

## 利用高斯函数拟合自体荧光光谱诊断乳腺癌的研究

陈文静, 教召航, 齐东丽, 沈龙海, 许诺, 谢东非, 李宜霏, 尤佳华, 李奇, 冯瑜\*

沈阳理工大学理学院, 辽宁 沈阳 110158

**摘要** 乳腺组织自体荧光光谱法通常采用 420~512 nm 波段的光源, 本次实验采用波长为 405 nm 的激光源获取乳腺组织切片的自体荧光光谱, 避免了活体检测时脂肪和血液对光谱的干扰。荧光光谱是多种发光基团的叠加光谱, 基于荧光光谱对乳腺癌细胞的生化信息进行定性和定量分析时, 光谱特征峰不易分辨。本文提出了高斯函数拟合乳腺组织切片荧光光谱的方法, 该方法可有效分离重叠峰。与正常组织相比, 癌变组织的荧光特征峰出现了明显的红移现象。通过面积比值法分析了正常与癌变乳腺组织中荧光物质的差异, 癌变组织在 517 nm 和 635 nm 处的峰面积与 492 nm 处的峰面积之比( $A_{517}/A_{492}$  和  $A_{635}/A_{492}$ )分别是正常组织的 2.4~5.5 倍和 5.4~8.5 倍, 可作为诊断乳腺癌的标准。最后, 本文分析了支持向量机(SVM)算法对乳腺组织荧光光谱进行分类的可行性, 分类准确率为 87.50%, 为乳腺癌的快速诊断提供了新方向。

**关键词** 医用光学; 高斯函数拟合; 乳腺癌; 荧光光谱; 支持向量机

中图分类号 O433.4

文献标志码 A

DOI: 10.3788/CJL202249.2007106

## 1 引言

乳腺癌是最常见的恶性疾病, 具有较高的死亡率和发病率, 是全球特别是发展中国家女性死亡的主要原因之一<sup>[1-3]</sup>。乳腺癌筛查方法有 X 光摄影、钼靶检查、乳腺超声和乳腺增强磁共振等, 但这些方法作为临床辅助手段, 无法作出最终诊断, 须以临床组织学结论为诊断金标准。由于组织学诊断易受医师经验等外界因素的影响, 临床上迫切需要客观、快速、高准确率的诊断方法。

光谱学为临床检测正常和癌变组织提供了一种新方法<sup>[4]</sup>。作为获取与生物化学信息相关的代谢性质以及组织中细胞外基质结构成分的工具, 荧光光谱法具有无损、安全、高效的优点<sup>[5-7]</sup>。丁建华等<sup>[8]</sup>使用 488 nm 激光源获取离体乳腺正常组织、良性瘤组织和癌组织的自体荧光光谱, 分析后发现三种组织的荧光光谱具有明显差异。荧光光谱是多个显著荧光团的叠加光谱, 直接分辨谱峰具有较高难度。已有研究表明, 可以使用高斯函数拟合方法、Gram-Schmidt 方法以及光谱反褶积等方法得到荧光光谱的基谱, 实现对疾病的诊断。秦艳利等<sup>[9]</sup>使用高斯曲线拟合方法分别对健康牙齿以及患有牙结石牙齿的荧光光谱进行拟合, 结果发现, 676 nm 和 486 nm 处的荧光峰强度比可以作为诊断牙结石的标准。Pu 等<sup>[10]</sup>使用 Gram-Schmidt

方法得到了乳腺组织中关键荧光团的光谱, 有效检测了乳腺组织中荧光团的组成随癌症发展的变化。Chithra 等<sup>[11]</sup>使用光谱反褶积方法计算不同内源荧光团的荧光强度、峰值波长和全波半峰全宽最大值, 得到了乳腺癌患者血浆中的多种荧光基团, 并发现荧光基团信息可以将治疗和未治疗患者区分开来。

乳腺组织自体荧光光谱法多采用 420~512 nm 波段的光源, 而本团队采用波长为 405 nm 的激光源采集乳腺组织切片的自体荧光光谱, 避免了活体检测时脂肪和血液对光谱的干扰, 提高了光谱强度。由于部分荧光物质所发射的荧光波长距离较近, 会出现重叠现象, 直接使用荧光光谱峰面积进行量化分析, 会导致量化不客观、不准确等问题。为了解决这一问题, 本文使用高斯函数拟合乳腺组织切片的自体荧光光谱, 然后利用拟合后的峰面积比分析正常组织和癌变组织的荧光光谱信息(峰面积、中心波长、半峰全宽)。研究表明, 使用高斯函数拟合荧光光谱可以简化提取的光谱信息, 并且可以准确、客观地分析正常组织和癌变组织的荧光光谱, 最终诊断是否为乳腺癌。对较大数量样本研究后, 本文决定采用支持向量机(SVM)算法对荧光光谱进行分类, 从而实现对乳腺癌的快速诊断。

