

· 论 著 ·



李庆利 华东师范大学信息科学技术学院上海市多维度信息处理重点实验室副主任，教授，博士生导师。从事医学光谱成像机理及其生物医学应用相关研究，师从中国科学院上海技术物理研究所薛永祺院士。曾先后成功研制出推帚式显微高光谱成像系统、凝采式分子高光谱成像系统并将其应用于生物组织样本的检测分析。首次实现未染色神经束的分析，最早提出高光谱舌诊概念，将舌诊现代化推广到多维检测领域。主持国家自然科学基金面上项目 2 项、青年项目 1 项，教育部博士点基金 1 项，华东师范大学科研创新基金 1 项；作为子课题负责人参与上海市基础重点项目、国家重点基础研究发展计划（“973”计划）、上海市基础重点项目、国家自然科学基金面上项目、上海市科技成果转化和产业化项目。发表 SCI 或 EI 收录论文 60 余篇，编写专著 1 章节，已获专利授权 11 项。获 2015 年度教育部技术发明奖二等奖、上海市技术发明奖二等奖。现为 *Int J Comput Ass Rad*、*Recent Pat Med Imaging* 期刊编委，*IEEE Trans Med Imaging*、*IEEE Trans Biomed Eng*、*IEEE J Biomed Health Inform*、*Biosens Bioelectron* 等多本 SCI 收录期刊和 ICDIP 2018 等国际会议论文集审稿专家。担任 The 9th International Congress on Image and Signal Processing 等国际会议大会主席。

DOI: 10.16781/j.0258-879x.2018.08.0886

基于神经网络的显微高光谱乳腺癌组织图像研究

郑欣^{1,2}, 周梅¹, 孙力¹, 邱崧^{1,2}, 于观贞³, 李庆利^{1*}

1. 华东师范大学多维度信息处理上海市重点实验室, 上海 200241
2. 华东师范大学空间信息与定位导航上海市高校工程技术研究中心, 上海 200241
3. 上海中医药大学附属龙华医院肿瘤七科, 上海 200032

[摘要] **目的** 探索神经网络结合显微高光谱成像识别乳腺癌组织的可行性和应用价值。**方法** 采用显微高光谱成像技术采集乳腺癌组织的图像数据，使用基于神经网络的显微高光谱乳腺癌组织图像分析方法，实现乳腺癌组织的自动分类和区域划分。提出数据预处理方法以提高图像的信噪比，利用神经网络训练图谱信息识别乳腺组织病变区域并突显以利于可视化。**结果** 基于神经网络的显微高光谱的乳腺组织识别分析方法同时利用了图谱两个方面的特征，获得了比传统彩色病理图像更好的识别结果。**结论** 基于神经网络的显微高光谱乳腺组织图像分析方法可以提供特征性的样本信息，是传统彩色病理图像的有效补充。在神经网络分析方法的支持下，将显微高光谱成像技术应用于乳腺癌组织的分析具有一定的应用前景。

[关键词] 乳腺肿瘤；神经网络；显微高光谱图像；组织病理学

[中图分类号] R 737.9 **[文献标志码]** A **[文章编号]** 0258-879X(2018)08-0886-06

Micro-hyperspectral breast cancer tissue image analysis based on neural network

ZHENG Xin^{1,2}, ZHOU Mei¹, SUN Li¹, QIU Song^{1,2}, YU Guan-zhen³, LI Qing-li^{1*}

1. Shanghai Key Laboratory of Multidimensional Information Processing, East China Normal University, Shanghai 200241, China
2. Engineering Center of SHMEC for Space Information and GNSS, East China Normal University, Shanghai 200241, China
3. Department of Oncology (VII), Longhua Hospital Affiliated to Shanghai University of Traditional Chinese Medicine, Shanghai 200032, China

[Abstract] **Objective** To explore the feasibility and value of neural network combined with micro-hyperspectral imaging in identifying breast cancer tissue. **Methods** The micro-hyperspectral imaging technology was used to collect image data of breast cancer tissue, and the micro-hyperspectral breast tissue image analysis method based on neural network was used to realize the automatic classification and regional division of breast cancer tissue. Meanwhile, data preprocessing

[收稿日期] 2018-06-21 **[接受日期]** 2018-07-26

[基金项目] 国家自然科学基金(61377107). Supported by National Natural Science Foundation of China (61377107).

