

近红外激光拉曼技术在体探测胃癌腹膜播散

马 君¹ 徐 明¹ 巩龙静¹ 高 媛¹ 毛伟征² 郑荣儿¹

(¹中国海洋大学光学光电子实验室, 山东 青岛 266100)

²青岛市市立医院普外科, 山东 青岛 266071

摘要 以腹膜接种人胃癌细胞 SGC-7901 的裸鼠为胃癌腹膜播散的动物模型, 进行模拟外科手术, 在体探测不同种植期的裸鼠腹膜癌结节及正常腹膜组织的激光拉曼光谱, 对比光谱差异, 采用支持向量机(SVM)算法对光谱进行分类和分期判决。结果表明, 癌结节和正常组织拉曼光谱差异显著, 用支持向量机算法进行分类的灵敏度、特异度和诊断准确度分别为 95.73%、70.73% 和 90.73%; 不同生长期的癌结节组织拉曼光谱也存在明显差异, 用支持向量机算法进行分期的结果分别为 98.82%、98.73% 和 98.78%。从分类结果可以看出, 此方法对指导外科手术中癌变组织的识别有重要的意义。

关键词 光谱学; 胃癌; 激光拉曼光谱; 裸鼠; 支持向量机; 在体

中图分类号 Q631; R735.2 文献标识码 A doi: 10.3788/CJL201138.0915001

Near-Infrared Raman Spectroscopy for Detection of Gastric Cancer Peritoneal Dissemination *in vivo*

Ma Jun¹ Xu Ming¹ Gong Longjing¹ Gao Yuan¹ Mao Weizheng² Zheng Rong'er¹

¹Optics and Optoelectronics Laboratory, Ocean University of China, Qingdao, Shandong 266100, China

²Department of General Surgery, Qingdao University Medical College Hospital, Qingdao, Shandong 266071, China

Abstract The nude mice injected with human gastric cancer cells (SGC-7901) in their peritoneums are chosen as animal models of gastric cancer peritoneal dissemination in this research. The Raman spectra at 785 nm excitation of both these nude mice which are in different tumor planting periods and the normal counterpart are taken *in vivo* in the imitate laparotomy. 205 spectra are collected. The spectra of different tissue types are compared and classified by support vector machine (SVM) algorithm. The results show significant differences between normal and malignant tissues. For normal and malignant tissues, the sensitivity, specificity and accuracy are 95.73%, 70.73% and 90.73%, respectively, while for different tumor planting periods, they are 98.82%, 98.73% and 98.78%. The experimental results show that Raman spectra differ significantly between cancerous and normal gastric tissues, which provides experimental basis for the diagnosis of gastric cancer by Raman spectroscopy technology. And SVM algorithm can give well generalized classification performance for the samples, which expands the application of mathematical algorithms in classification.

Key words spectroscopy; gastric cancer; laser Raman spectroscopy; nude mice; support vector machines; *in vivo*

OCIS codes 170.2680; 170.4580; 170.5660; 300.6330

1 引 言

腹膜播散是胃癌扩散的主要形式, 手术中准确判定腹膜转移及程度是医生选择手术方式的重要依据, 然而凭肉眼判断的准确率不高, 传统冰冻病理检验耗时、有创。目前国内外对拉曼光谱在肿瘤诊断

方面的报道主要集中在对离体的肿瘤手术标本进行拉曼光谱检测, 以区分癌变组织与正常组织^[1,2]; 利用光纤通过内窥镜将激光导入胃肠内从粘膜面对胃肠癌进行筛选, 辅助寻找可疑的癌灶^[3]; 对体表癌症(如乳腺癌、皮肤癌)从体表外进行探测^[4]。而激光

