

## ◆ 专论

## Research progresses in feature extraction of radiomics

XIE Kai, SUN Hongfei, LIN Tao, GAO Liugang, SUI Jianfeng, NI Xinye\*

(Department of Radiation Oncology, the Second People's Hospital of Changzhou

Affiliated to Nanjing Medical University, Changzhou 213003, China)

[Abstract] Radiomics refers to the comprehensive quantification of information such as tumor biological feature and heterogeneity through assessing a large number of quantitative image features from ROI of CT, MRI and PET images. With the development of data science, radiomics has been paid more and more attention. Feature extraction is an important step of Radiomics. The processes in feature extraction of radiomics were reviewed in this paper.

[Key words] Radiomics; Heterogeneity; Diagnosis; Feature extraction

DOI:10.13929/j.1003-3289.201707022

## 影像组学中特征提取研究进展

谢 凯, 孙鸿飞, 林 涛, 高留刚, 眭建锋, 倪昕晔\*

(南京医科大学附属常州第二人民医院放疗科, 江苏 常州 213003)

[摘要] 影像组学是从 CT、MRI、PET 等医学影像的 ROI 提取大量影像特征, 来描述肿瘤生物学特征和异质等信息的一种定量分析方法。随着数据科学发展, 影像组学的研究方法日益受到国内外的重视。影像组学的特征提取是其中的重要步骤。本文对影像组学的特征提取进行综述。

[关键词] 影像组学; 异质性; 诊断; 特征提取

[中图分类号] R445 [文献标识码] A [文章编号] 1003-3289(2017)12-1792-05

随着 CT、PET、MRI 设备及后处理、分析软件等的发展, 医学影像学为临床医学提供了更为精准的信息, 并有望在分析肿瘤表型差异及内涵特征方面取得进展<sup>[1]</sup>。但长久以来, 影像学家仅从中提取了定性和半定量的信息。目前活组织检查是诊断肿瘤良恶性的金标准。由于肿瘤空间和时间的异质性, 活组织检查不能全面评估肿瘤信息, 且患者难以承受连续多次活检。医学影像可无创提供肿瘤形态学和部分功能的信息, 并可借助图像分析监测疾病的发生、发展及治疗

的反应<sup>[2]</sup>。影像组学<sup>[3]</sup>认为, 通过对现有医学影像进行深入挖掘和分析, 有望获得基因组学或蛋白组学宏观表达上的影像特征。影像学家可通过对医学影像的量化分析, 推断出肿瘤的各种表型<sup>[4]</sup>。特征提取是影像组学的一个重要步骤, 本文对其进行综述。

## 1 影像组学的流程

目前, 常将放射影像的建设及应用框架分为图像的获取及重建、图像分割及 ROI 勾画、特征的提取和量化、特征的选择、预测模型及数据分析 5 部分。

影像组学的工作流程见图 1<sup>[5]</sup>。第一步是获得高质量和标准化的影像(MR、CT 和 PET 等), 用于诊断和评估。为获得更理想的分析结果, 选择病例时应避免过于异质性, 防止手术、放疗、化疗中不同 ROI 干扰影像组学特征分析结果。同时采集图像时应尽可能选用同个或同种扫描机器, 并选择合理的层厚、像素大小和管电压等参数<sup>[6]</sup>。第二步是通过自动、半自动分割方法或由经验丰富的放射肿瘤学家, 从获取的图像中

