

# 结合组织芯片技术的人体胃管状腺癌组织太赫兹光谱检测

张逸<sup>1,2</sup>, 黄平捷<sup>1,2\*</sup>, 葛维挺<sup>1,2</sup>, 曹雨齐<sup>1,2</sup>, 侯迪波<sup>1,2</sup>, 张光新<sup>1,2</sup>

1. 浙江大学控制科学与工程学院, 工业控制技术国家重点实验室, 浙江 杭州 310027

2. 浙江大学医学院附属第二医院肿瘤研究所, 教育部恶性肿瘤预警与干预重点实验室, 浙江 杭州 310009

**摘要** 太赫兹波技术在生物医学领域具有广泛应用前景, 结合人体胃管状腺癌组织检测, 针对太赫兹时域光谱 (THz-TDS) 组织检测点位与实际病理难以配准的问题, 在处理石蜡包埋的胃组织的过程中引入组织芯片技术, 为实验样本的精准定位检测提供了可行方法。开展了人体胃正常组织和癌变组织太赫兹检测实验, 结合主成分分析方法, 对比分析了组织芯片技术配准的实验组与未配准的对照组的太赫兹吸收系数和折射率谱。此外对比研究了采用支持向量机和逻辑回归的太赫兹胃管状腺癌判别方法。研究结果表明, 使用组织芯片方法有助于提升组织取材的精准性从而提高癌变样本和正常样本光谱数据的区分效果。

**关键词** 太赫兹时域光谱; 胃管状腺癌; 组织芯片技术; 病变检测

**中图分类号:** R735 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)02-0397-09

## 引言

胃癌<sup>[1]</sup>诊断的常用方式主要有病理学检查、胃镜检查、生物标志物检查以及基因诊断等<sup>[2]</sup>, 其中最为通用的是准确率较高的病理学检查, 病理学检查也因此被作为胃癌诊断的“金标准”, 病理学检查需要将组织进行石蜡包埋并制作病理切片, 而后进行显微镜下观察, 这项工作比较依赖病理科医生的经验, 并且需要制作较多的病理切片以降低误判率。太赫兹时域光谱 (Terahertz time-domain spectroscopy, THz-TDS) 技术是一种无损检测手段, 其可以反映癌变与正常细胞中核酸、氨基酸、糖类、蛋白质等生物分子的整体差异, 该项技术近年来备受国内外科研人员的关注, 相关研究内容主要集中于各类癌组织切片的成像方法探索以及太赫兹时域光谱分析, 其目的主要是通过各类光谱参数整定方法, 对于癌变组织进行特异性地区分。

在研究方法上, 目前大多采用太赫兹波直接检测组织切片。根据手术中组织的取材区域或组织来源机构的鉴定结果将检测对象宏观地分为癌变组织与正常组织, 并利用太赫兹波进行检测。例如: Wahaia 等<sup>[3]</sup>利用太赫兹时域光谱技术检测了手术后石蜡包埋的癌变组织与正常组织切片, 其结果显示癌变组织的太赫兹波吸收系数和折射率均高于正常组织; Yngvesson 等<sup>[4]</sup>利用频域太赫兹成像系统对乳腺癌组织切片

进行了由吸收系数和折射率组成的反射系数的测量, 发现在不同频率处癌组织的反射系数较正常组织的都更高, 并且与水的反射系数接近, 由此得出了癌症组织的高水含量在太赫兹检测中具有一定区分性的结论; Wahaia 等<sup>[5]</sup>对结肠癌组织切片进行了太赫兹时域光谱透射检测以及连续波成像, 实验对象分别进行了干燥石蜡包埋和福尔马林浸泡, 其结果显示: 这两种处理方式下, 组织的癌变与正常区域都能够通过折射率进行鉴别, 同时, 在 0.76 THz 的频率点上的成像结果可以较好地地区分癌变和正常组织。

另外一些研究则对比了组织的太赫兹检测结果和其整体 H&E 染色结果。例如 Li 等<sup>[6]</sup>以脱水的胃癌组织为研究对象, 利用太赫兹时域光谱技术检测了组织切片的中心区域, 对照该区域 H&E 染色结果, 显示在 0.2~0.5 和 1~1.5 THz 两个频段区间中, 癌变组织和正常组织显示出了一定的差异, 利用 K-MEANS 算法和 SVM 方法, 可以有效地区分出癌变和正常的胃部组织; Chen 等<sup>[7]</sup>对来自于 31 位病人的 20 份正常结肠组织和 38 份癌变结肠组织切片进行了检测, 对组织切片进行整体 H&E 染色鉴定后, 发现癌变组织的太赫兹吸收系数明显高于正常组织, 检测结果的敏感性和特异性都达到 100%; Ji 等<sup>[8]</sup>利用太赫兹时域光谱技术的反射测量法检测了在内窥镜手术中切除的 8 片胃癌组织, 成像结果中 6 片组织的癌变区域可以较好地吻合病理检测结果, 研究中将太赫兹检测区域在组织的整体 H&E 染色图像上进行了

收稿日期: 2017-10-03, 修订日期: 2018-05-05

基金项目: 国家自然科学基金项目(61473255, 61873234)资助

作者简介: 张逸, 1993年生, 浙江大学控制科学与工程学院硕士研究生 e-mail: 4413153@qq.com

\* 通讯联系人 e-mail: huangpingjie@zju.edu.cn

定位。

先前的研究展示了太赫兹光谱技术对于癌组织的区分能力，然而多倾向于宏观地区分癌变与正常组织，存在不少局限性，例如：(1)组织类别的针对性问题。不少病变现象关联特定的组织类别，如胃癌发生于胃粘膜，而胃组织切片中极可能混合多种组织类型，因此，利用太赫兹波直接检测组织切片容易引入干扰。(2)病理对照的精准性问题。太赫兹光斑的直径一般仅为 1~4 mm，若对于组织切片直接进行检测，在比对组织的 H&E 染色结果时难以精确地确定太赫兹光斑实际照射的区域，从而较难将太赫兹检测结果与该组织点位的实际病理形成精确对照。即使单大类癌症中，基于发病机制的差异可能存在不同病理分型，病理信息是癌组织检

