

## · 人工智能影像学专题 ·

## 基于三维卷积神经网络肺结节深度学习算法模型临床效能初步评估

王祥, 李清楚, 邵影, 邹勤, 孙安, 陈彦博, 陈如谭, 高耀宗, 刘士远, 萧毅

**【摘要】** 目的:为评价人工智能模型的应用价值,本研究在专家共识的基础上建立了肺结节标准测试集,对前期建立的一种基于三维卷积神经网络肺结节深度学习算法模型进行验证,评价该模型临床效能和限度。**方法:**基于胸部 CT 肺结节数据标注与质量控制专家共识建立标准测试数据集,对前期建立的基于三维卷积神经网络的肺结节深度学习算法模型及传统 CAD 系统(Siemens syngo. via VB 3.0 和 Philips ISP V8)进行检验,在肺结节检出灵敏度、精准度以及平均每例假阳性个数等多个指标方面进行优效验证。**结果:**针对测试数据集中的肺结节,Syngo. via 工作站检出灵敏度为 36%,精准度为 69%,平均每例假阳性 1.2 个;Philips ISP 工作站肺结节检出灵敏度为 34%,精准度为 73%,平均每例假阳性 0.9 个;三维卷积神经网络的肺结节深度学习算法模型检出灵敏度为 90%,精准度为 71%,平均每例假阳性 2.8 个。**结论:**该三维卷积神经网络算法模型相较于传统 CAD 系统,肺结节检出灵敏度显著提升。由于训练数据集的偏倚等问题,灵敏度仍有进一步提升的空间。通过针对性地补充训练数据集,如增加磨玻璃结节的比重,可进一步提升肺结节检出灵敏度。改进之后的模型有望成为影像医生肺癌筛查工作的得力助手。

**【关键词】** 肺结节;人工智能;卷积神经网络;计算机辅助筛查;体层摄影术,X 线计算机;磨玻璃结节

**【中图分类号】** TP18;R734.2;R814.42 **【文献标识码】** A

**【文章编号】** 1000-0313(2019)09-0942-05

DOI:10.13609/j.cnki.1000-0313.2019.09.002

开放科学(资源服务)标识码(OSID):



### A preliminary clinical evaluation of a 3D convolutional neural network based deep learning system

WANG Xiang, LI Qing-chu, SHAO Ying, et al. Department of Radiology, Changzheng Hospital of Naval Military Medical University, Shanghai 200003, China

**【Abstract】 Objective:** To assess the application value of artificial intelligence model, a standardized testing set constructed according to the expert consensus were used to validate a prior developed 3D convolutional neural network (3D CNN) algorithm and evaluate its clinical efficacy and limitation. **Methods:** A standardized testing dataset was constructed based on the expert consensus on the rule and quality control of pulmonary nodule annotation based on thoracic CT and used to test a prior developed 3D CNN algorithm and two traditional CAD systems (Siemens syngo. via VB 3.0 and Philips ISP V8. Several metrics). Several evaluation indicators, such as detection sensitivity, precision and average number of false positives per subject (average FP), were calculated. **Results:** Testing with the standardized chest CT dataset, the sensitivity, precision and average FP of Syngo. via VB 3.0 were 36%, 69%, and 1.2, respectively. The three indicators of Philips ISP V8 were 34%, 73%, and 0.9, respectively. The studied 3D CNN algorithm obtained a sensitivity of 90%, a precision of 71%, and an average FP of 2.8. **Conclusion:** Compared to the traditional CAD systems, the 3D CNN algorithm shows a significant improvement on detection sensitivity of lung nodules. However, the detection sensitivity still needs further improvement due to the unbalanced training dataset. Studying with a more diversified training dataset, e. g., more included ground-glass nodules, may further improve the detection sensitivity. The

作者单位:200003 上海,海军军医大学(原第二军医大学)附属长征医院影像科(王祥,李清楚,邹勤,孙安,陈如谭,刘士远,萧毅);201210 上海,上海联影智能医疗科技有限公司(邵影,陈彦博,高耀宗)

