

DOI: 10.19650/j.cnki.cjsi.J1905739

基于高波数拉曼光谱的口腔癌检测方法研究^{*}

胡 阳¹, 董明利¹, 于明鑫¹, 张 韬^{2,3}, 朱智慧^{2,3}

(1. 北京信息科技大学 生物医学检测技术及仪器北京实验室 北京 100192; 2. 中国医学科学院
北京协和医学院 北京协和医院 口腔科 北京 100730; 3. 中国医学科学院北京
协和医学院 北京协和医院 协和转化医学中心 北京 100730)

摘 要:目前口腔癌触诊检查依赖于医生经验,术中冰冻切片的检测方式为有创检测,准确性较低,为满足临床手术对口腔癌检测的实际需求,研究了基于高波数($2\,400\sim 4\,000\text{ cm}^{-1}$)光纤激光拉曼光谱的口腔癌检测方法。研究光纤拉曼光谱实现口腔癌检测的原理,分析了主成分-线性判别分析(PCA-LDA)统计算法的可行性,搭建了便携口腔癌组织检测光纤拉曼系统,制备了口腔癌组织和正常组织病理切片,采集了口腔癌组织和正常组织的高波数拉曼光谱,利用 PCA-LDA 算法对光谱进行处理和分析。光谱分析结果表明,癌组织和正常组织的平均光谱在 $2\,780$ 、 $2\,890$ 、 $2\,936$ 、 $3\,180$ 、 $3\,285$ 、 $3\,300$ 、 $3\,650\text{ cm}^{-1}$ 有较大区别,利用 PCA-LDA 分析,对正常组织和口腔癌组织拉曼光谱进行降维分类,检测准确率为 96.6%,与共聚焦显微拉曼光谱检测相比,提高了检测效率,较指纹拉曼光谱检测准确率提高约 0.6%,研究表明高波数拉曼光谱结合统计学方法可以较好的区分正常组织和癌组织,有望成为一种高灵敏度,高时效性和高准确率的口腔癌检测方法,对术中口腔癌检测有重要意义。

关键词: 光纤拉曼光谱;高波数;口腔癌;术中在体检测

中图分类号: TH787 **文献标识码:** A **国家标准学科分类代码:** 460.4035

Study on oral cancer detection method based on high wavenumber Raman spectroscopy

Hu Yang¹, Dong Mingli¹, Yu Mingxin¹, Zhang Tao^{2,3}, Zhu Zhihui^{2,3}

(1. Beijing Laboratory of Biomedical Detection Technology and Instruments, Beijing Information Science and Technology University, Beijing 100192, China; 2. Department of Stomatology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China; 3. Translational Medical Center, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China)

Abstract: At present, the palpation examination of oral cancer relies on doctor's experience; the detection method of intraoperative frozen section is invasive detection and the accuracy is low. In order to meet the actual needs of clinical surgical oral cancer detection, the oral cancer detection method based on high-wavenumber ($2\,400\sim 4\,000\text{ cm}^{-1}$) fiber laser Raman spectroscopy is studied. The principle using fiber Raman spectroscopy to realize oral cancer detection is studied. The feasibility of the Principal Component Analysis-Linear Discriminant Analysis (PCA-LDA) statistical algorithm is analyzed. A portable oral cancer tissue detection fiber Raman system was built, and the histopathological sections of the oral cancer tissues and normal tissues were prepared. The high-wavenumber Raman spectra of the oral cancer tissues and normal tissues were collected, and the spectra were processed and analyzed with PCA-LDA algorithm. The spectral analysis results show that the average spectra of the cancer tissues and normal tissues are significantly different at $2\,780$, $2\,890$, $2\,936$, $3\,180$, $3\,285$, $3\,300$ and $3\,650\text{ cm}^{-1}$. The Raman spectroscopy of the normal tissues and oral cancer tissues were dimension-reduced and classified with the PCA-LDA analysis, and the detection accuracy is 96.6%. Compared with the result of confocal Raman spectroscopy, the detection efficiency is improved. Compared with the result of fingerprint Raman spectrum detection,

收稿日期: 2019-10-21 Received Date: 2019-10-21

^{*} 基金项目: 北京市自然科学基金-海淀原始创新联合基金(L182066)项目资助

the detection accuracy is improved by about 0.6%. The study result indicates that combining statistical method, the high-wavenumber Raman spectroscopy method can nicely distinguish between normal tissues and cancerous tissues, which is expected to be an oral cancer detection method with high sensitivity, high time-efficiency and high accuracy, and has important significance for intraoperative oral cancer detection.

