



DOI:10.13364/j.issn.1672-6510.2014.02.011

基于模式识别的人体肺癌电阻抗检测方法研究

陈晓艳, 赵秋红

(天津科技大学电子信息与自动化学院, 天津 300222)

摘要: 根据阻抗频谱理论及模式识别理论, 对手术中切除的恶性肿瘤组织及非肿瘤组织进行阻抗测量并进行肺癌辨识. 利用 Agilent4294A 阻抗分析仪测量手术中切除样本的阻抗值, 获得 100 Hz ~ 100 MHz 频率激励下的阻抗频谱, 经最小二乘法拟合, 得到频谱特征参数. 将测量样本分成训练集与测试集, 设计 LMSE 及 Fisher 两种线性分类器; 分类器经训练后获得判别函数, 利用分类器对测试集进行分类实验. 研究表明, LMSE 以及 Fisher 算法对测试样本均能进行有效识别, 并且识别结果与病理切片分析结果吻合, 验证了基于模式识别的方法对肺癌阻抗检测进行辨识的可行性与可靠性, 为肿瘤早期筛查提供有效检测方法.

关键词: 模式识别; 肺癌; 阻抗测量; LMSE 算法; Fisher 算法

中图分类号: R363; TP274⁺. 3 文献标志码: A 文章编号: 1672-6510(2014)02-0050-04

Impedance Detection of Human Lung Cancer Based on Pattern Recognition

CHEN Xiaoyan, ZHAO Qihong

(College of Electronic Information and Automation, Tianjin University of Science & Technology, Tianjin 300222, China)

Abstract: According to the impedance spectrum theory and pattern recognition theory, the impedance of the malignant tissue and the non-tumor tissue were measured for identification of lung cancer. With the Agilent4294A impedance analyzer measuring the impedances, the impedance values were obtained under 100 Hz-100 MHz excitation currents. Fitted by the least squares, the parameters in Cole-Cole equation were calculated. All the measured data were divided into a training set and a testing set. Two different linear classifiers based on LMSE and Fisher algorithm were designed. After training, the classification functions were obtained to distinguish the testing set. The research results indicated that LMSE and Fisher algorithms can classify the malignant tissue and non-tumor tissue effectively, which is also consistent with the pathological results. It is verified that the method is feasible and reliable in identifying the nature of the tissue, and it can provide an effective screening way for early cancer detection.

Key words: pattern recognition; lung cancer; impedance measurement; LMSE; Fisher

肺癌发病率及病死率居全球癌症首位, 已成为威胁人类健康和癌症死亡的主要病因之一. 据统计, 肺癌 5 年生存率为 13% ~ 15%, 早期诊断率仅 15%, 因此对肺癌患者早发现、早诊断、早治疗, 以减少肺癌晚期转移与恶化的可能性是非常重要的^[1]. 目前, 肺癌早期诊断方法主要有: 影像学检查、肿瘤标志物检查、痰液涂片脱落细胞检查、支气管镜检查、胸腔镜和纵隔镜等^[2], 痰液涂片脱落细胞检查是确诊肺癌的

最简便手段, 但受多种因素的影响, 其阳性率较低, 不作为早期肺癌的检查手段; 支气管镜能直接观察病变形态特征以及支气管狭窄情况, 但是会对气管造成物理伤害; 影像学检查是诊断肿瘤最重要的方法之一, 包括 X 射线、CT、MRI、PET、EIT 等. 通过这些影像学检查可以获取癌变的准确位置和大小, 但在肺癌早期尚未形成结节之前, 现有影像技术并不能预警病变, 并且频繁使用会对人体造成更大的伤害. 因

收稿日期: 2013-11-07; 修回日期: 2014-01-06

基金项目: 国家自然科学基金青年基金资助项目(61301246); 天津市自然科学基金资助项目(12JCYBJC19500)

作者简介: 陈晓艳(1973—), 女, 四川人, 教授, cxywxr@163.com.

