

近红外光谱技术应用于癌症检测的研究进展

朗韦玮¹ 李嘉晨^{1,2} 董林毅¹ 韩丽^{*1}

¹(天津医科大学药学院, 天津 300070) ²(天津市第五中心医院生态城医院, 天津 300467)

摘 要 近红外(Near-infrared, NIR)光谱主要反映了含氢基团的振动信息,可用于获取由疾病引起的生物样品中蛋白质、脂质等成分的含量和结构变化信息。近年来,近红外光谱技术由于其快速、无损和非侵入性等优势,在疾病诊断领域发挥了重要作用。在癌症检测方面,恶性肿瘤细胞与正常细胞的化学结构和成分存在明显差异,并且癌症引起的异常代谢功能会导致生物组织和体液成分发生变化,从而影响光谱吸收,这使得近红外光谱有望成为探测生物体内癌变信息的有效工具。然而,近红外光谱的谱峰重叠严重,因此通常需要借助化学计量学方法对光谱进行解析,以有效提取癌症诊断信息。本文总结了近年来近红外光谱技术应用于癌症检测的研究进展,包括使用光谱预处理方法用于获取高分辨率光谱信号和建立模型方法用于癌症诊断,以及近红外高光谱成像结合深度学习方法的医学诊疗应用趋势。通过对光谱进行解析,可以探测生物体系中分子的定量信息和结构变化,特别是利用近红外光谱对水的敏感性,分析了水在癌变过程中的含量和结构变化及其在癌症检测中的作用机理。

关键词 近红外光谱; 癌症诊断; 化学计量学; 定量分析; 结构分析; 评述

近红外(Near-infrared, NIR)光谱主要反映含氢基团(X—H)的振动信息,可获得由疾病引起的生物组织或者体液中蛋白质和脂质等的组成和结构的变化信息,因此,通过光谱变化可以获取潜在的诊断信息^[1-3]。相比于其它光谱技术,如红外光谱、拉曼光谱和激光诱导击穿光谱,近红外光具有更强的组织穿透能力,并且对样品无损,因此在无创检测中表现出更大的应用潜力。此外,近红外光谱对水的吸收强度适中,更适合检测含水量高的生物样本^[4],能够反映癌变样本中溶剂水分子和生物分子水合作用的变化。凭借其组织穿透性强、非侵入性、操作简便和检测快速等优点,近红外光谱技术早在 1977 年就被证明在疾病诊疗方面具有应用潜力^[5-6]。目前,近红外光谱已被用于研究一系列不同的疾病,如通过检测近红外光在人体内的漫反射光谱,建立多变量模型进行分析,预测人体内的血糖浓度,实现糖尿病的无创监测^[7]。通过检测血清的近红外光谱,建立光谱与血清中含氮量之间的定量关系模型,准确测定尿素浓度,从而有助于诊断肾功能疾病^[8]。

恶性肿瘤细胞与正常细胞的化学结构和成分含量都存在明显差异,异常的代谢功能会使生物组织和体液成分发生变化。近红外光谱技术可识别癌细胞生长和扩散时释放的化学物质,通过探测生物样品中水分子、蛋白质和脂质等物质的含量和结构的变化,利用多波长数据与肿瘤生长之间的关系建立多元校正模型,对癌症类型和分期进行判别^[9-10]。然而,生物样品的近红外光谱主要由水的宽吸收峰和生物分子的弱响应信号组成,正常样品和癌症样品的光谱形状相似,难以从原始光谱中直接获取水和生物分子的变化信息。因此,需要借助于化学计量学方法提高光谱分辨率或对光谱进行解析^[11-13],区分不同物质的光谱特征,有效提取癌症引起的数量、结构以及分子间相互作用变化信息,从而有效获取癌症的诊断信息。

本文总结了近年来近红外光谱技术应用于癌症检测的研究进展,包括化学计量学方法及其在定量和结构分析中的应用。首先,通过对光谱进行预处理,获取高分辨光谱信号,进而获取癌症相关信息,再利用合适的建模方法对癌症进行判别,并介绍了深度学习方法在近红外高光谱成像中的应用。然后,从光

2024-05-15 收稿; 2024-07-31 接受

天津市教委科研计划项目(No. 2022ZD053)资助。

* E-mail: lihan19@tmu.edu.cn

谱中解析出癌症引起的生物分子含量和结构变化信息,以及从水的光谱变化中反映出生物样品的状态信息和相互作用,从而探测癌症的诊断信息。最后,对该研究领域未来的发展趋势进行了展望。

1 诊断信息的提取与识别方法

1.1 光谱信号分辨率的提高

相比于红外或拉曼等分子光谱,近红外光谱的谱峰通常存在较严重的重叠现象。因此,在信号提取过程中,须选取合适的机器学习方法解析重叠谱峰。生物体系中的成分复杂多样,并且生物组织或者体液样品含大量水,其近红外光谱主要由水的宽吸收峰组成,掩盖了其它生物分子的变化信息,导致难以从原始光谱中直接获取正常样品和癌变样品的差异信号。因此,需要借助化学计量学方法对光谱进行预处理^[14],提高光谱的分辨率,放大病变样品的光谱特征,从而有效提取癌症诊断信息,建立癌症的判别模型(图1)。常用的增强光谱分辨率的方法包括求导、小波变换(Wavelet transform, WT)、多元曲线分辨和正交信号校正(Orthogonal signal correction, OSC)等。在众多预处理方法中,研究者需要综合考虑多种因素,如数据的噪声水平、变量的冗余性以及模型的复杂性等,以选择最合适的分析工具和算法用于提取有效的光谱信号。