Keywords: fiber Raman spectroscopy; high wavenumber; oral cancer; intraoperative in vivo detection

0 引 言

口腔癌是常见癌症的一种,中国口腔癌发病率为 8.7/10 万人,在癌症发病率中排第 6 位^[1]。原位癌治愈率接近 100%,I 期患者术后 5 年生存率为 60%~90%,II 期和 IV 期患者术后 5 年生存率仅为 5%~20%^[2],口腔癌检测的及时性和准确性对治疗效果和患者生存率影响显著,研究口腔癌检测方法具有十分重要的意义。

口腔癌组织术中检测方式主要有触诊和病理切片。触诊的诊断方式高度依赖医生的经验,不能准确定位癌组织区域,常用于口腔癌的早期诊断,不符合术中癌组织快速精确定位的需求。术中冰冻切片是目前口腔癌切除手术中用于确定癌组织切除彻底与否的手段,由于取材有限,不对切片组织进行染色,检测准确性较低^[3],同时,冰冻切片检查是一种有创方式,会增加患者痛苦。所以急需研究一种安全简便的口腔癌检测术中检测方法,减少患者痛苦,提高癌组织检测精确度。

随着光学技术的发展,光谱分析技术在医学领域的应用越来越广泛^[4-5]。拉曼光谱与物质的分子成分或者转动模式相关,用不同频率的入射光照射同一物质时,其拉曼频率的改变量相同,因此拉曼光谱可以用于确定物质分子水平的化学结构,进而对待测物质进行定性、定量的分析,由于水的拉曼信息极其微弱,因此拉曼光谱技术可不受人体组织中水的干扰,适合进行在体成分检测,通过选择合理的数据处理和寻峰算法,可以实现对特征物质的拉曼峰的识别^[6]。目前,拉曼检测已逐渐进入医学诊断领域,林学亮等^[7]将表面增强拉曼(surface-enhanced raman scattering, SERS)应用于鼻咽癌的唾液检查中,获得了较好的灵敏度和特异性结果;刘燕玲等^[8]将 SERS 用于诊断肠胃癌和结直肠癌的临床诊断中,使用银溶胶作为增强基底,利用共聚焦显微拉曼光谱仪,证明拉曼光谱有应用于胃癌诊断的临床应用能力;Lu 等^[9]将增强显微拉曼用于前列腺癌的诊断中,实现了前列腺癌的快速诊断;Jermyn 等^[10]将手持拉曼应用于人体脑癌的检测,检测结果的灵敏度为 93%,特异性为 91%。另外,拉曼检测在肺癌、鼻咽癌、乳腺癌等疾病检测方面均有应用^[11]。相对于传统组织病理学检查方式,拉曼检测不需要手术切除组织,无需对组织进行预处理,检测速度快,可以实现无创在体实时检测^[12],为口腔癌的检测提供了新思

路。Sahu 等^[13]通过对低波数拉曼光谱进行分类,分类准确率约为 78%;闫冰等^[14]将血清 SERS 光谱应用于口腔鳞状细胞癌患者临床分期的研究,通过分析低波数拉曼光谱,对光谱数据进行分类,总体准确率达到 80%以上,但该方法需要提前制备血清样本,时效性较差;Hole 等^[15]使用共聚焦显微拉曼光谱对犬类口腔癌组织和正常组织进行检测,准确率约 96%,共聚焦显微镜体积较大,需要提前制备组织切片,不能充分满足术中快速获取癌组织边界的需求。

由于拉曼光谱指纹区域($800\sim 1\,800\text{ cm}^{-1}$)包含了大量组织生物化学信息,如核酸、蛋白质、脂类、葡萄糖和其他碳水化合物的许多振动频率属于该区域^[16],因此目前拉曼光谱在生化领域的应用主要集中在指纹区域,但是指纹区域存在强烈的荧光背景,同时,二氧化硅材质的光纤作为激光传输介质,其拉曼信号位于低波数区域,荧光背景和光纤激发的拉曼信号会对微弱的组织拉曼信号造成干扰,增加后续数据处理和信号分析的复杂程度。高波数拉曼光谱与指纹区域拉曼光谱相比,同样包含组织的生化特征,高波数拉曼光谱具有更强的拉曼信号,减少了荧光背景和光纤拉曼信号的影响,减少了后续信号处理的复杂程度,缩短了数据分析时间。核酸在生物代谢中具有重要作用,与细胞分裂和细胞生长密切相关,细胞生长旺盛时核酸含量增加,口腔癌组织生长快,分裂时间短,因此口腔癌组织中核酸含量更丰富。口腔癌组织中的蛋白质、核酸和氨基酸含量增加,导致光谱中-OH、-CH 和-CH₂ 等对应的拉曼峰强度高于正常组织^[17]。主成分分析是一种常见的用于光谱信息分析的方法,通过分析光谱变量之间相关性,将原有变量进行重构,保留较少的变量,实现原有数据的降维,减少数据处理的运算量,提高光谱分析速度。线性判别分析通过计算组间差异和组内差异,对降维后的口腔组织光谱数据进行分类,区分正常组织和口腔癌组织。

本文针对目前拉曼光谱口腔癌检测中存在的问题,提出了一种基于高波数光纤拉曼光谱的口腔癌检测方法,使用便携拉曼系统采集癌组织光谱数据并进行处理分析。在实验中,采集了 15 位患者的正常组织和癌组织拉曼光谱共 59 组,对测得的光谱进行了归一化处理,然后利用主成分-线性判别分析(principal components analysis-linear discriminant analysis, PC-LDA)算法进行统计分析。