此,临床需要无创监测肺癌早期病变的新方法和新技术。

电阻抗对病变较敏感,组织或器官在尚未形成结构性结节之前,其阻抗就已经发生了变化,通过测量组织或器官的电阻抗,可以对肺癌早期进行筛查,起到预警作用;而且电阻抗测量具有无毒素、无射线、无损伤等突出优势,是一种理想的、具有诱人应用前景的无损医学检测手段^[3-4]。目前,国内外关于电阻抗进行癌症组织的辨识方法研究尚少,国内开展人体活性组织介电特性与表征方法研究的有第四军医大学,其研究侧重乳腺与脑出血病变^[5-6],宁波大学马青教授研究了动物及人血液细胞的阻抗谱分析方法^[7-8];天津大学王化祥教授研究小组进行肺部活性组织介电特性与表征方法研究,已发表1篇阻抗测量结果^[9],但均未采用模式识别方法对病变组织进行辨识。

本文根据阻抗频谱理论及模式识别理论,对手术中切除的恶性肿瘤组织及非肿瘤组织进行阻抗测量和辨识,研究内容属于初期探索。

1 原理与方法

1.1 阻抗特征参数

根据 Cole-Cole 理论,生物阻抗可等效为内、外电阻和膜电容生物阻抗模型^[10]

$$Z = R_{\infty} + \frac{R_0 - R_{\infty}}{1 + (j\omega\tau)^{\alpha}}$$

式中:特征参数 R_0 为低频电阻的极限值; R_{∞} 为高频电阻的极限值; α 为松弛因子; τ 为时间常数; ω 为角频率。用于计算生物阻抗特征参数的 Cole-Cole 阻抗圆模型见图 1。其中: (x_0, y_0) 为圆弧的圆心坐标; r 为圆弧的半径。

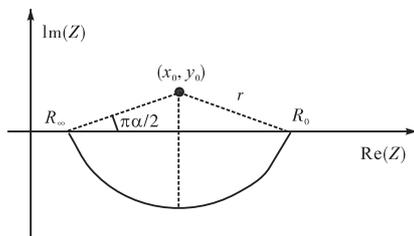


图 1 Cole-Cole 阻抗圆

Fig. 1 Cole-Cole plots

阻抗特征参数的计算公式如下:

$$R_0 = x_0 + \sqrt{r^2 - y_0^2} \quad (1)$$

$$R_{\infty} = x_0 - \sqrt{r^2 - y_0^2} \quad (2)$$

$$\alpha = 1 - \frac{\pi}{2} \arcsin\left(\frac{|y_0|}{r}\right) \quad (3)$$

$$\tau = \frac{\left(\frac{R_0 - Z_i(\omega_i)}{Z_i(\omega_i) - R_{\infty}}\right)^{\frac{1}{\alpha}}}{j\omega_i} \quad (4)$$

这里采用最小二乘法拟合曲线,使目标函数 F 最小。

$$F(x_0, y_0, r) = \sum_{i=1}^N [\sqrt{(x_i - x_0)^2 + (y_i - y_0)^2} - r]^2 \quad (5)$$

1.2 模式识别

模式识别诞生于 20 世纪 20 年代,至今已有很多种模式识别分类方法^[11-12]。由于统计模式识别方法比较成熟,能够考虑干扰噪声等影响,识别模式基元能力强,所以对统计模式识别法中的线性分类方法,即 LMSE 算法和 Fisher 算法进行了研究。

1.2.1 LMSE 算法

LMSE 算法是对准则函数引进最小均方误差这一条件而建立起来的,适用于线性可分的情况,其把对满足 $\mathbf{XW} > \mathbf{0}$ 的求解改为对满足 $\mathbf{XW} = \mathbf{B}$ 的求解。LMSE 算法的出发点就是选择一个准则函数 $J(\mathbf{W}_i, \mathbf{X}) = \frac{1}{2} E\{[r_i(\mathbf{X}) - \mathbf{W}_i^T \mathbf{X}]^2\}$,使其达到极小值时,可以得到 $\mathbf{XW} = \mathbf{B}$ 的最小二乘近似解^[13-14]。算法的实现步骤如下:

(1) 设各个权向量的初始值 $\mathbf{W}_1 = \mathbf{W}_2 = \mathbf{0}$ 。

(2) 输入第 k 次样本 $\mathbf{X}(k)$, 计算判别函数 $d_i(k) = \mathbf{W}_i^T(k) \mathbf{X}(k), i=1, 2$ 。