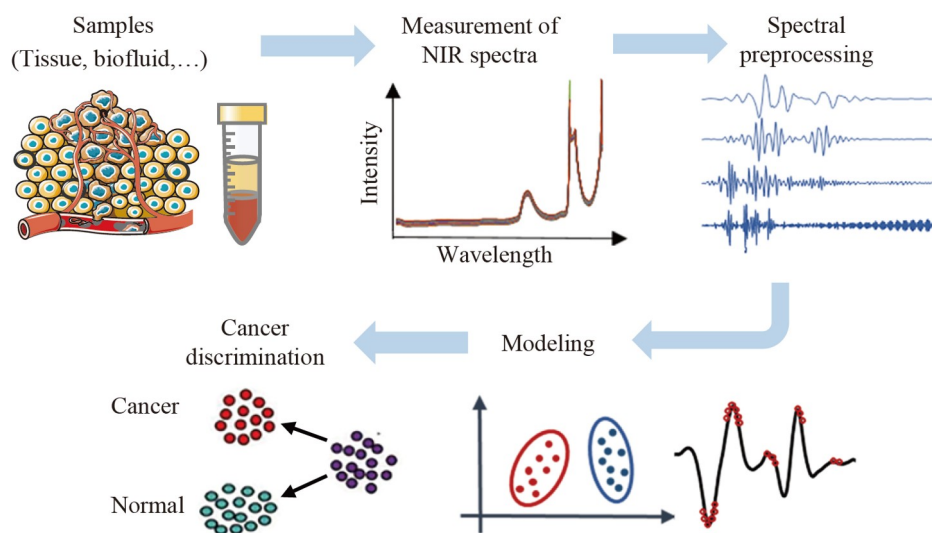


图1 利用生物样品的近红外(NIR)光谱结合化学计量学方法对癌症进行诊断

Fig.1 Cancer diagnosis by near-infrared (NIR) spectroscopy combined with chemometric methods

求导是一种常见的提高光谱分辨率的方法,同时还可用于校正光谱基线,有效消除背景干扰。Savitzky-Golay(S-G)求导^[15]是处理近红外光谱数据的常用方法,利用移动窗口多项式对光谱进行拟合,提高信噪比。Ehlen等^[16]在利用近红外光谱检测结直肠癌的研究中,使用S-G方法对光谱求二阶导数,放大了可区分结直肠癌和正常组织的主要光谱区域,在水的吸收波段表现出了明显的差异。Hurskainen等^[17]通过对纯水、岩藻糖溶液和脯氨酸溶液的近红外光谱进行导数处理,获取了岩藻糖和脯氨酸的特征吸收峰,并将这两种化合物作为口腔癌生物标志物,与口腔癌患者唾液的导数光谱进行对比,在1530~1650 nm之间显示出了与两种生物标志物水溶液相似的光谱特征,由此可以对癌症患者和健康人样本进行区分。

WT作为一种近似求导方法^[18]已被广泛用于信号处理,是增强信号分辨率的有效手段^[19-21]。该方法通过信号和滤波器的卷积计算导数,通过选择不同的滤波器可以获得任意阶导数,通过调整尺度参数可以达到信号平滑的效果。Chen等^[22]在利用结肠组织的近红外光谱对结肠癌进行诊断的研究中,采用连续小波变换有效提高了光谱信号分辨率,并从高分辨信号中选取了40个重要的特征波长变量,建立了判别模型,实现了结肠癌患者的准确识别。由此可知,通过WT得到的不同频率下的光谱特征可以呈现原

始光谱中所隐含的与癌症相关的关键信息。最近, Han等^[23]发展了小波包变换技术, 获取了超高分辨光谱信号, 用于分析含水样本中相互作用的变化信息, 在提取具有不同分辨率信号过程中获取了不同分子基团的结构信息。该方法有望用于复杂生物样品体系分析, 提取癌变引起的不同生物分子的光谱变化特征信息, 实现精准的癌症筛查和诊断。

为了剔除光谱中与癌症诊断不相关的信息, 提高有效信号的分辨率, 张卓勇研究组^[24]使用正交信号校正 OSC 对原始光谱中噪声信息进行扣除, 尽可能保留了光谱中的有效信息。与多元散射校正 (Multiplicative scatter correction, MSC) 等其它预处理结果相比, OSC 方法增强了诊断模型的预测能力, 区分了癌变、增生和正常的子宫内膜组织切片。在此基础上, 该研究组发展了加强正交信号校正 (EOSC) 方法^[25], 与传统方法相比, EOSC 可以更好地去除背景基线漂移, 很好地保留了光谱信号峰, 有效避免了在信号校正时出现过拟合现象。Liu等^[26]采用主成分分析 (Principal component analysis, PCA) 方法同样实现了噪声和无关信息扣除, 提高了有效信号的分辨率, 并同时实现了数据降维。在该研究中, 前两个主成分的载荷光谱呈现了样品之间的显著差异, 从中选择 6 个波长作为最佳波长用于建立分类模型, 实现了胃癌和正常胃组织的区分。以上方法在晶型信号分辨解析和滤噪时, 同时提取了与癌症信息相关的有效光谱信号, 提高了光谱分辨率。

