

甲状腺疾病

DOI: 10.13406/j.cnki.cyx.001931

甲状腺超声半自动语义标注软件的可重复性评估

高璐滢, 高 琼, 姜玉新, 李建初, 张 波, 王 莹, 席雪华, 刘如玉, 刘 佳

(中国医学科学院北京协和医学院北京协和医院超声医学科, 北京 100730)

【摘要】目的:为实现计算机辅助诊断下的甲状腺结节超声图像自动识别,本课题组研发了多特征甲状腺超声图像半自动语义标注软件工具,以完成大样本甲状腺超声图像的高质量人工标注。本研究以探讨超声医生应用甲状腺超声图像标注工具的可重复性。**方法:**随机选取我院 2013 年甲状腺结节图像 50 例,分别由 2 名超声医生独立利用标注软件工具进行甲状腺超声图像识别;2 名超声医生独立通过直观超声图像进行超声图像识别;1 名超声医生分别利用标注软件工具、直观超声图像进行随机图像识别,并分别利用 κ 值进行观察者一致性分析。**结果:**本研究中,2 名超声医生分别应用标注软件工具对各项超声特征及超声分级的观察者差异 κ 值为:边缘/边界 0.94,形态 0.94,回声 0.63,内部结构 0.87,钙化 0.89,2015 版美国甲状腺协会(American Thyroid Association, ATA)指南危险分层 0.84($P<0.01$);2 名超声医生通过直观超声图像的观察者差异 κ 值如下:边缘/边界 0.75,形态 0.80,回声 0.53,内部结构 0.62,钙化 1.00,ATA 危险分层 0.42($P<0.01$)。1 名医生分别利用图像标注系统、直观超声图像的观察者差异 κ 值为:边缘/边界 1.00,形态 1.00,回声 1.00,内部结构 0.88,钙化 1.00,ATA 危险分层 0.96($P<0.01$)。**结论:**多特征甲状腺超声图像半自动语义标注软件工具为甲状腺结节超声诊断的标准化提供方法,为其细粒度智能识别研究提供高质量的训练和验证数据集。

【关键词】甲状腺结节;超声;图像半自动语义标注软件工具

【中图分类号】R736.1

【文献标志码】A

【收稿日期】2018-10-18

Evaluation on repeatability of the semi-automatic semantic annotation software tools for multi-feature thyroid ultrasound images

Gao Luying¹, Gao Qiong¹, Jiang Yuxin¹, Li Jianchu¹, Zhang Bo¹, Wang Ying¹, Xi Xuehua¹, Liu Ruyi¹, Liu Jia¹
(Department of Ultrasound, Chinese Academy of Medical Sciences & Peking Union Medical College Hospital)

【Abstract】Objective: In order to achieve high-quality manual annotation of large-sample thyroid ultrasound images, we developed a semi-automatic semantic annotation software tool for multi-feature thyroid ultrasound images. In the study, we aimed to explore the reproducibility of doctors' application of the thyroid ultrasound image annotation tool. **Methods:** Fifty cases of thyroid nodules in our hospital were randomly selected. Two ultrasound doctors independently used the annotation software tool to perform thyroid ultrasound image recognition. Two ultrasound doctors independently performed ultrasound images recognition through visual images. One ultrasound doctor randomly used annotation software tools and visual images for image recognition, and then observer consistency analysis was performed. **Results:** In this study, two ultrasound doctors used annotation software tools to observe the differences between ultrasound features and ultrasound grading (κ value): boundary 0.94, morphology 0.94, echo 0.63, internal structure 0.87, calcification 0.89, 2015 American Thyroid Association (ATA) risk stratification was 0.84 ($P<0.01$). The difference between the two ultrasound doctors through the visual ultrasound image was as follows (κ value): boundary 0.75, morphology 0.80, echo 0.53, internal structure 0.62, calcification 1.00, ATA risk stratification 0.42 ($P<0.01$). The difference between the observers using the image annotation system and the visual ultrasound image was (κ value): boundary 1.00, morphology 1.00, echo 1.00, internal structure 0.88, calcification 1.00, ATA risk stratification 0.96 ($P<0.01$). **Conclusion:** The multi-feature thyroid ultrasound image semi-automatic semantic annotation software tool provides a standardized method for the diagnosis of thyroid nodule ultrasound diagnosis, and provides high-quality training and verification data sets for its fine-grained intelligent recognition research.