1 原理与算法

1.1 光纤拉曼检测原理

拉曼光谱是一种散射光谱,当入射光照射到样品上时,绝大部分可以透过,大约有 0.1% 的入射光与样品分子之间发生碰撞时非弹性碰撞,产生能量交换。如图 1 所示,拉曼散射包括斯托克斯散射和反斯托克斯散射,由于斯托克斯散射的强度高于反斯托克斯散射,在本文采用斯托克斯散射光谱^[18]。

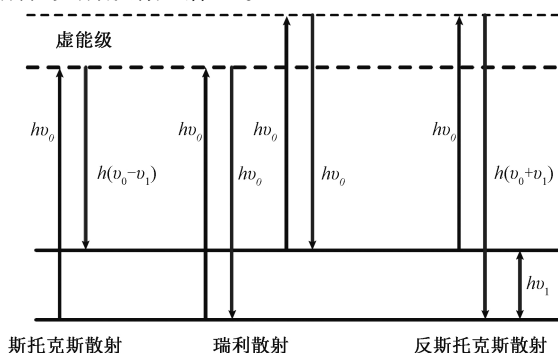


图 1 拉曼散射能级

Fig.1 Raman scattering energy level

在口腔癌检测方面,即使口腔癌病变组织中的细胞尚未发生显微镜下可见的形态学改变之前,由于细胞增殖、分化或恶变以及一些活性因子的分泌等都会引起组织中 DNA、RNA、蛋白质和脂类的结构成分和含量的改变,这些物质对应的分子成分和含量的改变,导致组织拉曼光谱波形出现差异,实现对正常组织和口腔癌组织的区分,拉曼光谱检测具有传统病理学诊断不具备的优势^[19]。

口腔癌发病与核酸和蛋白质含量变化有十分紧密的关系,其拉曼峰主要来源于 C-C、C-N、C=C、C=N 和 -NH₂ 等分子的运动^[18]。根据不同的碱基堆积、碱基配对和配体结合情况,这些拉曼峰会表现出很大的强度变化,由于基团取向或对象的不同,峰频也会发生偏移。蛋白质的大多数拉曼峰归属于氨基酸的侧链振动,由于侧链的不同性质,可以用于研究蛋白质成分的变化。与指纹区拉曼光谱相比,高波数区拉曼光谱表征组织中成分种类和浓度的变化,同时减少背景荧光和光纤本身拉曼信号的干扰,在本文中采用高波数拉曼光谱区分正常组织和口腔癌组织。

1.2 PCA-LDA 算法

为评估实验获取的拉曼光谱对正常组织和口腔癌组织的特异性区分能力,对整个数据集(59 个平均光谱)进行线性判别分析以鉴定不同种类光谱之间的最大区分方

向。由于获取的光谱数量与数据维度相比较小,线性判别分析具有过度拟合的趋势,因此,使用主成分分析提取光谱数据主成分后,再进行线性判别分析。

主成分分析(principal components analysis, PCA)将原有变量在新的数据空间重新组合成一组新的相互无关的几个综合变量,在不影响原有变量信息的情况下根据实际需提取较少的综合变量,选取数据集内对方差影响最大的特征,通过保留低阶主成分,忽略高阶成分实现数据简化,实现数据的降维处理。原有高维数据降维后,可以减少随机因素的干扰,避免维数过高导致的数据过拟合,降低数据分析的复杂度。

原始指标数据的标准化采集 p 维随机向量 $\mathbf{x} = (x_1, x_2, \dots, x_p)^T$, n 个样品 $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$, $i = 1, 2, \dots, n$, 构造样本矩阵进行标准化变换。

$$Z_{ij} = \frac{x_{ij} - \frac{\sum_{i=1}^n x_{ij}}{n}}{\sqrt{\frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1}}}, i = 1, 2, \dots, p \quad (1)$$

Z 标准化矩阵求相关系数矩阵。

$$\mathbf{R} = \left[\frac{\sum_{k=1}^n z_{kj} \cdot z_{kj}}{n-1} \right]_{p \times p} = \frac{\mathbf{Z}^T \mathbf{Z}}{n-1} \quad (2)$$

解样本相关矩阵 $\mathbf{R}$ 的特征方程 $|\mathbf{R} - \lambda \mathbf{I}_p| = 0$ 得 p 个特征根,确定主成分。按照每个主成分的方差贡献率确定主成分的权数,对主成分进行加权求和,得到最终评价。

线性判别分析(linear discriminant analysis, LDA)是一种用来实现两个或者多个对象特征分类方法,基于特征值与特性向量实现降维处理,将数据投影到不同分类,基于最大类间方差与最小类内方差,减少分类内部之间差异,扩大不同分类之间差异,将经过主成分分析降维处理的数据进行线性判别分析,实现对 PCA 重构光谱数据的二分类。

2 实验系统与试样

2.1 光纤拉曼检测系统

本文中设计的系统光路如图 2 所示,785 nm 激光器发出的光束通过光纤直接照射组织样本,后向拉曼散射信号被收集光纤收集后,通过两片长通滤光片耦合进光谱仪中进行探测,被电荷耦合元件(charge-coupled device, CCD)接收。