## 2 实验系统和方法

实验装置如图 1 所示。波长为 405 nm 的激光经

收稿日期: 2022-05-16; 修回日期: 2022-06-21; 录用日期: 2022-07-20

基金项目: 科技部外专局项目(G2021006007L)、科技部重点研发项目(KJHZ-2019-01L)、超硬材料国家重点实验室开放课题资助项目(202004)

通信作者: \*14025921@qq.com

透镜 1 聚焦后被二向色镜反射,通过透镜 2 后聚焦到乳腺组织切片样品上。样品固定在精密二维平移台上,移动平移台就可以获取样品不同位置处的自体荧光光谱。样品被激发出的荧光透过二向色镜后被透镜

3 聚焦,聚焦后的光信号被光纤探头耦合到光谱仪,最后信号经光谱仪内置的 CCD 探测器完成光电转换后传输到计算机,通过软件对采集的光谱数据进行观察和分析。实验所用仪器的详细参数如表 1 所示。

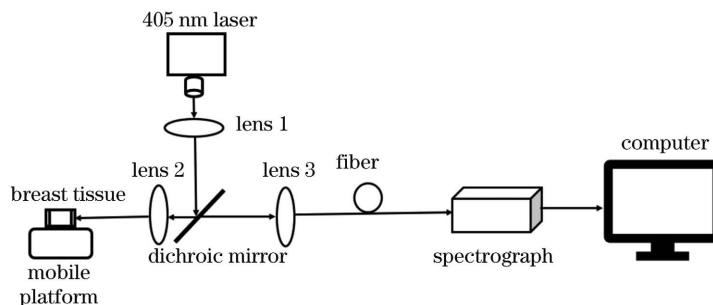


图 1 实验装置示意图

Fig. 1 Schematic of experimental device

表 1 实验仪器信息表

Table 1 Experimental instrument information table

Name	Instrument information
Laser	Wavelength: 405 nm; power: 150 mW
Dichroic mirror	Critical wavelength: 410 nm; transmission wavelength: > 419 nm; transmittance: > 80%; reflection wavelength: 360-405 nm; reflectance: > 99%
Fiber probe	Diameter: 400 μm; wavelength range: 300-800 nm
Fiber spectrometer	Spectral range: 200-1100 nm; scanning speed: 4500 scan·s <sup>-1</sup> ; resolution: 0.8 nm

用于实验的乳腺组织切片样本共 80 片,包括 28 片 SP 染色正常组织切片,10 片苏木精-伊红(HE)染色癌变组织切片,10 片未染色癌变组织切片,32 片 SP 染色癌变组织切片。SP 染色法中使用的染色剂由蒸馏水、DAB A、DAB B、DAB C 按体积比为 5:1:1:1 混合配制而成,HE 染色法中的染色剂由苏木精与伊红按体积比为 1:1 混合配制而成。染色剂配比时效为 0.5 h,需在该时间内完成光谱采集。

本次实验中对每个组织样本分别检测 10 个不同的位置点,每一个位置点均采集三次光谱,共获取 4 类样本的 2400 组数据。采集光谱积分时间为 112 ms,波长采集范围为 450~750 nm。为避免日光等照明光源对实验结果造成干扰,以上实验操作均在暗室条件下完成。

数据处理采用 Origin 软件,先对 4 类乳腺组织样本的荧光光谱进行平均处理,然后使用 Savitzky-Golay 函数进行平滑处理,接着使用最值法将光谱峰值归一化到(0,1)范围内,最后使用高斯函数对归一化后的荧光光谱进行拟合。

### 3 分析与讨论

#### 3.1 背景荧光

实验中使用的载玻片和染色剂产生的背景荧光会影响组织样品的荧光光谱,因此本次实验分别采集了载玻片和染色剂的荧光光谱,如图 2(a)、(b)所示。由图 2(a)可知,实验中使用的载玻片不产生背景荧光。由图 2(b)可知,染色剂存在背景荧光,荧光峰位置在 551~555 nm 范围内。

