚佳峰, 姜祝鹏, 徐梓菲, 刘夏移, 陈柏, 吴洪涛 2018 *生物化工* **4** 4]
- [14] Sun T P, Ching T S, Cheng C S, Huang S H, Chen Y J, Hsiao C S, Chang C H, Huang S Y, Shieh H L, Liu W H 2010 *Cancer Epidemiol.* **34** 207
- [15] Mellert F, Kai W, Schneider C, Dudykevych T, Preusse C J 2011 *IEEE Trans. Biomed. Eng.* **58** 1511
- [16] Czerniec S A, Ward L C, Lee M J, Refshauge K M, Beith J, Kilbreath S L 2011 *Support. Care Cancer* **19** 703
- [17] Shen J Y 2020 *Inf. Technol. Informatization* **5** 228 (in Chinese) [沈继云 2020 *信息技术与信息化* **5** 228]
- [18] Ke P F, Cai M G, Wu T 2020 *Comput. Eng.* **46** 262 (in Chinese) [柯鹏飞, 蔡茂国, 吴涛 2020 *计算机工程* **46** 262]
- [19] Kimberly S, Yiqiang S, Lisa G, Heather J, Maureen S, Mendonca E A 2020 *Health Inform. J.* **26** 388
- [20] Liu Q S, Wang D, Zhu Y G, Yang Z B, Bo Y 2020 *Rock Soil Mech.* **41** 319 (in Chinese) [刘泉声, 王栋, 朱元广, 杨战标, 伯音 2020 *岩土力学* **41** 319]
- [21] Keprate A, Ratnayake R M C 2017 *2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)* Singapore, December 10-13, 2017 p1331
- [22] Wang Z Z, Zhang W G, Sun D B, Zhang Xin 2020 *Comput. Digital Eng.* **48** 546 (in Chinese) [王征征, 张卫钢, 孙道斌, 张鑫, 王培丞 2020 *计算机与数字工程* **48** 546]
- [23] Joshi R P, Qin H, Schoenbach K H 2004 *3rd International Symposium on Nonthermal Medical/Biological Treatments Using Electromagnetic Fields and Ionized Gases* San Antonio, TX, the United States of America, June 11-13, 2003 p1677
- [24] Jiang Z, Yao J, Wang L, Wu H, Huang J, Zhao T, Takei M 2019 *IEEE Sens. J.* **19** 5979
- [25] Li C, Xu P 2021 *Neural Comput. Appl.* **33** 613
- [26] Smola A J, Schölkopf B 2004 *Stat. Comput.* **14** 199
- [27] Zhang J G, Li Y Y, Cao J 2011 *Proceedings of 2011 IEEE International Conference on Computer Science and Automation Engineering* Shanghai China, June 10-12, 2011 p47

A method of identifying cell suspension concentration based on bioimpedance spectroscopy*

Liu Sheng-Long¹⁾ Yang Lu¹⁾ Zhu Cheng-Jun²⁾ Liu Kai¹⁾
Han Wei³⁾ Yao Jia-Feng^{1)†}

1) (*College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China*)

2) (*Department of Oncology, the First Affiliated Hospital with Nanjing Medical University, Nanjing 210029, China*)

3) (*Nanjing Stomatological Hospital Medical School of Nanjing University, Nanjing 210008, China*)

(Received 4 October 2021; revised manuscript received 6 December 2021)

Abstract

Based on bioimpedance spectroscopy technology, a method of automatically identifying the cell suspension concentration is proposed. This method combines multiple linear regression algorithm and bioimpedance spectroscopy technology, which can identify the concentration of cell suspension quickly and accurately. Firstly, a strategy of random distribution of cell locations is proposed to simulate the true existence of cells. Secondly, 2400 groups of normal, cancerous and mixed cell models with different concentrations are generated by numerical simulation and their bioimpedance spectroscopy data are calculated. Thirdly, the multiple linear regression algorithm (MLR), support vector machine (SVM), and gradient boosting regression algorithm (GBR) are used to identify the concentration of cancerous cells. The simulation results show that the MLR is the best regression model for cell suspension concentration identification and its average goodness of fit and mean square error are 0.9997 and 0.0008 respectively. Finally, the MLR is applied to the identification of red blood cell suspensions with different concentrations, the experimental results show that the average goodness of fit and mean square error are 0.9998 and 0.0079, respectively, indicating that this method has a greater ability to identify cell suspension concentrations.

Keywords: cell concentration measurement, bioimpedance spectroscopy (BIS), multiple linear regression (MLR), random distribution strategy of cell

PACS: 87.55.dk, 87.19.xj, 07.50.Qx

DOI: [10.7498/aps.71.20211837](https://doi.org/10.7498/aps.71.20211837)

* Project supported by the National Natural Science Foundation of China (Grant No. 62071224), and the Open Fund Project of State Key Laboratory of Precision Measuring Technology and Instruments, China (Grant No. pilab2107).

† Corresponding author. E-mail: jiaf.yao@nuaa.edu.cn