该系统中激光器(FC-D-785,中国长春新产业光电技术有限公司)为具有高效率和小型化特点的半导体激光器,发出窄带激光(波长 785.062 nm,最大输出功率位

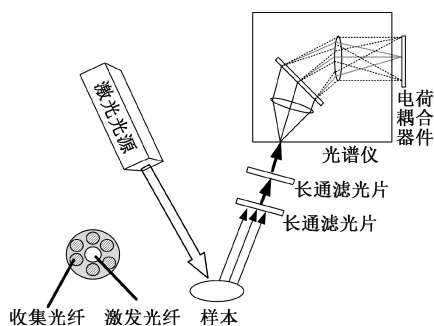


图 2 仪器光路系统

Fig.2 Instrument optical path system

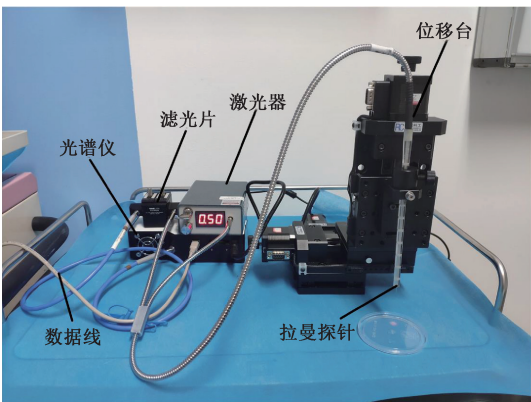


图 3 实验系统实物

Fig.3 Experiment system photo

300 mW,光谱线宽 $<0.1\text{ nm}$),该波长的激光具有低吸收、深穿透、较低荧光和合理的量子效率的特点,被广泛应用于生物和化学领域;激光出射后耦合进激发光纤(纤芯直径 $105\text{ }\mu\text{m}$,包层直径 $125\text{ }\mu\text{m}$),照射组织样本。拉曼信号被环绕激发光纤环形布置的6根收集光纤束(纤芯直径 $50\text{ }\mu\text{m}$,包层直径 $125\text{ }\mu\text{m}$)收集,经两块长通滤波器(FOFMS/M型号,THORLABS公司)滤除干扰光信号后,进入光谱仪(I0785MM0350MF型号,Ocean公司)进行分光,光谱仪将不同波长的光分散开并将它们成像于像平面的不同位置上,在像平面上装有CCD可以同时测得不同波长的光束的强度,光谱仪波数测量范围为 $0\sim 4\,000\text{ cm}^{-1}$,分辨率 6 cm^{-1} ,散射光由CCD探测器接收,将光信号转换成电信号传输到上位机,在上位机进行光谱显示及处理。

实验中采集的组织样本均为手术切取的癌组织和癌旁正常组织,尽量保证每块组织不小于 $5\text{ mm}\times 5\text{ mm}\times 5\text{ mm}$,但由于患者自身条件和癌症发生位置不同,组织大小不完全相同,根据实际条件确定测量点数量。样本组织运输过程中使用内部温度低于 -80°C 的液氮罐冷冻转运,可以保证组织成分在较长时间内不发生改变。光谱测量前在室温环境下使组织自然解冻,完成后进行光谱测量。

使用如图3所示光纤拉曼设备采集组织高波数光谱,光谱测量范围 $2\,400\sim 4\,000\text{ cm}^{-1}$,激发光功率为 100 mW ,激光光斑直径为 1 mm 。组织上的每个位点累计测量30次,每次 20 s ,以保证获取到拉曼信号,通过取光谱平均值,降低环境噪声对结果的影响,提高信噪比。对于标准大小样本,每块组织在不同位点测量3次,然后取平均光谱,所有谱线都在相同条件下测量。

2.2 实验组织样本

本实验中的组织样本来自北京协和医院,共采集15位患者癌组织和正常组织标本,其中男性10例,女性5例,年龄 $52\sim 79$ 岁,最大者79岁,最小者52岁,中位年龄65岁,如表1所示。

表 1 患者临床信息

Table 1 The clinical information of the patient

性别	癌组织提供者	正常组织提供者
男	7	8
女	2	3

石蜡切片苏木素-伊红染色是病理诊断的“金标准”,切片之前需要对组织进行固定、脱水和包埋等处理,然后使用切片机切取厚度为 $50\text{ }\mu\text{m}$ 的病变组织,使用苏木素-伊红染色,在显微镜下观察样本的组织结构和细胞形态,明确样本是否癌变及癌变程度。

