

基于 X 射线吸收谱的不同生物组织的辨识

王倩, 杨正, 方正
(厦门大学仪器与电气系, 厦门 361005)

摘要 为了验证 X 射线吸收光谱法对生物组织的辨识能力, 选取猪的心脏、肝、肾、胃、瘦肉以及肥肉作为实验样本, 以 55 kV 的电压激发 X 射线管, 利用 X 射线探测器获取这 6 类生物组织样本的 X 射线吸收光谱. 将采集到的光谱数据分为测试集与训练集, 利用主成分分析法提取光谱主成分, 以训练集为输入建立径向基函数(RBF)神经网络分类预测模型, 对测试集样本进行分类预测. 通过交叉验证法对所有样本进行辨识的识别正确率达到 90.22%. 实验结果表明, X 射线光谱技术结合主成分分析法和 RBF 神经网络能够很好地用于猪的组织分类, 对将 X 射线光谱技术应用于生物组织辨识具有重要的指导意义.

关键词 X 射线吸收光谱; 生物组织; 分类识别; 径向基函数神经网络

中图分类号 O657.39 **文献标志码** A

与传统医学相比, 现代医学的一个重大变革就是由传统的基于症状的治疗模式转变为以信息为依据的治疗模式. 当今一些重大的医学课题, 着重考虑的是探索导致疾病的生物信息规律, 从各个学科探索医学诊断和治疗的新方法^[1]. 关于人类疾病组织生化性质的研究是现代医学的一个中心宗旨, 组织发生疾病的过程可能会导致组织生物化学性质的变化, 对这种变化的分析为疾病的诊断提供了重要线索^[2]. 基于此, 光谱技术作为一种有效的化学分析方法被应用于医学领域, 主要是在分子水平上对生物组织进行分类识别^[3]. 目前, 已应用于医学诊断的光谱技术有拉曼光谱和红外光谱, 可对肝^[4]、胃^[5]、胰腺^[6]、肺^[7]及宫颈等器官^[8~12]的正常组织与非正常组织进行分类识别. 但拉曼散射的能量强度低, 对探测器灵敏性与精度要求较高; 红外光的穿透力有限, 无法实现生物组织较深部位的探测^[2]. X 射线光谱技术是一种成熟的分析方法, 具有能量高、穿透力强和无损检测等特点^[13]. X 射线光谱技术在医学、化学、材料学和物理学等领域都有着重要的应用^[14]. 因此, 将 X 射线光谱技术应用于生物组织的分类识别不失为一种辅助医学诊断和治疗的新方法.

本文使用 X 射线光谱检测系统采集新鲜的猪的心脏、肝、肾、胃、瘦肉和肥肉组织的入射 X 射线光谱和透射 X 射线光谱, 再通过数据预处理得到其 X 射线吸收谱(XAS), 并比较它们在光谱上的差异. 进一步建立了基于 XAS 的生物组织分类与识别方法, 先利用主成分分析法(PCA)对 XAS 数据进行特征提取, 再利用径向基函数(RBF)神经网络进行分类预测. 对实验结果进行分析发现, 利用 X 射线光谱技术并结合 PCA 方法和 RBF 神经网络可以较好地对猪的各种组织进行辨识.

1 实验部分

1.1 仪器与材料

实验采用由本实验室设计开发的 X 射线光谱检测系统^[15], 如图 1 所示, 其中核心部件为 X 射线管和 X 射线探测器. X 射线管采用上海科颐维电子科技有限公司生产的 KYW800 型正负电压 X 射线管, 额定功率为 80 W, 阳极电压上限为 80 kV, 阳极靶材料为钨(⁷⁴W); X 射线探测器选用美国 AmpTek 公司生产的 X-123 型 CdTe 探测器, 能量分辨率为 850 eV(FWHM)@122 keV, ⁵⁷Co. 实验过程中, X 射线管电压设置为 55 kV, 管电流设置为 10 μ A, 灯丝电流设置为 1000 mA, X 射线探测器采集

收稿日期: 2017-12-18. 网络出版日期: 2018-06-21.

基金项目: 国家自然科学基金(批准号: 61571381)资助.

联系人简介: 方正, 男, 博士, 教授, 主要从事生物医学和高端科学仪器开发方面的研究. E-mail: fangzheng@xmu.edu.cn

时间为 60 s.