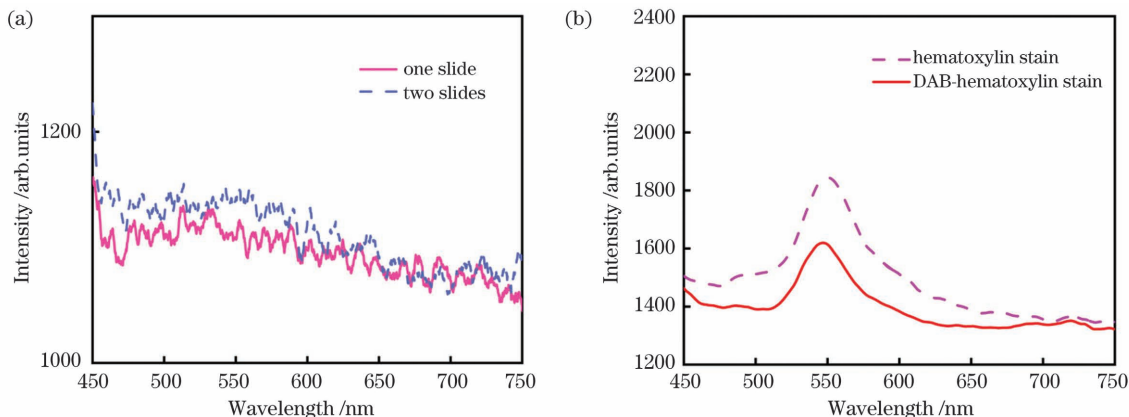


图 2 载玻片和染色剂的荧光光谱。(a)载玻片的荧光光谱;(b)染色剂的荧光光谱

Fig. 2 Fluorescence spectra of slide and stain. (a) Fluorescence spectra of slide; (b) fluorescence spectra of stain

### 3.2 乳腺组织的荧光光谱分析

图 3 是乳腺组织切片的归一化平均荧光光谱图。数据经平滑和归一化处理,减少了噪声干扰,提高了信噪比,增加了不同组织光谱数据间的可比性。如图 3 所示,正常和癌变乳腺组织的荧光光谱有较多重叠峰,癌变组织的荧光峰出现红移现象。

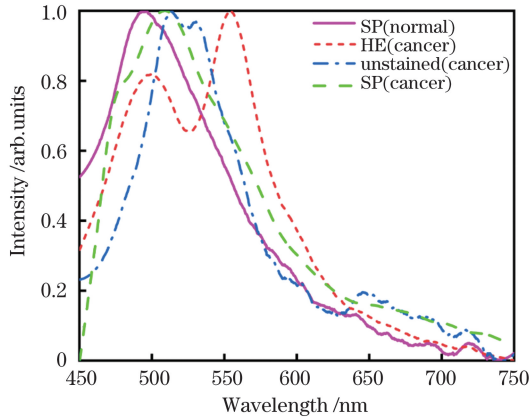


图 3 乳腺组织的归一化平均荧光光谱

Fig. 3 Normalized mean fluorescence spectra of breast tissues

图 4 为高斯函数拟合的 SP 染色正常乳腺组织、未染色癌变乳腺组织、HE 染色癌变乳腺组织、SP 染色癌变乳腺组织的归一化平均荧光光谱图,高斯峰的中心波长分别为  $492 \text{ nm} \pm 3 \text{ nm}$ 、 $517 \text{ nm} \pm 4 \text{ nm}$ 、

$553 \text{ nm} \pm 3 \text{ nm}$ 、 $635 \text{ nm} \pm 10 \text{ nm}$ , 决定系数 ( $R^2$ ) 分别为 0.998、0.987、0.992、0.996。位于 492 nm 处的荧光峰归属于烟酰胺腺嘌呤二核苷酸 (NADH)。NADH 是一种存在于所有活细胞中的天然辅酶,该物质通过一种新的信号通路调节免疫反应和维持内环境稳定<sup>[12-13]</sup>。在癌变过程中,细胞需要消耗大量能量才能完成代谢和繁殖活动。NADH 参与多项人体生理代谢过程,起到电子传递体的作用,所以癌变组织中含有的 NADH 较正常组织更少。核黄素包含黄素腺嘌呤二核苷酸 (FAD) 和黄素单核苷酸 (FMN),其荧光峰位于 510~530 nm 之间,因此 517 nm 处的荧光物质为核黄素<sup>[14]</sup>。核黄素是一种水溶性 B 族维生素,位于线粒体内膜上,核黄素参与线粒体的能量代谢。线粒体膜呈现弱电性,会使核黄素聚集<sup>[15]</sup>;同时,癌变乳腺细胞中组蛋白赖氨酸特异性去甲基化酶 1 (LSD1) 的表达水平较低,也会造成核黄素聚集<sup>[16]</sup>。551~555 nm 处的荧光峰为染色剂的荧光峰。630~650 nm 波段的荧光峰归属于乳腺组织细胞质中含有的卟啉物质。癌变过程中,细胞中的胆色素脱氢酶会产生大量卟啉物质,而亚铁螯合酶活力下降,无法及时将卟啉转换为血红素,也会导致卟啉积累<sup>[17-18]</sup>。表 2 是高斯拟合峰的特征参量信息,其中  $A_1$ 、 $A_2$ 、 $A_3$  分别指 NADH、FAD 与 FMN、卟啉拟合峰的峰面积, $\lambda_1$ 、 $\lambda_2$ 、 $\lambda_3$  分别指 NADH、FAD 与 FMN、卟啉拟合峰