[基金项目] 江苏省自然科学基金(BK20151181)、常州市社会发展项目(CJ20160029)、常州市卫生高层次人才资助项目(2016CZLJ004)。

[第一作者] 谢凯(1990—), 男, 安徽安庆人, 硕士, 医学物理师。研究方向: 影像剂量学。E-mail: 229623265@qq.com

[通信作者] 倪昕晔, 南京医科大学附属常州第二人民医院放疗科, 213003。E-mail: nxy2000@aliyun.com

[收稿日期] 2017-07-06 [修回日期] 2017-11-03

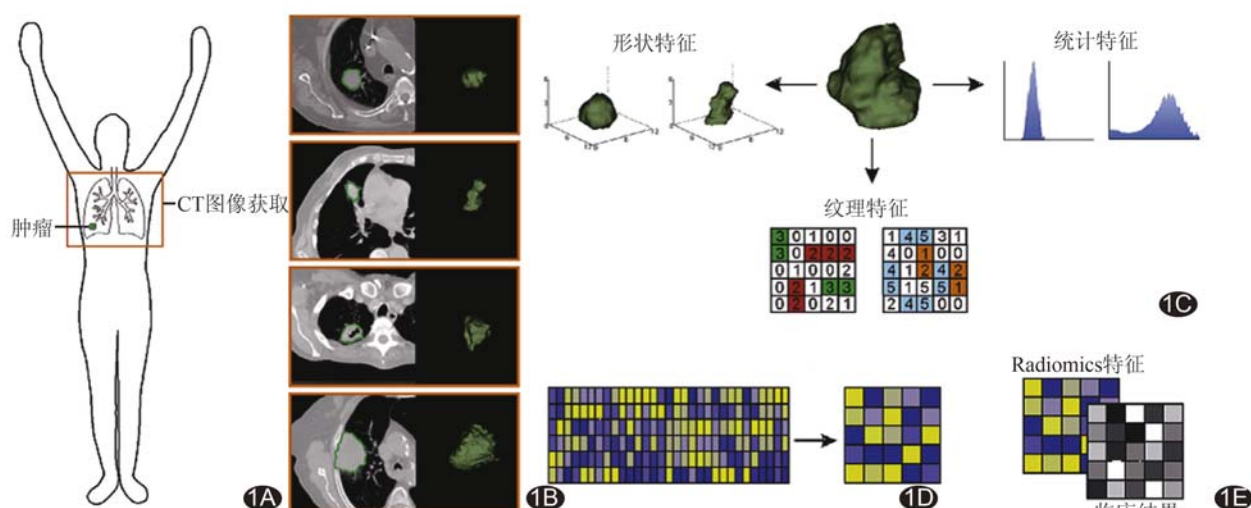


图1 影像组学流程示意图<sup>[5]</sup> A. 图像的获取与重建; B. 图像分割及 ROI 勾画; C. 特征的提取和量化; D. 特征的选择; E. 预测模型及数据分析

勾画肿瘤实质,即 ROI。目前常用的分割算法有图形切割法、活动轮廓法、区域生长法<sup>[7]</sup>和水平集法等。勾画 ROI 时注意图像是否需配准,同时 ROI 不应过小,通常不能小于  $5 \text{ cm}^3$ 。在手动、半自动、自动勾画及可用软件中,应针对精度、时间等综合选择。第三步是从 ROI 中提取影像组学特征,包括强度分布的描述、不同强度水平间的空间关系、纹理图案的形状和异质性以及肿瘤与周围组织关系的描述。第四步是根据数据的可重复性、其他特征的相关性等,对影像特征进行筛选,并分析选定特征与治疗结果或基因表达的关系。最后,通过将影像组学特征纳入治疗结果的预测模型,并评估影像组学特征相对常用预测因子提供的附加信息,从而提供准确的风险预测。

## 2 影像组学特征介绍

影像组学的核心步骤是提取高维特征数据用于定量分析 ROI 的属性。基于统计划分,常将影像组学特征分为两大类,一类认为纹理是二阶特征<sup>[8]</sup>,另一类认为纹理包括一阶、二阶和高阶特征,同时也包括基于模型和基于变换的纹理<sup>[9-10]</sup>。本文主要对后者进行阐述。

**2.1 形态特征及语义特征** 针对给定的肿瘤,首先需初步了解其三维大小和形状等,形态特征是通过计算 ROI 描述肿瘤特征,其中常用的有紧密度、三维直径、球形不均匀度、球形度、表面积、体积,其中最大的三维直径、表面积和体积可提供病变组织的尺寸信息。紧密度、球形不均匀度、球形度等表明与体积的比例,可描述球形、圆形或细长形状的肿瘤。Vallières 等<sup>[11]</sup>在

影像组学特征计算中也考虑硬度和偏心率。

另外语义特征<sup>[8]</sup>在分析中也有广泛应用,是指在放射学中常用来描述 ROI 的相关词汇,在计算机辅助下,放射学家也可定量分析病变组织。常用的语义特征有尺寸、形状、位置、血管分布、毛刺和坏死等。