实验材料为实验当天购买的新鲜猪心、猪肝、猪肾、猪胃、猪瘦肉和猪肥肉 6 类材料.

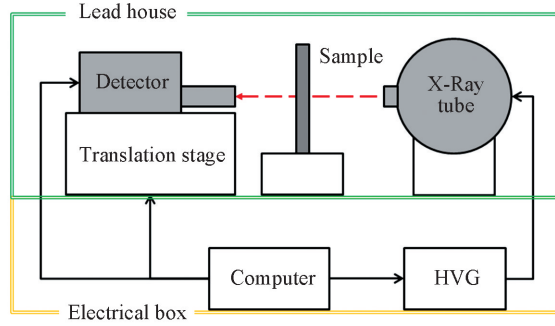


Fig.1 Schematic diagram of X-ray spectral detection system

1.2 实验过程

对于 6 类实验材料, 每类材料制备 3 个厚度约为 1 cm 的样本, 每个样本采集 25 次, 每次采集样本不同部位的光谱数据, 共得到 450 组光谱数据.

数据分析包括预处理、分类和评估 3 部分. 预处理过程包括计算吸收系数与归一化. 当 X 射线穿透物质时, 其强度会发生衰减, 衰减程度取决于被穿透的物质^[16]. 假设入射强度为 I_0 的 X 射线经过厚度为 d 的物质时, 透射 X 射线强度 I 满足下式的衰减关系:

$$I = I_0 e^{-\mu d} \quad (1)$$

式中: μ 为该物质的 X 射线衰减系数, 是一个常数; d 为样品厚度. μ 由散射系数和吸收系数两部分组成. 但由于 X 射线是一种波长极短的高能电磁波, 具有极强的穿透性. 所以散射系数比吸收系数小得多, 因此通常将散射系数忽略不计, 直接将 μ 称为 X 射线吸收系数^[14].

在连续能量的 X 射线穿透样品的过程中, 样品对不同能量 X 射线的吸收程度不同, X 射线吸收谱 (XAS) 表示的是物质对于不同能量 X 射线的 μ 值的变化. 实验中采集到的原始数据为入射 X 射线光谱和透射 X 射线光谱, 因此由式(1)推导出下式, 可以得到样品的 XAS:

$$\mu = \frac{1}{d} \ln \frac{I_0}{I} \quad (2)$$

在实验过程中, 制备样本时无法满足样本厚度精确为 1 cm, 故每次 X 射线穿透的样本厚度会有所不同, 导致每次采集的谱线在幅度上有一定的区别. 因此对由式(2)得到的 XAS 数据进行 0-1 归一化处理, 以消除每次数据采集时样本被穿透厚度不同所带来的影响.

分类过程包括利用 PCA 提取光谱特征和利用 RBF 神经网络进行分类. PCA 的作用是估算样本数据的协方差矩阵或自相关矩阵, 通过求解特征方程得到各个主成分方向. 选择适当数目的主成分作为样本的新特征, 从而达到特征提取和数据降维的效果, 再将样本投影到这些主成分方向上进行分类^[17]. 人工神经网络是模拟生物神经网络进行数据分析处理的数学模型, 也简称神经网络. 人工神经网络最突出的特点是并行分布式处理、自学习功能和非线性处理^[18]. 本文采用 RBF 神经网络对主成分分析结果进行分类预测.

RBF 神经网络通常由一个输入层、一个以非线性径向基函数为激发函数的隐含层和一个线性输出层组成^[19]. 输入层神经元个数等于输入维度数, 即特征个数, 输出层神经元个数为样本种类数. 隐含层神经元个数等于训练集样本个数, 网络权值和阈值由根据训练集建立的线性方程组直接解出, 无需迭代过程. 与传统常用的误差反向传播 (BP) 神经网络相比, RBF 神经网络的优点在于自组织、自适应功能, 更加适合小样本的学习情况, 克服了 BP 神经网络存在的局部极小值问题. 在隐层节点数相同的情况下, RBF 神经网络在逼近能力、学习速度及分类效果等方面均优于 BP 神经网络^[20].