收稿日期: 2011-01-17; 收到修改稿日期: 2011-04-04

基金项目: 青岛市科技发展计划(08-1-3-31-jch)资助课题。

作者简介: 马 君(1963—), 女, 博士, 教授, 主要从事激光光谱学和计算机模式识别等方面的研究。

E-mail: majun@ouc.edu.cn

拉曼光谱技术在腹部手术中实时、原位探查肿瘤尚属空白。

拉曼光谱是通过光子与物质分子发生非弹性碰撞,产生物质固有的分子振动光谱,它反映了分子的结构信息。研究表明,生物组织癌变后核酸、蛋白质、类胡萝卜素、脂类等物质在含量及结构上会发生变化,这些变化会相应地反映在拉曼光谱上^[1~4]。同时,如何运用先进的算法对数据进行处理和分析,提高信息的识别能力,是把激光拉曼光谱技术应用于生物医学领域的关键问题之一。支持向量机(SVM)算法具有能够较好地解决小样本学习的能力,当计算复杂性与数据的维数不成正比时,分类的准确率一般要高于人工神经网络算法,在癌症诊断方面有很大的发展空间^[5~7]。

本文以腹膜接种人胃癌细胞 SGC-7901 的裸鼠为胃癌播散的动物模型,进行全麻剖腹的模拟外科手术^[8],在体探测不同生长期的裸鼠腹膜癌结节及正常腹膜组织的激光拉曼光谱。采用 SVM 算法对光谱进行分类,为将来研制高效准确的“胃癌腹膜播散探测器”,协助医生手术中实时准确地鉴别和诊断胃癌打下实验基础。

2 材料与方法

2.1 人胃癌裸鼠模型及分组

20 只雄性裸鼠每只腹腔内注入 SGC-7901 人胃癌细胞 5×10^6 个,按种植时间长短,分成不同肿瘤生长期的实验组:A~D 组癌细胞种植天数分别为 4,10,14,30 天,除 D 组为 8 只裸鼠外,其余三组各为 4 只。

2.2 在体拉曼光谱探测系统

图 1 为在体拉曼光谱的探测系统。选用 785 nm 近红外连续激光为激发光,该波长是生物组织的传输“窗口”,可避免传输中的强吸收问题,并且能有效避开蛋白质、NADH、核酸、卟啉等强荧光物质在 236~680 nm 范围内的荧光光谱的影响。激光经 Y 型光纤探头(美国 InPhotonics 公司)的激发光纤传入组织表面,输出功率约为 30 mW。组织后向散射光经收集光纤耦合进入 QE65000 型光谱仪(美国 Ocean Optics 公司),由 Ocean Optics 软件进行光谱记录。

2.3 模拟手术及在体拉曼光谱采集

取 1 mL 质量分数为 5% 的盐酸氯胺酮与 4 mL 生理盐水混合后,制成麻醉剂。每只裸鼠腋下注射该麻醉剂 0.6 mL,待其进入麻醉状态后置于样品台

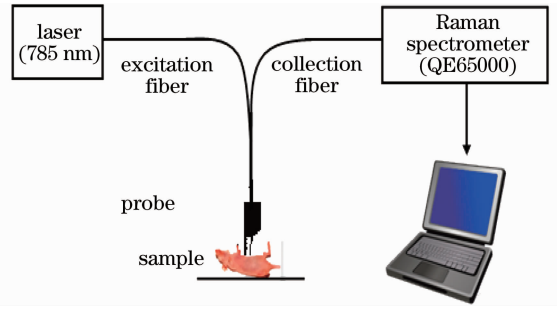


图 1 在体拉曼光谱探测系统

Fig. 1 Experimental system of the Raman spectroscopy 固定。行剖腹手术,翻开腹膜发现,A 组:肉眼未见肿瘤;B 组:偶尔可见小结节;C 组:见 1 mm 米粒状结节;D 组:肿瘤个数增多,大小不一。总之,种植时间越长,肿瘤腹膜播散程度越严重。每只裸鼠选择不同结节及距离结节 1 cm 以外的腹膜处采集光谱多条,4 组裸鼠共得光谱 205 条。切取光谱探测处腹膜组织送病理检验,确定其性质。病理检验结果表明,进行光谱探测的结节处组织均为癌组织,其余为正常组织。