作者介绍:高璐滢, Email: gaoluying@pumch.cn,

研究方向:浅表及血管超声及超声新技术。

通信作者:张 波, Email: thyroldus@163.com。

基金项目:国家自然科学基金资助项目(编号:61672077);北京市卫生科技成果与适宜技术推广资助项目(编号:2018-TG-39)。

优先出版: <http://kns.cnki.net/kcms/detail/50.1046.R.20181106.1138.012.html>
(2018-11-07)

tion software tool provides a standardized method for the diagnosis of thyroid nodule ultrasound diagnosis, and provides high-quality training and verification data sets for its fine-grained intelligent recognition research.

【Key words】 thyroid nodule; ultrasonography; semi-automatic semantic annotation software tools for multi-feature

甲状腺结节人群患病率为 19%~68%,其中 7%~15%为甲状腺癌^[1-2]。2012 年北京市卫生与人群健康状况报告显示,北京户籍居民甲状腺癌发病率为 11.0/10 万,高于我国肿瘤登记地区平均水平^[3-4]。病变早、中期缺乏特异的临床表现,良恶性特征重叠,易造成漏诊、误诊。超声检查具有无创、简便、费用低、无辐射等优点,是甲状腺结节的主要诊断手段,缺点是医生主要通过超声图像进行判断,具有较大的观察者依赖性,缺乏经验的医师误诊率高,准确度很难保证^[5]。因此,过度治疗和治疗不足同时存在,如何提高甲状腺结节的整体诊治效果,是目前迫切需要解决的问题。

计算机辅助诊断 (computer aided diagnosis, CAD) 为解决以上问题供了可能性。将医学、计算机技术、图像处理技术、人工智能技术和数据分析理论进行紧密结合,比较不同组织的回声以及良恶性病变超声图像特征间的差别,来实现病变的自动检测及良恶性判别,辅助医生对病变进行更为客观的定性乃至定量分析,可降低人为判断和医生主观经验对诊断结果的影响,可有效提高甲状腺病变诊断的准确性和效率。近年来机器学习技术发展迅速,在图像分类,图像检索,以及物体识别等方面有广泛应用,取得了显著的成果^[6]。前期作者受到国家自然科学基金的资助,通过医工结合的方式,在图像局部

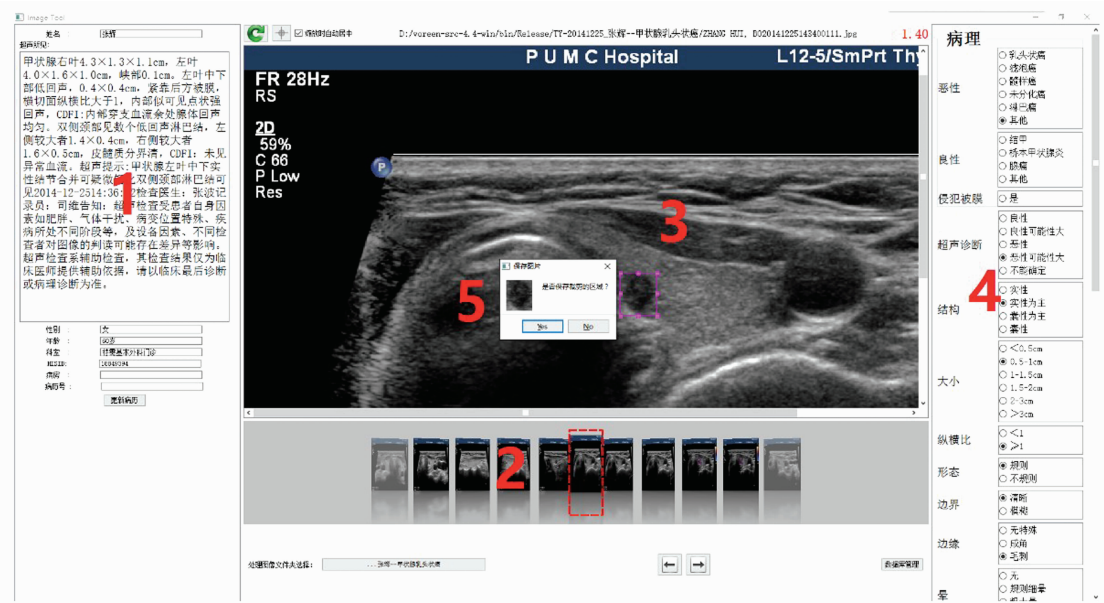
和全局视觉特征统一建模的基础上,最终实现基于广义集成学习的超声图像甲状腺结节自动识别。为实现大样本甲状腺超声图像的高质量人工标注,课题组研发了多特征甲状腺超声图像半自动语义标注软件工具(专利登记号:2017SR570560)。在超声图像中标记出疑似病变区域并给出超声诊断,具有相对主观的观察者依赖性,缺乏经验的医师误诊率高,观察者差异的评估和改进可能有助于解决这一问题^[7-10]。本研究拟已建立甲状腺超声图像标注软件为研究对象,探讨超声医生应用甲状腺超声图像标注工具的可重复性。