评估过程是指对分类结果的效果进行评估的过程. 本文采用正确率、拒绝率、阳性预测值、阴性预测值、准确性及受试者工作特征曲线^[21]来评估分类效果, 从而减少评估的随意性和主观性. 正确率

(Sensitivity)和拒绝率(Specificity)定义如下:

$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN}) \quad (3)$$

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP}) \quad (4)$$

式中:对于某一类样本,TP指正确识别次数,即该类样本被识别为该类的次数;FP指错误识别次数,即不是该类的样本被识别为该类的次数;TN指正确拒绝次数,即不是该类的样本未被识别为该类的次数;FN指错误拒绝次数,即该类样本未被识别为该类的次数。

阳性预测值(PPV)、阴性预测值(NPV)和准确性(Accuracy)分别定义如下:

$$\text{PPV} = \text{TP} / (\text{TP} + \text{FP}) \quad (5)$$

$$\text{NPV} = \text{TN} / (\text{TN} + \text{FN}) \quad (6)$$

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (7)$$

最后,再利用受试者工作特征曲线(ROC曲线)评估分类结果.ROC曲线可由Sensitivity和Specificity计算获得,ROC曲线下区域面积(AUC)可以简单、直观地展示分类结果的好坏^[22].若分类器具有完美的分类效果,则AUC=1;若分类器完全不具备分类效果,则AUC≤0.5.ROC曲线越靠近左上方则说明AUC越大,分类效果越好。

2 结果与讨论

2.1 生物组织的X射线吸收光谱

在不放置样本的情况下,测得的人射X射线光谱如图2所示,纵坐标为入射X射线的强度.实验中X射线管的阳极靶材料为钨(⁷⁴W),该光谱即是钨在55 kV电压激发下发射的L系特征X射线光谱,包括L_{α1}, L_{β1}和L_{γ1}3个特征峰,分别出现在能量为8.396, 9.670和11.283 keV处.本实验所有生物样本的XAS数据均在该入射光谱条件下获得。

入射X射线光谱经生物样本吸收后的透射X射线光谱如图3所示.通过入射光与透射光可以计算得到样本对入射光的吸收程度,再经归一化处理得到的样本归一化XAS如图4所示.图4横坐标为X射线能量,能量范围为8~45 keV,纵坐标为归一化后的X射线吸收系数.可见,不同种类生物样本的XAS的差异性主要出现在能量11.283 keV后连续的10 keV内,其中猪肝的XAS与其它5类样本的光谱差异最大,其它5类样本之间的差异较小。

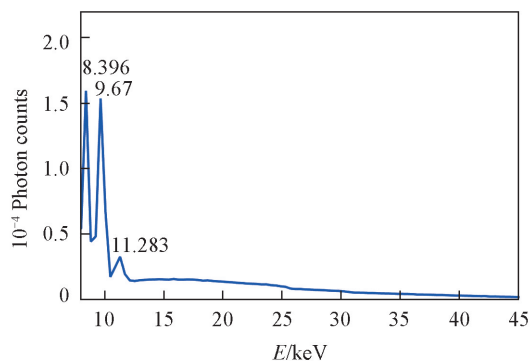


Fig.2 Incident X-ray spectrum emitted by X-ray tube under the excitation voltage of 55 kV