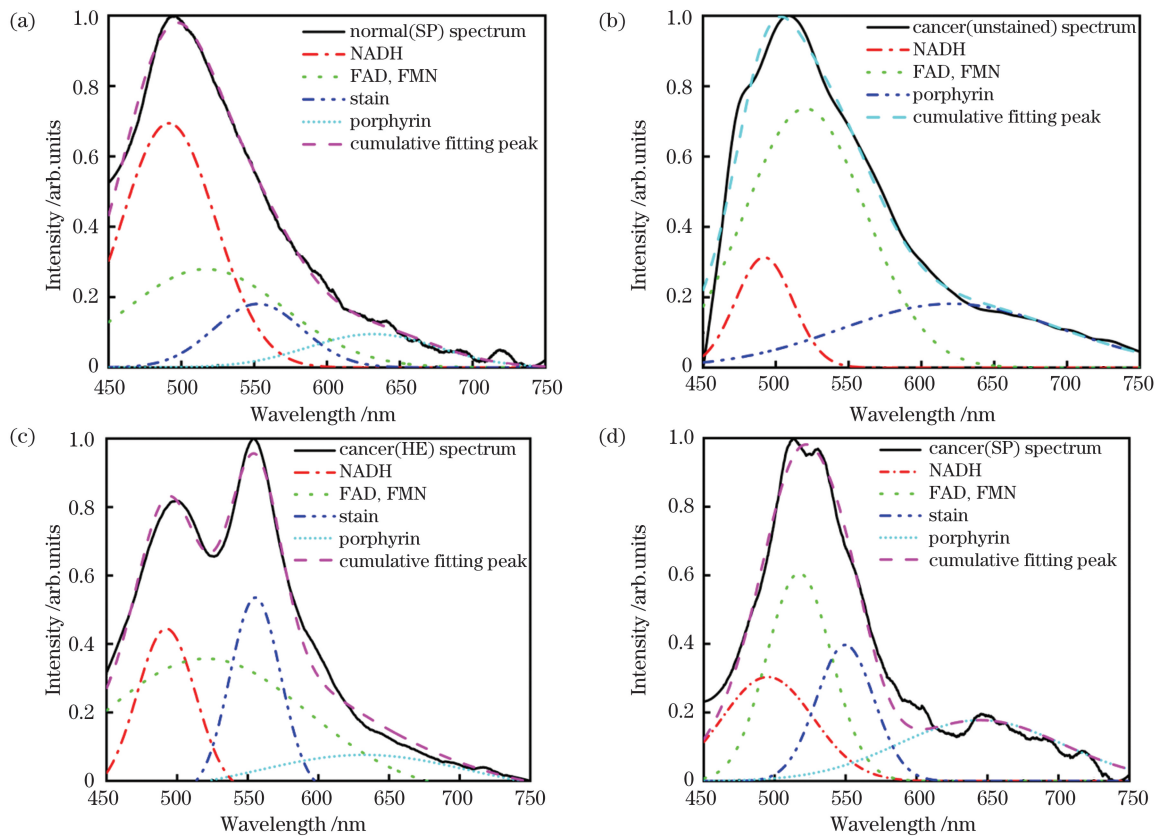


图 4 高斯函数拟合的归一化平均荧光光谱图。(a)SP 染色正常乳腺组织;(b)未染色癌变乳腺组织;(c)HE 染色癌变乳腺组织;(d)SP 染色癌变乳腺组织

Fig. 4 Normalized mean fluorescence spectrograms fitted by Gaussian function. (a) Normal breast tissue by SP staining; (b) unstained cancerous breast tissue; (c) cancerous breast tissue by HE staining; (d) cancerous breast tissue by SP staining



表 2 高斯拟合峰的特征参量信息  
Table 2 Characteristic parameter information of Gaussian fitting peak

Sample	$A_1$	$A_2$	$A_3$	$\lambda_1$	$\lambda_2$	$\lambda_3$	$\omega_1$	$\omega_2$	$\omega_3$	$R^2$
SP(normal)	55.73	37.59	10.61	491.39	516.76	631.68	64.06	107.37	89.16	0.998
Unstained(cancer)	19.48	72.23	31.49	492.74	520.77	630.59	39.41	82.59	147.65	0.987
HE(cancer)	22.94	70.61	23.81	492.35	520.89	631.24	41.38	135.55	147.13	0.992
SP(cancer)	24.47	38.57	27.86	495.14	517.25	640.84	64.34	45.05	116.20	0.996