2.4 支持向量机算法

SVM 算法折衷考虑最少错分样本和最大分类间隔,在高维空间中寻找一个超平面作为二类的分割。对于一组训练集

$$T = \{(x_1, y_1), \dots, (x_l, y_l)\} \in (\chi \times \gamma)^l,$$

$$x_i \in R^n, y_i \in \gamma = \{1, -1\}, i = 1, \dots, l, \quad (1)$$

存在分类面 $(w \cdot x) + b = 0$ 使得两类点的间隔最大,此即最优分类面。求最大间隔分类面等价于

$$\min_{w, b, \xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i, s. t. y_i [(w \cdot x_i) + b] \geq 1 - \xi_i, \quad (\xi_i \geq 0, i = 1, \dots, l) \quad (2)$$

式中 ξ 为 Vapnik 和 Cortes 等^[9]引入的非负的松弛变量,由 $\sum_{i=1}^l \xi_i$ 作为一种度量,描述训练集被错划的程度,同时引进一个惩罚参数 C , C 越大,表示对错误的惩罚越重。利用优化理论中的对偶理论可以求得(2)式的最终分类方程为

$$f(x) = \text{sgn} \left[\sum_{i=1}^l y_i \alpha_i^* K(x_i, x_j) + b^* \right]. \quad (3)$$

目前,最常用的核函数形式主要有以下几种:线性核函数 $K(x, x') = x \cdot x' + 1$;多项式核函数 $K(x, x') = [(x \cdot x') + c]^d$,其中 $c \geq 0$;高斯核函数 $K(x, x') = \exp(-\|x - x'\|^2 / \sigma^2)$ 。对于一个已知的数据集,仅由核函数和惩罚参数 C 就可确定一个 SVM 分类器。

用 Matlab6.5 进行编程,分别选用线性内积核函数(Linear)、高斯径向基核函数(RBF)、多项式核函数(Poly)进行分类分析。惩罚因子 C 的取值范围为 $10^{-3} \sim 10^4$, 径向核函数的核参数 σ 的取值范围为 $0.1 \sim 100$, 多项式核函数的核参数 d 取值范围为 $2 \sim 9$ 。以 1-特异度为横坐标,灵敏度为纵坐标,绘制受试者工作特征曲线(ROC 曲线)并求得其曲线下面积(AUC),对诊断效果进行评估。根据 Swets 的评价标准^[10],AUC 在 $0.7 \sim 0.9$ 之间有一定的准确率,可以用于临床诊断;AUC 在 0.9 以上时,诊断准确性比较高,在临床诊断中可以作为重要的诊断依据。

3 实验结果

由于生物组织的拉曼光谱主要集中在 $600 \sim 1800 \text{ cm}^{-1}$ 范围内,因此,主要选取这一范围内的组织光谱进行处理和分析。对光谱进行平滑、去背景及面积归一化处理,以消除噪声、荧光背景及光谱收集效率波动等对光谱的影响。

3.1 光谱特性分析

图2为正常和癌结节组织典型的拉曼光谱,图

中各特征峰的归属及生物分子来源如表 1 所示^[1,11,12]。分析可得,相比正常组织,癌结节组织拉曼光谱在 $870, 1330, 1450$ 和 1660 cm^{-1} 处强度较低,而在 $1007, 1050, 1093$ 和 1209 cm^{-1} 处强度较高。此外,癌结节组织存在 1297 和 1330 cm^{-1} 双峰,而正常组织在 1297 cm^{-1} 处有一个单峰, 1330 cm^{-1} 处仅存在一不明显的肩峰或者无峰。这表明,在组织癌变过程中,与拉曼光谱相关的生物分子的含量或结构发生了改变。