作者简介:王祥(1992-),男,山东临沂人,硕士,住院医师,主要从事胸部影像诊断工作。

通讯作者:萧毅, E-mail: cz-xiaoyi@smmu.edu.cn

基金项目:上海市科学技术委员会基金项目(NO:17411952400);国家重点研发计划政府间合作项目(NO:2016YFE0103000);上海市卫计委智慧医疗项目(NO:2018ZHYL0101);科技部国家重点研发计划(NO:2018YFC0116404);

AI system can help the radiologists in lung cancer screening in a more efficient way after improvement.

**【Key words】** Lung nodule; Artificial intelligence; Convolutional neural network; Computer-aided detection; Tomography, X-ray computed; Ground-glass nodule

肺癌是全球范围内发病率及死亡率最高的恶性肿瘤<sup>[1,2]</sup>,患者早期往往没有任何临床症状而被忽视,研究表明肺癌患者五年生存率约为10%~16%<sup>[3]</sup>,大多数患者检出时已经失去了早期根治的机会。早期发现、早期诊断及早期治疗可以显著提高肺癌患者五年生存率<sup>[4]</sup>。低剂量胸部CT具有较高的空间分辨率、扫描时间短及可对病变进行三维重建等优点,广泛应用于肺结节的检出及鉴别诊断。然而,CT筛查数量急剧增加及扫描层厚越来越薄,影像医生工作强度显著提升,临床工作中漏诊及误诊现象时有发生<sup>[5,6]</sup>。因此一个准确的肺结节自动检测系统将给影像科医生提供巨大帮助,有望真正成为医生的好帮手。

传统肺结节计算机辅助检查技术虽然可以提高肺结节检测的效率和准确性,但是由于该方法对磨玻璃结节检出率低、假阳性率高等问题,因此在临床应用并未常规使用<sup>[7,8]</sup>。随着医疗大数据库的建立及计算机硬件水平的发展,深度学习算法在医疗诊断领域优势明显,为实现医学影像的自动智能诊断提供了新的契机<sup>[9]</sup>。目前,虽然部分研究者采用深度学习算法对肺结节进行自动检测<sup>[10,11]</sup>,但大部分算法还是基于2D或2.5D影像<sup>[12]</sup>,没有有效利用CT影像中的三维信息。因此,笔者建立了基于三维卷积神经网络的肺结节自动检出模型,并对其在独立测试集上进行验证。

## 材料与方法

### 1. 病例资料

本研究经长征医院伦理委员会批准并免除了患者知情同意书。本研究属于回顾性研究,所涉及到的患者信息均进行了脱敏处理。所有数据分为训练集及测试集两部分,训练集用于训练三维卷积神经网络,测试集用于验证三维卷积神经网络的效能。训练数据集来自2018~2019年多家医院以及临床机构收集到的回顾性数据。训练数据集分别来自体检、门诊及住院患者中采集的胸部CT影像。

病例纳入标准:①行薄层CT(层厚 $\leq 1$  mm)检查;②术前未有治疗史;③术后病理报告完善;④所有结节大小均 $\leq 3$  cm;⑤薄层CT上无明显空洞。

病例排除标准:①肺叶扫描不全;②图像存在严重伪影;③图像存在缺层或断层;④图像不符合DICOM 3.0协议。

通过纳入及排除标准的筛选,最终7144例患者共45662个结节纳入训练集。7144例患者中女4000例(56%),年龄20~96岁,平均(51.0 $\pm$ 14.7)岁;男3144例(44%),年龄19~94岁,平均(52.2 $\pm$ 14.3)岁。

测试数据集来自长征医院2012年7月至2017年3月回顾性纳入的400例患者(男173例,女227例,年龄27~80岁)。采用与训练集相同的纳入标准与排除标准,最终经过筛选96例患者共733个肺结节纳入测试集。所有人组病例均为多发结节,每个结节视为独立结节,本研究没有考虑不同结节之间的相关性影响。训练集与测试数据集中肺结节大小和类型的分布见表1。