的中心波长,  $\omega_1$ 、 $\omega_2$ 、 $\omega_3$  分别指 NADH、FAD 与 FMN、吡啶拟合峰的半峰全宽。

由表 2 可知,与正常乳腺组织相比,乳腺癌组织在 492 nm 处的峰面积增大,在 517 nm 和 635 nm 处的峰面积减小,峰面积比值差异明显。因此,本文使用  $A_{517}/A_{492}$  和  $A_{635}/A_{492}$  参数衡量正常和癌变组织。表 3 为正常和癌变乳腺组织的峰面积比值。

表 3 正常乳腺和癌变乳腺组织的峰面积比值

$A_{517}/A_{492}$  和  $A_{635}/A_{492}$   
Table 3 Peak area ratios of  $A_{517}/A_{492}$  and  $A_{635}/A_{492}$  of normal and cancerous breast tissues

Breast tissue	$A_{517}/A_{492}$	$A_{635}/A_{492}$
Normal tissue with SP staining	0.6745	0.1904
Unstained cancerous tissue	3.7079	1.6105
Cancerous tissue with HE staining	3.0780	1.0379
Cancerous tissue with SP staining	1.5762	1.1385

由表 3 可以看出,癌变组织的峰面积比  $A_{517}/A_{492}$  和  $A_{635}/A_{492}$  分别是正常组织的 2.4~5.5、5.4~8.5 倍。因此,可把 517 nm、640 nm 处与 492 nm 处峰面积的比值( $A_{517}/A_{492}$  和  $A_{635}/A_{492}$ )作为检测乳腺癌的标准。

表 3 表明,与染色处理的组织相比,未染色处理的组织具有较高的峰面积比,染色会影响乳腺组织的荧光光谱信息。这可能是染色机理导致的。在 HE 染色过程中,细胞的不同部位具有不同的酸碱敏感程度(细胞质的等电点为 4.7~5.0,具有嗜酸性,而细胞核的等电点为 3.3~3.6,具有嗜碱性),因此,细胞质被碱性伊红染液染色后呈现明亮的鲜红色,而细胞核被酸性苏木精染液染色后呈现有光泽的蓝色,可用于区分一般细胞的形态<sup>[19]</sup>。有研究证明,苏木精对伊红染液有荧光淬灭效应,苏木精伊红混合染液的荧光强度较低,对组织自体荧光产生的干扰较小。所以,HE 染色组织切片的峰面积比与石蜡包埋未染色组织切片的峰面积比略有不同,这与文献[20]所得结果一致。SP 染色的机理是:位于细胞核、细胞质、细胞膜上的抗体与被检测组织中的目的蛋白抗原特异性结合后,通过 DAB 显色<sup>[21]</sup>。在 SP 染色法中,为去除细胞中残留的杂蛋白、内源性酶等,避免非特异性着色,常使用过氧化物酶封闭液破坏酶活性,这样做同时也破坏了组织内源性荧光物质,导致荧光强度降低,因而造成了两种

染色组织切片的荧光光谱差异。

综上所述,HE 染色中使用的苏木精对伊红染液具有荧光淬灭效应,对组织中内源性荧光物质的影响较小,而 SP 染色过程中使用的封闭液对组织中内源性荧光物质的影响较大,在采集乳腺组织的自体荧光光谱时,应尽量选用石蜡包埋或 HE 染色切片。

### 3.3 荧光光谱分类算法实验

采用高斯函数对乳腺组织切片的自体荧光光谱进行拟合,可以提取光谱中的关键荧光基谱,然后通过峰面积比就可以判断组织是否癌变。为客观、准确、高效地筛查乳腺癌变组织,本次实验将采集的乳腺组织切片自体荧光光谱作为数据集,分析了 SVM 算法对乳腺组织荧光光谱进行分类、判别的可行性。SVM 算法是一种有监督的分类算法,利用最优超平面对预测样本点的类别进行判断<sup>[22-24]</sup>,具有适用于多样本数据集的优势。

正常和癌变乳腺组织的荧光光谱在 490~540 nm 波段内有较大重叠,不利于判断乳腺组织是否癌变。SVM 算法使用核函数将正常和癌变乳腺组织的荧光光谱数据映射为可分数据,可以快速区分癌变组织。采用 SVM 算法对乳腺组织的荧光光谱进行判别,判别结果的准确率是 87.50%,说明 SVM 算法对乳腺组织荧光光谱分类具有潜在的可行性。后续研究可以从电压、细胞 pH 值等影响组织样品荧光光谱的外界环境因素出发,通过抑制噪声干扰来提高信噪比,进而提高准确率。判别结果的召回率和精度分别为 93.94%、88.57%,说明 SVM 算法可以将癌变乳腺组织荧光光谱的特征信息查全、查准,对乳腺癌变组织的荧光光谱信息具有较高的敏感度和识别能力。判别结果的 F1 分数为 91.18%,说明 SVM 算法在检测乳腺组织荧光光谱中平衡召回率和精度的能力较强。为了与 SVM 算法的判别结果进行对比,后续工作拟考虑应用 K 最近邻(KNN)、卷积神经网络(CNN)等算法分析乳腺组织的荧光光谱,进而进行分类判别。