1 对象与方法

1.1 多特征甲状腺超声图像半自动语义标注软件工具

北京航空航天大学虚拟现实国家重点实验室与本中心联合研发了甲状腺超声图像半自动语义标注软件工具。该工具的交互界面大致可分为 A、B、C 3 个区域:A 区域为病例信息区,自动读取到的病例信息会在该区域显示,供医生标注时参考,并可以由医生修改;B 区域为图像显示区,读取到的病例图像会在该区域显示,医生可根据需要选择图像、对图像进行缩放、分割病灶区域;C 区域为特征标注区,医生在此区域选择框定病灶的特征。

如图 1 所示,该工具对甲状腺图像预处理过程大致可分为 5 步,分别对应图中红色数字:1)阅读病例中的超声所见,



注释:图中红色数字 1. 阅读病例中的超声所见,为标注提供参考;2. 为选择病例中的一幅图像,选择的图像将在上方放大显示;3. 可通过鼠标滚轮缩放、鼠标按键平移等方式寻找病灶,并用紫色框将其框定;4. 可选择该病灶的各项特征,不存在的特征可为空;5. 可双击紫色框,保存病灶图像及其特征

图 1 甲状腺图像预处理过程示例图

为标注提供参考;2)选择病例中的一幅图像,选择的图像将在上方放大显示;3)通过鼠标滚轮缩放、鼠标按键平移等方式寻找病灶,并用紫色框将其框定;4)选择该病灶的各项特征,不存在的特征可为空;5)双击紫色框,保存病灶图像及其特征。

同时,该工具还支持病灶数据库管理,如图 2 所示,点击 B 区域右下角“数据库管理”按钮即可打开数据库管理界面。在 A 区域中选择病例和超声图像,对应超声图像会显示在 B 区域,而标注后的病灶信息则显示在 C 区域,点击病灶信息将弹出对话框在 D 区域显示原始病灶图像及其标注信息,而且临床医生可在此对标注信息进行继续修改。

1.2 2015 版美国甲状腺协会(American Thyroid Association, ATA)指南分级^[1]

超声恶性风险分级包括高度可疑恶性、中度可疑恶性、低度可疑恶性、极低度可疑恶性和良性结节(以下简称高危、中危、低危、极低危及良性)。高危:实性低回声或囊实性结节中的实性成分为低回声的结节,同时具有以下一项或多项超声特征:①不规则边缘(小分叶、毛刺、浸润性);②微钙化;③纵横比>1;④边缘钙化中断,低回声突出钙化外;⑤甲状腺被膜外侵犯。中危:实性低回声结节,边缘光滑、规则,无微钙化、纵横比>1 及腺体外侵犯。低危:等回声或高回声实性结节或囊实性结节的实性部分偏心,无微钙化、边缘不规则、纵横比>1 及腺体外侵犯。极低危(恶性风险<3%):a“海绵样”的结节;b 囊实性结节实性部分不偏心,无微钙化、边缘不规则、纵横比>1 及被膜外侵犯。良性:主要为囊性结节。