口腔癌变组织切片如图4所示,与正常组织相比,口腔癌组织的切片没有明显的分层,增殖的上皮侵入结缔组织内,形成许多互相连接的癌巢。

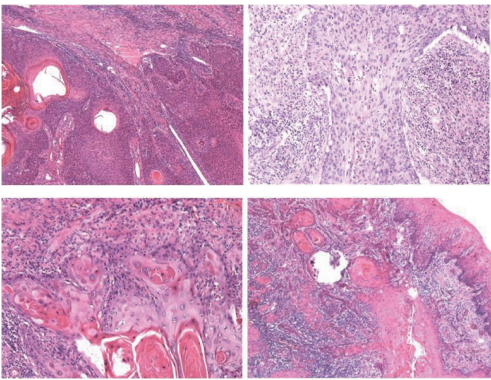


图 4 患者癌组织病理切片

Fig.4 The cancer tissue pathological section of the patient

3 实验分析与讨论

3.1 光纤拉曼检测数据预处理与光谱分析

为了去除拉曼谱线背景噪声,采用五阶多项式拟合

背景曲线后从原始谱线中扣除,同时去除宇宙射线背景和 高频噪声,为了便于比较不同样品光谱形状和谱峰的 相对强度,对曲线下的光谱面积进行归一化处理,通过归 一化处理,消除由于探针高度、组织表面角度和患者个体 差异造成的光谱光强差。光谱的平滑采用 Savitzky-Golay 算法,基线修正采用 Vancouver 算法,在时域内基于局域 多项式最小二乘法拟合的滤波方法,在滤除噪声的同时 可以确保信号的形状、宽度不变。

口腔癌组织病理表现为蛋白质、脂质、核酸、氨基酸、 碳水化合物等分子含量变化,其中蛋白质、核酸和氨基酸 含量尤为突出,也有研究表明,DNA、蛋白质和脂质含量 随恶性肿瘤发生而增加。

正常组织和癌组织的平均谱图如图 5 所示。在健康 皮肤与肿瘤皮肤的谱中,拉曼峰的主要区别在于波数为 2 780、2 890、2 936、3 180、3 285、3 300、3 650 cm^{-1} ,详细 的谱峰归属如表 2 所示。口腔癌组织和正常组织平均拉 曼光谱在 2 780、2 890 和 2 936 cm^{-1} 处肿瘤组织拉曼光谱 强度高于健康皮肤,在 3 180、3 300 和 3 650 cm^{-1} 弱于健 康组织。高波数区域(2 800~3 600 cm^{-1})基于 N-H、O-H 和 C-H 等化学键振动。测量的癌组织与正常组织拉曼 光谱的峰值在 2 931 cm^{-1} 处差异十分明显,表明蛋白质及 脂类成分及浓度有较大差异,其余波数区谱峰则代表氨 基酸、蛋核酸、胶原等发生变化。

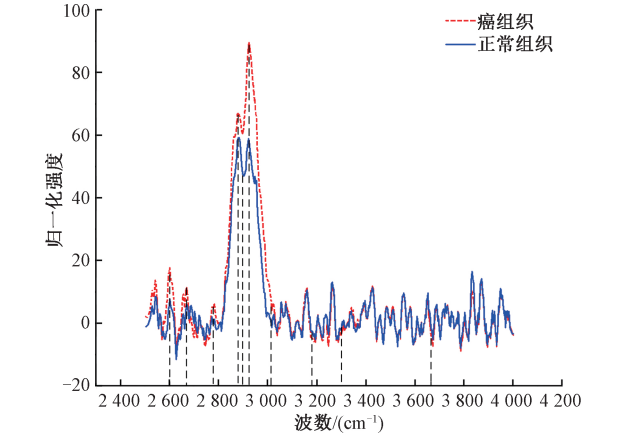


图 5 口腔癌组织和正常组织平均拉曼光谱
Fig.5 The mean Raman spectra of the oral cancer tissues and normal tissues

3.2 基于 PCA-LDA 算法的统计分析

为了验证高波数拉曼光谱区分口腔癌组织和正常组 织的能力,对光谱进行基于 PCA 和 LDA 组合的联合统计 学分析。

利用 PCA 算法对光谱数据进行降维处理,选择最具 有显著性差异的主成分,衡量主成分显著性的指标包括 特征根和主成分的方差贡献率,特征根代表各阶主成分

表 2 拉曼谱峰的振动模式与谱峰归属
Table 2 The vibration mode and peak assignment of the Raman spectrum peaks

峰值(cm^{-1})	峰值归属
2 780	C-H 伸缩振动
2 890	CH_2 变形
2 936	CH_2 的不对称拉伸
3 180	OH 约束
3 285	NH_2 的不对称拉伸
3 300	C-H 伸缩振动
3 650	O-H 伸缩振动

对原始数据影响力度的大小,引入的主成分经过原有数 据的重构,具有相似信息的原始变量合并,通常主成分的 特征根应大于 1,表示重构的主成分对结果的影响高于 原有变量。表 3 所示为高波数光纤拉曼光谱经过 PCA 分析的特征值,前 11 阶主成分特征值均大于 1,将对应 的主成分重构光谱数据。前 11 位主成分的累计方差为 95.172%,表明前 11 位主成分包含了原始数据的大部分 信息,使用前 11 维主成分重构光谱对原有光谱进行降维 处理。

表 3 数据初始特征值
Table 3 Data initial eigenvalue

成分	合计	方差百分比/%	总计/%
1	286.614	55.545	55.545
2	89.508	17.347	72.892
3	50.781	9.841	82.733
4	25.273	4.898	87.631
5	12.368	2.397	90.028
6	8.275	1.604	91.632
7	5.365	1.040	92.671
8	3.970	0.769	93.441
9	3.500	0.678	94.119
10	2.842	0.551	94.670
11	2.589	0.502	95.172