1.2 诊断模型的建立

近红外光谱数据预处理后, 采用化学计量学方法建立光谱数据与化学成分之间的数学模型, 建立的模型可以从复杂的光谱数据中提取与化学成分相关的信息。构建合适的模型并对近红外光谱进行分析, 可以有效提高光谱的精准度和准确度, 提取有效数据, 提高癌症诊断的准确性。

主成分聚类分析是常见的无监督模式识别方法, 常用于考察样本之间的分类效果。Yi等^[27]采用傅里叶变换近红外光谱对胃癌组织样品进行筛查, 采用 PCA 提取一组正交主成分, 从前两个主成分的得分图中可以观察到正常样品和癌症样品数据清晰聚类, 并且发现癌症样本之间存在相对较大的差异, 可能是由于癌症的分期不同所致。然后, 根据光谱的相似性或差异性, 利用层次聚类分析 (Hierarchical cluster analysis, HCA) 将样本进行簇类划分, 判别癌症组织, 并获得了 100% 的灵敏度, 表明近红外光谱可作为胃癌早期筛查的高灵敏度工具。Jang等^[28]在近红外光谱检测胆汁的研究中, 为了考察所有样品的近红外光谱特征差异, 对光谱进行了 PCA, 并观察了得分的分布情况, 发现 PC1 即可区分胆囊癌与其它胆囊疾病 (胆结石和胆息肉), 因此可用于基于胆汁液的胆囊疾病快速筛查, 特别是早期胆囊癌的识别。

为了对生物样本的癌变情况进行准确判别, 通常先将 PCA 聚类作为初步方法观察样品之间的分离度, 然后建立诊断模型。Chen等^[29]在利用近红外光谱区分正常组织和恶性结肠直肠组织的研究中, 发现两类样品之间的 PCA 得分分离不清晰, 异常样本和正常样本之间存在重叠, 说明数据结构或关系可能是复杂和非线性的; 随后采用连续投影算法 (Successive projection algorithm, SPA) 筛选出 3 个共线性最小的特征波长, 并建立线性判别模型鉴别出癌变组织。Zufry等^[30]在利用近红外光谱鉴别甲状腺结节恶性肿瘤的研究中, 使用前两个主成分有效区分了甲状腺良性结节和恶性肿瘤, 然后利用线性判别分析 (Linear discriminant analysis, LDA) 对 7 个主成分建立了高准确的分类模型, 但仍需进一步考察模型的灵敏度、特异性等评价指标, 避免模型发生过拟合现象。

为了建立多分类的诊断模型, Zhu等^[31]比较了包括偏最小二乘、随机森林和支持向量机 (Support-vector machine, SVM) 等在内的 5 种机器学习模型, 对肺癌、宫颈癌、鼻咽癌和健康对照血清样本进行区分。结果表明, 所有方法都可对癌症与健康对照组进行显著区分, 但是, 相比于其它分类模型, SVM 可高精度地实现四元分类, 即同时区分 3 种不同类型的癌症以及健康人群, 其原因可能是该方法适于分析高维数据。Yim等^[32]在分析膀胱切片的近红外光谱时, 使用 PCA 方法提取数据的最大方差, 然后结合 K 近邻算法 (K-Nearest neighbor, KNN)、SVM、随机梯度下降、梯度提升和神经网络的结果, 构建了机器学习组合 (或“堆积”) 模型, 有效区分了不同等级和不同分期的膀胱癌组织。

1.3 深度学习方法

传统的化学计量学方法在光谱分析过程中通常步骤繁琐, 需要进行大量的光谱预处理操作。深度学习是近年迅速发展起来的一类机器学习方法, 主要依赖于人工神经网络模拟人脑处理数据的方式, 通过学

习大量数据的内在规律,实现端到端的高精度光谱分析,进而达到识别复杂模式的目的^[33-35]。Shang 等^[36]基于一维卷积神经网络(1D-convolutional neural network, 1D-CNN)建立了判别模型,提出了近红外光谱与深度学习相结合的乳腺癌原位检测和分类方法,能够快速识别不同癌变位置甚至不同阶段的乳腺组织,并且模型的准确度、灵敏度和特异性等均明显优于 KNN 和 Fisher 等传统的判别分析模型算法,为原位非侵入性癌症的自动诊断提供了一种前景良好的新方法。Zhang 等^[37]使用近红外光谱结合反向传播人工神经网络检测 BRAF V600 基因突变型和野生型结直肠癌,通过分析不同光谱波段对模型结果的影响,表明判别模型的建立主要基于缬氨酸和谷氨酸之间的光谱差异,该研究结果将近红外光谱应用拓展到人类癌症基因突变辅助诊断。

深度学习方法在处理大批量数据方面具有优越性,在医学图像分析中,尤其是多维的高光谱成像(Hyperspectral imaging, HSI)数据分析中,具有独特优势^[38-39],如图 2 所示。Muniz 等^[40]在利用傅里叶变换红外 HSI 区分健康、炎症性和肿瘤性结肠组织过程中,设计了全连接神经网络(Fully connected neural network, FCNN)模型,对图像上每个像素的病理状态进行判别,结果的准确度明显高于 SVM 等传统机器学习方法。为了克服数据量大以及深度网络计算成本高的问题,Liu 等^[41]设计了浅层残差网络(Shallow residual network, SR-Net),并用于胃癌等级分类,通过直接下采样、非对称滤波器和全局平均池化对常规的残差网络进行改进,减少了参数和浮点运算,简化了分类过程,同时实现了较高的分类准确率。Ortega 等^[42]利用可见-近红外高光谱成像技术分析乳腺癌组织的病理学切片,使用 2D-CNN 对正常和肿瘤细胞进行自动识别,由于高光谱波长点的增多提供了丰富的化学组成信息,使得判别准确度优于 RGB 图像的识别结果,表明高光谱技术结合深度学习为数字化病理图像自动检测乳腺癌细胞提供了有力的工具。