测中的重要信息。因此，太赫兹检测结果和病理的精准对照对于太赫兹癌组织检测研究具有重要意义。

基于上述讨论，本文尝试引入组织芯片方法进行胃癌组织检测样本制备和实验分析。

(1)胃癌组织样本的选取

如图 1 为胃壁的结构示意图，胃壁包括粘膜组织、粘膜下组织、肌肉组织以及浆膜等，目前太赫兹胃癌检测研究的实验对象多为整片活检取材得到的胃壁组织<sup>[9,11]</sup>，很可能混合粘膜、粘膜下组织和肌肉等多种组织类型，胃腺癌包括胃管状腺癌发生于胃粘膜组织中的胃腺，并逐渐向下扩散(见图 2)，因此，对胃癌的诊断和早期预警，胃粘膜组织是需要首要关注的对象。

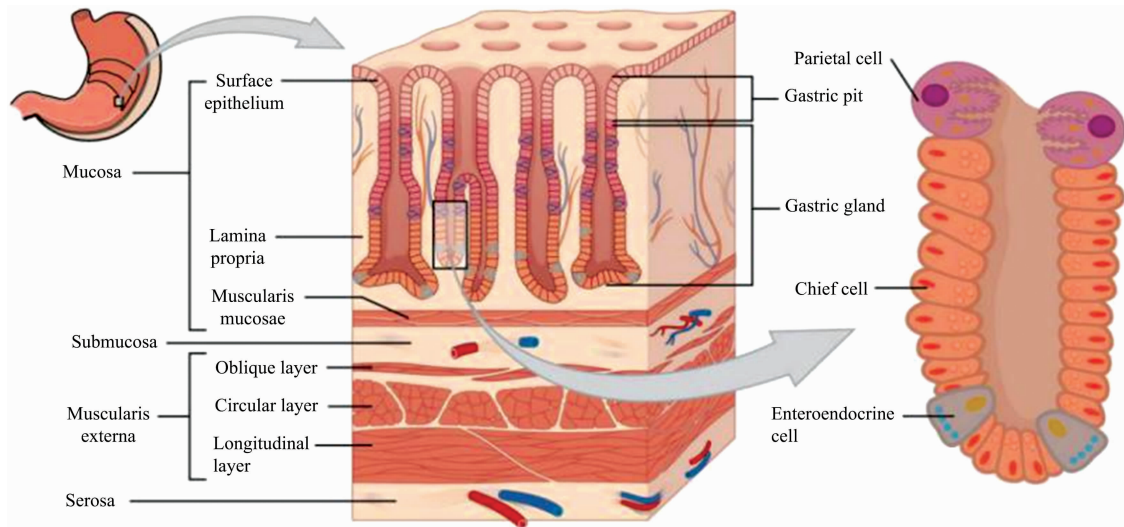


图 1 胃壁及胃腺结构示意图<sup>[1]</sup>

Fig. 1 Schematic diagram of gastric wall and gastric glands<sup>[1]</sup>

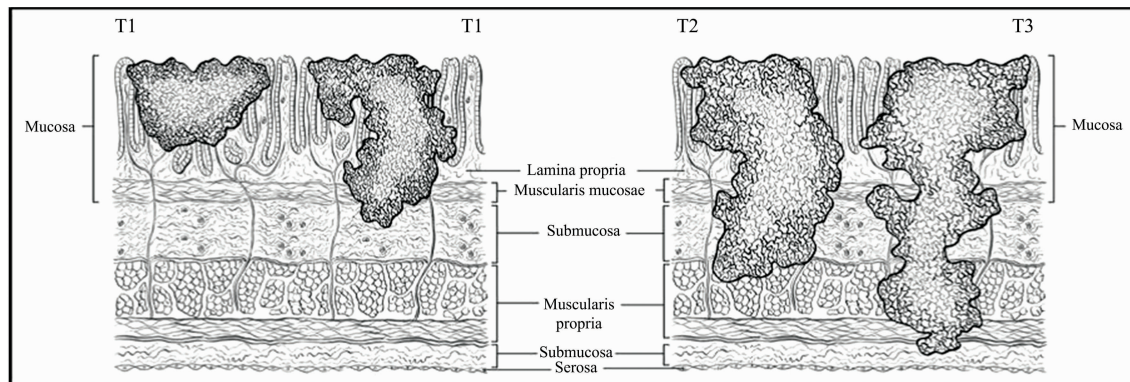


图 2 胃癌的发生及扩散<sup>[10]</sup>

Fig. 2 The occurrence and diffusion of gastric cancer<sup>[10]</sup>

在各类胃癌病症中，管状腺癌<sup>[11]</sup>是最为常见的类型之一，其肿瘤由大小不同、裂隙样有分枝或扩张形的管腔组成。肿瘤细胞形态呈立方或柱状，腺腔内黏液较多时瘤细胞被挤压，可呈扁平状，有时可有透明细胞样变异。

为此，我们的研究主要选取胃管状腺癌组织作为检测样本，并将 THz 检测区域限制在胃粘膜上。

(2)组织芯片技术的引入

为了实现病理分析与太赫兹检测结果的精确对照，本文引入了组织芯片技术。组织芯片(tissue chip)，也称组织微阵列(tissue microarrays)，是生物芯片技术的一个重要分支<sup>[12]</sup>，该方法将许多不同个体组织标本以规则阵列方式分布于同一载体上，进行同一指标的组织学研究。其应用范围

非常广泛<sup>[13]</sup>, 如: 细胞高通量免疫表型分析、联合基因芯片寻找治疗肿瘤的新靶点、制作缩微组织学和病理学图谱、筛选疾病相关抗体和探针等。

本文主要利用组织芯片中组织点的病理信息可明确获取, 且其大小与太赫兹光斑相匹配的特点, 达到组织病理与太赫兹检测结果配准的目的。

## 1 实验部分

### 1.1 样品制备

研究中的组织样本由浙江大学肿瘤研究所提供。其中病变样本均指向胃管状腺癌, 原始实验样本为经石蜡包埋的癌组织和癌旁正常组织(图 3), 在制作组织芯片前, 对各组织块都制作了 H&E 染色切片。