形态特征主要提供肿瘤的三维大小和形状等相关信息,并不涉及肿瘤的异质性。为获得更多关于肿瘤异质性的信息,纹理特征的应用较广泛。

**2.2 纹理特征** 纹理分析是指于一个图像导出的特征,用各种数学方法提供图像不同灰度级相对位置信息,从而量化肿瘤异质性<sup>[12]</sup>。其计算方法可分为基于统计、基于模型和基于变换 3 类。其中基于统计方法最常用,用于描述图像中灰度值的分布和关系,按照计算中涉及的体素数目可分为一阶(单体素)、二阶(双体素)、高阶(三体素或更多)特征。

**2.2.1 一阶特征** 一阶特征也称强度特征,与 ROI 内的灰度级频率分布有关,依赖于单个体素值,而非相邻交互的体素,可从体素强度直方图中获得,包括能量、熵、锐利度、最大值、均值、平均绝对偏差、中值、最小值、范围、均方根、偏度、标准差、一致性、方差等,其中标准差、方差和平均绝对偏差用于测量直方图密集程度,即灰度级与均值偏离程度的度量;方差、偏度和锐利度常用于中心测度,偏度可度量直方图关于均值的不对称程度,锐利度是直方图锐度的度量;直方图的一致性和熵可度量直方图随机性<sup>[13]</sup>;熵可反映图像中平均信息量的多少<sup>[14]</sup>。

一阶特征反映强度直方图的整体信息,但未考虑

相邻体素的相互影响,因此需更深入地提取特征,在实际应用中基于空间灰度共生矩阵的二阶特征有广泛应用。

**2.2.2 二阶特征** 二阶特征通过空间灰度共生矩阵(gray level co-occurrence matrix, GLCM)、灰度游程步长矩阵(gray level run length matrix, GLRLM)等获得。这些矩阵决定一个体素的强度  $i$  以某种关系通过另一个体素的强度  $j$  寻找自身的程度。

GLCM 描述体素的两两排列,基于 GLCM 的二阶统计包括熵(矩阵的随机性)、能源/角二阶矩(体素重复/规律和测量图像的同质性)、同质性(共生矩阵的一致性)、不同度(测量矩阵中每个元素的不同程度)和相关度(灰度线性依赖的测量)。

GLRLM 描述具有相同强度体素的排列<sup>[15]</sup>。游程是预先设定方向上具有相同灰度级别的连续体素长度。游程长度间的关系产生纹理。细纹理的游程较短,且灰度强度相似,而粗纹理的游程与之相反。GLRLM 的定义参数包括短游程强调(short run emphasis, SRE;测量图像中的短游程分布)、长游程强调(long run emphasis, LRE)、灰度非均匀性(gray level nonuniformity, GLNU;测量灰度值相似性)、游程长度的不均匀性(run length nonuniformity, RLNU;测量游程步长相似性)。

**2.2.3 高阶特征** 利用邻域灰度差矩阵计算高阶统计量,研究 3 个或更多体素间的空间关系,并被认为与人类的图像经验非常相似<sup>[16]</sup>。通常利用邻域灰度差值矩阵(neighborhood gray-tone difference matrix, NGTDM)和灰度区域大小矩阵(gray-level size zone matrix, GLSZM)等计算。NGTDM 描述的是每个体素与相邻体素的不同,包括对比度(图像的局部变化数)、粗糙度(测量边缘密度)和忙碌度(测量灰度值变化的空间速度)。GLSZM 描述的是同质性区域的特征,包括短区域强调和长区域强调等。

根据纹理特征对肿瘤的描述,通常认为一阶特征

是在全局尺度上、基于 GLCM 和 NGTDM 的特征是在本地尺度上、基于 GLRLM 和 GLSZM 的特征是在局部尺度上反映肿瘤的异质性<sup>[17]</sup>。

为说明不同阶特征统计量之间的区别,图 2<sup>[10]</sup>显示了 4 个不同四色矩阵图像的模拟情况。基于一阶统计量,因为是使用单个体素值计算,4 种情况均会给出相同的特征。基于 GLCM 的二阶统计量将在图 2A 和图 2B 间给出不同的特征,因为计算考虑了体素在同一层面的相对分布。在 X 方向偏移一个体素(即偏移量),图 2B、2C 和 2D 获得的二阶特征相同。高阶统计量(如 NGTDM)由于空间体素相对于另外体素的特殊分布,4 种情况可能会得出不同的特征结果。