[作者简介] 郑欣, 硕士生. E-mail: zxxxxx416@163.com

*通信作者(Corresponding author). Tel: 021-54345199, E-mail: qlli@cs.ecnu.edu.cn

method was proposed to improve the signal to noise ratio of the image, and map information was trained by neural network to identify breast tissue lesions and highlight them for visualization. **Results** The micro-hyperspectral breast tissue image analysis method based on neural network utilized two characteristics of the images at the same time, and it was better than traditional color pathological images in identifying breast tissue. **Conclusion** The micro-hyperspectral breast tissue image analysis method based on neural network can provide more characteristic sample information compared with traditional color pathology images, and may serve as an effective complement to traditional color pathological images. With the support of neural network, the micro-hyperspectral imaging technology has prospects in analyzing breast cancer tissue.

[Key words] breast neoplasms; neural networks; micro-hyperspectral image; histopathology

[Acad J Sec Mil Med Univ, 2018, 39(8): 886-891]

近年来乳腺癌的发病率和死亡率逐年上升,严重影响女性身心健康,其早期诊断和早期治疗尤为重要^[1]。随着影像诊断技术的飞速发展,X线、计算机断层扫描(computed tomography, CT)、磁共振成像(magnetic resonance imaging, MRI)等的使用不仅可以提高乳腺癌的诊断准确率,也在一定程度上减轻了患者诊断时的痛苦^[2-3]。计算机技术和图像处理技术的日趋成熟使计算机影像诊断技术在病理诊断中的应用前景广阔,也使其成为当下病理研究中的一个重要方向^[4]。

乳腺癌的临床确诊主要是病理医师根据组织切片中的组织结构和细胞形态做出最终诊断,病理诊断是乳腺癌确诊的金标准。诊断结果的准确性依赖医师的经验,需要临床医师长期培训和学习,人力和物力成本都比较高。随着图像识别、深度学习^[5-6]、神经网络^[6]等关键技术的突破,“人工智能+医疗”的概念应运而生,其在自诊断、高效率、低成本等方面有着极大的竞争力。由于具有自学习、非线性映射、可塑性高等优点,神经网络在各个领域应用广泛,其在生物医学领域也取得了具有里程碑意义的成果,如皮肤癌诊断^[7]、预测阿尔茨海默病^[8]等。现有的生物组织病理诊断主要是基于灰度图像或彩色图像,其在计算机中表现出的信息有限,导致对生物组织的分析算法要求较高。

显微高光谱成像技术结合了显微成像技术和高光谱技术,成像图像不仅包含目标的空间信息,还包含了光谱信息^[9],其为生物医学临床诊断提供了一种新的有效的辅助手段^[10]。本研究结合组织病理学和计算机视觉,使用反向传播(back propagation, BP)神经网络对乳腺癌组织的显微高光谱图像进行识别分析,发现该技术能够区分组织切片中的肿瘤组织和非肿瘤组织,获得了较为准确的分类结果。

1 材料和方法

1.1 研究对象 乳腺癌患者 30 例,均经病理证实,均为女性,年龄为 42~82 岁,平均(54.48±12.75)岁。切取每例患者的病灶组织,分割为 5 mm×15 mm×15 mm 大小,使用 10% 甲醛固定液进行组织固定。使用 75%、85%、95% 和 100% 的乙醇进行梯度脱水脱水处理,而后将组织块置于二甲苯中透明,用石蜡包埋机将组织标本包埋成病理石蜡块,用全自动石蜡切片机制成 4 μm 的病理组织切片,行苏木精-伊红染色(H-E 染色)、脱水、透明,以获得乳腺癌组织样本。使用自主研发的显微高光谱成像系统(MHSI)^[11]采集组织样本得高光谱图像 90 组,选取其中 72 组(占总数的 80%)作为训练样本,另外 18 组作为测试样本,采用 5 次交叉验证,进行后续分析。