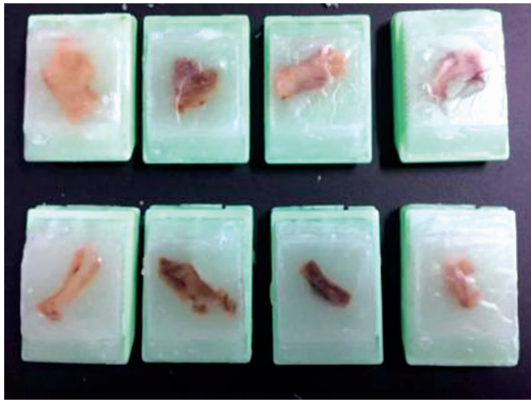


图 3 石蜡包埋的组织样本

Fig. 3 Paraffin embedding tissue samples

直接检测组织切片的实验方式中, 实际太赫兹光斑照射点与 H&E 染色切片的观察区域易存在偏差, 因此, 设置实验组和对照组, 两者样本形式均为组织芯片, 实验组样本采用标准流程取样制作, 对照组样本组织取样点位于实验组取样点邻域, 用以模拟直接检测组织切片方式下实际光斑照射区域与病理观察区域存在偏差的情况, 实验组和对照组区别如表 1 所示。

表 1 实验组和对照组样本区别

Table 1 The difference between the experimental group and the control group

	实验组	对照组
原始组织样本	10 位病人的癌组织和正常组织	
取样区域	经 H&E 染色切片标记确认的胃粘膜区域	实验组取样点邻域, 中心点距离不超过 5 mm
目的	精准检测胃粘膜区域组织	模拟直接检测组织切片
样本最终形式	组织芯片	组织芯片

组织芯片的制作步骤如下:

(1) 基于组织块的 H&E 染色切片, 显微镜下定位目标胃粘膜区域, 于 H&E 染色切片上进行标记, 基于此标记石

蜡组织块上同一部位, 以便后续操作中正确取样。

(2) 实验组按照图 4 所示的方式提取上一步骤中的目标组织, 对照组于实验组取样点邻域取样, 取样中心点距离不超过 5 mm。采用组织芯片阵列仪将提取的组织注入到受体石蜡块阵列中, 并对每个组织点都进行编号, 从而对应 H&E 染色结果。本研究采用的受体阵列共包含 12 个孔洞, 单个孔洞直径为 2 mm。

(3) 将注入组织后的受体石蜡块置于 50 °C 的恒温箱中 2 h, 使得组织与石蜡紧密连接, 避免后续切片过程中发生组织脱落。

(4) 切割受体石蜡块, 以获得组织芯片切片, 用于太赫兹检测, 各切片厚度均为 1.5 mm, 且厚度均匀, 如图 5 所示, 左下角两个孔洞未注入组织, 用以标记切片正面。

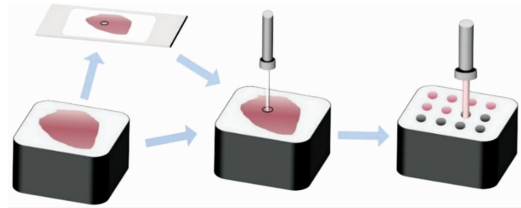


图 4 制作组织芯片的流程示意图

Fig. 4 Schematic diagram of the process of making tissue microarray



图 5 组织芯片的切片

Fig. 5 Tissue microarray slice

### 1.2 数据采集

为达到太赫兹检测结果与病理的精准对照, 需确保检测过程中太赫兹光斑准确照射在组织芯片切片的对应样本点, 并且获取样本点的病理图片, 实验中的数据采集方法如下:

(1) 样本点的太赫兹准确检测

本研究所采用的太赫兹时域光谱实验装置是美国 Zomega 公司的 Z3 系统, 具体介绍可参照文献[14]。太赫兹检测采用透射模式, 为防止实验过程中空气中水分的太赫兹波吸收干扰测量结果, 采用对系统充入高纯度氮气的方式使湿度基本保持在 0%。同时, 组织芯片切片贴附于开有直径为 1.5 mm 圆形孔洞的铜片上, 从而避免检测中太赫兹光斑照射组织点周围的石蜡部分, 铜片固定于定制的样品架上, 利用样品架上相应标记, 确保每次检测中样本点固定于相同位置。检测过程中, 铜片的圆形孔洞对准目标组织点位, 组

织点的直径均为 2 mm, 检测中仅通过圆形孔洞的太赫兹波透射相应组织点, 由此确保太赫兹检测结果完全对应组织样本点信息。

### (2) 样本点病理图片的获取

在切割受体石蜡块获得组织芯片切片后, 利用石蜡切片机切削余下受体石蜡块, 步进距离设置为 5  $\mu\text{m}$ , 切片制作为 H&E 染色切片, 该染色切片中样本点与组织芯片切片上样本点一一对应。

## 1.3 光学参数提取

构成肿瘤标志物的核酸、蛋白质、糖类等生物分子的振动和转动模式频率都位于太赫兹频段。因此, 基于肿瘤组织与正常组织的医学病理机制的区别以及太赫兹的检测特性, 理论上太赫兹技术能够对两类组织进行区分。组织的太赫兹吸收系数和折射率对应于组织内部成分和分子结构信息, 最能反应病变组织的太赫兹响应特征。

折射率和吸收系数根据 Dorney 和 Duvillaret 等提出的太赫兹时域光谱技术光学常数提取模型计算<sup>[15-16]</sup>。设未放置样品时采集到的太赫兹时域信号傅里叶变换后为参考信号  $E_{\text{ref}}(\omega)$ , 透射经过厚度为  $d$  的样本后的太赫兹信号为  $E_{\text{sam}}(\omega)$ 。经过模型简化, 样本的复折射率  $\tilde{n}(\omega) = n(\omega) - j\kappa(\omega)$  与采集信号之间存在如式(1)的关系, 其中  $n(\omega)$  为折射率, 表征样品的色散特性;  $\kappa(\omega)$  为消光系数,  $\alpha(\omega) = 2\omega\kappa(\omega)/c$  为吸收系数, 表征样品的吸收特性,  $c$  为光速。

$$H(\omega) = \frac{E_{\text{sam}}(\omega)}{E_{\text{ref}}(\omega)} = \frac{4\tilde{n}}{(1+\tilde{n})^2} e^{-\frac{\omega d}{c}(\tilde{n}-1)} \quad (1)$$