**2.2.4 基于模型和基于变换的方法** 基于模型的方法是使用复杂的数学模型如分形分析表示纹理。分形分析是一种模式或几何识别的形式。分形维数是测量表面的不规则性或粗糙度;分形维数越大,纹理越粗糙。

基于变换的方法是使用如傅里叶、Gabor、小波变换和高斯滤波器的拉普拉斯变换(LoG)分析在频率或尺度空间的纹理。傅里叶变换分析无空间定位的频率内容,因此不经常使用;Gabor 变换是引入高斯函数的傅里叶变换窗口,允许频率和空间定位,但受单个滤波器分辨率的限制;小波特征经小波分解计算获得原始图像的强度和纹理特征,将特征集中在肿瘤体积内的不同频率范围<sup>[18]</sup>;LoG 特征从具有越来越粗糙纹理图案的区域中提取。

**2.3 影像组学特征分析的影响因素** 影响影像组学特征分析的因素主要包括图像采集、ROI 勾画、算法等<sup>[19]</sup>。

图像采集因素包括 CT 图像采集的金属伪影<sup>[20]</sup>、噪音和扰动、图像重建算法等,PET 图像的光滑和量化、SUV 的离散程序<sup>[21]</sup>等,MR 图像采集的算法协议、参数的鲁棒性<sup>[22]</sup>等。在 ROI 勾画中,手动勾画与自动勾画的比较<sup>[23]</sup>、图像配准和轮廓生成等均可能影响

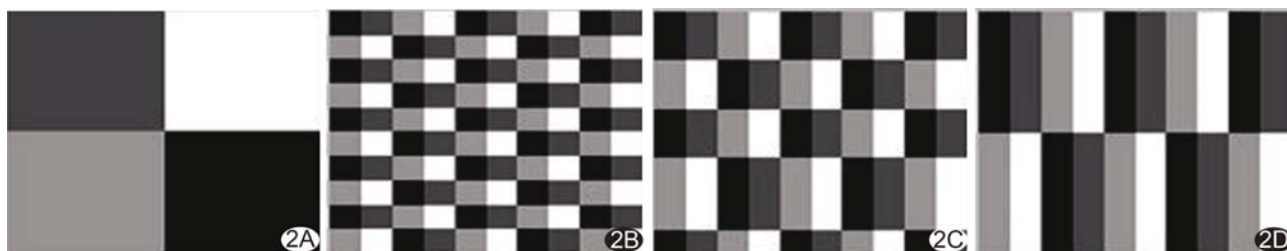


图 2 基于统计的纹理特征 4 种不同情况下的模拟<sup>[10]</sup> A.  $2 \times 2$  四色矩阵图像; B.  $10 \times 10$  矩阵图像; C.  $5 \times 10$  矩阵图像; D.  $2 \times 10$  矩阵图像

影像组学特征的计算。影像组学算法中,图像的重新取样、纹理矩阵中的方向性、算法公式中的特殊参数<sup>[24]</sup>以及不同集成软件的选择等,也会对特征计算有影响。

2.4 影像组学特征的使用与局限性 目前,在影像组学分析中,大量的临床特征(主要包括性别、年龄、肿瘤阶段、症状表现、肿瘤大小等)也会被进行分析比较,此外,分析肺癌时还需考虑有无吸烟史<sup>[25]</sup>。

在上述特征中,由于计算和理解方便,GLCM 计算的二阶特征获得广泛应用<sup>[26]</sup>。同时,为更好地描述肿瘤及其性质,通常综合上述特征提取数百个特征进行筛选分析。而且针对不同的医学影像(CT、MR、PET)手段,所使用的影像组学特征也有所不同。