## 4 结 论

本团队搭建了 405 nm 激光荧光光谱采集系统,并采用该系统采集了乳腺组织切片的自体荧光光谱;之后提出了使用高斯函数拟合自体荧光光谱的方法,计算了正常和癌变组织在 517 nm 和 635 nm 处的荧光光谱峰面积分别与 492 nm 处荧光光谱峰面积的比

值;最后分析了 SVM 算法对荧光光谱的分类能力。研究表明:癌变后,细胞中的荧光基因受代谢影响,导致癌变乳腺组织较正常组织的自体荧光光谱出现了红移现象;使用高斯函数拟合荧光光谱可将重叠峰分离,得到癌变组织在 517 nm 和 635 nm 处的峰面积与 492 nm 处峰面积之比分别为正常组织的 2.4~5.5 倍和 5.4~8.5 倍,说明峰面积比对正常与癌变组织具有较强的区分能力。SVM 算法的判别结果(准确率为 87.50%,召回率为 93.94%,精度为 88.57%,F1 分数为 91.18%)证明该方法对癌变乳腺组织具有查全、查准能力,对召回率和精度的平衡能力较强,为乳腺癌的快速诊断提供了研究方向。

综上所述,采用高斯函数拟合荧光光谱可以得到乳腺癌组织中荧光基因的光谱特征信息,峰面积比可以作为诊断乳腺癌的标准,荧光光谱与 SVM 分类算法模型的结合可以为多样本乳腺癌的快速诊断提供可行途径。