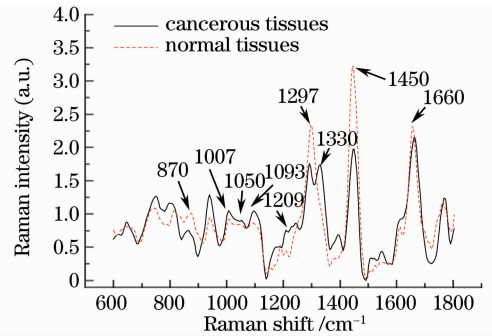


图 2 正常和癌结节组织的典型拉曼光谱
Fig. 2 Typical Raman spectra of normal and cancerous tissues

表 1 光谱特征峰的归属及生物分子来源

Table 1 Tentative biochemical assignment for Raman peaks

Raman peak / cm^{-1}	Tentative biochemical assignment	Biological molecules
870	C-C stretching vibration of Hydroxyproline	nucleic acid, protein
1007	C-C symmetric ring breathing of Phenylalanine	protein/carotenoid
1050	C-O stretching vibration of Deoxyribose	nucleic acid
1093	PO_2^- symmetric stretching vibration of Phosphodiester groups	nucleic acid
1209	C-C ₆ H ₅ stretching vibration of Phenylalanine	nucleic acid
1330	C-H bond and CH ₂ deformation vibration in the base	phospholipid/nucleic acid
1450	CH ₂ stretching or bending vibration	protein/lipid
1660	C=O stretching vibration of Amide I and H-O-H bending vibration of the water	protein/water

3.2 SVM 对癌结节及正常组织拉曼光谱的识别

用 SVM 算法对归一化的 164 条癌组织和 41 条正常组织的光谱进行分类,标记正常组织为 -1,癌结节组织为 +1。采用 10 重交叉验证法(10-Fold Cross-Validation)^[13]进行训练和测试。SVM 在不同核函数下测试结果见表 2。

表 2 SVM 在不同核函数下对癌结节和正常组织进行分类的 10 重交叉验证测试结果

Table 2 Classification of normal and neoplastic gastric tissues by SVM of 10-Fold Cross-Validation at different kernels

Kernel	Fold	Number of support vectors	Sensitivity / %	Specificity / %	Accuracy / %
Linear	1	50	93.75	60.00	85.71
	2	56	100.00	50.00	90.00
	3	50	100.00	50.00	90.00
	4	58	100.00	75.00	95.24
	5	56	88.24	100.00	90.48
	6	58	88.24	75.00	85.71
	7	48	93.75	75.00	90.00

Kernel	Fold	Number of support vectors	Sensitivity /%	Specificity /%	Accuracy /%
RBF	8	54	93.75	75.00	90.00
	9	55	93.75	75.00	90.00
	10	50	88.24	100.00	95.00
	totality		93.90	73.17	89.76
	1	58	100.00	60.00	90.48
	2	58	100.00	100.00	100.00
	3	55	100.00	50.00	90.48
	4	53	100.00	75.00	95.24
	5	54	88.24	50.00	80.95
	6	58	88.24	75.00	85.71
	7	60	93.75	75.00	90.00
Poly	8	55	93.75	100.00	95.00
	9	61	100.00	75.00	95.00
	10	56	93.75	50.00	85.00
	totality		95.73	70.73	90.73
	1	57	93.75	60.00	85.71
	2	54	100.00	50.00	90.00
	3	52	100.00	75.00	95.24
	4	58	100.00	50.00	90.48
	5	56	82.35	100.00	85.71
	6	58	82.35	25.00	71.43
	7	52	93.75	75.00	90.00
Linear	8	54	93.75	100.00	95.00
	9	50	93.75	100.00	95.00
	10	56	87.50	75.00	85.00
	totality		92.68	70.73	88.29

当采用线性核函数为基核,惩罚因子 $C=1$ 时,诊断灵敏度、特异度和准确度分别为 93.90%、73.17% 和 89.76%;当采用径向核函数为基核, $C=10$,核参数 $\sigma=5$ 时,诊断灵敏度、特异度和准确度分别为 95.73%、70.73% 和 90.73%;当采用多项式核函数为基核, $C=0.01$,核参数 $d=2$ 时,诊断灵敏度、特异度和准确度分别为 92.68%、70.73% 和 88.29%。