1.2 显微高光谱成像及预处理 获取图像和光谱信息的高光谱成像技术将由物质成分决定的光谱与反映物质存在格局的影像完整结合起来,其中每张图像都包含了物质在不同波长下的信息^[9,11]。高光谱立方体数据如图 1 所示,其中空间维 $M \times N$ 代表每张单波段图像的大小,光谱维 K 表示采集的波段总数,在每个波段下都有 1 幅图像。本研究采用的高光谱成像系统的光谱范围为 373~809 nm,显微镜目镜倍数为 10 倍(固定),物镜倍数为 10 倍(可调节),图像数据空间分辨率为 1 800×1 300,总波段数为 60。

高光谱图像的质量受系统光源、光学器件、电流噪声等的影响。由于生物样本对光谱的吸收特性相对较小,未处理的图像不能反映生物组织的特异性,图像的预处理不可或缺。根据高光谱图像的性质,预处理的关键是在空间和光谱维度上分别进行校正^[11]。平行单色入射光依次经过盖玻片、生物组织和载玻片,定义透光率 $T(m, n; \lambda)$ 为:

$$T(m, n; \lambda) = \frac{D(m, n; \lambda) - I(m, n; \lambda)}{B(m, n; \lambda) - I(m, n; \lambda)}$$

其中 $D(m, n; \lambda)$ 为生物组织的第 λ 波段图像的

n 行 m 列的像素点值, $B(m, n; \lambda)$ 为空白区域的第 λ 波段图像的 n 行 m 列的像素点值, $I(m, n; \lambda)$ 为该像素点在关闭光源情况下的噪声。