(3) 若 $\mathbf{X}(k) \in \omega_i$, 则 $r_i[\mathbf{X}(k)] = 1$; 否则 $r_i[\mathbf{X}(k)] = 0$ 。

(4) 计算 $\mathbf{W}_i(k+1) = \mathbf{W}_i(k) + \alpha_k \mathbf{X}(k) \{r_i[\mathbf{X}(k)] - \mathbf{W}_i^T \mathbf{X}(k)\}$, 其中 $\alpha_k = 1/k$ 。

(5) 循环执行步骤(2),直到属于 ω_i 类的所有样本都满足 $d_i(\mathbf{X}) > d_j(\mathbf{X}), \forall j \neq i$ 。

1.2.2 Fisher 算法

Fisher 准则^[15]的基本原理是:找到一个最合适的投影轴,使两类样本在该轴上投影之间的距离尽可能远,而每一类样本的投影尽可能紧凑,从而使分类效果为最佳。算法的实现步骤如下:

(1) 计算各类样本的均值向量

$$\mathbf{m}_i = \frac{1}{n_i} \sum_{\mathbf{X} \in \omega_i} \mathbf{X} \quad i=1, 2 \quad (6)$$

其中, n_i 表示 ω_i 类的样本个数。

(2) 计算样本类内离散度矩阵 \mathbf{S}_i 和总类内离散度

矩阵 S_w .

$$S_i = \sum_{X \in X_i} (X - m_i)(X - m_i)^T \quad i=1,2 \quad (7)$$

$$S_w = S_1 + S_2 \quad (8)$$

(3) 计算类间离散度矩阵 S_b .

$$S_b = (m_1 - m_2)(m_1 - m_2)^T \quad (9)$$

(4) 若希望投影后,在一维空间里各类样本尽可能地分开,也就是两类样本均值之差越大越好,同时希望各类样本内部尽量密集,即希望类内离散度越小越好. 因此,定义 Fisher 准则函数为

$$J(W) = \frac{W^T S_b W}{W^T S_w W} \quad (10)$$

使得 $J(W)$ 取得最大值的 W^* 为

$$W^* = S_w^{-1} (m_1 - m_2) \quad (11)$$

式中, W^* 是样本投影到投影线上的投影向量.

(5) 将训练集内所有样本进行投影.

$$Y = (W^*)^T X \quad (12)$$

(6) 计算在投影空间上的分割阈值 Y_0 .

$$Y_0 = \frac{\tilde{m}_1 + \tilde{m}_2}{2} + \frac{\ln[P(\omega_1)/P(\omega_2)]}{N_1 + N_2 - 2} \quad (13)$$

投影后,各类样本均值

$$\tilde{m}_i = \frac{1}{N_i} \sum_{Y \in \omega_i} Y_j \quad i=1,2, \quad j=1,2,\dots,N_1+N_2 \quad (14)$$

(7) 根据决策规则分类

$$\begin{cases} X \in \omega_1 & \forall Y_j > Y_0 \\ X \in \omega_2 & \forall Y_j < Y_0 \end{cases} \quad j=1,2,\dots,N_1+N_2 \quad (15)$$

2 实验

实验数据均来自天津医科大学总医院肺外科手术室,样本离体时间均不超过 30 min. 采集仪器是 Aglient4294A 阻抗分析仪,夹具是 16092A,测试舱内的温度为 37 °C,相对湿度为 90%以上. 获取的待测样本经过修剪后放入标本测试盒,采用二电极法,在 100 Hz ~ 100 MHz 范围内,取成对数递增的 31 个频率点测量样本的阻抗谱.