绘制 SVM 在不同核函数下的 ROC 曲线(图 3),并求得 AUC_{Linear} 为 0.83, AUC_{RBF} 为 0.85, AUC_{Poly} 为 0.78。综合考虑不同核函数的准确性和 AUC 值发现,当核函数采用径向核函数时,SVM 的分类诊断结果最佳。

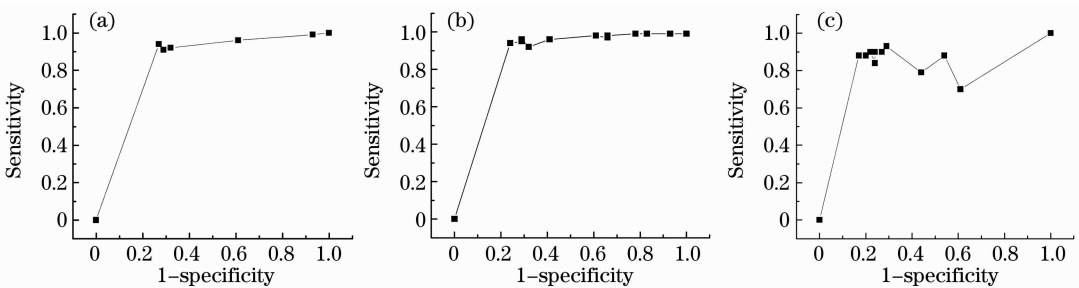


图 3 (a)线性核函数;(b)径向核函数;(c)多项式核函数的 ROC 曲线

Fig. 3 ROC curves of different kernel functions. (a) Linear; (b) radial basis; (c) polynomial

其中当惩罚因子 $C=10$,核参数 $\sigma=5$ 时,诊断结果最优,灵敏度、特异度和诊断准确度分别为 95.73%、70.73% 和 90.73%。

3.3 SVM 对癌细胞不同生长期拉曼光谱的分类

根据裸鼠生长状态,可进一步将 A~D 组裸鼠分为,甲组:肿瘤生长时间较长的 D 组;乙组:肿瘤生长时

间较短的 A、B、C 组。用 SVM 算法对甲、乙两组光谱进行分类, 标记甲组为 +1, 乙组为 -1。

采用 5 重交叉验证法(5-Fold Cross-Validation)^[13] 对归一化的 85 条甲组和 79 条乙组的癌结节组织拉曼光谱进行训练和测试。测试结果中, 甲组裸鼠的组织光谱判断正确的概率用灵敏度表示; 乙组裸鼠的组织光谱判断正确的概率用特异度表示。结果见表 3。

表 3 SVM 在不同核函数下对癌结节组织进行分期的 5 重交叉验证测试结果

Table 3 Classification of different growth stages of cancer by SVM of 5-Fold Cross-Validation at different kernels

Kernel	Fold	Number of support vectors	Sensitivity /%	Specificity /%	Accuracy /%
Linear	1	29	94.12	93.75	93.94
	2	32	94.12	93.75	93.94
	3	29	100.00	93.75	96.97
	4	27	94.12	100.00	96.97
	5	23	94.12	93.33	93.75
	totality			95.29	94.94
RBF	1	33	94.12	100.00	96.97
	2	35	100.00	100.00	100.00
	3	32	100.00	93.75	96.97
	4	38	100.00	100.00	100.00
	5	30	100.00	100.00	100.00
	totality			98.82	98.73
Poly	1	25	94.12	100.00	96.97
	2	30	94.12	100.00	96.97
	3	29	94.12	100.00	96.97
	4	26	94.12	100.00	96.97
	5	25	100.00	100.00	100.00
	totality			95.29	100.00