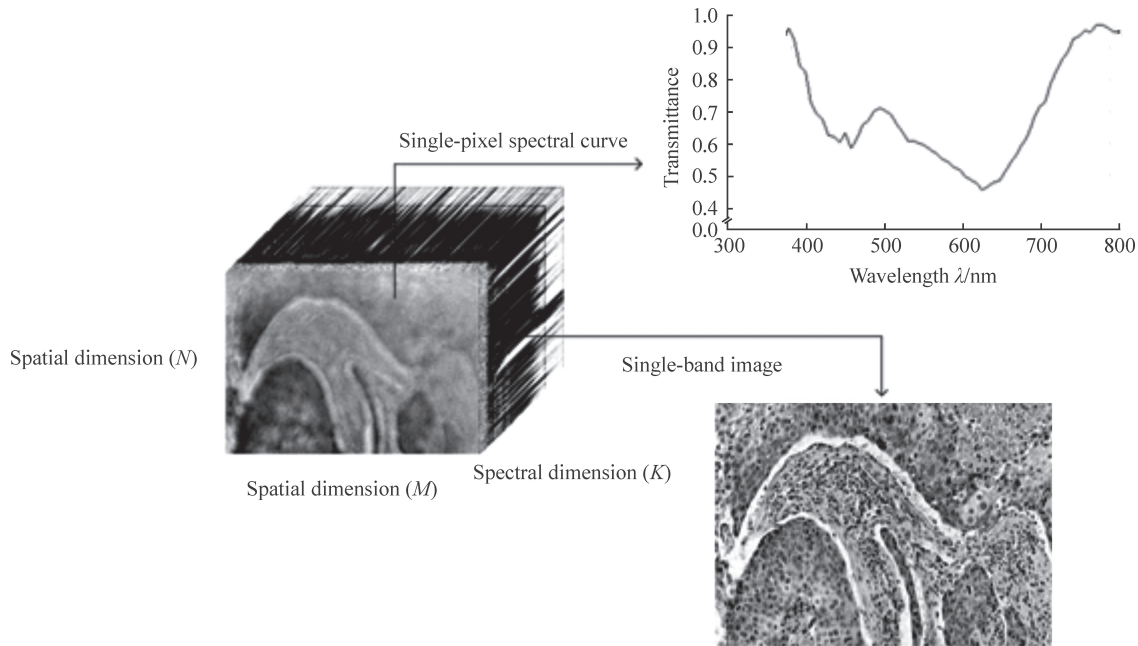


图1 高光谱立方体数据

Fig 1 Hyperspectral data cube

1.3 BP 神经网络 BP 神经网络是一种按照反向传播算法训练的多层前馈神经网络^[6,12], 由于具有很强的非线性映射能力和自适应能力, 它在文字识别^[13]、语音识别^[14]、图像识别^[15]等领域都有着广泛的应用。其结构如图2所示, 一般由输入层、隐含层和输出层组成, 其训练过程就是不断更新各层神经元的参数。以3层BP神经网络为例, 输入向量为 $X=(X_1, \dots, X_n)^T$, 输出向量为 $Y=(Y_1, \dots, Y_n)^T$, 输入层与隐含层的连接权值为 w_{ih} , 隐含层与输出层的连接权值为 w_{ho} , 隐含层各神经元的阈值为 b_h , 输出层各神经元的阈值为 b_o , $D=(D_1, \dots, D_q)^T$ 为期望的输出向量, 误差函数 $e=1/2\sum_i(D_i - Y_i)^2$ 。训练的具体过程如下^[13]。

(1) 选取范围在 $(-1, 1)$ 内的随机数初始化误差函数 e , 权值 w_{ih} 、 w_{ho} 。

(2) 随机选取第 k 组输入样本 X_i 和期望输出 D_i , 并重复以下步骤直到误差达到要求。

(3) 计算隐含层各神经元的输入和输出:

$$h_i(k) = \sum_{j=1}^n (w_{ij} X_j - b_h)^2$$

$$h_o(k) = f(h_i(k))$$

$$y_i(k) = \sum_{h=1}^q (w_{ho} h_o(k) - b_o)^2$$

$$y_o(k) = f(y_i(k))$$

其中 h_i 是隐含层的输入, h_o 是隐含层的输出, y_i 是输出层的输入, y_o 是输出层的输出, $f(\cdot)$ 是激活函数, 常用的激活函数有 sigmoid、tanh 函数等。

(4) 计算误差函数 e 对输出层各神经元的偏导数 $\Delta w_{ih}(k)$, 并利用偏导数更新隐含层的连接值:

$$w_{ih}^{N+1} = w_{ih}^N + \eta \Delta w_{ih}(k)$$

其中 η 为学习效率, 其值越大训练速度越快, 但过大可能导致训练结果不收敛。

(5) 计算 n 个样本 (输入)、 q 个类别 (输出) 的全局误差:

$$E = 1/2 \sum_{k=1}^n \sum_{o=1}^q (D_o(k) - y_o(k))^2$$

(6) 从训练样本中继续选取 1 组输入和期望输出, 不断更新各节点的权值和阈值, 直至全局误差小于预先设定的一个值, 也就是达到收敛状态。

BP 神经网络的训练过程主要分为 2 个阶段, 第 1 阶段是信号的前向传播, 第 2 阶段是误差的反向传播, 依次调节隐含层到输出层、输入层到隐含层的权重和偏置, 使训练样本的输出和期望值在误差范围内接近, 通过自学习完成样本的分类。

应用 BP 神经网络对显微高光谱乳腺组织图像

中的不同区域进行分割, 提取感兴趣区域。将算法分割结果与人工分割结果进行比较, 包括每个区域正确的分类, 被错误分割成其他类以及未被正确分割成对应组织。以 T 代表肿瘤、 S 代表基质、 B 代表空白, 定义参数 TT (正确标记为肿瘤的像素点)、 TS (正确标记为基质的像素点)、 TB (正确标记为空白的像素点, 若样本不含空白区域, 该值标记为 0)、 FT (错误标记为肿瘤的像素点)、 FS (错误标记为基质的像素点)、 FB (错误标记为空白的像素点, 若样本不含空白区域, 该值标记为 0)。计算 BP 神经网络分割每组样本的准确率和精密度, 计算公式: 准确率 (%) = $(TT+TS+TB) / (TT+TS+TB+FT+FS+FB) \times 100\%$, 精密度 (%) = $TT / (TT+FT) \times 100\%$ 。