临床获取的所有样本均做了病理切片,临床医生根据切片进行病理分析,确诊是否为恶性肿瘤组织,并判断恶性肿瘤类型. 根据医生诊断结果,选取 123 个训练样本,其余 162 个作为测试样本. 在 4 个参数 R_0 、 R_∞ 、 τ 、 α 中任选 2 个作为模式识别的特征参数,共有 6 种组合方式,分别是 (R_0, R_∞) 、 (R_∞, τ) 、 (τ, α) 、 (R_0, α) 、 (R_0, τ) 、 (R_∞, α) . 由于 LMSE

算法以及 Fisher 算法属于线性分类算法,对于线性不可分的情况不适用. (R_∞, α) 、 (τ, α) 、 (R_∞, τ) 这 3 种组合的样本均属于线性不可分的情况,因此 2 种算法对此都不适用. 实验表明,对于 (R_0, α) ,2 种算法虽能适用,却得不到辨识结果.

2.1 LMSE 算法识别结果

采用 LMSE 算法对 (R_0, R_∞) 和 (R_0, τ) 进行辨识,结果如下:

(1) 当输入量是 (R_0, R_∞) 时,分界线为

$$y = -75.3012x + 7.1168 \times 10^{-4} \quad (16)$$

判别函数为

$$y = 0.0018x_1 + 2.3904 \times 10^{-5}x_2 - 1.7012 \quad (17)$$

123 个训练样本及从待测试样本中随机选取 23 个样本的分布及模式识别结果如图 2 所示.

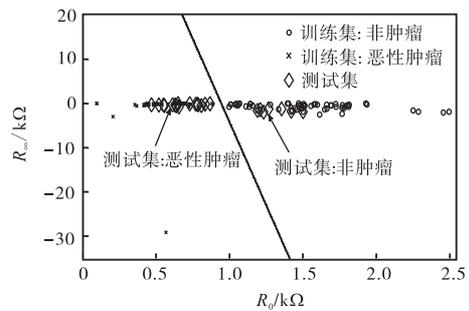


图 2 (R_0, R_∞) 的 LMSE 识别结果

Fig. 2 Results of (R_0, R_∞) identified by LMSE

(2) 当输入量为 (R_0, τ) 时,分界线为

$$y = -7.3899 \times 10^{-5}x + 0.699 \quad (18)$$

判别函数为

$$y = 0.0018x_1 + 24.3575x_2 - 1.7027 \quad (19)$$

123 个训练样本及 23 个测试样本的分布及模式识别结果如图 3 所示.

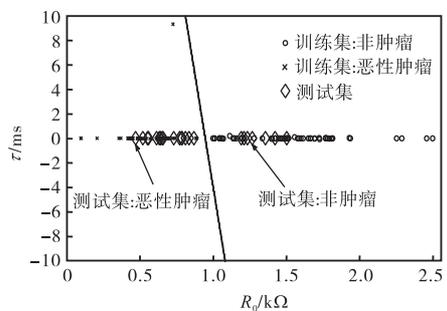


图 3 (R_0, τ) 的 LMSE 识别结果

Fig. 3 Results of (R_0, τ) identified by LMSE

2.2 Fisher 算法识别结果

采用 Fisher 算法对 (R_0, R_∞) 进行辨识,结果

如下:

$$y = x \frac{w(2)}{w(1)} \quad (20)$$

则投影线为

$$y = -\frac{1}{w(2)/w(1)}(x - w_1(35,1) + w_2(60,1)) + w_1(35,2) + w_2(60,2) \quad (21)$$

式中: w 为投影向量; w_1 为正常样本; w_2 为肿瘤样本. 采用式(13)计算在投影空间上的分割阈值, 计算过程中的各参数值见表 1.

表 1 阈值 y_0 的各参数意义及其数值

Tab. 1 The parameters and their values of threshold y_0

参数	意义	数值
\bar{m}_1	非肿瘤样本在投影向量上投影的均值	0.146 0
\bar{m}_2	恶性肿瘤样本在投影向量上投影的均值	0.053 9
N_1	训练集中非肿瘤样品个数	61
N_2	训练集中恶性肿瘤样品个数	62
$P(\omega_1)$	非肿瘤样本的先验概率	0.495 9
$P(\omega_2)$	恶性肿瘤样本的先验概率	0.504 1
Y_0	阈值	0.099 8

123 个训练样本分布及对 23 个测试样本进行模式识别的结果如图 4 所示.