将复折射率的表达式  $\tilde{n}(\omega) = n(\omega) - j\kappa(\omega)$  代入式(1), 并将  $H(\omega)$  表示成模和辐角的形式  $H(\omega) = A(\omega)e^{-j\varphi(\omega)}$ 。在样本消光系数远远小于折射率的前提下, 样本的折射率和吸收系数的计算公式为

$$n(\omega) = \frac{c\varphi(\omega)}{\omega d} + 1 \quad (2)$$

$$\alpha(\omega) = \frac{2\kappa(\omega)\omega}{c} = \frac{2}{d} \ln \frac{4n(\omega)}{A(\omega)(n(\omega)+1)^2} \quad (3)$$

由于癌细胞代谢水平的加强, 癌组织整体含水量有一定升高, 水分子间氢键对于太赫兹波的强吸收使得癌变组织和正常组织在太赫兹吸收系数上有所差异<sup>[4,7]</sup>。组织经石蜡包埋后, 排除了水分含量差异的影响, 一些癌变标志物的太赫兹波特异性吸收可能导致癌变和正常组织的太赫兹吸收特性差异, 如癌变组织中的色氨酸含量增高<sup>[17]</sup>会导致特定频率点上太赫兹吸收增强, 另外, 由于太赫兹波段囊括了大多生物分子的振动和转动模式频率, 癌组织旺盛的代谢引起的细胞内蛋白增多、细胞核内染色质增多等情况在理论上也会导致癌变组织对于太赫兹波的折射率和吸收系数强于正常组织。

## 2 结果与讨论

### 2.1 太赫兹时域光谱响应的对比分析

#### 2.1.1 实验组(配准)的太赫兹响应

图 6 左侧为实验组中病变情况明确且较为清晰的三位病人样本 H&E 染色结果, 样本全部区域均为胃粘膜, 具有较

好的对照基础, 其中, a, b 和 c 样本为三位病人的癌变胃粘膜组织, d, e 和 f 样本分别为其各自正常胃粘膜组织。

a 样本中可见管腔结构大小差异, 部分呈现分支状或者裂隙状, 属于典型的管状腺癌征象, b 和 c 样本的腺管大小差异较大, 同时存在明显的不规整性以及裂隙状或分支结构, 都表现出管状腺癌的典型病理特征。

d, e 和 f 样本中腺管结构和大小均匀, 无明显的分支状或者裂隙状, 为正常的胃粘膜组织, 由于其中 e 样本切片周边有少量污染, 因此截取了其较为规整清晰的部分。

图 6 右侧展示了这些样本的太赫兹吸收谱及折射率谱, 癌变粘膜组织对应太赫兹谱用红线表示, 正常粘膜组织对应太赫兹谱则用蓝线表示。可见, 三位病人癌组织的太赫兹吸收系数和折射率相对于其正常组织样本都出现了明显升高。

在机理层面, 研究结果表明癌细胞代谢水平加强引起的水含量增高现象会导致癌组织对于太赫兹波的吸收水平强于正常组织, 然而组织经石蜡包埋后, 排除了水分含量差异的影响, 癌组织对应的太赫兹吸收水平依然高于正常组织, 说明其他影响因素也在这种太赫兹响应差异中有所贡献<sup>[18-19]</sup>, 如癌细胞快速不受控分裂引起的局部细胞密度增大; 癌细胞核的体积增大, 细胞核质比的增高<sup>[9]</sup>; 癌组织微环境与正常组织的差异, 包括癌细胞胞浆内核蛋白体增多<sup>[20]</sup>, 癌细胞的异常分泌物或代谢产物(如蛋白、粘液和色素等)。

图 8 中带误差棒的曲线为实验组样本太赫兹吸收系数和折射率的平均谱图。可见癌变后样本的太赫兹吸收系数和折射率明显高于正常样本, 变化趋势明显且稳定, 与上述机理层面的推测相符合。

#### 2.1.2 对照组(未配准)的太赫兹响应

对照组中, 即使组织取材点相对实验组的偏移不超过 5 mm, 仍有约 1/3 的样本点未能完全取到粘膜组织, 其中典型样本如图 7 所示。三样本都为正常组织, g 样本全部为肌肉组织, h 和 i 为典型的掺杂粘膜下层组织的样本, 其中 h 样本中除右下区域的正常胃粘膜组织外, 左上区域为粘膜下层组织, 两者间有明显的粘膜肌层分隔, i 样本相对于 h 样本, 整体掺杂有更多的粘膜下层组织, 分布有更多的血管、淋巴组织, 且图中样本点右下区域显示还混杂有少量肌肉组织。

图 8 中显示了图 7 对应样本的太赫兹谱图, 并以实验组中样本太赫兹吸收系数和折射率的平均谱图作为对照, 可见图 7 中对应样本虽然都为正常组织, 其太赫兹吸收系数和折射率相对正常粘膜组织明显偏高。整体为肌肉组织的 g 样本的太赫兹响应接近癌变后粘膜组织, 而掺杂有粘膜下层组织的 h 和 i 样本对应更高的太赫兹吸收系数和折射率, 在 0.6 THz 后的高频段更为明显。

由此可见组织类型对于组织的太赫兹吸收响应有较大影响。肌肉组织结构显著区别于粘膜组织, 其太赫兹响应也不同于粘膜组织, 此外, 胃粘膜下层组织的构成相对更为复杂, 血管和淋巴分布的差异对太赫兹检测结果的影响难以定量评估, 因此, 太赫兹胃组织检测的研究应当对于样本的组织类型加以重视。