各阶主成分的方差贡献率表明主成分的方差在总样 本方差中的比重,PC1~PC11 的累计方差贡献率超过 95%,表明全 11 阶主成分提取了主成分中 95%的主要信 息,在保证数据不失真的前提下减少了变量数量,便于后 续的 LDA 分析。如图 6 所示,PC1、PC2 和 PC3 的方差贡 献率较高,对原始数据的解释力度较大。结合波数分析, PC1 主要解释波数 3 100~4 000 cm^{-1} 的光谱数据,PC2 解 释的光谱在 2 550~2 900 cm^{-1} 范围内,PC3 对 2 200~

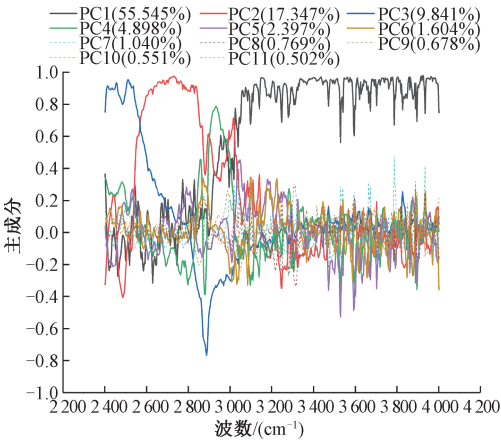


图 6 前 11 阶主成分方差累计

Fig.6 The first 11 order principal component variance accumulation

2 550 cm^{-1} 波数的拉曼光谱解释力度较大。PC4~PC11的主成分方差贡献率较小,但可能包含一些有助于光谱识别的信息,如果丢弃可能导致判别信息的丢失。

碎石图用于降序显示与因子关联的特征值及分量或因子的数量,用于主成分分析中,可以直观地评估分量在主成分分析后重构数据变异中的作用,如图 7 所示,主成分在第 6 点区域平缓。对 PCA 的分类准确性进行分析,提取的主成分由 1 增加到 6 时,解释的分类准确度得到了极大改善,继续增加主成分数量,将主成分数量增加到 11 个,并不会对分类准确度有影响。结合碎石图和 PCA 分类准确性分析,选择 PC1~PC6 用于 LDA 分析。

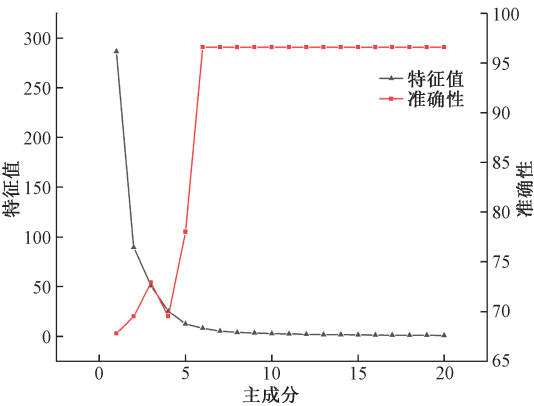


图 7 主成分分析碎石图和分类准确度折线

Fig.7 Principal component analysis scree plot and classification accuracy line chart

表 4 所示为 PCA-LDA 分类的结果,将拉曼光谱分析结果与病理组织分析结果对比,正常组织拉曼光谱全部 28 组光谱被正确分类,识别准确率为 100%。对于口腔

癌组织 2 组光谱被错误归类为正常组织,其余 29 组口腔癌拉曼光谱被正确识别,口腔癌拉曼光谱正常组织准确率 93.3%。PCA-LDA 算法总准确率为 96.6%。

表 4 PCA-LDA 分类结果

Table 4 The classification result of PCA-LDA		
类别	实际正常组织	实际口腔癌组织
正常组织拉曼光谱分类结果	28	0
口腔癌组织拉曼光谱分类结果	2	29
准确性/%	93.3	100

典则判别结果显示线性判别分析结果,通过 LDA 判别函数可以计算光谱的判别分数,判别分数代表算法判别的程度。图 8 所示为算法分组结果,0 线为判别标准,红色为口腔癌组织,黑色代表正常组织,对于口腔癌组织和正常组织的拉曼光谱以 0 线为界发生了比较明显的分离。所有的正常组织都位于 0 线下方,表示所有正常组织都被正确的识别,31 组口腔癌组织光谱有 2 组未被正确分类到 0 线上方。在算法结果图中,有许多样本得分接近 0 线,可能是因为一些患者的某些组织成分发生了改变,导致光谱判别分数区别较小。

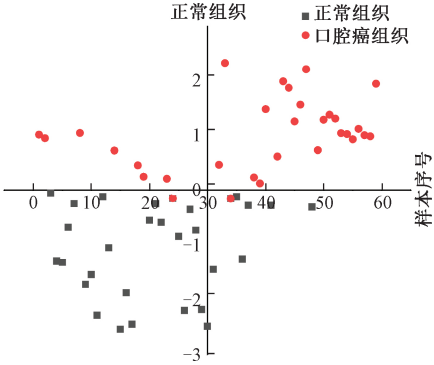


图 8 PCA-LDA 算法分组结果

Fig.8 The grouping result of PCA-LDA algorithm

接受者操作特性曲线(receiver operating characteristic curve,ROC),反映算法敏感性和特异性,PCA-LDA 分析后,使用留一法对 59 组数据进行验证,得到算法的 ROC 曲线,如图 9 所示曲线下面积为 0.878,置信区间大于 85%。