实验表明, 当采用线性核函数为基核, 惩罚因子 $C=10$ 时, 诊断灵敏度、特异度和准确度分别为 95.29%、94.94% 和 95.12%; 当采用径向核函数为基核, 惩罚因子 $C=100$, 核参数 $\sigma=15$ 时, 诊断灵敏度、特异度和准确度分别为 98.82%、98.73% 和 98.78%; 当采用多项式核函数为基核, 惩罚因子 $C=1$, 核参数 $d=2$ 时, 诊断灵敏度、特异度和准确度分别为 95.29%、100% 和 97.56%。

绘制 SVM 在不同核函数下的 ROC 曲线(图 4), 求得 ROC 曲线下的面积 AUC_{Linear} 为 0.94, AUC_{RBF} 为 0.96, AUC_{Poly} 为 0.80。这表明, 当采用径向核函数为基核时, SVM 的分类诊断结果最佳。

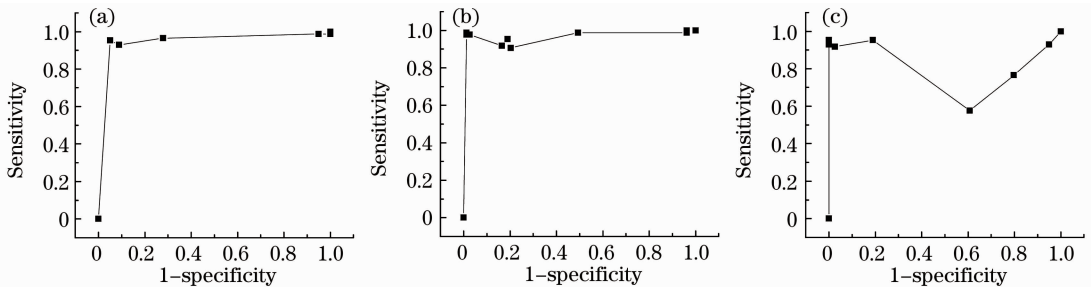


图 4 (a)线性核函数;(b)径向核函数;(c)多项式核函数的 ROC 曲线

Fig. 4 ROC curves of different kernel functions. (a) Linear; (b) radial basis; (c) polynomial

其中当惩罚因子 $C=100$, 核参数 $\sigma=15$ 时, 诊断结果最优, 灵敏度、特异度和诊断准确度分别为 98.82%、98.73% 和 98.78%。

4 讨 论

研究表明, 与正常组织相比, 癌结节组织的拉曼

光谱在 870、1330、1450、1660 cm^{-1} 处强度较低, 在 1007、1050、1093、1209 cm^{-1} 处强度较高, 并且在 1330 cm^{-1} 附近出现双峰。对癌和正常组织拉曼光谱间的差异进行分析, 结果提示这些差异主要体现在核酸、蛋白质及水的振动谱带强度上。医学研究表明, 细胞癌变过程中细胞核增大, DNA 含量增多,

这可能是造成癌结节组织拉曼光谱中 1050 cm^{-1} 、 1093 cm^{-1} 谱线增强的原因^[11]。相对于正常组织,癌结节组织在 1007 cm^{-1} 处的拉曼光谱强度升高^[12],可能是由于癌组织中蛋白质或类胡萝卜素含量较正常组织要高的缘故。另外,细胞质的黏液素损耗和金属蛋白酶浓度的升高可以在异常组织基质层分解胶原蛋白,或者与不典型增生进展有关的上皮细胞增厚可能减弱激发光能量并掩盖深层胶原蛋白基底膜的胶原蛋白拉曼发射,这可能是造成癌组织拉曼信号在 870 cm^{-1} 处减小的原因^[1]。可以把这些特征作为判断胃组织是否癌变的标志。此外,癌结节组织在 1330 cm^{-1} 附近出现双峰,也是组织癌变的重要标志,但何种原因造成这一现象,还需进一步研究。

5 结 论

用SVM算法对癌结节和正常组织在体激光拉曼光谱进行分类诊断,灵敏度、特异度、准确度分别为95.73%、70.73%和90.73%。对不同生长期的癌结节组织激光拉曼光谱进行分期,诊断灵敏度、特异度和诊断准确度分别为98.82%、98.73%和98.78%。实验证明了癌结节和正常组织拉曼光谱差异显著,这为拉曼光谱在胃癌诊断中的应用提供了实验基础。而SVM算法与激光拉曼光谱相结合,扩展了数学算法在组织光谱分类中的应用。