对照组实验结果说明: (1) 直接检测胃组织切片的实验方式下, 即使实际检测点与病理观察区域存在的偏差不得超过

5 mm, 也可能造成检测区域实为非粘膜组织而未能发现的情况。(2)组织类型差异对组织的太赫兹响应影响较大, 在

进行太赫兹癌组织检测研究过程中, 检测样本内混合有不同的组织类型将带来干扰。

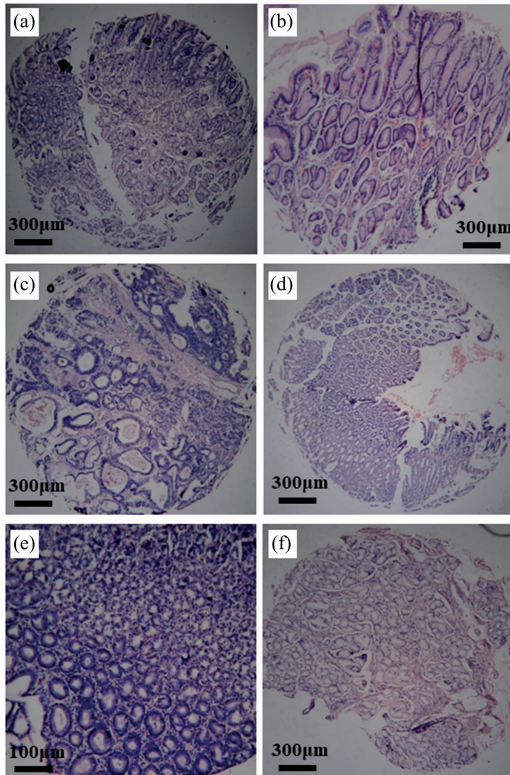


图 6 (a), (b), (c)为三位病人癌变胃粘膜组织; (d), (e), (f)依次为其正常胃粘膜组织, 各样本太赫兹吸收系数和折射率谱图

Fig. 6 (a), (b), (c) represent tumor tissue of three patients with normal tissue of (d), (e), (f), terahertz absorption spectrum and refractive index spectrum of each sample

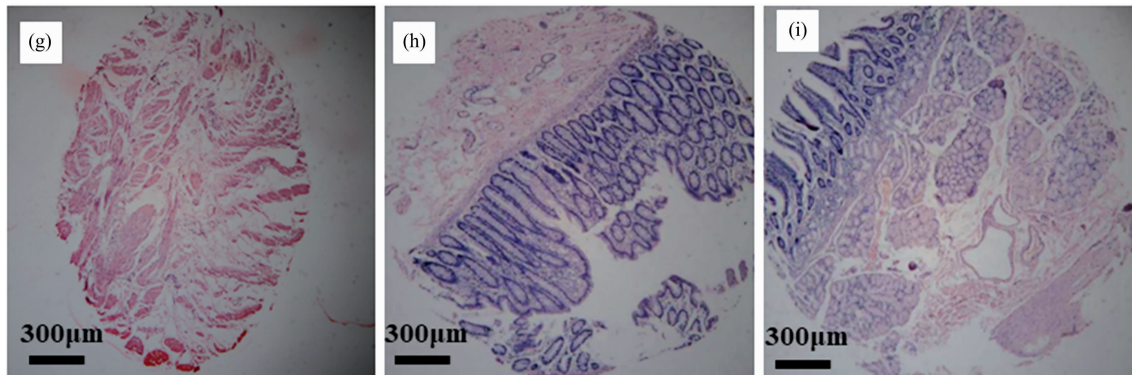
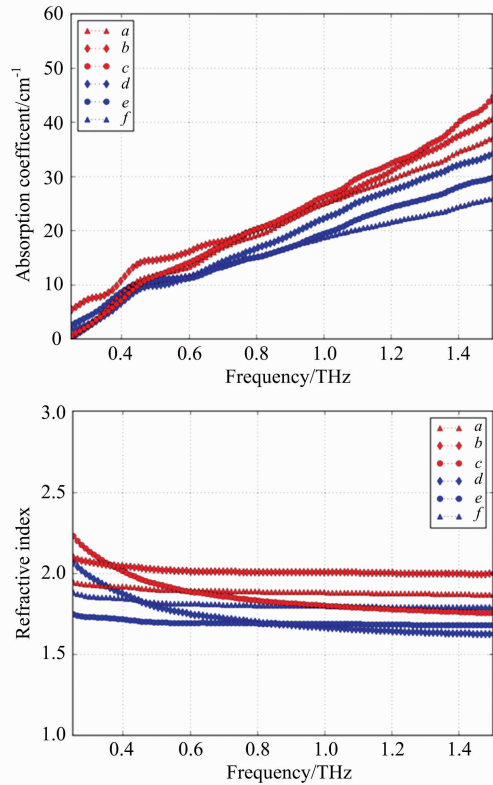


图 7 对照组中的非粘膜组织样本

(g): 肌肉; (h): 粘膜+粘膜下层组织; (i): 粘膜+粘膜下层组织+肌肉

Fig. 7 Non-mucosal tissue samples in the control group

(g): Muscle tissue; (h): Mucosa tissue+submucosa tissue;

(i): Mucosa tissue+submucosa tissue+muscle tissue

### 2.1.3 实验组、对照组的太赫兹谱主成分分析

特征吸收峰是传统红外光谱分析技术中重要的参数之一, 但由于生物大分子在 THz 频段的总响应谱带范围较宽, 吸收带重叠, 因此对于无明显吸收峰的样品, 难以采用传统

归属峰位置及强度的视角进行分析。主成分分析(PCA)方法可望解决这个问题。PCA 方法的主要原理是将原变量进行转换, 使数目较少的新变量成为原变量的线性组合, 而且新变量能最大程度的表征原变量的数据结构特征, 以消除众多

共存信息中相互重叠的部分，从而达到数据降维和特征提取的目的。

对实验组、对照样本太赫兹吸收系数和折射率谱分别进行主成分分析，得到主成分得分如图 9 所示，可见第一、二主成分(PC1 和 PC2)的累计贡献率均超过 96%，因此，前两个主成分即可很好地概括数据信息。实验组中采用吸收系

数和折射率都可以较为有效地区分癌变及正常组织样本，对照组中，使用相同的参数，区分效果明显弱于实验组，归因于样本中部分非粘膜组织的太赫兹响应不同于粘膜组织，其数据引入了干扰。因此使用组织芯片方法提升组织取材的精准性有助于提高癌变样本和正常样本光谱数据的区分性。

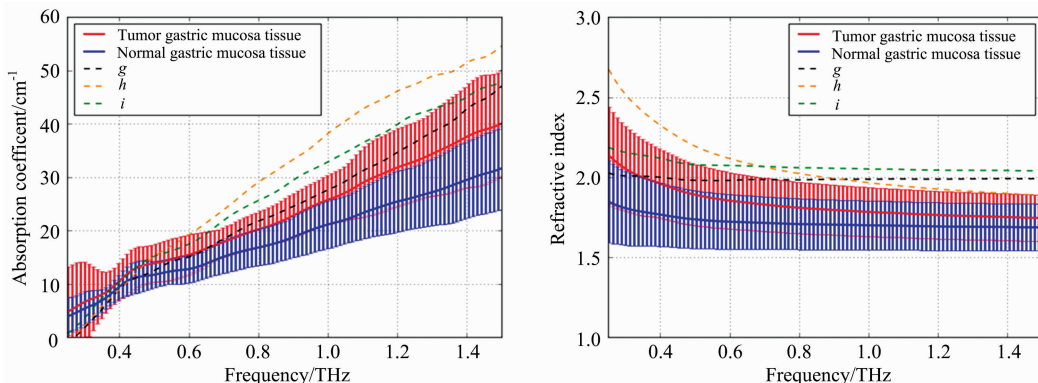


图 8 图 7 对应样本和带误差棒的实验组样本平均太赫兹吸收系数和折射率谱图

Fig. 8 Average terahertz absorption spectrum and refractive index spectrum of experimental group samples with error bar and samples of Figure 7