表1 数据集中不同大小、不同类型结节的分布

类型	训练数据集		测试集	
	结节个数(个)	患者例数(例)	结节个数(个)	患者例数(例)
实性	32223	6510	277	80
<5mm	26925	6132	240	74
5~10mm	4390	2590	35	25
>10mm	908	830	2	1
磨玻璃	3589	2291	390	95
<5mm	1677	1264	153	71
5~10mm	1394	1098	178	82
>10mm	518	496	59	42
胸膜	4681	2824	39	24
钙化	5169	2320	27	19
汇总	45662	7144	733	96

### 2. 图像采集

测试集数据扫描范围自肺尖至后肋膈角下缘,包括两侧胸壁、腋窝。患者吸气末屏气扫描,避免呼吸运

表2 测试集5台CT机器详细扫描参数

相关参数	CT16(1)	CT128(2)	GE64(3)	CT128(4)	CT256(5)
厂家	Toshiba	Philips	General Electric	Philips	Philips
管电压(kV)	120	120	120	120	120
管电流(mA)	50~150	50~150	Auto	50~150	50~150
探测器准直(mm)	16 $\times$ 0.5	64 $\times$ 0.625	64 $\times$ 0.625	64 $\times$ 0.625	128 $\times$ 0.625
矩阵	512 $\times$ 512	512 $\times$ 512	512 $\times$ 512	512 $\times$ 512	512 $\times$ 512
重建层厚(mm)	1.000	1.000	1.000	1.000	0.625

动伪影。所有测试病例均行胸部薄层 MSCT 扫描, CT 图像采集来自 5 台 CT 机器, 详细扫描参数见表 2。

### 3. 标准测试数据集

参照胸部 CT 肺结节数据标注与质量控制专家共识<sup>[13]</sup>, 建立本次肺结节测试数据集。对入组的 733 个结节的 CT 影像学特征在标准肺窗(窗宽 1500 HU, 窗位 -450 HU)进行标注, 每个结节均由 3 位胸部影像诊断医师(5 年以上胸部亚专业工作经验)同步盲法诊断并标注, 判断结节类型, 当诊断结果不一致时, 由第 4 位高年资亚专业国内权威影像医师作为仲裁专家进行判定, 最后汇总意见作为结节诊断及标注金标准。本测试数据库根据结节大小分为 3 组: <5 mm、5~10 mm 和 >10 mm。结节密度分为钙化结节与非钙化结节(实性结节、亚实性结节)。其中, 亚实性结节进一步分为纯磨玻璃密度结节与混杂磨玻璃结节, 纯磨玻璃密度结节指肺实质内圆型或类圆形边界清楚的密度增高影, 其内血管及支气管显示清晰, 通过纵隔窗判断有无实性肿瘤成分, 无实性成分者为纯磨玻璃密度结节, 反之判断为混杂磨玻璃密度结节<sup>[14]</sup>。特殊部位的结节主要纳入胸膜结节, 定义为与胸膜广基底相连的圆形、类圆形或不规则形的局灶性密度增高影。结节尺寸测量以肺部 CT 肺结节数据标注与质量控制专家共识为标准, 在标准肺窗测量, 选取结节最大横截面长径及短径计算其平均直径, 平均直径为(长径+短径)/2<sup>[13]</sup>, 并记录检出结果。本次测试集总纳入 733 个肺结节, 其中实性结节 277 个, 磨玻璃结节 390 个, 胸膜结节 39 个, 钙化结节 27 个。肺结节大小范围为 3~30 mm, 平均肺结节直径为 6.5 mm, 其中恶性结节 145 个, 均经病理证实。