1.3 研究对象

随机选取我院 2013 年甲状腺结节图像 50 例,(1)2 名甲状腺超声经验丰富的超声医生,独立利用图像标注系统进行甲状腺超声图像识别;(2)2 名甲状腺超声经验丰富的超声医生,独立通过直观超声图像进行甲状腺超声图像识别;(3)1 名甲状腺超声经验丰富的超声医生分别利用图像标注系统、直观超声图像进行超声图像识别。甲状腺超声图像识别的具体内容包括:对边缘/边界(清晰、毛刺、小分叶、浸润性)、形态(规则、不规则)、纵横比、回声(高、中、低、极低、无)、内部结构(囊性、囊实性、实性)、有无微钙化、ATA 指南分级诊断等,并进行观察者一致性分析。

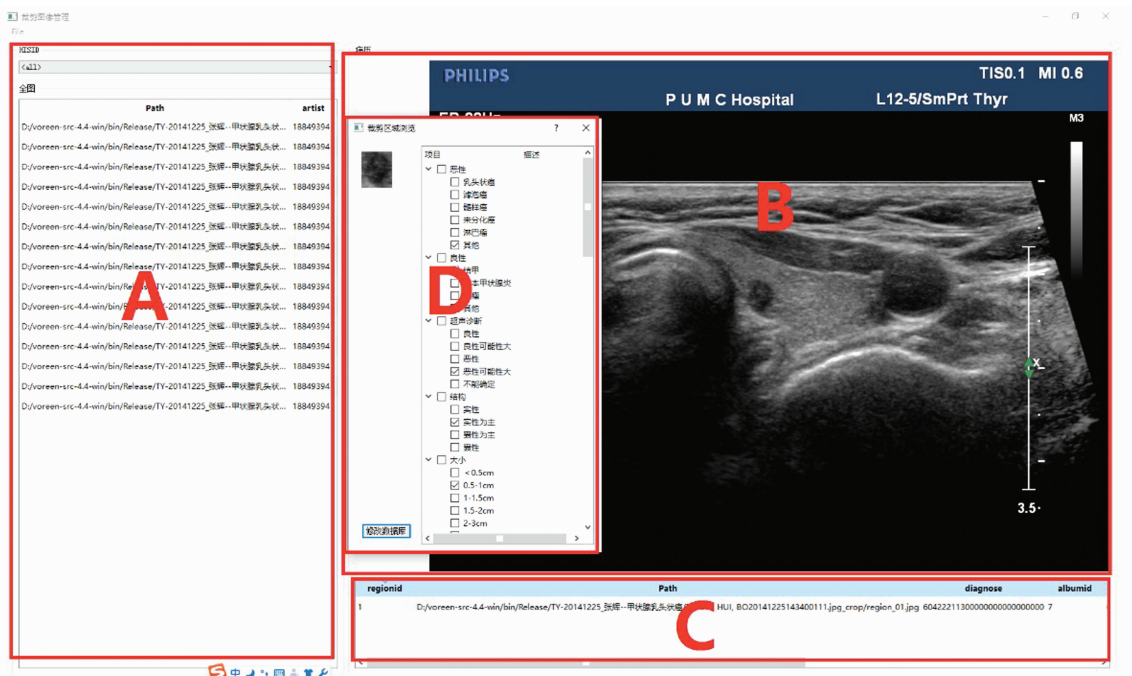
1.4 统计学分析

采用 SPSS 22.0 及 SAS 统计软件进行统计学分析。计数资料采用 χ^2 检验, κ 进行一致性检验。其中 κ 值>0.80 为“优秀”,0.60~0.80 为“较好”,0.40~0.60 为“良好”,0.20~0.40 为“较差”, ≤ 0.20 为“差”。检验水准 $\alpha=0.05$ 。

2 结果

2.1 2 名医生利用图像标注系统进行甲状腺超声图像识别

本研究显示 2 名超声医生分别应用图像标注系统对各项超声特征及超声分级的观察者差异如下:边缘/边界 $\kappa=0.94$ ($P<0.01$),形态 $\kappa=0.94$ ($P<0.01$),回声 $\kappa=0.63$ ($P<0.01$),内部结构 $\kappa=0.87$ ($P<0.01$),钙化 $\kappa=0.89$ ($P<0.01$),ATA 危险分层 $\kappa=0.84$ ($P<0.01$)(表 1)。