影像组学特征分析在放疗分析的应用越来越广泛,但仍面临着诸多困难<sup>[19]</sup>,主要有:①影像组学特征分析主要在回顾性研究设计中进行,与实际临床的预测性需要还存在一定差别<sup>[27]</sup>;②通常提取的特征数量会大于所考虑的患者数量,而且数据间存在相关性,故常假阳性率过高<sup>[27]</sup>,因此,有必要减少特征数量以避免过度拟合分析的风险,通常一个特征对应 15 个左右的案例,并构建分类器或预测模型;③在分析的患者中,概括和验证统计学特征还需深入探讨;既往研究<sup>[24]</sup>中患者多数源于同一机构,今后需更多机构的合作,进行交叉验证;④对于常用的影像组学特征的生物学解释<sup>[28]</sup>,如常用的二阶特征 GLCM 等;⑤影像组学特征的可重复性、鲁棒性等是否可靠,还需探讨。

### 3 小结

医学影像未能提供更多关于肿瘤异质性内容,同时活体检查由于肿瘤空间和时间上的基因异质性,肿瘤的信息不能被全面评估。影像组学可高通量地从 CT、PET、MRI 中提取并分析大量高级定量影像特征,从而指导肿瘤的治疗、预后以及疗效<sup>[29]</sup>,有广阔的应用前景。

总之,影像组学作为肿瘤生物信息来源的一种手段,可多方位提供肿瘤生物学特性,为肿瘤精准治疗提供帮助<sup>[30]</sup>。同时,影像组学的特征提取也面临一系列的机遇与挑战。如何更好地通过影像组学特征提取与分析,需要继续努力,提出更多的理论与方法。

### [参考文献]

[1] 苏会芳,周国锋,谢传森,等.放射组学的兴起和研究进展.中华医学杂志,2015,95(7):553-556.

[2] 王敏,宋彬,黄子星,等.大数据时代的精准影像医学:放射组学.中国普外基础与临床杂志,2016,23(6):752-755.

[3] Kumar V, Gu Y, Basu S, et al. Radiomics: The process and the challenges. Magn Reson Imaging, 2012,30(9):1234-1248.

[4] 胡盼盼,王佳舟,胡伟刚,等.影像组学在精准放疗中的应用.中华放射肿瘤学杂志,2017,26(1):103-106.

[5] Huynh E, Coroller T P, Narayan V, et al. CT-based radiomic analysis of stereotactic body radiation therapy patients with lung cancer. Radiother Oncol, 2016,120(2):258-266.

[6] Court LE, Fave X, Mackin D, et al. Computational resources for radiomics. Transl Cancer Res, 2016,5(4):340-348.

[7] Hojjatoleslami SA, Kittler J. Region growing: A new approach. IEEE T Image Process, 1998,7(7):1079-1100.

[8] Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. Radiology, 2016,278(2):563-577.

[9] Davnall F, Yip CS, Ljungqvist G, et al. Assessment of tumor heterogeneity: An emerging imaging tool for clinical practice? Insights Imaging, 2012,3(6):573-589.

[10] Chicklore S, Goh V, Siddique M, et al. Quantifying tumour heterogeneity in <sup>18</sup>F-FDG PET/CT imaging by texture analysis. Eur J Nucl Med Mol Imaging, 2013,40(1):133-140.

[11] Vallières M, Freeman CR, Skamene SR, et al. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. Phys Med Biol, 2015,60(14):5471-5496.

[12] Castellano G, Bonilha L, Li LM, et al. Texture analysis of medical images. Clin Radiol, 2004,59(12):1061-1069.

[13] 王超.基于影像组学的乳腺癌病灶分割及预后分析.哈尔滨:哈尔滨理工大学,2017:24-27.

[14] 汪俊.基于定量影像组学的肺结节良恶性预测.哈尔滨:哈尔滨理工大学,2016:25-26.

[15] Loh HH, Leu JG, Luo RC. The analysis of natural textures using run length features. IEEE T Ind Electron, 1988,35(2):323-328.

[16] Srinivasan GN, Shobha G. Statistical texture analysis. Proceedings of World Academy of Science Engineering & Technology, 2008,36:1264-1269.

[17] Tixier F, Le RC, Hatt M, et al. Intratumor heterogeneity characterized by textural features on baseline <sup>18</sup>F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. J Nucl Med, 2011,52(3):369-378.

[18] Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun, 2014,5:4006-4013.

[19] Scalco E, Rizzo G. Texture analysis of medical images for radiotherapy applications. Brit J Radiol, 2016,90(1070):20160642.

[20] Leijenaar RT, Carvalho S, Hoebbers FJ, et al. External validation of a prognostic CT-based radiomic signature in oropharyngeal squamous cell carcinoma. Acta Oncol, 2015,54(9):1423-1429.