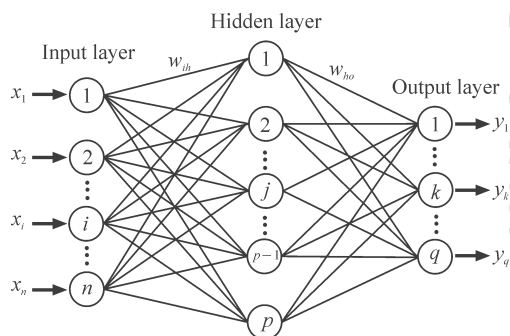


图 2 反向传播 (BP) 神经网络结构图

Fig 2 Structure of back propagation (BP) neural network

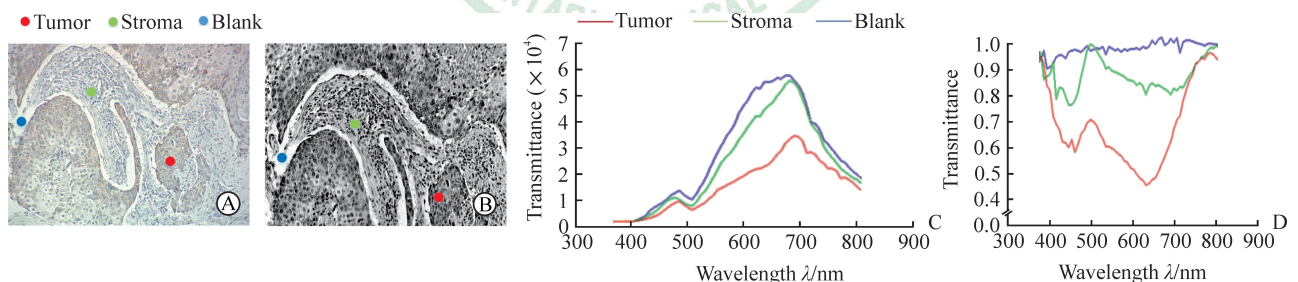


图 3 显微高光谱图像预处理

Fig 3 Preprocessing of micro-hyperspectral image

A: RGB image (H-E staining); B: 40-band hyperspectral grayscale image (H-E staining); C: Spectral curve of different tissues before preprocessing; D: Spectral curve of different tissues after preprocessing. Original magnification: $\times 100$ (A, B)

2.3 实验结果分析 预处理后通过 BP 神经网络完成组织分割。设计 3 层 BP 神经网络结构, 即只含有 1 个隐含层, 激活函数采用 logistic, 设置最大迭代次数为 1 000, 学习速率为 0.2, 输出个数为 3, 即三分类, 其结果如图 4 所示。第 40 波段下的高光谱灰度图像和 BP 神经网络分割结果分别见图 4A 和 4B, 其中红、绿、蓝色分别代表分割出

2 结果

2.1 实验环境 本实验的计算机操作系统为 Microsoft Windows 10, Professional (SP2), 处理器为 Intel®Core™i5-4590 CPU 3.30 GHz, 内存为 4 G。软件平台为 ENVI 5.1。首先对数据统一行预处理, 然后使用 ENVI 5.1 中的神经网络工具进行分析。