3.3 方法可行性分析与讨论

使用光纤激光拉曼检测肿瘤边界,基于拉曼效应在分子水平上检测人体正常组织和癌变组织中物质成分和物质含量的变化。拉曼光谱高波数区域包含脂质和蛋白质的成分和含量信息,同时癌变组织与正常组织中部分蛋白质和脂质成分和含量有明显区别,因此根据正常组织和癌变组织高波数拉曼光谱差别可以区分正常组织和癌组织。

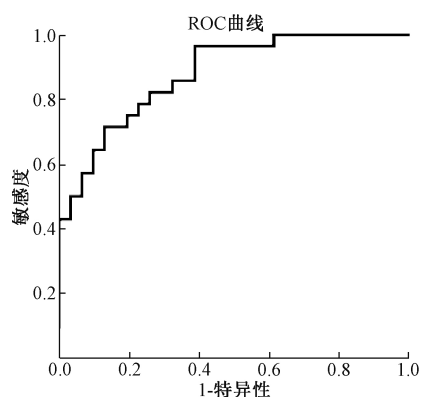


图9 PCA-LDA 算法的 ROC 曲线

Fig.9 The ROC curve of PCA-LDA algorithm

采用人体安全的 785 nm 激光光源,具有较低的荧光效应,选用合理的滤波方式和试验方案,提高了拉曼信号的信噪比,设计了合理的拉曼探头结构,将收集光纤环绕激发光纤布置,控制拉曼光散射范围,提高了组织拉曼光的收集效率,在探头后端连接集成化的滤波装置,减少了反射激光和环境光的干扰,采用 PCA-LDA 统计学分析方法,进一步提取光谱中的有效信息,同时对光谱信号降维处理,提高了数据处理的速度。

4 结 论

为减少口腔癌检测中对患者造成的痛苦,提高口腔癌术中检测的时效性和准确率,本文提出了一种基于高波数光纤激光拉曼的口腔癌在体检测方法,通过便携光纤拉曼系统获取了患者正常组织和口腔癌组织的高波数拉曼光谱,结合口腔癌组织物质成分特点对高波数区光谱数据进行光谱分析,并利用 PCA-LDA 算法对光谱进行数据重构和分类,分类结果与病理切片检查结果对比表明,使用高波数光纤激光拉曼光谱检测口腔癌的准确率为 96.6%,相比低波数光纤拉曼检测方法,检测准确率提高约 0.6%,另外,采用该方法不需要制备血清样本,可以实现癌组织的在体检测,于口腔癌组织在体诊断方面有很好的应用前景。

本文提出的基于高波数光纤拉曼检测方法为今后实现口腔癌组织术中实时检测奠定了基础,未来将致力于在此方法的基础上进一步提高诊断准确率,提高系统探测深度等方面的研究。

参考文献

[1] 范贤光,王海涛,王昕,等. 基于非均匀 B 样条的拉曼光谱基线校正算法[J]. 光谱学与光谱分析, 2016, 36(3): 724-728.
FAN X G, WANG H T, WANG X, et al. Baseline

correction algorithm for Raman spectroscopy based on non-uniform B-spline[J]. Spectroscopy and Spectral Analysis, 2016, 36(3): 724-728.

- [2] RIVERA C. Essentials of oral cancer[J]. International Journal of Clinical and Experimental Pathology, 2015, 8(9): 11884.
- [3] 于晓红. 术中快速诊断口腔癌淋巴结转移方法的建立及优化[D]. 济南: 山东大学, 2008.
YU X H. Establishment and optimization of rapid diagnostic methods for cervical lymphatic metastasis during oral carcinoma operation [D]. Jinan: Shandong University, 2008.
- [4] 马海涛, 丁东, 崔继承, 等. 近红外光谱在无创血糖监测中的应用[J]. 仪器仪表学报, 2002, 24(S3): 231-233.
MA H T, DING D, CUI J CH, et al. Near infrared spectroscopy for noninvasive glucose determination [J]. Chinese Journal of Scientific Instrument, 2002, 24(S3): 231-233.
- [5] 杨宁, 张荣标, 赵雨琦, 等. 基于荧光显微拉曼光谱的细菌快速解耦检测检测方法[J]. 仪器仪表学报, 2012, 33(2): 435-441.
YANG N, ZHANG R B, ZHAO Y Q, et al. Bacteria fast decoupling detection based on raman spectra and fluorescence microscopy[J]. Chinese Journal of Scientific Instrument, 2012, 33(2): 435-441.
- [6] 夏嘉斌, 祝连庆, 姚齐峰, 等. 基于 Kolmogorov-Smirnov 检验的远程拉曼光谱寻峰算法[J]. 仪器仪表学报, 2018, 39(3): 141-147.
XIA J B, ZHU L Q, YAO Q F, et al. Remote raman spectral peak searching algorithm based on kolmogorov-smirnov test [J]. Chinese Journal of Scientific Instrument, 2018, 39(3): 141-147.
- [7] 林学亮, 林多, 邱素芳, 等. 鼻咽癌患者唾液的表面增强拉曼光谱研究[J]. 光谱学与光谱分析, 2018, 38(8): 2430-2434.
LIN X L, LIN D, QIU S F, et al. Detection of nasopharyngeal carcinoma based on human saliva surface-enhanced raman spectroscopy[J]. Spectroscopy and Spectral Analysis, 2018, 38(8): 2430-2434.
- [8] 刘燕玲. 表面增强拉曼光谱在胃肠道肿瘤血清检测中的应用研究[D]. 广州: 南方医科大学, 2018.
LIU Y L. Research on surface enhanced raman spectroscopy in serum detection of gastrointestinal cancer[D]. Guangzhou: Southern Medical University, 2018.
- [9] LU W, SINGH A K, KHAN S A, et al. Gold nanopopcorn-based targeted diagnosis, nanotherapy treatment, and in situ monitoring of photothermal therapy response of