### 4. CAD 性能统计学指标

本研究从肺结节检出的灵敏度(Sensitivity), 精准度(Precision)以及平均每例假阳性个数来衡量肺结节筛查算法的综合性能。真阳性检出(True Positive)指当 CAD 系统检出的肺结节定位框与金标准中任意一个肺结节定位框有重合时, 则这个肺结节检出被认为是一个真阳性检出; 反之, 则认为这个肺结节检出是假阳性检出。CAD 系统的肺结节灵敏度定义为 CAD 系统真阳性个数/金标准肺结节总个数; 精准度定义为 CAD 系统真阳性个数/CAD 系统检出肺结节的总个数; 平均每例假阳性个数定义为总共假阳性检出数量/总共测试集的病例个数。肺结节检出的灵敏度反映了 CAD 系统对肺结节的检出率, 精准度和平均每例假阳性个数体现了 CAD 系统肺结节检出的假阳性。

分别使用本研究建立的三维卷积神经网络模型, Siemens syngo.via(VB 3.0)及 Philips ISP(V8)CAD

软件进行标准测试数据集中肺结节的检出。通过与医生确立的肺结节“金标准”对比, 计算三类方法在肺结节检出的灵敏度、精准度以及平均每例患者假阳性检出的个数, 同时比较对于不同大小、不同类型肺结节检出的灵敏度。

## 结果

在测试集 733 个肺结节中, 三维卷积神经网络算法模型共检出 658 个结节, 另外有 273 个为假阳性检出。整体肺结节检出灵敏度为 90%, 精准度为 71%, 平均每例患者假阳性检出 2.8 个。验证的三维卷积神经网络的自由响应受试者工作特征(free-response receiver operating characteristics, fROC)曲线<sup>[15]</sup>见图 1, 绿色框在 fROC 曲线所对应的点是该三维卷积神经网络的临床应用的操作点。

Siemens syngo.via 工作站 CAD 软件共检出 263 个结节, 另外有 117 个为假阳性检出, 整体肺结节检出灵敏度为 36%, 精准度为 69%, 平均每例患者假阳性检出 1.2 个。在同样的假阳性检出下, fROC 曲线显示三维卷积神经网络的灵敏度为 67%, z 检验表明该灵敏度显著优于 syngo.via 工作站( $P < 0.05$ )。

Philips ISP 工作站 CAD 软件共检出 249 个结节, 另外有 90 个为假阳性检出, 整体肺结节检出灵敏度为 34%, 精准度为 73%, 平均每例患者假阳性检出 0.9 个(表 3)。在同样的假阳性检出下, 三维卷积神经网络的灵敏度为 59%, z 检验表明该灵敏度显著优于 Philips ISP( $P < 0.05$ )。

表 3 基于标准测试集的肺结节检测各项指标

检出方法	灵敏度 (%)	精准度 (%)	平均每例假阳性检出
三维卷积神经网络	90	71	2.8
Siemens syngo.via	36	69	1.2
Philips ISP	34	73	0.9

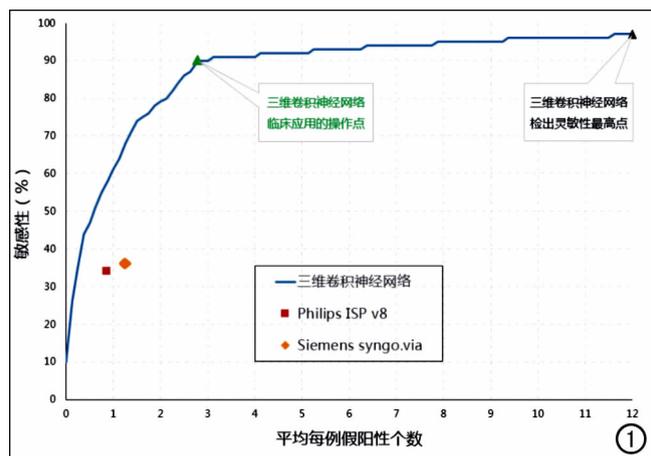


图 1 三维卷积神经网络的 fROC 曲线。

表4 测试集中不同大小、不同类型的肺结节检出个数及灵敏度

检出方法	实性			磨玻璃			胸膜	钙化	汇总
	<5mm	5~10mm	>10mm	<5mm	5~10mm	>10mm			
三维卷积神经网络	237(99)	35(100)	2(100)	109(71)	157(88)	52(88)	39(100)	27(100)	658(90)
Siemens Syngo. via	38(16)*	35(100)	1(50)	21(14)*	99(56)*	40(68)*	9(23)*	20(74)*	263(36)*
Philips ISP	35(15)*	32(91)*	2(100)	16(11)*	84(47)*	49(83.1)	13(33)*	18(67)*	249(34)*