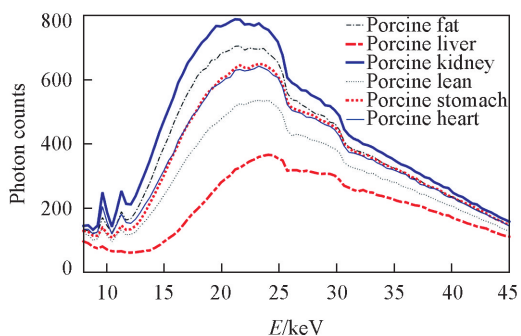


Fig.3 Transmission X-ray spectra after sample absorption

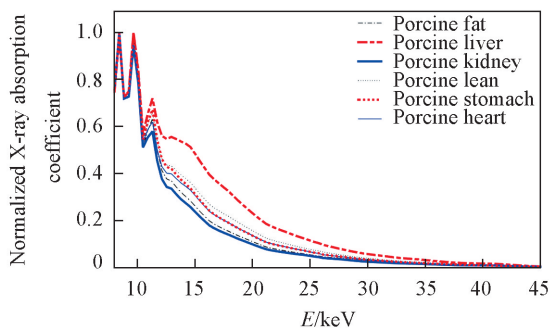


Fig.4 Normalized XAS of biological tissues

2.2 光谱的分类预测

利用PCA对预处理后的450组生物组织归一化XAS(以下简称XAS)进行特征提取计算,前10个

主成分的贡献率如表 1 所示. 贡献率指主成分解释原始信息的能力, 因此 PC1 解释了 XAS 中 92.39% 的信息, 而前十主成分几乎解释了 XAS 中的所有信息. 其中 PC1 与 PC2 的分布如图 5 所示, 可见不同组织的样本可以较好地聚集在一定的范围内. 其中, 猪肝样本的聚集效果最好, 未与其它样本发生重叠, 而其它 5 类样本都发生了与不同类样本间的重叠. 结合各主成分的贡献率大小, 本文选用 XAS 的前 3 个主成分作为特征参数进行下一步的辨识计算, 在实际的分类计算过程中也发现, 取前 3 个主成分进行计算可以达到最佳的分类效果.

Table 1 Contribution of the first ten principle components of normalized XAS

Principle component	Contribution (%)	Cumulative contribution (%)	Principle component	Contribution (%)	Cumulative contribution (%)
PC1	92.388	92.388	PC6	0.251	98.892
PC2	3.737	96.126	PC7	0.186	99.078
PC3	1.570	97.696	PC8	0.160	99.238
PC4	0.590	98.286	PC9	0.124	99.362
PC5	0.355	98.641	PC10	0.105	99.467

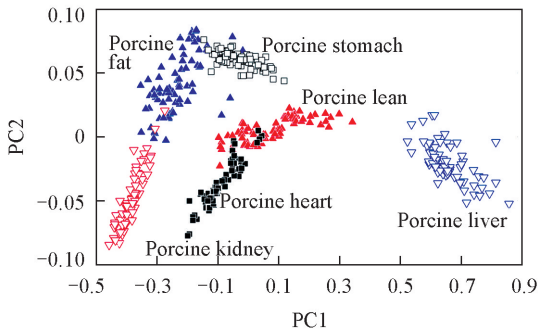


Fig.5 PC1 vs. PC2 of normalized XAS

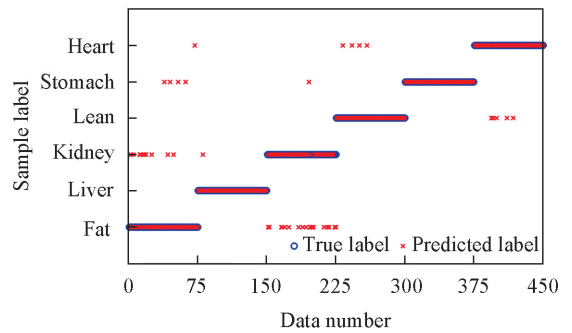


Fig.6 Prediction results of RBF neural network trained by the first three principal components

再引入 RBF 神经网络并使用交叉验证法依次对所有样本进行辨识. 每次选取每类组织中 2 个样本的 50 组数据作为训练集, 另一个样本的 25 组数据作为测试集, 因此训练集大小为 300 组, 测试集大小为 150 组. 对训练集进行 PCA 计算, 提取训练集的前 3 个主成分并得到对应的变换系数矩阵, 利用该变换系数矩阵提取测试集的 3 个特征. 再根据神经网络输入输出的维度与训练集的大小, 建立结构为 3-300-6 的 RBF 神经网络结构. 利用训练集的前 3 个主成分训练 RBF 神经网络, 并利用训练完成的 RBF 神经网络对测试集进行辨识. 如此重复 3 次完成对每类组织 3 个样本 75 组数据的分类预测, 预测结果如图 6 所示. 对 450 组数据进行识别的过程中共出现 44 例误判情况, 总识别率为 90.22%, 其中猪肥肉与猪肾之间的相互误判情况比较严重, 此外猪肥肉还出现了部分被误判为猪胃的情况, 猪瘦肉与心之间也出现少量相互误判情况.