- prostate cancer cells using surface-enhanced Raman spectroscopy[J]. Journal of the American Chemical Society, 2010, 132(51): 18103-18114.
- [10] JERMYN M, MOK K, MERCIER J, et al. Intraoperative brain cancer detection with Raman spectroscopy in humans [J]. Science Translational Medicine, 2015, 7(274): 274ra19-ra19.
- [11] MOTZ J, FITZMAURICE M, MILLER A, et al. In vivo Raman spectral pathology of human atherosclerosis and vulnerable plaque [J]. Journal of Biomedical Optics, 2006, 11(2): 021003.
- [12] GANDRA N, HENDARGO H C, NORTON S J, et al. Tunable and amplified Raman gold nanoprobe for effective tracking (TARGET): In vivo sensing and imaging[J]. Nanoscale, 2016, 8(16): 8486-8494.
- [13] SAHU A, NANDAKUMAR N, SAWANT S, et al. Recurrence prediction in oral cancers: a serum Raman spectroscopy study [J]. Analyst, 2015, 140 (7): 2294-2301.
- [14] 闫冰, 骆献阳, 谭迎赞, 等. 血清表面增强拉曼光谱应用于口腔鳞状细胞癌患者临床分期的研究 [J]. 国际口腔医学杂志, 2019, 46(3): 277-281.
- YAN B, LUO X Y, TAN Y Y, et al. Research on the surface-enhanced Raman spectrum of blood serum for clinical staging of oral squamous cell carcinoma [J]. International Journal of Stomatology, 2019, 46 (3): 277-281.
- [15] HOLE A, BHUJBAL M, BENDALE K, et al. Preliminary study of canine oral cancer by Raman spectroscopy [C]. Proceedings of the Optics and Biophotonics in Low-Resource Settings V, F, International Society for Optics and Photonics, 2019, 10869: 1086917.
- [16] MOVASAGHI Z, REHMAN S, REHMAN I U. Raman spectroscopy of biological tissues [J]. Applied Spectroscopy Reviews, 2007, 42(5): 493-541.
- [17] BERGHOLT M S, LIN K, WANG J, et al. Simultaneous

fingerprint and high-wavenumber fiber-optic Raman spectroscopy enhances real-time in vivo diagnosis of adenomatous polyps during colonoscopy [J]. Journal of Biophotonics, 2016, 9(4): 333-342.

- [18] 杨序纲, 吴琪琳. 拉曼光谱的分析与应用[M]. 北京: 国防工业出版社, 2008.
- YANG X G, WU Q L. Analysis and Application of Raman Spectroscopy [M]. Beijing: National Defense Industry Press, 2008.
- [19] WILSON B C, MARPLE E, SHIM M G, et al. Study of fiber-optic probes for in vivo medical raman spectroscopy [J]. Applied Spectroscopy, 2016, 53 (6): 619-627.

作者简介



胡阳, 2017 年于北京信息科技大学获得学士学位, 现为北京信息科技大学硕士研究生, 主要从事拉曼物质成分检测方面的研究。

Email: hy9411@yeah.net

Hu Yang received his B.Sc. degree in 2017 from Beijing Information Science and Technology University. Now, he is a M. Sc. candidate in Beijing Information Science and Technology University. His main research interests include Raman material composition detection.



董明利 (通信作者), 1989 年于合肥工业大学获得硕士学位, 2009 年于北京理工大学获得博士学位, 现为北京信息科技大学教授, 主要研究方向为生物医学检测技术与仪器、视觉测量技术。

E-mail: dongml@ sina.com

Dong Mingli (Corresponding author) received her M. Sc. degree in 1989 from Hefei University of Technology, and Ph. D. degree in 2009 from Beijing Institute of Technology. Now, she is a professor in Beijing Information Science and Technology University. Her main research interests include biomedical detection technology and instrument, vision measurement technology.