注:表中数据分别表示肺结节检出个数及灵敏度(括号);\*表示在该类型的肺结节检出上,三维卷积神经网络显著优于所比较方法,差异有统计学意义(Z检验, $P<0.05$ )。

针对不同大小、不同类型的肺结节进行分层统计,分别计算三维卷积神经网络、Siemens syngo. via 以及 Philips ISP 对于各类结节检出的灵敏度。对于三维卷积神经网络,<5 mm 的实性结节检出 237 个(99%),5~10 mm 实性结节检出 35 个(100%),>10 mm 实性结节检出 2 个(100%);<5 mm 的磨玻璃结节检出 109 个(71%),5~10 mm 的磨玻璃结节检出 157 个(88%),>10 mm 的磨玻璃结节检出 52 个(88%);胸膜结节检出 39 个(100%);钙化结节检出 27 个(100%)(表 4)。

通过与 syngo. via 以及 Philips ISP 进行对比,可以看出三维卷积神经网络几乎在所有类别的肺结节检出上都有更高的灵敏度,尤其在<5 mm 的实性结节、磨玻璃结节、胸膜和钙化结节上,其灵敏度的优势更为明显。

## 讨论

肺结节早期检出并明确诊断对肺结节的临床管理至关重要。我们前期基于回顾性数据建立了基于三维卷积神经网络算法的肺结节深度学习算法模型,所建立的肺结节检出算法是基于特征金字塔网络的检测框架<sup>[16]</sup>,采用三维卷积操作来替代传统的二维卷积操

作。整个算法的流程图如图 2 所示,通过多个检测网络对胸部 CT 肺结节进行检出。基于深度学习计算的肺结节影像特征,每个检测网络会对图像中的任意一个位置是否存在肺结节进行判断。只有所有检测网络一致认为存在肺结节的位置,才会被算法输出给医生进行审阅。通过结合多个深度学习检测网络,该算法模型可以有效并准确地检出肺结节的位置。通过对模型的准确性进行独立验证,并与传统 CAD 系统(Siemens syngo. via VB 3.0 和 Philips ISP V8)在肺结节检出灵敏度、精准度以及平均每例假阳性个数等多个指标方面进行优效验证,结果显示该模型基于三维卷积神经网络,可以充分反映每枚结节的三维空间信息,较传统 CAD 模型具有较明显的优势。基于三维卷积神经网络算法的肺结节自动检测模型较传统 CAD 肺结节检出系统具有更高的灵敏度,整体肺结节检出灵敏度为 90%,精准度为 71%,平均每例患者假阳性检出 2.8 个,尤其对于<5mm 的肺结节的检出提升效果更为显著。在不考虑结节大小对模型假阳性影响的情况下,对结节假阳性率的控制稍优于 Li 等<sup>[17]</sup>的研究结果,该算法模型在实性结节,胸膜结节及钙化结节检出上有着较高的灵敏度,但在磨玻璃结节检出上灵敏度仍有不足。主要原因是由于磨玻璃结节在训练集中