2.2 数据预处理 根据显微高光谱图像预处理方法处理采集到的图像数据, 实验结果如图 3 所示。图 3A 所示为系统采集的彩色图像, 包含感兴趣的乳腺癌肿瘤组织。由于高光谱系统采集彩色图像和高光谱图像使用的是两个摄像头, 所以在观察上会存在位置上的偏移, 但彩色图像不是本研究主要对象, 其作用是便于观察和对比。图 3B 所示为高光谱灰度图像中第 40 波段对应的图像, 该波段下噪声较小, 图片质量较高, 便于与分类结果进行对比。图中红、绿、蓝 3 种颜色标记分别表示肿瘤区域、基质区域和空白区域, 这 3 类组织是本研究的主要成分 (部分样本不存在空白区域)。通过观察图像预处理前后 3 类组织对应的光谱曲线可见, 预处理前各类组织的光谱曲线趋势类似 (图 3C), 这是由于系统噪声掩盖了生物的本来特性; 预处理削弱或消除了条带噪声等系统噪声, 各类组织的光谱曲线开始呈现出各自的特性 (图 3D), 为下一步神经网络特征学习奠定了基础。

来的肿瘤、基质和空白区域。与人工分割结果 (图 4C) 相比, BP 神经网络算法分割样本的准确率为 $(85.13 \pm 4.73)\%$, 精密率为 $(79.68 \pm 5.94)\%$ 。由于部分样本组织具有比较复杂的生物特性, 以及不同医师的人工分割结果有所不同, 都可能导致实验结果的误差, 故将原始图像与分割结果合成到 1 张图像上, 可见 BP 神经网络的分类效果比较理

想 (图 4D)。本研究中 BP 神经网络训练的均方误差 (root mean square, RMS) 不断趋近 0.2, 当

迭代次数达到 200 次后, 误差更新速率变缓, 具有良好的收敛性。见图 5。

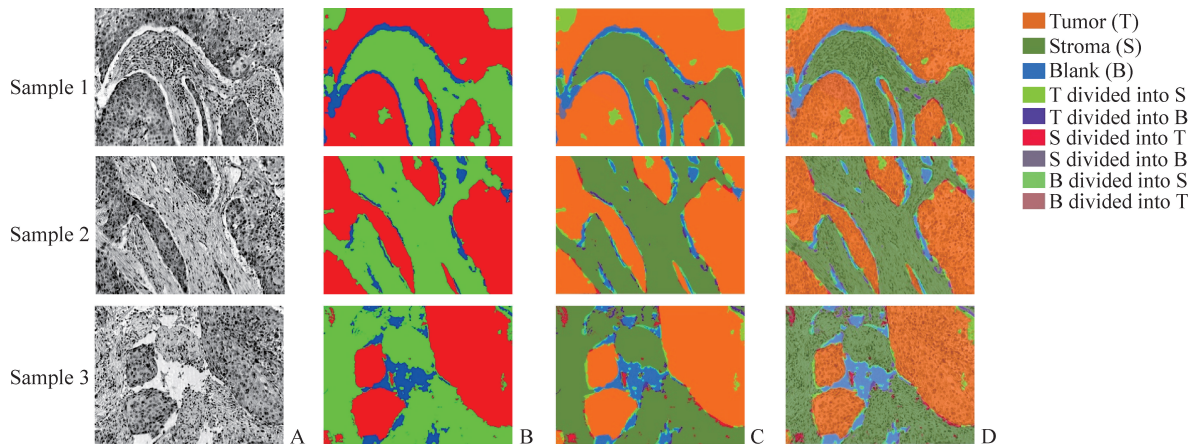


图 4 显微高光谱图像分割

Fig 4 Segmentation of micro-hyperspectral images

A: 40-band hyperspectral grayscale image; B: Segmentation result by back propagation (BP) neural network; C: Manual segmentation result; D: Composition of original image and segmentation result by BP neural network. Original magnification: × 100