### 参 考 文 献

- Akram M, Iqbal M, Daniyal M, et al. Awareness and current knowledge of breast cancer[J]. *Biological Research*, 2017, 50(1): 33.
- Siegel R L, Miller K D, Jemal A. Cancer statistics, 2021[J]. *CA: A Cancer Journal for Clinicians*, 2021, 71(1): 7-30.
- Sung H, Ferlay J, Siegel R L, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. *CA: A Cancer Journal for Clinicians*, 2021, 71(3): 209-249.
- Peng X R, Dai R Y, Ma Y Q, et al. Early diagnosis and bioimaging of lung adenocarcinoma cells/organs based on spectroscopy machine learning[J]. *Journal of Innovative Optical Health Sciences*, 2022, 15(2): 2250011.
- Andreeva V, Aksamentova E, Muhachev A, et al. Preoperative AI-driven fluorescence diagnosis of non-melanoma skin cancer[J]. *Diagnostics*, 2021, 12(1): 72.
- Litvinova K S, Rafailov I E, Dunaev A V, et al. Non-invasive biomedical research and diagnostics enabled by innovative compact lasers[J]. *Progress in Quantum Electronics*, 2017, 56: 1-14.
- 陈毛宁, 吴玲玲. 荧光分析法在早期龋齿检测中的应用研究[J]. *激光与光电子学进展*, 2022, 59(13): 1330001.
- Chen M N, Wu L L. Application of fluorescence analysis in the detection of early caries[J]. *Laser & Optoelectronics Progress*, 2022, 59(13): 1330001.
- 丁建华, 余虹, 张志麟, 等. 离体乳腺正常组织、良性瘤及癌的自体荧光光谱[J]. *应用激光*, 2000, 20(1): 37-39.
- Ding J H, Yu H, Zhang Z L, et al. Laser induced fluorescence spectra of normal, benign and malignant tissues of breast *in vitro*[J]. *Applied Laser*, 2000, 20(1): 37-39.
- 秦艳利, 刘奕, 徐送宁, 等. 利用高斯函数拟合荧光光谱诊断龈下结石的研究[J]. *光电子·激光*, 2009, 20(7): 988-990.
- Qin Y L, Liu Y, Xu S N, et al. Detection of subgingival calculus by Gaussian curve fitting of laser-induced fluorescence emission[J]. *Journal of Optoelectronics-Laser*, 2009, 20(7): 988-990.
- Pu Y, Sordillo L A, Yang Y L, et al. Key native fluorophores analysis of human breast cancer tissues using Gram-Schmidt subspace method[J]. *Optics Letters*, 2014, 39(24): 6787-6790.
- Chithra K, Aruna P, Einstein G, et al. Monitoring breast cancer response to treatment using Stokes shift spectroscopy of blood plasma[J]. *Journal of Fluorescence*, 2019, 29(3): 803-812.
- Xue J P, Pu Y, Smith J, et al. Machine learning based analysis of human prostate cancer cell lines at different metastatic ability using native fluorescence spectroscopy with selective excitation wavelength[J]. *Proceedings of SPIE*, 2018, 10504: 105040L.
- 宋晓康, 赵强, 张元志, 等. 利用三维荧光光谱与平行因子分析法测定细胞培养基中多类代谢成分的含量[J]. *中国激光*, 2022, 49(9): 0911001.
- Song X K, Zhao Q, Zhang Y Z, et al. Utilizing three-dimensional fluorescence spectra and parallel factor analysis algorithm to quantify concentration of multiple metabolic fluorophores in a cell culture medium[J]. *Chinese Journal of Lasers*, 2022, 49(9): 0911001.
- 王可人, 叶玉琴, 朱珠, 等. 荧光光谱学在乳腺疾病诊治中的研究进展[J]. *中国实验诊断学*, 2013, 17(2): 386-388.
- Wang K R, Ye Y Q, Zhu Z, et al. Research progress of fluorescence spectroscopy in diagnosis and treatment of breast diseases[J]. *Chinese Journal of Laboratory Diagnosis*, 2013, 17(2): 386-388.
- 罗贤懋, 林培中, 刘雨菁, 等. 核黄素预防恶性肿瘤的研究进展[J]. *癌症进展*, 2020, 18(4): 325-330.
- Luo X M, Lin P Z, Liu Y J, et al. Research progress in prevention of riboflavin malignant tumor [J]. *Oncology Progress*, 2020, 18(4): 325-330.
- 王豪杰. 基于辅酶 FAD 设计的新型 5-氧基-6-苯基嘧啶类 LSD1 抑制剂及其生物活性研究[D]. 郑州: 郑州大学, 2020.
- Wang H J. Design and activity evaluation of novel 5-cyano-6-phenyl pyrimidine LSD1 inhibitors based on coenzyme FAD[D]. Zhengzhou: Zhengzhou University, 2020.
- Balsukuri N, Das S, Gupta I. Carbazole-corrrole and carbazole-prophyrin dyads: synthesis, fluorescence and electrochemical studies[J]. *New Journal of Chemistry*, 2015, 39(1): 482-491.
- 陈家璧, 彭润玲. 激光原理及应用[M]. 北京: 电子工业出版社, 2019.
- Chen J B, Peng R L. Principle and application of laser[M]. Beijing: Publishing House of Electronics Industry, 2019.
- 翁丹枫, 唐坚清, 黎三艳. 苏木精与伊红染液在病理小标本制片中的应用比较[J]. *临床与实验病理学杂志*, 2019, 35(5): 610-611, 613.
- Weng D F, Tang J Q, Li S Y. Comparison of hematoxylin and eosin staining in the preparation of small pathological specimens[J]. *Chinese Journal of Clinical and Experimental Pathology*, 2019, 35(5): 610-611, 613.
- 赵箭光, 吴小玲, 杨佩青, 等. 肝癌与正常肝组织激光荧光光谱的比较研究[J]. *中国医学物理学杂志*, 1996, 13(4): 220-222.
- Zhao J G, Wu X L, Yang P Q, et al. Comparative research on laser fluorescence spectrum of liver cancer and normal human liver tissues[J]. *Chinese Journal of Medical Physics*, 1996, 13(4): 220-222.
- 林娟, 晋龙, 吴晶晶, 等. 不同脱黑色素方法对 HE 和免疫组化染色结果的影响[J]. *临床与实验病理学杂志*, 2021, 37(6): 751-753.
- Lin J, Jin L, Wu J J, et al. Effects of different melanin removal methods on HE and immunohistochemical staining results[J]. *Chinese Journal of Clinical and Experimental Pathology*, 2021, 37(6): 751-753.
- Yang F C, Chen W, Wei H F, et al. Machine learning for histologic subtype classification of non-small cell lung cancer: a retrospective multicenter radiomics study [J]. *Frontiers in Oncology*, 2021, 10: 608598.
- 储宝, 黄尧, 倪敬书, 等. 基于 GS-SVM 的脑组织差分路径因子定量方法研究[J]. *中国激光*, 2022, 49(5): 0507008.
- Chu B, Huang Y, Ni J S, et al. Quantitative methods of brain tissue differential pathlength factor based on GS-SVM [J]. *Chinese Journal of Lasers*, 2022, 49(5): 0507008.
- 陈鹏, 齐超, 刘人玮, 等. 基于支持向量机回归的 LIBS 飞灰含碳量定量分析[J]. *光学学报*, 2022, 42(9): 0930003.
- Chen P, Qi C, Liu R W, et al. Quantitative analysis of carbon content in fly ash using LIBS based on support vector machine regression[J]. *Acta Optica Sinica*, 2022, 42(9): 0930003.