注释:图中红色数字 A. 区域中选择病例和超声图像;B. 对应超声图像会显示在 B 区域;C. 标注后的病灶信息则显示在 C 区域;D. 点击病灶信息将弹出对话框在 D 区域显示原始病灶图像及其标注信息,临床医生可在此对标注信息进行继续修改。

图 2 语义标注工具的病灶数据库管理功能示例图

2.2 2 名医生通过直观超声图像进行甲状腺超声图像识别

本研究显示 2 名超声医生分别通过直观超声图像对各项超声特征及超声分级的观察者差异如下:边缘/边界 $\kappa=0.75$ ($P<0.01$),形态 $\kappa=0.80$ ($P<0.01$),回声 $\kappa=0.53$ ($P<0.01$),内部结构 $\kappa=0.62$ ($P<0.01$),钙化 $\kappa=1.00$ ($P<0.01$),ATA 危险分层 $\kappa=0.84$ ($P<0.01$) (表 2)。

2.3 1 名医生分别利用图像标注系统、直观超声图像进行超声图像识别

本研究显示 1 名医生分别利用图像标注系统、直观超声图像进行各项超声特征及超声分级的观察者差异:边缘/边界 $\kappa=1.00$ ($P<0.01$),形态 $\kappa=1.00$ ($P<0.01$),回声 $\kappa=1.00$ ($P<0.01$),内部结构 $\kappa=0.88$ ($P<0.01$),钙化 $\kappa=1.00$ ($P<0.01$),ATA 危险分层 $\kappa=0.96$ ($P<0.01$)。边缘/边界 1.00,形态 1.00,回声 1.00,内部结构 0.88,钙化 1.00,ATA 危险分层 0.96 (表 3)。

3 讨论

超声是细针穿刺活检和进一步治疗的主导和首选筛查方式。但是,甲状腺结节的超声诊断依赖于操作者的经验和技能。因此,缺乏经验的从业人

员误诊率很高。计算机辅助诊断最近被开发用于根据超声特征来诊断甲状腺结节。计算机辅助诊断需要利用甲状腺超声图像标注系统提取兴趣区域,本研究建立了甲状腺超声图像标注系统,并验证了该系统在不同观察者间的一致性,验证了甲状腺结节计算机辅助诊断系统的准确性。

本研究显示两名医生通过直观超声图像进行甲状腺超声图像识别的观察者差异性为优秀—良好的水平,而两名超声医生分别应用图像标注系统的图像识别观察者差异性为优秀—较好的水平。提示 CAD 系统的图像标注系统有较好的可靠性,优于传统的直观超声图像的观察者一致性。另外,本研究结果表明一名医生分别利用图像标注系统、直观超声图像进行超声图像识别的观察者一致性均为优秀水平,提示超声医生的观察者内差异小、有良好的稳定性。

尽管计算机辅助诊断系统被广泛应用于评估肺癌和乳腺癌,但使用计算机辅助诊断系统诊断甲

表 1 2 名医生利用图像标注系统进行图像识别的观察者一致性检验

	一致(例数)	不一致(例数)	κ 值	一致性水平	P 值
边缘/边界	45	5	0.94	优秀	<0.01
形态	45	5	0.94	优秀	<0.01
回声	37	13	0.63	较好	<0.01
内部结构	44	6	0.87	优秀	<0.01
钙化	44	6	0.89	优秀	<0.01
ATA 危险分层	39	11	0.84	优秀	<0.01

表 2 2 名医生通过直观超声图像进行图像识别的观察者一致性检验

	一致(例数)	不一致(例数)	κ 值	一致性水平	P 值
边缘/边界	43	7	0.75	较好	<0.01
形态	45	5	0.81	优秀	<0.01
回声	36	14	0.53	良好	<0.01
内部结构	45	5	0.62	较好	<0.01
钙化	50	0	1.00	优秀	<0.01
ATA 危险分层	31	19	0.42	良好	<0.01

表 3 1 名医生分别利用图像标注系统、直观超声图像进行图像识别的观察者一致性检验

	一致(例数)	不一致(例数)	κ 值	一致性水平	P 值
边缘/边界	50	0	1.00	优秀	<0.01
形态	50	0	1.00	优秀	<0.01
回声	50	0	1.00	优秀	<0.01
内部结构	48	2	0.88	优秀	<0.01
钙化	50	0	1.00	优秀	<0.01
ATA 危险分层	48	2	0.96	优秀	<0.01