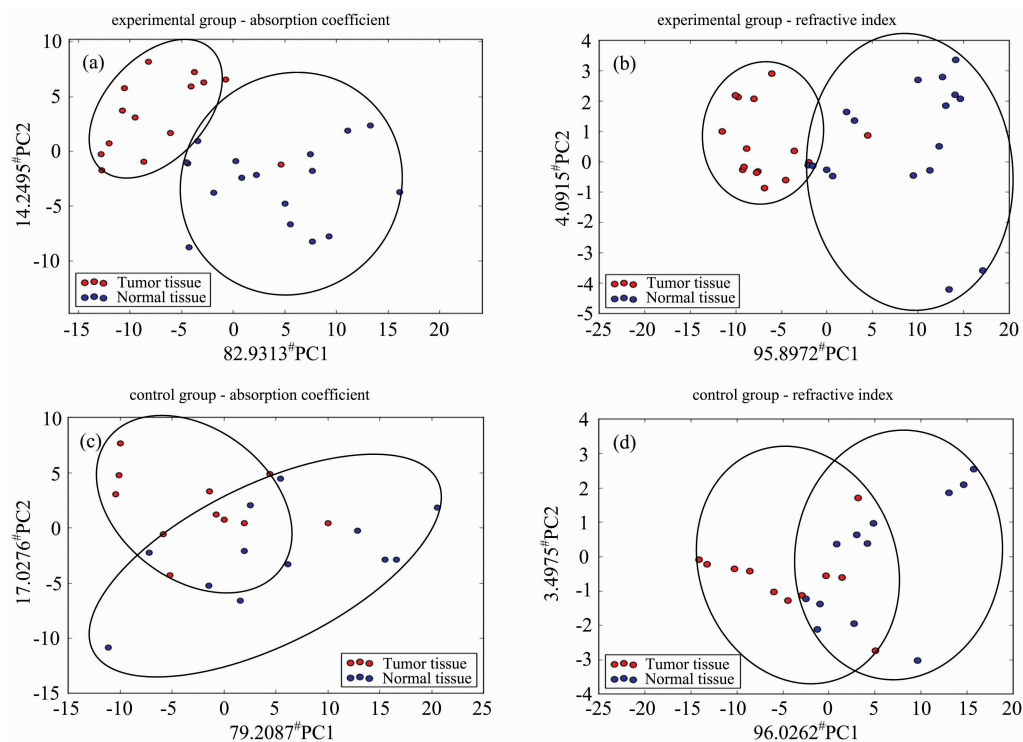


图 9 实验组各参数主成分得分图[(a): 吸收系数, (b): 折射率]; 对照组各参数主成分得分图[(c): 吸收系数, (d): 折射率]  
Fig. 9 Principal component scores of experimental group by parameters [(a): Absorption coefficient, (b): Refractive index]; Principal component scores of control group by parameters [(c): Absorption coefficient; (d): Refractive index]

### 2.2 基于 Thz 时域光谱的胃管状腺癌判别方法

基于在机理层面对于一些癌变征象导致组织太赫兹响应差异化的解释，以及本研究中病理确认的癌变粘膜组织与正常粘膜组织的太赫兹响应稳定的差异性，本文进一步研究了

利用太赫兹光谱建立胃癌判别模型的方法。

在判别方法的选取方面，由于研究涉及的组织样本量小，需要着重考虑分类算法的泛化能力，防止过拟合。因此，适用于小样本分类的支持向量机(support vector machine,

SVM)算法较为合适。另外,由于胃癌检测在实际情况中对于样本癌变概率的信息有较大需求,因此本身能输出样本类别概率且可解释性较强的逻辑回归(logistic regression, LR)方法也被纳入比较。

图 10 为本研究采取的基于太赫兹谱的胃癌组织分类检测算法流程,其中对于胃组织的吸收系数谱和折射率谱分别进行了 SG 平滑去噪和 z-score 归一化,以消除各频率点吸收系数和折射率取值范围的较大差异,防止单点值比重过大,提高分类模型的稳健性,在此基础上利用主成分分析方法提取主要特征后,与时域信号特征相结合,共同构成分类模型的输入特征。时域信号特征的提取方法参考了 Grootendorst 等<sup>[21]</sup>太赫兹癌组织检测研究中设计的相关特征,分别为信号最大值、最小值、峰峰值、13.15~13.65 ps 的主峰二次拟合值、14.10~14.40 ps 的主脉冲最后部分线性拟合值以及

12.20~14.50 ps 的主脉冲积分值,图 11 为其可视化表示。

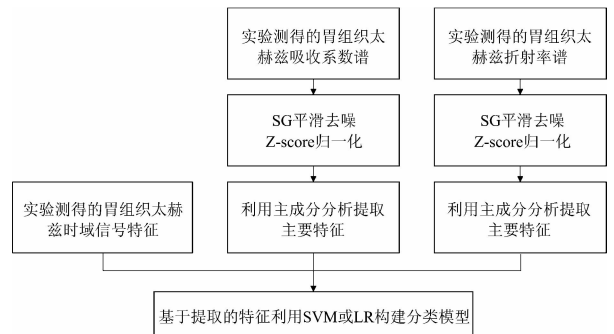


图 10 基于太赫兹谱的胃癌组织分类检测算法流程  
Fig. 10 The algorithm flow chart of gastric cancer detection based on terahertz spectrum