2.3 预测结果评估

6 类生物组织样本预测结果评估参数统计如表 2 所示. 可见, 所有类别的预测准确率均达到 90% 以上. 猪肾的正确率最低, 75 组数据仅有 56 组预测正确, 出现 19 组误判. 猪肥肉样本的 PPV 最低,

Table 2 Statistical parameters calculated for prediction results of all classes

No.	Sample name	TP	TN	FP	FN	Sensitivity	Specificity	PPV	NPV	Accuracy
1	Porcine fat	60	357	18	15	0.8000	0.9520	0.7692	0.9597	0.9267
2	Porcine liver	74	375	0	1	0.9867	1.0000	1.0000	0.9973	0.9978
3	Porcine kidney	56	364	11	19	0.7467	0.9707	0.8358	0.9504	0.9333
4	Porcine lean	71	370	5	4	0.9467	0.9867	0.9342	0.9893	0.9800
5	Porcine stomach	75	370	5	0	1.0000	0.9867	0.9375	1.0000	0.9889
6	Porcine heart	70	370	5	5	0.9333	0.9867	0.9333	0.9867	0.9778

预测为猪肥肉的样本仅有 76.92% 的概率是实际属于猪肥肉的. 分析其原因是猪肥肉与猪肾的光谱具有较高的相似性, 导致它们的主成分有重叠部分, 从而在分类时出现了互相误判的情况.

预测结果的 ROC 曲线如图 7 所示, 图例括号中的数字代表曲线的 AUC 值. 由图 7 可见, 猪肝的分类效果最佳, AUC 值达到 1; 猪胃、猪瘦肉和猪心的分类效果也较好, AUC 值分别为 0.9921, 0.9895 和 0.9753; 猪肥肉和猪肾的分类效果相对一般, AUC 值分别为 0.9407 和 0.9322.

3 结 论

通过对比猪肥肉、瘦肉、心、肝、肾和胃组织的 XAS 数据, 发现光谱差异性主要集中在能量 11.283 keV 后连续的 10 keV 内. 利用主成分分析法提取光谱特征, 并利用 RBF 神经网络进行分类预测. 实验结果表明, 利用 XAS 能够对猪的组织样本进行分类, 并获得较好的分类效果, 总体识别率达到 90.22%. 本文结果对将 XAS 技术应用于生物组织的自动识别领域具有重要的指导意义. 进一步的工作会优化实验设备与方法, 提升光谱数据的稳定性, 后续将对小鼠的正常与致病组织进行实验, 并结合计算机断层扫描(CT)图像重建的方法^[23]进行光谱 CT 重建与自动识别, 以达到无损检测并识别生物体内部组织光谱的目的.

参 考 文 献

- [1] Zhang Z. X., *Acta Laser Biology Sinica*, **2004**, (3), 159—160(张镇西. 激光生物学报, **2004**, (3), 159—160)
- [2] Manoharan R., Wang Y., Feld M S., *Spectrochim. Acta A*, **1996**, 52(2), 215—249
- [3] Siqueira L. F. S., Araujo R. F., De Araujo A. A., Morais C. L. M., Lima K. M. G., *Chemometr. Intell. Lab.*, **2017**, 162, 123—129
- [4] Lambrakos S. G., Trzaskoma-Paulette P. P., *Appl. Spectrosc.*, **2000**, 54(2), 305—315
- [5] Yi W. S., Cui D. S., Li Z., Wu L. L., Shen A. G., Hu J. M., *Spectrochim. Acta A*, **2013**, 101, 127—131
- [6] Cheng C. G., Tian Y. M., Zhang C. J., *Chinese J. Anal. Chem.*, **2008**, 36(8), 1051—1055(程存归, 田玉梅, 张长江. 分析化学, **2008**, 36(8), 1051—1055)
- [7] Bard M. P. L., Amelink A., Hegt V. N., Graveland W. J., Sterenborg H. J. C. M., Hoogsteden H. C., Aerts J. G. J. V., *Am. J. Resp. Crit. Care*, **2005**, 171(10), 1178—1184
- [8] Liu G., Liu J. H., Zhang L., Yu F., Sun S. Z., *Spectrosc. Spect. Anal.*, **2005**, 25(5), 723—725(刘刚, 刘剑虹, 张林, 俞帆, 孙世中. 光谱学与光谱分析, **2005**, 25(5), 723—725)
- [9] Zhang H. P., Zheng C., Lu L., Zou Y. B., Wang C. H., Xu S. P., Xu W. Q., Fan Z. M., Lu L. J., Han B., *Chem. J. Chinese Universities*, **2014**, 35(9), 1877—1882(张海鹏, 郑超, 路璐, 邹亚斌, 王翠花, 徐抒平, 徐蔚青, 范志民, 路来金, 韩冰. 高等学校化学学报, **2014**, 35(9), 1877—1882)
- [10] Rehman S., Movasaghi Z., Tucker A. T., Joel S. P., Darr J. A., Ruban A. V., Rehman I. U., *J. Raman Spectrosc.*, **2007**, 38(10), 1345—1351
- [11] Cicchi R., Anand S., Crisci A., Giordano F., Rossari S., *Proceeding of SPIE*, **2015**, 9537, 95370I
- [12] Liu H. Y., Gu W. Q., Li Q. L., Wang Y. T., Chen Z. G., Xu Q. T., *Spectrosc. Spect. Anal.*, **2015**, 35(1), 38—43(刘洪英, 顾文荃, 李庆利, 王依婷, 陈增溢, 徐沁同. 光谱学与光谱分析, **2015**, 35(1), 38—43)
- [13] Savin S. L. P., Berko A., Blacklocks A. N., Edwards W., Chadwick A. V., *C. R. Chim.*, **2008**, 11(9), 948—963
- [14] Ma L. D., *Shanghai Measurement and Testing*, **2007**, 34(6), 2—11(马礼敦. 上海计量测试, **2007**, 34(6), 2—11)
- [15] Hu B., Zhang X. L., Ouyang Q. N., Wu X. M., Fang Z., *Measurement*, **2016**, 93, 252—257
- [16] Yano J., Yachandra V. K., *Photosynth. Res.*, **2009**, 102, 241—254
- [17] Jolliffe I. T., Cadima J., *Philos. T. R. Soc. A*, **2016**, 374(2065), 20150202