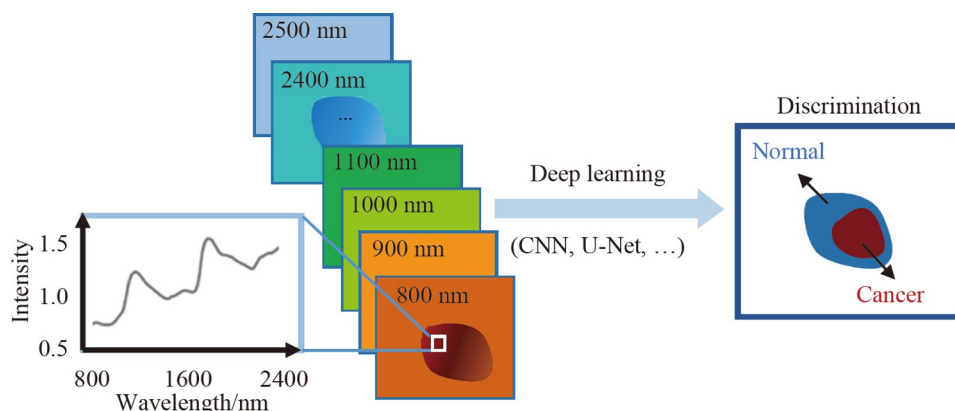


图2 利用深度学习(HIS)方法处理 NIR 高光谱实现对癌症的识别

Fig.2 Recognition of cancer by NIR hyperspectral imaging (HSI) with deep learning

分割是一种对图像进行像素级别分类的有效方法,对于检测肿瘤边界具有重要意义。临床上,如果专业医生人为标记每个像素的类别,将耗费其大量的时间和精力,并且分割质量受医生的专业水平影响。为了提高医学影像的分割质量,缩短分割时间,已有大量研究将深度学习用于医学影像的高精度分割。Ma 等^[43]开发了自适应深度学习方法检测头颈癌,通过高光谱图像训练自动编码器网络,提取深度信息。该算法根据错误分类的像素自适应地调整其权重,提高癌组织和良性组织的识别能力,从而提高检测性能。动物模型实验结果表明,该自适应学习方法在肿瘤边界检测方面具有很高的准确性,特别是对于边缘模糊和不规则的肿瘤。Trajanovski 等^[44]首次使用深度学习语义分割解析用于肿瘤检测的可见-近红外高光谱数据,结果表明,基于 U-Net 神经网络算法对光谱图像进行分割,其预测性能明显优于传统的机器学习和像素级别(Pixel-wise)方法,并且近红外光谱波段对于肿瘤检测至关重要。因此,近红外 HSI 技术结合深度学习方法在癌症等疾病的诊断和病变区域分割方面具有应用潜力,有望成为实时手术的辅助工具。

2 定量分析和结构分析

2.1 生物分子变化

由于近红外光谱主要反映含氢基团的振动信息,因此蛋白质、脂质、核酸和葡萄糖等重要生物分子的含量和结构信息均可以体现在近红外光谱上。癌细胞由于核酸发生突变,表达的生物分子与正常细胞不同,例如,与正常细胞或组织相比,肿瘤通常具有较高的水/脂质之比^[31,45-46]。因此,癌变组织或患者体液的近红外光谱在峰形、峰强度和峰波数位置方面可能存在较大差别。图 3 为细胞密度均为 1000 cell/mL 的正常肺上皮细胞和两种肺癌细胞的培养液近红外光谱。图 3A 的原始光谱主要由水的吸收峰组成,很难区分正常细胞与癌细胞。采用连续小波变换(Symmlet 小波基,消失矩为 4,尺度系数为 30)对光谱进行处理,提高了光谱的分辨率,得到图 3B。三种细胞培养液在 4400 cm⁻¹ 附近的光谱强度存在差异,该波段主要包含了蛋白质酰胺基团的吸收信息^[10,47]。两种癌细胞的光谱强度均低于正常细胞,这是由于癌细胞的强代谢能力使得培养液中蛋白质的消耗较多,含量降低,反映在光谱上为强度降低。因此,通过从近红外光谱中提取生物分子的变化信息,可以反映生物体系的癌变情况。对文献中报道的癌症相关近红外光谱变化信息进行的总结见表 1。

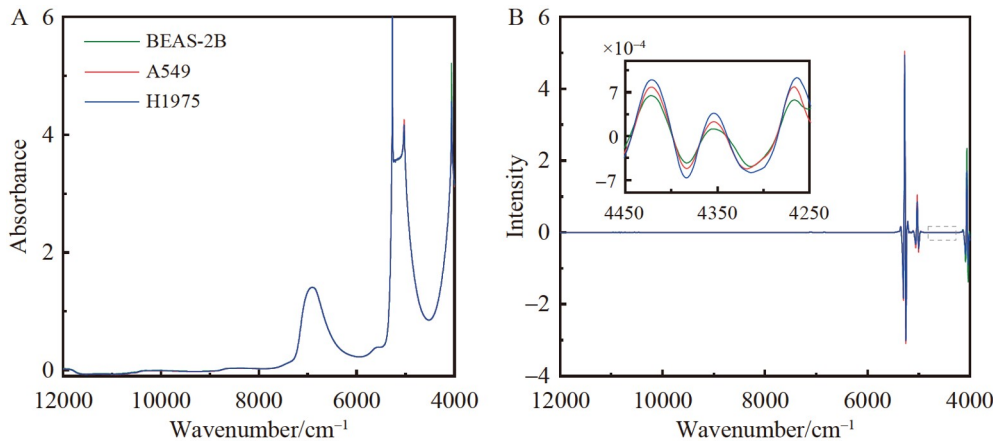


图3 正常肺上皮细胞(BEAS-2B)和不同类型肺癌细胞(A549 和 H1975)培养液的近红外光谱(A)和分辨率增强光谱(B)