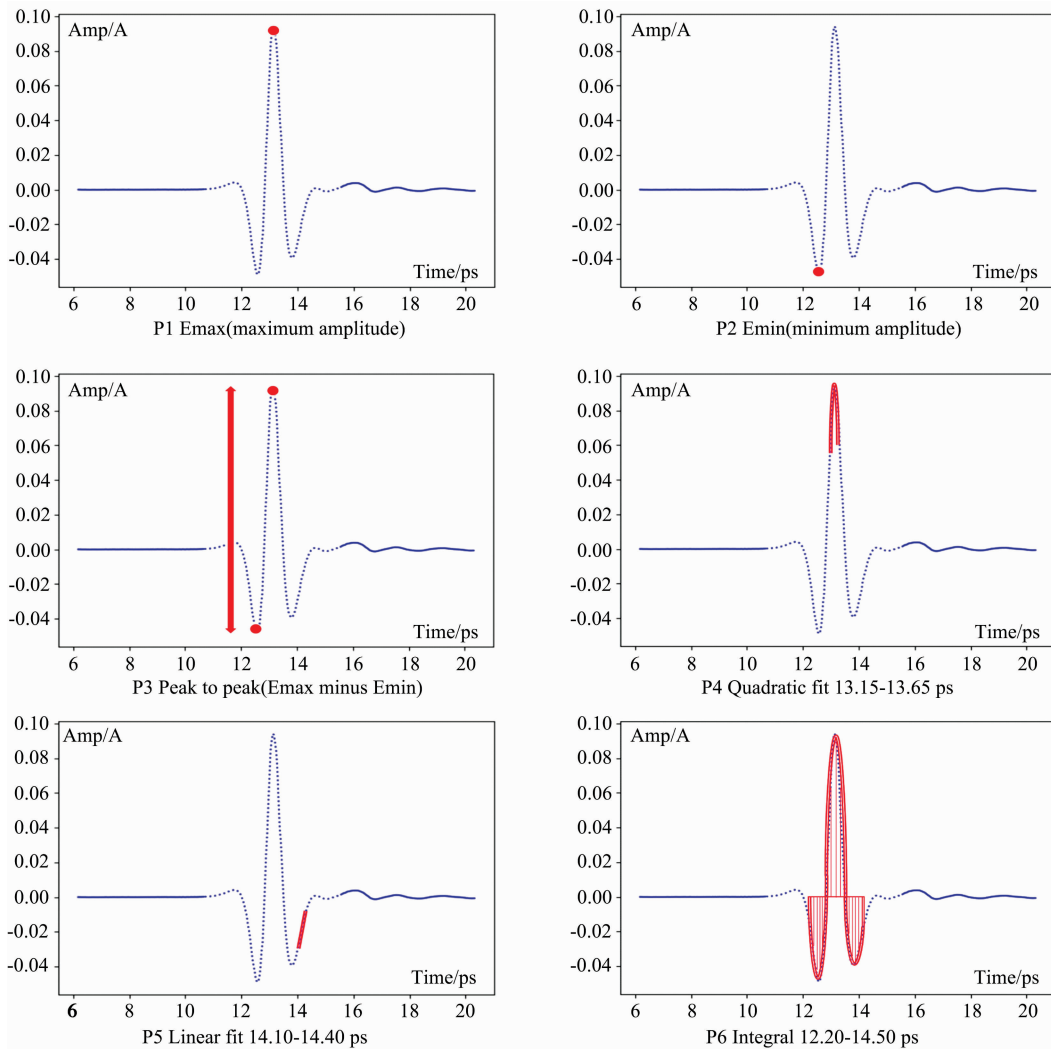


图 11 时域信号特征的可视化表示,各参数以红色表示

Fig. 11 The visual representation of time domain signal characteristics, parameters marked in red

两种分类方法各自经五折交叉验证后的平均灵敏度和准确率如表 2 所示,其中支持向量机采用的是径向基核函数。可见无论使用支持向量机或逻辑回归,实验组的灵敏度和准

准确率都明显高于对照组,其原因主要是组织芯片方法带来的粘膜组织取材精准性减少了由组织类型差异引入的干扰。另外,对于全部组织取材于胃粘膜的实验组,支持向量机方法

的灵敏度和准确率都优于逻辑回归方法,其原因可能是少量异常点在逻辑回归的对数损失函数中相对其在支持向量机的合页损失函数中有更大的影响,另外,所有训练样本都影响逻辑回归方法分类超平面的确立,而支持向量机分类超平面仅取决于少量核心支持向量,有利于获得小样本数据的关键特征。

表 2 五折交叉验证后的平均灵敏度和准确率

Table 2 Average sensitivity and accuracy of five-fold cross-validation

方法	样本组别	灵敏度/%	准确率/%
SVM	实验组	92.6	89.1
	对照组	86.7	83.9
LR	实验组	86.1	84.8
	对照组	81.1	77.7

表 3 逻辑回归和支持向量机方法输出的图 6 中相应样本的癌变概率

Table 3 Probability of cancer of the samples in figure 6 which is output by logistic regression and support vector machine

方法	a	b	c	d	e	f
LR	0.78	0.66	0.56	0.07	0.05	0.34
SVM	0.90	0.73	0.81	0.13	0.03	0.09

另外,鉴于癌变概率信息对于组织的癌变情况有更进一步的信息量,本研究尝试利用支持向量机和逻辑回归输出检测样本的癌变概率,其中支持向量机方法虽然常作为分类方法,本研究中依据 JC Platt 提出的方法<sup>[22]</sup>,通过将样本点在高维空间中与分类超平面的距离进行 sigmoid 拟合后,获取该样本癌变的概率信息。对于图 6 当中的病理特征较为明确的样本,两种方法输出的癌变概率信息如表 3 所示。可见若以值为 0.5 作为分界,两种方法对于这些样本都没有出现类别的误判,对比而言,除了样本 d,支持向量机相对逻辑回归做出的概率信息输出更为接近样本真实病理类别,给出的信息更为明确,这与两种方法在前述分类灵敏度及准确率的效果对比结果一致。

## References

- [1] Ferlay J, Shin H R, Bray F, et al. *Int. J. Cancer*, 2010, 127(12): 2893.
- [2] YANG Jia-li, LÜ Ming(杨佳莉, 吕 铭). *World Health Digest(中外健康文摘)*, 2012, 9(19): 78.
- [3] Wahaia F, Kasalynas I, Seliuta D, et al. *Journal of Molecular Structure*, 2014, 1079: 391.
- [4] Yngvesson S K, Peter B S, Siqueira P, et al. *Proceedings of SPIE*, 2012, 8221: 82210N.
- [5] Wahaia F, Valusis G, Bernardo L M, et al. *Journal of Molecular Structure*, 2011, 1006: 77.
- [6] Hou D, Li X, Cai J, et al. *Physics in Medicine & Biology*, 2014, 59(18): 5423.
- [7] Chen H, Ma S, Wu X, et al. *Journal of Biomedical Optics*, 2015, 20(3): 36017.
- [8] Ji Y B, Park C H, Kim H, et al. *Biomedical Optics Express*, 2015, 6(4): 1398.
- [9] Goryachuk A A, Khodzitsky M K, Borovkova M A, et al. *Journal of Physics Conference Series*, 2016.
- [10] Frederick L G M D, David L P M D, Irvin D F M D, et al. *AJCC Cancer Staging Manual*. *AJCC Cancer Staging Manual*. Springer, 2010. 803.