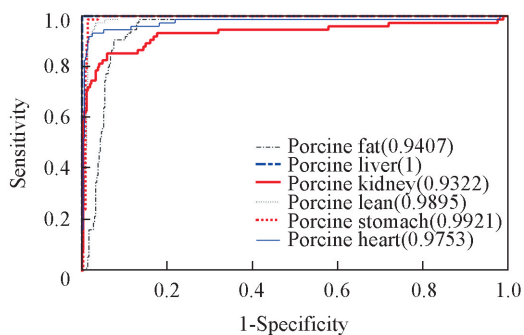


Fig.7 ROC curves for prediction results of all classes

- [18] Schmidhuber J., *Neural Networks*, **2014**, 61, 85—117
- [19] Schwenker F., Kestler H. A., Palm G., *Neural Networks*, **2001**, 14(4/5), 439—458
- [20] Ramedani Z., Omid M., Keyhani A., Shamshirband S., Khoshnevisan B., *Renew. Sust. Energ. Rev.*, **2014**, 39, 1005—1011
- [21] Fawcett T., *Pattern Recogn. Lett.*, **2006**, 27(8), 861—874
- [22] Unnikrishnan V. K., Choudhari K. S., Kulkarni S. D., Nayak R., Kartha V. B., Santhosh C., *RSC Adv.*, **2013**, 3(48), 25872—25880
- [23] Zhang Z., Han X., Pearson E., Pelizzari C., Sidky E. Y., Pan X. C., *Phys. Med. Biol.*, **2016**, 61(9), 3387—3406

Biological Tissue Recognition Based on X-Ray Absorption Spectral Detection[†]

WANG Qian, YANG Zheng, FANG Zheng*

(Department of Instrumental and Electrical Engineering, Xiamen University,
Xiamen 361005, China)

Abstract In order to verify the ability of X-ray absorption spectroscopy to recognize biological tissues, porcine heart, liver, kidney, stomach, lean meat and fat are selected as specimens. The X-ray absorption spectra of these specimens are obtained by X-ray detector at the excitation voltage of 55 kV. The collected spectra are divided into training set and test set, and principal component analysis is used to extract spectral principal component. Using the training set as input, a radial basis function(RBF) neural network model is established to predict the samples of the test set. The recognition rate of all samples through cross-validation method reached 90.22%. The experimental results show that X-ray spectral technique coupled with statistical analysis methods can be used for the classification of pig tissue, which is of great significance to the application of X-ray spectral technique in organism identification.

Keywords X-Ray absorption spectrum; Biological tissue; Material recognition; Radial basis function(RBF) neural network

(Ed.: N, K)

[†] Supported by the National Natural Science Foundation of China(No.61571381).