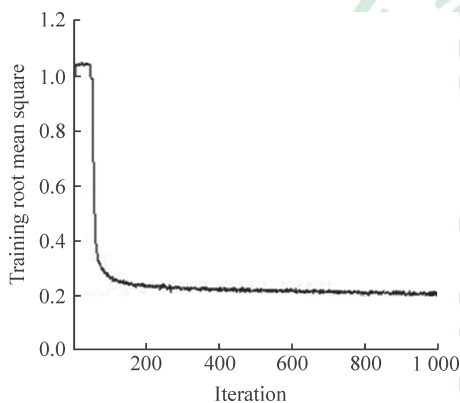


图 5 反向传播 (BP) 神经网络训练的误差变化

Fig 5 Training root mean square of back propagation (BP) neural network

3 讨论

乳腺癌是女性发病率最高的肿瘤, 随着医学技术的发展, 乳腺癌患者的治疗效果和总体生存率均得到改善。近年来, 随着计算机医学的发展, 人工智能技术辅助诊断和治疗在乳腺癌中也取得了长足进步。目前人工智能技术在乳腺癌领域的研究包括乳腺癌风险评估^[16]、计算机辅助诊断^[17]、疾病复发和预后模型^[18]、免疫分型^[19]、治疗指导^[20]和病理诊断^[21]等。医学影像和病理图像是乳腺癌人工智能技术进展最快的领域, 目前多项研究表明基于特定的算法和病理图片数字化处理, 通过训练和算法优化, 可开发出高精度、高效率的病理识别算法模型^[21-22]。

本研究旨在探求一种由神经网络与生物医学结合的图像分割方法。首先, 使用显微高光谱成像仪器采集乳腺癌样本图像; 然后对显微高光谱图像进行预处理, 消除系统中产生的一些干扰和噪声, 突显生物组织的特异性; 最后将预处理后的图像作为神经网络的输入, 经过自学习后得到组织分类的图像。神经网络的参数设置对分类结果有直接影响, 如学习速率过大虽然可以加速训练过程, 但也可能导致不收敛, 失去训练的意义, 这要求从大量样本中找到适合训练的框架。本实验结果证明了利用 BP 神经网络进行乳腺癌样本中各区域的分割, 避免了人工操作的耗时、耗力, 且具有良好的准确率, 可辅助病理医师进行精确的临床诊断。除临床病理诊断外, 该技术结合图像提取还有助于对肿瘤内部的不同特征进行分类, 以及对免疫组织化学染色进行提取和分类, 同时对上述特征予以量化分析。

本研究初步证明了神经网络结合生物医学图像在临床中的应用价值, 但本研究也存在不足, 如纳入病例数较少、肿瘤类型单一、缺乏免疫学指标等。因此, 需进一步展开后续研究: (1) 扩大样本量, 进一步证实和优化现有的算法; (2) 实现样本多样化, 纳入更多的病理类型, 包括导管内癌、小叶原位癌、导管内乳头状癌、乳头湿疹样乳腺癌、乳头状癌、髓样癌 (伴大量淋巴细胞浸润)、鳞状细胞癌、浸润性导管癌、浸润性小叶

癌等; (3) 加入乳腺癌分子分型, 纳入雌激素受体、孕激素受体、人表皮生长因子受体 2 和 Ki-67 等指标。生物医学图像与人工智能技术的有效结合, 必将促进医学发展, 造福患者和医师。

[参考文献]