Fig.3 NIR spectra (A) and resolution-enhanced spectra (B) of the culture medium of normal cell (BEAS-2B) and two types of lung cancer cells (A549 and H1975)

表1 癌症相关的近红外光谱变化

Table 1 NIR spectral changes related to cancer

癌症类型 Cancer type	生物分子变化 Change of biological molecules	特征波长 Feature wavelength/nm	参考文献 Ref.
乳腺癌 Breast cancer	DNA ↑	1467	[47]
	脂质 Lipid ↓	1212, 1274, 1763	[47]
	自由水 Free water ↑	935	[48]
胆囊癌 Gallbladder cancer	蛋白质 Protein ↑	4050	[28]
前列腺癌 Prostatic cancer	蛋白质糖基化异常 Abnormal N-glycosylation	1062, 1454	[49]
	自由水 Free water ↓	1444, 1944	[2]
皮肤癌 Skin cancer	水 Water ↓	~1450	[50]

注: ↑代表含量增多, ↓代表含量减少。
Note: ↑ represents an increase in content, ↓ represents a decrease in content.

Hirosawa 等^[47]对雌性大鼠的乳腺组织进行了原位近红外光谱检测,经过二阶导数变换得到高分辨光谱,通过比较发现乳腺癌组织中的 DNA 和水的含量高于正常组织,而脂质含量低于后者,进一步,利用脂

质谱带的光谱强度变化定量评估了大鼠体内乳腺组织的恶化程度。Jang 等^[28]在采用胆汁近红外光谱识别不同胆囊疾病的研究中,通过对比胆汁中磷脂和白蛋白等主要成分的纯物质光谱,发现胆囊癌样本的近红外光谱在 4050 cm^{-1} 处的峰强度明显高于其它类别样本,该吸收峰来源于白蛋白的 C—N—C 酰胺振动,表明在胆囊癌患者的胆汁中蛋白质含量较高。Vermassen 等^[49]测量了患有良性和恶性肿瘤患者的前列腺组织的近红外光谱,通过观察导数光谱发现,在 C—H 和 O—H 的振动吸收波段范围内,良性和恶性组织之间的差异尤为明显,这主要与蛋白质的糖基化比例相关。因此,通过从近红外光谱中提取蛋白质、脂质和糖类等化合物的变化信息,将其作为生物标志物,可用于指示生命体的癌变情况。

2.2 水光谱变化

在过去很长一段时间内,水的吸收峰常作为近红外光谱分析生物样品中的干扰因素。然而,水具有复杂的氢键网络结构,水分子之间的氢键作用容易受到分子周围环境变化的影响^[51],所以水在不同体系中表现出来的功能也存在差异。由于近红外光谱对水的变化信息敏感,疾病所引起的体液组成变化通常会影响到水的氢键结构,这导致水的光谱会随疾病的发生而产生变化。基于此原理, Tsenkova 研究组^[4]提出了“水光谱组学”的概念,将水的光谱吸收模式作为一种生物指标,从而理解生命体系中水的作用,并实现对疾病或异常状态的诊断。例如,利用随温度变化的水特征光谱研究朊病毒蛋白纤维化的过程机理^[52],通过水化层的结构信息实现对大豆花叶病潜伏期的诊断^[53]等。Raypah 等^[54]分析了 HeLa 等 3 种细胞的光谱,发现水的不同氢键结构以及与蛋白质、碳水化合物的相互作用对于区分癌细胞和正常细胞具有重要价值。

癌细胞的代谢活动通常比正常细胞活跃^[55],肿瘤微环境具有 pH 值低、氧化应激强以及基质金属蛋白酶活性高等特征^[56],这些情况都会改变水分子的含量、结构和功能,进而影响水的近红外光谱^[57-58]。Ali 等^[2]采用近红外光谱研究了前列腺癌组织和正常前列腺组织中水含量的差异,结果表明,癌组织光谱中水的吸收峰降低,说明水含量比正常组织中的较少。同时,水的吸收峰向较长波长的位置移动,表明前列腺癌组织中的水分子氢键结构更加有序,这可能是水分子与细胞或组织中的生物大分子相互作用所致。类似地,在皮肤癌检测的研究^[50]中,研究者发现黑色素瘤皮肤中的水分含量低于健康皮肤。然而,在 Chung 等^[49]关于近红外光谱检测乳腺癌的研究中发现,乳腺癌组织中水分子含量更多,并且自由水的比例高于健康组织,由此证明了水分子具有作为一种非侵入性体内癌症指标的潜力。