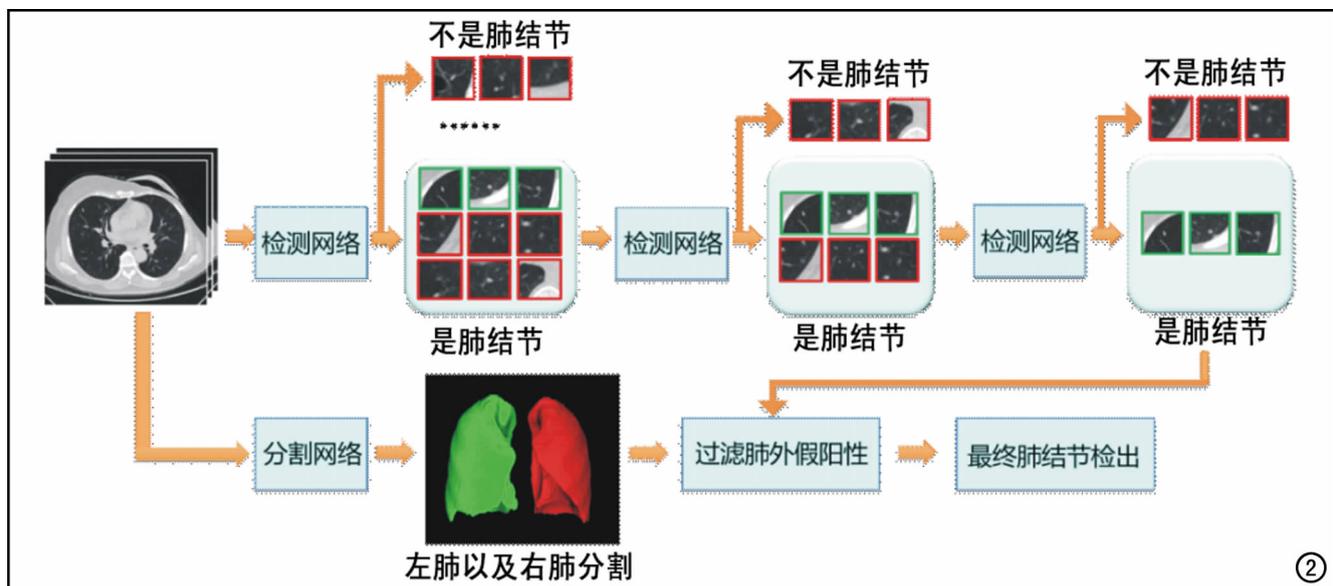


图2 基于级联学习的肺结节辅助筛查流程图,绿色方块表示肺结节,红色方块表示非肺结节。

仅占 7.8%，大量标注结节为实性结节，因此训练得到的算法模型偏向于学习实性结节的影像特征，对实性结节表现灵敏，而对样本量较少的磨玻璃结节则相对不灵敏。

通过对测试集假阳性结节进行回顾性研究，笔者发现导致误诊的主要原因包括扩张支气管内分泌物、迂曲增粗的血管、正常小叶核心结构、马赛克样局限气体滞留及一些不同病因导致的炎性感染病变等。同样对于密度较淡磨玻璃结节、隐藏在血管旁结节、肺门旁结节、气管内结节及胸膜旁结节是主要的漏检原因。对于一款肺癌筛查软件来说，较高的假阳性率在肺结节临床筛查中尚可接受，但假阴性率需尽力避免，因此，本研究模型虽然灵敏度较高，但对于临床应用来说仍有较大的提升空间。

通过对本研究的初级模型的检测结果可以看出，由于深度学习算法需要大量带有肺结节标注的胸部 CT 影像数据来进行训练，数据的多样性以及标注质量直接决定了算法的效能；从而反映出一个深度学习的算法模型，其训练集数据的量和数据的标注质量至关重要，同时训练集的数据分布也直接影响着模型的鲁棒性<sup>[18]</sup>。

本研究有以下局限性：①本研究属于回顾性研究且测试集样本量相对较少，存在选择偏倚；②没有对结节特殊位置进行详细分类；③不同类别的样本量不均衡，会导致算法训练性能下降，这是机器学习中常见的问题<sup>[19]</sup>。下一步我们将建立经过质控达标的标准测试集来验证深度学习模型的效能，并对漏诊结节分析其漏诊原因，有针对性地增加模型训练集分类权重进一步迭代优化。