参 考 文 献

- 1 S. K. Teh, W. Zheng, K. Y. Ho *et al.*. Near-infrared Raman spectroscopy for optical diagnosis in the stomach; identification of *Helicobacter-pylori* infection and intestinal metaplasia[J]. *Int. J. Cancer*, 2010, **126**(8): 1920~1927
- 2 Gao Zehong, Yu Jingong, Liu Fuxiang *et al.*. Micro-Raman

- spectra for lipids C-H in colorectal tissue[J]. *Chinese J. Lasers*, 2010, **37**(2): 605~608
- 高泽红,于晶功,刘福祥等. 结直肠癌组织中脂类伸缩振动的拉曼光谱[J]. *中国激光*, 2010, **37**(2): 605~608
- 3 Z. Huang, M. S. Bergholt, W. Zheng *et al.*. *In vivo* early diagnosis of gastric dysplasia using narrow-band image-guided Raman endoscopy[J]. *J. Biomed. Opt.*, 2010, **13**(5): 037017
- 4 S. H. Abigail, V. Zoya, A. G. Joseph *et al.*. *In vivo* margin assessment during partial mastectomy breast surgery using Raman spectroscopy[J]. *Cancer Res.*, 2006, **66**(6): 3317~3322
- 5 Zhang Xuegong. Introduction to statistical learning theory and support vector machines [J]. *Acta Automatica Sinica*, 2000, **26**(1): 32~42
- 张学工. 关于统计学习理论与支持向量机[J]. *自动化学报*, 2000, **26**(1): 32~42
- 6 S. K. Majumder, N. Ghosh, P. K. Gupta. Support vector machine for optical diagnosis of cancer[J]. *J. Biomed. Opt.*, 2005, **10**(2): 024034
- 7 Liu Lu, Liu Wanyu, Chu Chunyu *et al.*. Classification of tumid lymph nodes metastases and non-metastases from lung cancer in CT image [J]. *Journal of Electronics & Information Technology*, 2009, **31**(10): 2476~2482
- 刘露,刘丸予,楚春雨等. CT图像中肿大淋巴结肺癌转移分类方法[J]. *电子与信息学报*, 2009, **31**(10): 2476~2482
- 8 Ma Jun, Zhang Xining, Xu Ming *et al.*. Detection of gastric cancer peritoneal dissemination in nude mouse model by laser-induced fluorescence spectroscopy in vivo[J]. *Chinese J. Lasers*, 2009, **36**(10): 2566~2570
- 马君,张奚宁,徐明等. 激光诱导荧光技术在体探测裸鼠腹膜胃癌播散[J]. *中国激光*, 2009, **36**(10): 2566~2570
- 9 C. Cortes, V. Vapnik. Support vector networks[J]. *Machine Learning*, 1995, **20**: 273~297
- 10 J. A. Swets. Measuring the accuracy of diagnostic systems[J]. *Science*, 1988, **240**(4857): 1285~1293
- 11 Tang Weiyue, Wang Jiefang, Xu Ping. Research of stomach cancer tissue by Raman spectroscopy[J]. *Laser Journal*, 2004, **25**(1): 82~83
- 唐伟跃,王杰芳,徐平. 胃癌组织拉曼光谱的研究[J]. *激光杂志*, 2004, **25**(1): 82~83
- 12 Hu Yaogai, Shen Aiguo, Jiang Tao *et al.*. Classification of normal and malignant human gastric mucosa tissue with confocal Raman microspectroscopy and wavelet analysis[J]. *Spectrochim. Acta*, 2008, **69**: 378~382
- 13 F. Sebastiani. Machine learning in automated text categorization [J]. *ACM Comput. Surv.*, 2002, **34**(1): 1~47