肿瘤细胞通过分泌较多的乳酸影响局部的酸碱度^[59],进而导致水分子之间的氢键结构发生变化。Li 等^[60]使用细胞培养基和反胶束分别模拟细胞外和细胞内的液体环境,分析了环境 pH 值发生改变时,细胞内外流体的光谱变化,结合连续小波变换和偏最小二乘回归建立了定量分析模型,发现利用水的特征吸收可以准确测定细胞内外的 pH 值。通过变量选择方法,选取了与水分子结构相关的重要波数变量,如图 4 所示,这些波数与不同氢键的水分子结构相对应。其中,自由水分子的光谱强度与 pH 值呈较

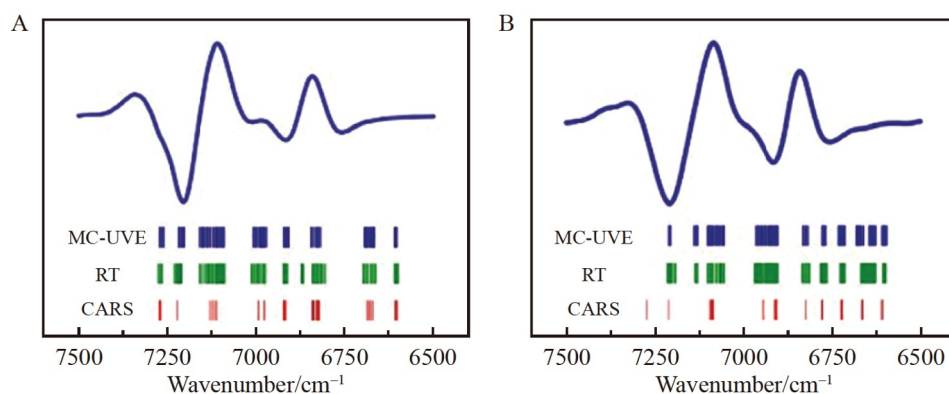


图4 利用不同变量选择方法对细胞外(A)和细胞内(B) pH 定量模型选取水光谱波数变量^[60]

Fig.4 Selected wavenumbers for the extra-cellular pH (A) and intra-cellular pH (B) models^[60]

MC-UVE: 蒙特卡洛无信息变量消除(Monte Carlo uninformativ variable elimination); RT: 随机检验(Randomization test);

CARS: 竞争自适应重加权采样(Competitive adaptive reweighted sampling)

强的线性关系,相比于其它水结构具有更高的准确性和灵敏度。因此,具有不同氢键结构的水分子可作为生物体系的探针,通过测定 pH 值监测实际生物样品的癌变情况。该方法为理解和监测细胞内外环境组成变化提供了新的视角。利用水的光谱变化可以探测肿瘤细胞与正常细胞之间的差异,这对于癌症的非侵入性、高灵敏度检测具有潜在的应用价值。

3 总结与展望

采用近红外光谱技术对生物组织或者体液样本中的化学组成进行分析已成为识别不同疾病状态的重要手段,特别是在癌症诊断中,相较于传统诊断方法,近红外光谱的无损、快速等优势使其具有巨大的应用潜力。结合化学计量学方法,能够有效地提升近红外光谱癌症诊断模型的性能。通过深度学习方法挖掘高光谱成像数据,可有效展现病变组织的空间信息,不仅能判别生物体是否癌变,还能准确识别肿瘤区域的边界。目前,近红外光谱技术在癌症诊断方面的研究多用于生物组织样本,然而,相对于病理组织,血液和尿液等生物流体更易获取,更能体现近红外光谱非侵入性检测的优势,因此,基于体液的癌症检测具有广阔的应用前景。基于近红外光谱对水的高灵敏度,利用水光谱变化提取癌症相关的化学组成和分子之间相互作用的信息,为近红外光谱在生物和生命体系分析中的应用提供了新的研究思路。随着研究的不断深入,癌症相关的水光谱特征将进一步得到挖掘,为探索和理解化学和生物过程的作用与功能提供重要的信息来源。