以上胃癌判别的结果表明:(1)在组织取材方法上,实验组相对对照组更优异的准确率和灵敏度体现了利用组织芯片方法进行精准检测在太赫兹组织研究中的优势和必要性。(2)在判别方法上,支持向量机相对逻辑回归表现出更高的准确率和灵敏度,且对于病理明确的典型组织点的癌变概率信息输出与真实病理更为接近,说明本研究中组织样本量不大的情况下,支持向量机更好的数据概括能力有其明显优势。

## 3 结 论

主要利用组织芯片制备方法开展了人体胃管状腺癌组织的太赫兹时域光谱检测研究。

针对太赫兹时域光谱直接检测组织切片方法中检测点位与实际病理难以配准的问题,将石蜡包埋的胃组织样本分为利用组织芯片技术配准的实验组与未配准的对照组,并进行了如下分析:

(1)对比了组织芯片技术配准的实验组与未配准的对照组的太赫兹吸收系数和折射率谱的差异,采用主成分分析方法分析了实验组与对照组中癌变组织与正常组织光谱数据的差异程度。

(2)基于癌变组织细胞密度、细胞核体积、组织微环境等层面的差异导致太赫兹响应变化的理论解释,以及实验结果中癌变组织与正常组织稳定明显的差异性,研究了应用逻辑回归和支持向量机进行组织癌变情况判别和癌变概率计算的方法。

实验结果中,对照组由于取材定位不精确,易引入胃粘膜下层组织或肌肉组织等非目标组织,不同组织类型的太赫兹响应规律不同,从而易产生干扰,对比而言,实验组中,癌变组织与正常组织的光谱数据具有更良好稳定的区分性。另外,在癌变情况判别研究中,支持向量机方法在灵敏度和准确率等方面展现出了更好的效果,同时,对于相同判别方法,实验组数据表现出更高的分类准确率和灵敏度,说明组织芯片方法带来的取材精准性的优势有助于提升太赫兹胃癌检测效果。



- [11] “Gastric Cancer Molecular Fractal and Individualized Diagnosis” in the National “863” Major Project(国家“863”重大项目“胃癌分子分型与个体化诊疗”课题组). *World Health Digest(中外健康文摘)*, 2010, 39(4): 266.
- [12] Dancau A M, Simon R, Mirlacher M, et al. *Methods in Molecular Biology*, 2016, 1381: 53.
- [13] Barrette K, Oord J J V D, Garmyn M. *Journal of Investigative Dermatology*, 2014, 134(9): e24.
- [14] Liu Z, Li X, Yue F H, et al. *Advanced Materials Research*, 2013, 718-720: 1060.
- [15] Duvillaret L, Garet F, Coutaz J L. *IEEE Journal of Selected Topics in Quantum Electronics*, 1996, 2(3): 739.
- [16] Dorney T D, Baraniuk R G, Mittleman D M. *Journal of the Optical Society of America A-Optics Image Science and Vision*, 2001, 18(7): 1562.
- [17] Kamson D O, Mittal S, Buth A, et al. *Molecular Imaging*, 2013, 12(5): 327.
- [18] Bowman T, Wu Y, Gauch J, et al. *Journal of Infrared Millimeter & Terahertz Waves*, 2017, 38(6): 766.
- [19] Wahaia F, Kasalynas I, Venkevicius R, et al. *Journal of Molecular Structure*, 2016, 1107: 214.
- [20] Cheon H, Yang H J, Son J H. *IEEE Journal of Selected Topics in Quantum Electronics*, 2017, 23(4): 8600109.
- [21] Grootendorst M R, Fitzgerald A J, Sg B D K, et al. *Biomedical Optics Express*, 2017, 8(6): 2932.
- [22] Platt J C. *Advances in Large Margin Classifiers*, 2000, 10(4): 61.

## Human Gastric Tubular Adenocarcinoma Tissue Detection Based on Terahertz Time Domain Spectroscopy and Tissue Microarray Technology

ZHANG Yi<sup>1,2</sup>, HUANG Ping-jie<sup>1,2\*</sup>, GE Wei-ting<sup>1,2</sup>, CAO Yu-qi<sup>1,2</sup>, HOU Di-bo<sup>1,2</sup>, ZHANG Guang-xin<sup>1,2</sup>

1. State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China
2. Cancer Institute, Key Laboratory of Cancer Prevention and Intervention, the Second Affiliated Hospital, Zhejiang University, College of Medicine, Hangzhou 310009, China

**Abstract** Terahertz wave technique is very prospective in the biomedical application field, where there exists a practical problem that it is hard to register a terahertz time-domain spectroscopy (THz-TDS) detection point and its accurate corresponding position in pathology image. In this paper, complex human gastric tissue is taken as the experimental subject to search for accurate detection approach. This paper uses tissue microarray technology to process paraffin-embedded gastric tissue, and then contrasts the Terahertz absorption coefficient and refractive index spectrum of detection subjects which are divided into registered experimental group and unregistered control group using principal component analysis method. Meanwhile, this paper contrasts the gastric tubular adenocarcinoma THz-TDS discrimination methods using support vector machine and logistic regression respectively. The research result reveals that the detection accuracy brought by tissue microarray technology helps better discriminate tumor samples and normal samples.

**Keywords** Terahertz time domain spectroscopy; Gastral tubular adenocarcinoma; Tissue microarray technology; Lesion detection

(Received Oct. 3, 2017; accepted May 5, 2018)

\* Corresponding author