甲状腺结节的研究数量有限^[12-14]。并且既往的研究没有阐述如何提取兴趣区域并进行甲状腺超声图像的标注。甲状腺超声图像标注系统主要是通过对兴趣区域的识别解析,在大多数情况下,可根据已标注的训练集,通过将给定的未知图像与已知图像的特征进行比较,来将图像的兴趣区域划分为不重叠的语义区域。本研究详细描述了甲状腺超声图像标注系统,增加了计算机辅助诊断的可靠性。该工具为临床医生的甲状腺超声图像病例语义标注提供了极大的便利,可为甲状腺结节自动识别技术的研究提供高质量的训练和验证数据集,这从可靠数据源和数据预处理工具两个方面为本项目研究工作的开展奠定了良好的基础。

基于该工具协和医院甲状腺科已经对协和医院部分确诊病人的甲状腺影像进行了高质量人工标注,形成了 5 000 张左右可用于甲状腺结节良恶性智能识别的训练数据集。该工具为临床医生的多特征甲状腺超声图像病例语义标注提供了极大的便利,可在后续工作中继续用于甲状腺结节类型详细分类数据的标注,从而为甲状腺结节的细粒度智能识别研究提供高质量的训练和验证数据集。然而,这项研究受到空间和人数的限制,纳入的医生人数有限。

参 考 文 献

- [1] International Agency for Research on Cancer. GLOBOCAN 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012[R/OL]. <http://globocan.iarc.fr/Default.aspx>
- [2] Guth S, Theune U, Aberle J, et al. Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination [J]. European Journal of Clinical Investigation, 2010, 39(8), 699-706.
- [3] 赫杰, 陈万青. 2012 中国肿瘤登记年报[M]. 北京: 军事医学科学出版社, 2012.
- [4] 北京市人民政府. 北京市 2012 年度卫生与人群健康状况报告 [M]. 北京: 人民卫生出版社, 2013.
- [5] Cooper DS, Doherty GM, Haugen BR, et al. Revised American Thyroid Association. Management guidelines for patients with thyroid nodules and differentiated thyroid cancer; the American Thyroid Association (ATA) guidelines taskforce on thyroid nodules and differentiated thyroid cancer [J]. Thyroid, 2015, 19(11): 1167-1214.
- [6] Greff K, Srivastava RK, Schmidhuber J. Training very deep networks [J]. Advances in Neural Information Processing Systems, 2015: 2377-2385.
- [7] Tan GH, Gharib H. Thyroid incidentalomas; management approaches to nonpalpable nodules discovered incidentally on thyroid imaging [J]. Annals of Internal Medicine, 1997, 126(3): 226-231.
- [8] Hegedüs L. The thyroid nodule [J]. The New England Journal of Medicine, 2004, 351(17): 1764-1771.
- [9] Leenhardt L, Erdogan MF, Hegedus, L. European thyroid association guidelines for cervical ultrasound scan and ultrasound-guided techniques in the postoperative management of patients with thyroid cancer [J]. European Thyroid Journal, 2013, 2(3): 147-159.
- [10] Perros P, Boelaert K, Colley S, et al. Guidelines for the management of thyroid cancer [J]. Clinical Endocrinology, 2014, 81(S1): 1-122.
- [11] Haugen BR, Alexander EK, Bible KC, et al. American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer; The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer [J]. Thyroid, 2015, 26(1): 1-133.
- [12] Acharya UR, Faust O, Sree SV, et al. ThyroScreen system: high resolution ultrasound thyroid image characterization into benign and malignant classes using novel combination of texture and discrete wavelet transform [J]. Comput Methods Programs Biomed, 2012, 107(2): 233-241.
- [13] Acharya UR, Sree SV, Krishnan MMR, et al. Non-invasive automated 3D thyroid lesion classification in ultrasound: a class of ThyroScan systems [J]. Ultrasonics, 2012, 52(4): 508-520.
- [14] Chang Y, Paul AK, Kim N, et al. Computer-aided diagnosis for classifying benign versus malignant thyroid nodules based on ultrasound images: a comparison with radiologist-based assessments [J]. Med Phys, 2016, 43(1): 554-567.

(责任编辑: 唐秋姝)