综上所述，三维卷积神经网络算法模型相较于传统 CAD 系统，在肺结节检出灵敏度上有显著提升，通过对漏诊结节的补充训练迭代，有望在临床上提高医生的阅片效率并降低漏诊率，成为影像医生的助手。当然，这仅仅是用于结节筛查，肺结节鉴别诊断模型有待进一步研发验证。

#### 参考文献：

[1] Jacobs CD, Jafari ME. Early results of lung cancer screening and radiation dose assessment by low-dose CT at a community hospital[J]. Clin Lung Cancer, 2017, 18(5): 327-331.

[2] Torre LA, Siegel RL, Jemal A. Lung cancer statistics[J]. Adv Exp Med Biol, 2016, 893(1): 1-19.

[3] Henschke CI. Early lung cancer action project: overall design and

findings from baseline screening[J]. Cancer, 2000, 89(S11): 2474-2482.

[4] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018[J]. CA Cancer J Clin, 2018, 60(5): 277-300.

[5] 胡琼洁, 陈冲, 王玉锦, 等. 实习医师以共同阅片和第二阅片者模式使用计算机辅助检测系统在低剂量 CT 中的应用研究[J]. 放射学实践, 2018, 33(10): 1022-1028.

[6] Sui Y, Wei Y, Zhao D. Computer-aided lung nodule recognition by SVM classifier based on combination of random undersampling and SMOTE[J]. Comput Math Methods Med, 2015, 12(1): 1-13.

[7] Murphy K, Ginneken BV, Schilham AM, et al. A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification[J]. Med Image Anal, 2009, 13(5): 757-770.

[8] Messay T, Hardie RC, Rogers SK. A new computationally efficient CAD system for pulmonary nodule detection in CT imagery[J]. Med Image Anal, 2010, 14(3): 390-406.

[9] 金征宇. 人工智能医学影像应用: 现实与挑战[J]. 放射学实践, 2018, 33(10): 989-991.

[10] Ather S, Kadir T, Gleeson F. Artificial intelligence and radiomics in pulmonary nodule management: current status and future applications[J]. Clin Radiol, 2019 Jun 12. pii: S0009-9260(19)30210-7. DOI: 10.1016/j.crad.2019.04.017.

[11] Setio AAA, Traverso A, de Bel T, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge[J]. Med Image Anal, 2017, 42(12): 1-13.

[12] Gong L, Jiang S, Yang Z, et al. Automated pulmonary nodule detection in CT images using 3D deep squeeze-and-excitation networks[J]. Int J Comput Assist Radiol Surg, 2019, 26(4): 1-11.

[13] 中国食品药品检定研究院, 中华医学会放射学分会心胸学组. 胸部 CT 肺结节数据标注与质量控制专家共识(2018)[J]. 中华放射学杂志, 2019, 53(1): 9-15.

[14] 邹勤, 刘士远, 管宇, 等. 超高分辨率 CT 对 3cm 以下肺磨玻璃密度结节的诊断价值[J]. 临床放射学杂志, 2017, 36(4): 484-488.

[15] Bandos AI, Rockette HE, Gur SD. Area under the free-response ROC curve (FROC) and a related summary index[J]. Biometrics, 2009, 65(1): 247-256.

[16] Lin TY, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]. Las Vegas: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 936-944.

[17] Li X, Guo F, Zhou Z, et al. Performance of deep-learning-based artificial intelligence on detection of pulmonary nodules in chest CT[J]. Chin J Lung Cancer, 2019, 22(6): 336-340.

[18] 萧毅, 刘士远. 肺结节影像人工智能技术现状与思考[J]. 肿瘤影像学, 2018, 27(4): 8-11.

[19] Japkowicz N, Stephen S. The class imbalance problem: A systematic study[J]. Intell Data Anal, 2002, 6(5): 429-449

(收稿日期: 2019-07-29 修回日期: 2019-08-11)