References

- [1] BEĆ K B, GRABSKA J, HUCK C W. *Molecules*, 2020, 25(12): 2948.
- [2] ALI J H, WANG W B, ZEVALLOS M, ALFANO R R. *Technol. Cancer Res. Treat.*, 2004, 3(5): 491-497.
- [3] YU G. *J. Biomed. Opt.*, 2012, 17(1): 010901.
- [4] TSENKOVA R. *J. Near Infrared Spectrosc.*, 2009, 17(6): 303-313.
- [5] JÖBSIS F F. *Science*, 1977, 198(4323): 1264-1267.
- [6] SAKUDO A. *Clin. Chim. Acta*, 2016, 455: 181-188.
- [7] HAN T, LIU J, LIU R, CHEN W, YAO M, LIU X, GE Q, ZHANG Z, LI C, WANG Y, ZHAO P, SUN D, XU K. *Appl. Spectrosc.*, 2022, 76(9): 1100-1111.
- [8] YANG Y, PAN T, ZHANG J. *Am. J. Anal. Chem.*, 2019, 10(5): 143-152.
- [9] VITORINO R, BARROS A S, GUEDES S, CAIXETA D C, SABINO-SILVA R. *Photodiagn. Photodyn. Ther.*, 2023, 42: 103633.
- [10] KONDEPATI V R, HEISE H M, BACKHAUS J. *Anal. Bioanal. Chem.*, 2008, 390(1): 125-139.
- [11] WANG H, CHEN P, DAI J, LIU D, LI JING XU Y, CHU X. *TrAC, Trends Anal. Chem.*, 2022, 153: 116648.
- [12] BIAN X, ZHAO Z, LIU J, LIU P, SHI H, TAN X. *Anal. Methods*, 2023, 15(39): 5190-5198.
- [13] XU Z, TALPUR Z H, YANG W, XIONG Y, WU T, ZHANG Y, SHEN X, DU Y. *Talanta*, 2022, 248: 123608.
- [14] CHEN Pu, DAI Jia-Wei, LI Jing-Yan, XU Yu-Peng, LIU Dan, CHU Xiao-Li. *Chem. Reagents (Beijing, China)*, 2023, 45(6): 105-112.
陈瀑, 戴嘉伟, 李敬岩, 许育鹏, 刘丹, 褚小立. *化学试剂*, 2023, 45(6): 105-112.
- [15] SAVITZKY A, GOLAY M J E. *Anal. Chem.*, 1964, 36(8): 1627-1639.
- [16] EHLEN L, ZABARYLO U J, SPEICHINGER F, BOGOMOLOV A, BELIKOVA V, BIBIKOVA O, ARTYUSHENKO V, MINET O, BEYER K, KREIS M E, KAMPHUES C. *J. Surg. Res.*, 2019, 242: 349-356.
- [17] HURSKAINEN M O, SARIN J K, MYLLYMAA S, GONZÁLEZ-ARRIAGADA W A, KULLAA A, LAPPALAINEN R. *Appl. Sci.*, 2021, 11(20): 9662.
- [18] SHAO X, MA C. *Chemom. Intell. Lab. Syst.*, 2003, 69(1-2): 157-165.
- [19] LEUNG A K, CHAU F, GAO J. *Anal. Chem.*, 1998, 70(24): 5222-5229.
- [20] SHAO X G, LEUNG A K M, CHAU F T. *Acc. Chem. Res.*, 2003, 36(4): 276-283.
- [21] XU X, HAN L, ZHENG Z, ZHAO R, LI L, SHAO X, LI G. *Anal. Chem.*, 2023, 95(4): 2221-2228.
- [22] CHEN H, LIN Z, MO L, WU H, WU T, TAN C. *Spectrochim. Acta, Part A*, 2015, 151: 286-291.
- [23] HAN L, SUN Y, WANG Y, FU H, DUAN C, WANG M, CAI W, SHAO X. *Spectrochim. Acta, Part A*, 2023, 289: 122233.
- [24] ZHAI Wei, XIANG Yu-Hong, DAI Yin-Mei, ZHANG Jia-Jin, ZHANG Zhuo-Yong. *Spectrosc. Spectral Anal.*, 2011, 31(4): 932-936.
翟玮, 相玉红, 代荫梅, 张家进, 张卓勇. *光谱学与光谱分析*, 2011, 31(4): 932-936.