# Diagnosis of Breast Cancer Using Gaussian Function to Fit Autofluorescence Spectrum

Chen Wenjing, Jiao Zhaohang, Qi Dongli, Shen Longhai, Xu Nuo, Xie Dongfei,  
Li Yifei, You Jiahua, Li Qi, Feng Yu\*

*College of Science, Shenyang Ligong University, Shenyang 110158, Liaoning, China*

## Abstract

**Objective** Breast cancer, the most common malignant disease with high mortality and morbidity, is the leading cause of death in women worldwide, especially in developing countries. The current methods for breast cancer diagnosis are time-consuming, expensive, and have low sensitivity, and these should be urgently addressed. Spectroscopy appears to offer a new method for clinical detection of normal and cancerous tissues. Fluorescence spectroscopy is a tool for the noninvasive acquisition of biochemical information and extracellular matrix, and its use is rapidly expanding owing to its safety and efficiency.

**Methods** Breast tissue autofluorescence spectroscopy mostly uses a light source in the 420–512 nm band, and the tryptophan and tyrosine residues in the tissue are more sensitive to a laser with a wavelength of 405 nm. Therefore, this study used a 405 nm laser source to obtain autofluorescence spectra of breast tissue sections to avoid the interference of fat and blood on the spectrum during *in vivo* detection and to improve the spectrum intensity. Various biological macromolecules or molecular groups in human tissues emit fluorescence, and the fluorescence spectrum is a superimposed spectrum of various luminescent substances. Because of the short distance between the fluorescence wavelengths emitted by some fluorescent substances, a peak overlapping phenomenon occurs, and the spectral characteristic peaks are not easy to distinguish. On directly using the fluorescence spectral peak area for quantitative analysis, the quantification is neither objective nor accurate. Therefore, this study proposes a Gaussian function to fit the fluorescence spectrum of breast tissue slices. Because the Gaussian function describes the normal distribution, it effectively separates the overlapping peaks, simplifies the spectrum, and extracts information using the characteristic parameter peaks of the fitted spectrum. The area ratio method was used to analyze the differences in the spectra of the normal and cancerous breast tissues. Finally, after studying several samples, this paper proposes the use of an SVM algorithm to classify the fluorescence spectrum and analyzes the feasibility of this algorithm for the classification of the breast tissue fluorescence spectrum.

**Results and Discussions** After the breast tissue becomes cancerous, the content of some fluorescent substances also changes. The normalized mean fluorescence spectra of normal and cancerous breast tissues show obvious red shifts in the fluorescence peaks of cancerous tissues relative to normal tissues (Fig. 4). The difference in fluorescent substances in normal and cancerous breast tissues was analyzed using area ratio method. The peak area ratios ( $A_{517}/A_{492}$ ,  $A_{635}/A_{492}$ ) of cancerous tissues are 2.4–8.5 times that of normal tissues, which can be considered as criteria for the diagnosis of breast cancer (Table 3). The discrimination results of the SVM algorithm are as follows: accuracy rate of 87.50%, recall rate of 93.94%, precision of 88.57%, and F1-score of 91.18%. This proves that the method has high sensitivity and recognition ability for the fluorescence spectral information of breast cancer tissue, and can map normal and cancerous tissue fluorescence spectral signatures in different states for classification.

**Conclusions** In this study, the fluorescence spectrum of breast tissues under a 405 nm laser was collected. After fitting the fluorescence spectrum with a Gaussian function, the fluorescence spectrum peak area ratios for normal and cancerous tissues were calculated, and the effect of the SVM algorithm on the fluorescence spectrum was analyzed. The autofluorescence spectrum of cancerous breast tissues has a red-shift phenomenon compared with that of normal tissues; after separating the overlapping peaks by Gaussian function fitting fluorescence spectrum, the peak areas at 517 nm, 635 nm, and 492 nm of cancerous tissues have a significant ability to distinguish between normal and cancerous tissues. The SVM algorithm is feasible for the classification of breast tissue fluorescence spectrum, with an accuracy of 87.50%. This proves that the method can recall and pinpoint cancerous breast tissues, and has a strong balance between recall rate and precision, thus, providing a rapid diagnosis of breast cancer. In summary, the Gaussian function fitting the fluorescence spectrum can obtain the spectral characteristic information of the fluorophore in breast cancer and the peak area ratio can be used as the standard for diagnosing breast cancer. Moreover, the combination of the fluorescence spectrum and the SVM classification algorithm model can be used for multi-sample breast cancer, providing a feasible method for rapid diagnosis.

**Key words** medical optics; Gaussian function fitting; breast cancer; fluorescence spectrum; support vector machine