[21] Leijenaar RTH, Nalbantov G, Carvalho S, et al. The effect of

- SUV discretization in quantitative FDG-PET Radiomics: The need for standardized methodology in tumor texture analysis. Sci Rep-UK, 2015, 5:11075-11084.
- [22] Song YS, Park CM, Lee SM, et al. Reproducibility of histogram and texture parameters derived from intravoxel incoherent motion diffusion-weighted MRI of FN13762 rat breast Carcinomas. Anticancer Res, 2014, 34(5):2135-2144.
- [23] Orlhac F, Soussan M, Maisonneuve JA, et al. Tumor texture analysis in  $^{18}\text{F}$ -FDG PET: Relationships between texture parameters, histogram indices, standardized uptake values, metabolic volumes, and total lesion glycolysis. J Nucl Med, 2014, 55(3):414-422.
- [24] Hatt M, Tixier F, Pierce L, et al. Characterization of PET/CT images using texture analysis: The past, the present any future?. Eur J Nucl Med Mol Imaging, 2016, 44(1):1-15.
- [25] Coroller TP, Grossmann P, Hou Y, et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. Radiother Oncol, 2015, 114(3):345-350.
- [26] Panth KM, Leijenaar RTH, Carvalho S, et al. Is there a causal relationship between genetic changes and radiomics-based image features? An in vivo preclinical experiment with doxycycline-inducible GADD34 tumor cells. Radiother Oncol, 2015, 116(3):462-466.
- [27] Yip SS, Aerts HJ. Applications and limitations of radiomics. Phys Med Biol, 2016, 61(13):150-166.
- [28] Alobaidli S, Mcquaid S, South C, et al. The role of texture analysis in imaging as an outcome predictor and potential tool in radiotherapy treatment planning. Brit J Radiol, 2014, 87(1042):13.
- [29] 苏会芳, 周国锋, 谢传森, 等. 放射组学的兴起和研究进展. 中华医学杂志, 2015, 95(7):553-556.
- [30] 贾田颖, 余雯, 傅小龙, 等. 影像组学在非小细胞肺癌精准治疗中的应用进展. 中华放射医学与防护杂志, 2016, 36(12):947-950.

## 《磁共振成像临床应用入门(第2版)》已出版

《磁共振成像临床应用入门(第2版)》由王振常教授、孙波教授、徐建民教授担任名誉主编,靳二虎教授、蒋涛教授、张辉教授担任主编,范占明教授、刘佩芳教授、程晓光教授担任副主编,集全国21位磁共振成像临床应用专家的智慧,在《磁共振成像临床应用入门(第1版)》的基础上耗时1年修订而成。人民卫生出版社已于2015年2月出版。

本书第2版沿袭了第1版的整体布局和写作风格,共十五章,书中删减了一些陈旧或不常用的表述,使内容更加精练,更新的图片更具代表性。其中,第一章新增了不同部位CE-MRA检查时计算扫描延迟时间和注射钆对比剂的细节;第三章详细新增了肝细胞特异性对比剂的应用价值;第四章新增了脑小血管病MRI表现;第五章对主动脉疾病相关内容加以丰富,新增了特殊类型主动脉夹层、主动脉壁间血肿、穿透性动脉粥样硬化性溃疡、马方综合征和肺栓塞MRI表现;第七章新增了肝脓肿、肝包虫病的MRI表现;第八章新增了自身免疫性胰腺炎MRI表现;第九章新增了肾细胞癌及肾血管平滑肌脂肪瘤各亚型MRI表现;第十一章新增了剖宫产瘢痕妊娠MRI表现;第十三章新增了脊髓空洞与脊髓中央管扩张积水症的MRI表现;书末的三个附录也有更新和充实。

本书主要供放射科住院医师、进修医师、研究生、医学影像专业大学生和磁共振室技术员使用,对涉足MRI日常应用的临床医师、科研人员、大专院校讲师和其他人员也有所裨益。本书共117万字,书中插图包括线条图及示意图155幅,影像图片1647幅。全书582页,16开本,软精装,定价99元。各地新华书店和医学专业书店有售。购书电话010-59787592、010-59787584、010-65264830。