- [25] ZHANG J, ZHANG Z, XIANG Y, DAI Y, HARRINGTON P B. *Talanta*, 2011, 83(5): 1401-1409.
- [26] LIU N, GUO Y, JIANG H, YI W. *J. Biomed. Opt.*, 2020, 25(6): 066005.
- [27] YI W, CUI D, LI Z, WU L, SHEN A, HU J. *Spectrochim. Acta. Part A*, 2013, 101: 127-131.
- [28] JANG E, VU T D, CHOI D, JUNG Y K, LEE K G, CHUNG H. *Analyst*, 2019, 144(24): 7236-7241.
- [29] CHEN H, LIN Z, MO L, WU T, TAN C. *Biomed. Res. Int.*, 2015, 2015: 472197.
- [30] ZUFREY H, MUNAWAR A A. *Appl. Spectrosc.*, 2024, 78(6): 627-632.
- [31] ZHU J, YANG C, SONG S, WANG R, GU L, CHEN Z. *Anal. Biochem.*, 2023, 669: 115120.
- [32] YIM A, ALBERTO M, SHARMA V, GREEN A, MCLEAN A, DU PLESSIS J, WONG L M, WOOD B, ISCHIA J, RAMAN J, BOLTON D. *BJU Int.*, 2024, 133(S4): 44-52.
- [33] LIU X, AN H, CAI W, SHAO X. *TrAC, Trends Anal. Chem.*, 2024, 172: 117612.
- [34] YAN T, DUAN L, CHEN X, GAO P, XU W. *RSC Adv.*, 2020, 10(68): 41936-41945.
- [35] DONG F, BI Y, HAO J, LIU S, YI W, YU W, LV Y, CUI J, LI H, XIAN J, CHEN S, WANG S. *Food Chem.*, 2024, 440: 138040.
- [36] SHANG H, SHANG L, WU J, XU Z, ZHOU S, WANG Z, WANG H, YIN J. *Spectrochim. Acta, Part A*, 2023, 287: 121990.
- [37] ZHANG X, YANG Y, WANG Y, FAN Q. *Molecules*, 2019, 24(12): 2238.
- [38] HE H, YAN S, LYU D, XU M, YE R, ZHENG P, LU X, WANG L, REN B. *Anal. Chem.*, 2021, 93(8): 3653-3665.
- [39] KHAN U, PAHEDING S, ELKIN C P, DEVABHAKTUNI V K. *IEEE Access*, 2021, 9: 79534-79548.
- [40] MUNIZ F B, BAFFA M F O, GARCIA S B, BACHMANN L, FELIPE J C. *Comput. Methods Programs Biomed.*, 2023, 231: 107388.
- [41] LIU S, WANG Q, ZHANG G, DU J, HU B, ZHANG Z. *Anal. Methods*, 2020, 12(30): 3844-3853.
- [42] ORTEGA S, HALICEK M, FABELLO H, GUERRA R, LOPEZ C, LEJAUNE M, GODTLIEBSEN F, CALLICO G M, FEI B. *Proc. SPIE-Int. Soc. Opt. Eng.*, 2020, 11320: 113200V.
- [43] MA L, LU G, WANG D, QIN X, CHEN Z G, FEI B. *Vis. Comput. Ind. Biomed. Art.*, 2019, 2(1): 18.
- [44] TRAJANOVSKI S, SHAN C, WEIJTMANS P J C, DE KONING S G B, RUERS T J M. *IEEE Trans. Biomed. Eng.*, 2021, 68(4): 1330-1340.
- [45] FANTINI S, SASSAROLI A. *Ann. Biomed. Eng.*, 2012, 40(2): 398-407.
- [46] GROSENICK D, RINNEBERG H, CUBEDDU R, TARONI P. *J. Biomed. Opt.*, 2016, 21(9): 091311.
- [47] HIROSAWA N, SAKAMOTO Y, KATAYAMA H, TONOOKA S, YANO K. *Anal. Biochem.*, 2002, 305(2): 156-165.
- [48] CHUNG S H, CERUSSI A E, KLIFA C, BAEK H M, BIRGUL O, GULSEN G, MERRITT S I, HSIANG D, TROMBERG B J. *J. Phys. Med. Biol.*, 2008, 53(23): 6713-6727.
- [49] VERMASSEN T, DE BRUYNE S, HIMPE J, LUMEN N, CALLEWAERT N, ROTTEY S, DELANGHE J. *Int. J. Mol. Sci.*, 2019, 20(7): 1592.
- [50] TRUONG B C, TUAN H D, NGUYEN H T. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2013: 33-36.
- [51] KINOSHITA K, MIYAZAKI M, MORITA H, VASSILEVA M, TANG C, LI D, ISHIKAWA O, KUSUNOKI H, TSENKOVA R. *Sci. Rep.*, 2012, 2: 856.
- [52] TSENKOVA R N, IORDANOVA I K, TOYODA K, BROWN D R. *Biochem. Biophys. Res. Commun.*, 2004, 325(3): 1005-1012.
- [53] JINENDRA B, TAMAKI K, KUROKI S, VASSILEVA M, YOSHIDA S, TSENKOVA R. *Biochem. Biophys. Res. Commun.*, 2010, 397(4): 685-690.
- [54] RAYPAH M E, MUNCAN J, SUDIK S, OMAR A F, MAIL M H, TSENKOVA R, SEENI A. *Chemom. Intell. Lab. Syst.*, 2022, 227: 104611.
- [55] HEIDEN M G V, CANTLEY L C, THOMPSON C B. *Science*, 2009, 324(5930): 1029-1033.
- [56] LIU Y, ZHANG Q, XING B, LUO N, GAO R, YU K, HU X, BU Z, PENG J, REN X, ZHANG Z. *Cancer Cell*, 2022, 40(4): 424-437.e5.
- [57] APOSTOLOVA P, PEARCE E L. *Trends Immunol.*, 2022, 43(12): 969-977.
- [58] BARROSO E M, SMITS R W H, BAKKER S T C, HOVE I T, HARDILLO J A, WOLVIUS E B, BAATENBURG DE JONG R J, KOLJENOVIC S, PUPPELS G J. *Anal. Chem.*, 2015, 87(4): 2419-2426.
- [59] HOSONUMA M, YOSHIMURA K. *Front. Oncol.*, 2023, 13: 1175563.
- [60] LI J, LIANG F, HAN L, YU X, LIU D, CAI W, SHAO X. *Chemosensors*, 2023, 11(8): 425.

Advances of Near-Infrared Spectroscopy in Cancer Diagnosis

LANG Wei-Wei¹, LI Jia-Chen^{1,2}, DONG Lin-Yi¹, HAN Li^{*1}

¹(School of Pharmacy, Tianjin Medical University, Tianjin 300070, China)

²(The Eco-City Hospital of Tianjin Fifth Central Hospital, Tianjin 300467, China)

Abstract Near-infrared (NIR) spectroscopy reflects the vibrational information of hydrogen-containing groups, allowing the detection of changes in components such as proteins and lipids in biological samples caused by diseases. Due to the advantages such as rapidity, non-destructiveness and non-invasiveness, NIR spectroscopy has been widely used in disease diagnosis. Malignant tumor cells exhibit significant differences in structure and composition compared to normal cells, and abnormal metabolic functions cause changes in the composition of biological tissues and fluids, leading to the variation in spectra. This makes NIR spectroscopy a promising tool for detecting cancer. However, severe overlapping of spectral peaks in the spectra requires the help of chemometrics methods to analyze the spectra and extract diagnostic information. This review summarizes the recent advances in NIR spectroscopy technology in cancer detection, including the methods of spectral preprocessing, diagnosis modeling and the development in NIR hyperspectral imaging combined with deep learning methods. The quantitative information and structural changes of molecules in biological systems can be extracted. Furthermore, the sensitivity of NIR spectra to water is utilized to analyze the content and structural changes of water induced by cancer and its role in cancer detection.

Keywords Near-infrared spectra; Cancer diagnosis; Chemometrics; Quantitative analysis; Structural analysis; Review

(Received 2024-05-15; accepted 2024-07-31)

Supported by the Science & Technology Development Fund of Tianjin Education Commission for Higher Education (No. 2022ZD053).