- [1] VETA M, PLUIM J P, VAN DIEST P J, VIERGEVER M A. Breast cancer histopathology image analysis: a review[J]. *IEEE Trans Biomed Eng*, 2014, 61: 1400-1411.
- [2] ATIYA A,高燕,王培军. 乳腺癌 MRI 诊断进展[J]. 2018, 34:306-309.
- [3] 孙艳霞,丁敏. MRI、CT 及 X 线在乳腺癌术前诊断及治疗指导中的应用[J]. *中国 CT 和 MRI 杂志*,2017,15:43-46.
- [4] CHABI M L, BORGET I, ARDILES R, ABOUD G, BOUSSOUAR S, VILAR V, et al. Evaluation of the accuracy of a computer-aided diagnosis (CAD) system in breast ultrasound according to the radiologist's experience[J]. *Acad Radiol*, 2012, 3: 311-319.
- [5] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. *Nature*, 2015, 521: 436-444.
- [6] SCHMIDHUBER J. Deep learning in neural networks: an overview[J]. *Neural Net*, 2015, 61: 85-117.
- [7] ESTEVA A, KUPREL B, NOVOA R A, KO J, SWETTER S M, BLAU H M, et al. Dermatologist-level classification of skin cancer with deep neural networks[J]. *Nature*, 2017, 542: 115-118.
- [8] SUK H I, SHEN D. Deep learning-based feature representation for AD/MCI classification[J]. *Med Image Comput Comput Assist Interv*, 2013, 16(Pt 2): 583-590.
- [9] 李庆利,肖功海,薛永祺,张敬法. 基于显微高光谱成像的人血细胞研究[J]. *光电工程*,2008,35:98-101.
- [10] FREEMAN J, DOWNS F, MARCUCCI L, LEWIS E N, BLUME B, RISH J. Multispectral and hyperspectral imaging: applications for medical and surgical diagnostics[C]//19th Annual International Conference of the IEEE Engineering in Medicine & Biology Society. IEEE, 2002: 700-701.
- [11] 刘洪英,李庆利,顾彬,王依婷,薛永祺. 新型分子高光谱成像系统性能分析及数据预处理[J]. *光谱学与光谱分析*,2012,32:3161-3166.
- [12] ZHENG Q, MILTHORPE B K, JONES A S. Direct neural network application for automated cell recognition[J]. *Cytometry A*, 2004, 57: 1-9.
- [13] 苗琦龙,栾新. 基于遗传算法和 BP 网络的文字识别方法[J]. *计算机应用*,2005,25:330-332.
- [14] 戚龙,赵丹. 基于 BP 神经网络的非特定人语音识别算法[J]. *科学技术与工程*,2017,17:277-281.
- [15] 杨治明,王晓蓉,彭军,陈应祖. BP 神经网络在图像分割中的应用[J]. *计算机科学*,2007,34:234-236.
- [16] SEPANDI M, TAGHDIR M, REZAIANZADEH A, RAHIMIKAZEROONI S. Assessing breast cancer risk with an artificial neural network[J]. *Asian Pac J Cancer Prev*, 2018, 19: 1017-1019.
- [17] BURT J R, TOROSDAGLI N, KHOSRAVAN N, RAVIPRAKASH H, MORTAZI A, TISSAVIRASINGHAM F, et al. Deep learning beyond cats and dogs: recent advances in diagnosing breast cancer with deep neural networks[J/OL]. *Br J Radiol*, 2018: 20170545. doi: 10.1259/bjr.20170545.
- [18] SAHAA, HAROWICZ MR, WANG W, MAZUROWSKI M A. A study of association of oncotype DX recurrence score with DCE-MRI characteristics using multivariate machine learning models[J]. *J Cancer Res Clin Oncol*, 2018, 144: 799-807.
- [19] THAKUR S S, LI H, CHAN A M Y, TUDOR R, BIGRAS G, MORRIS D, et al. The use of automated Ki67 analysis to predict oncotype DX risk-of-recurrence categories in early-stage breast cancer[J/OL]. *PLoS One*, 2018, 13: e0188983. doi: 10.1371/journal.pone.0188983.
- [20] SOMASHEKHAR S P, SEPÚLVEDA M J, PUGLIELLI S, NORDEN A D, SHORTLIFFE E H, ROHIT KUMAR C, et al. Watson for Oncology and breast cancer treatment recommendations: agreement with an expert multidisciplinary tumor board[J]. *Ann Oncol*, 2018, 29: 418-423.
- [21] BEJNORDI B E, VETA M, VAN DIEST P J, VAN GINNEKEN B, KARSSEMEIJER N, LITJENS G, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer[J]. *JAMA*, 2017, 318: 2199-2210.
- [22] ROBERTSON S, AZIZPOUR H, SMITH K, HARTMAN J. Digital image analysis in breast pathology—from image processing techniques to artificial intelligence[J]. *Transl Res*, 2018, 194: 19-35.

[本文编辑] 曾奇峰