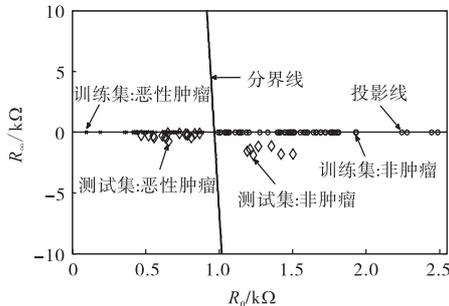


图 4 (R_0, R_∞) 的 Fisher 识别结果

Fig. 4 Results of (R_0, R_∞) identified by Fisher

采用同样的方法对 (R_0, τ) 进行投影计算, 得不到正确的投影线和分界线, 所以无法辨识.

3 结 语

在 LMSE 和 Fisher 此 2 种模式识别中, 训练样本量相同, 即 61 个恶性肿瘤组织, 62 个非肿瘤组织, 共计 123 个. 为了突出辨识效果, 从测试样本中随机选用 16 个恶性肿瘤组织, 7 个非肿瘤组织, 共计 23 个. 图 2—图 4 的识别结果表明: 测试集 23 个数据中有 7 个数据属于非肿瘤组织, 16 个数据属于恶性肿瘤组织, 与病理分析结果完全一致.

在实际测试中, 对更多的 (98 个恶性肿瘤组织和 64 个非肿瘤组织) 样本进行识别, 除经过初步筛选剔除掉的 3 种情况 (第 1 种为术前进行过化疗; 第 2 种为测试样本在测试盒中未充分填充; 第 3 种为测试样本的血液含量过多) 外, 辨识结果与病理分析结果仍是高度一致.

由于临床无法获得已发生癌变但尚未形成器质性结节的组织, 所以本研究中的恶性肿瘤样本仅局限于已经形成器质性结节的组织. 研究表明, 利用 LMSE 以及 Fisher 算法对术旁切除恶性肿瘤组织及非肿瘤组织能够获得良好的分类结果, 为下一步更加细化研究癌变过程中组织的阻抗变化奠定了基础. 在今后的研究中, 可以通过动物实验实现在体培养肿瘤, 分阶段测量组织阻抗, 以获得早期病变阻抗信息, 探寻阻抗变化与病理变化之间的对应关系, 以期对肿瘤的早预防、早诊断、早治疗提供一种新的检测方法.

参考文献:

- [1] 宁康, 王超. 肺癌早期诊断方法研究进展[J]. 山东医药, 2010, 50(21): 108-110.
- [2] 王玉新. 11 例肺炎型肺癌的误诊及其影像学特点分析[J]. 临床肺科杂志, 2013, 18(1): 166-167.
- [3] 董秀珍. 生物电阻抗成像研究的现状与挑战[J]. 中国生物医学工程学报, 2008, 27(5): 641-643, 649.
- [4] Kao T J, Boverman G, Kim B S, et al. Regional admittivity spectra with tomosynthesis images for breast cancer detection: Preliminary patient study[J]. IEEE Transactions on Medical Imaging, 2008, 27(12): 1762-1768.
- [5] 付峰, 臧益民, 董秀珍, 等. 部分离体动物组织复阻抗频率特性(100 Hz ~ 10 MHz)测量系统及初步测量结果[J]. 第四军医大学学报, 1999, 20(3): 220-222.
- [6] 季振宇, 史学涛, 尤富生, 等. 电阻抗扫描成像乳腺癌早期检测设备研制及初步评估[C]//中国科协第十四届年会第十七分会会场环境危害与健康防护研讨会论文集. 北京: 中国生物医学工程学会, 2012: 169-172.
- [7] 陈晓敏, 丰明俊, 王力, 等. 10 kHz ~ 100 MHz 人血液细胞介电性能[J]. 医学研究杂志, 2009, 38(3): 42-45.
- [8] 陈林, 马青, 王力, 等. 大鼠红细胞悬浮液阻抗谱 Cole-Cole 数学模型分析[J]. 航天医学与医学工程, 2010, 23(3): 182-187.
- [9] 王洁然, 王化祥, 徐晓. 人体肺部组织介电特性实验研究[J]. 中国生物医学工程学报, 2013, 32(2): 178-183.

(下转第 70 页)