

• 本刊特稿 •

基于卷积神经网络的深度学习算法对颅内出血的类型识别及血肿分割一致性的研究

李娟, 汤翔宇, 沈逸, 廖术, 石峰, 朱文珍

【摘要】目的:开发一种可以检测不同类型颅内出血并自动计算血肿体积的基于卷积神经网络的深度学习算法,探讨其识别的准确性及血肿分割的一致性。**方法:**数据集1纳入9594例颅脑CT平扫图像,随机选取223例颅内出血阳性患者作为颅内出血类型识别的测试集,剩余CT图像作为其训练集,评估测试集中算法识别五种不同类型颅内出血的效能。数据集2选取另外819例已人工勾画出血灶的CT图像,随机选取74例作为测试集,以人工手动分割为金标准,验证测试集中算法分割与人工分割的一致性。**结果:**在223例颅内出血阳性患者中,深度学习算法对五种类型颅内出血识别的曲线下面积均大于或接近0.85,特异度均大于0.95;在74例血肿分割测试数据中,算法自动测量的血肿体积与人工手动分割测量的血肿体积之间达到较高的一致性,脑实质内出血、硬膜外出血、脑室内出血及硬膜下出血体积测量的组内相关系数分别为1、0.990、0.996和0.878。**结论:**基于卷积神经网络的深度学习算法可以较好地识别不同类型的颅内出血,并能精确测量血肿体积,具有一定的临床应用前景。

【关键词】卷积神经网络;深度学习;颅内出血;血肿分割;体层摄影术,X线计算机

【中图分类号】R743.9;R814.42;TP18 **【文献标志码】**A

【文章编号】1000-0313(2021)01-0007-06

DOI:10.13609/j.cnki.1000-0313.2021.01.002

开放科学(资源服务)标识码(OSID):



Evaluation of intracranial hemorrhage subtype recognition and hematoma segmentation consistency using convolution neural network LI Juan, TANG Xiang-yu, SHEN Yi, et al. Department of Radiology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China

【Abstract】 Objective: To develop a deep learning algorithm based on convolutional neural network to detect different types of intracranial hemorrhage and automatically calculate the volume of hematoma. And to explore its accuracy and consistency of hematoma segmentation.**Methods:** 9594 cases of non-contrast head CT scan images were retrospectively included in Dataset 1. 223 positive cases were randomly selected as the testing dataset for intracranial hemorrhage subtype recognition, and the rest CT images were used as the training dataset. The effectiveness of the test set algorithm in recognizing five different types of intracranial hemorrhage were evaluated. Another 819 manually labeled CT scans were enrolled in Dataset 2, from which 74 cases were randomly selected as the testing dataset. Manual segmentation was taken as the gold standard to verify the consistency of algorithm segmentation and manual segmentation in the test set. **Results:** Among 223 patients with positive intracranial hemorrhage, the area under the curve (AUC) for five types of intracranial hemorrhage in the subtype recognition testing dataset were all above or approximate to 0.85, and the specificity were greater than 0.95. Among the 74 cases of hematoma segmentation test data, there was a high consistency between the volume of hematoma automatically measured by the algorithm and the volume of hematoma manually measured by the algorithm. The intraclass correlation coefficients of intraparenchymal hemorrhage, epidural hemorrhage, intraventricular hemorrhage and subdural hemorrhage were 1, 0.990, 0.996 and 0.878 respectively.

作者单位:430030 武汉,华中科技大学同济医学院附属同济医院放射科(李娟,汤翔宇,朱文珍);上海联影智能医疗科技有限公司(沈逸,廖术,石峰)

作者简介:李娟(1985—),女,硕士研究生,主治医师,主要从事中枢神经影像诊断工作。

通信作者:朱文珍,E-mail:zhuwenzhen8612@163.com

基金项目:国家自然科学基金青年基金(51907077)

0.878 respectively in the hematoma segmentation testing dataset. **Conclusion:** Deep learning algorithm based on convolution neural network has good performance in intracranial hemorrhage recognition and hematoma segmentation, which has a certain prospect of clinical application.

【Key words】 Convolution neural network; Deep learning; Intracranial hemorrhage; Hematoma segmentation; Tomography, X-ray computer

颅内出血是临床常见急重症,具有极高的致死率及致残率,严重危害患者的生命健康安全。颅内出血的病因多种多样,常见病因包括外伤、血管性疾病及先天发育异常等。根据出血部位的不同,颅内出血又可以分为脑实质内出血(intraparenchymal hemorrhage,IPH)、脑室内出血(intraventricular hemorrhage,IVH)、硬膜下出血(subdural hemorrhage,SDH)、硬膜外出血(extradural hemorrhage,EDH)及蛛网膜下腔出血(subarachnoid hemorrhage,SAH)五种类型。不同的出血类型及程度决定了患者是否需要外科手术处理等干预措施,而出血量反映了出血的严重程度,与患者的预后相关,并且是临床进行随访复查的重要指标。对不同类型颅内出血的早期检出对于挽救患者的神经功能乃至生命有重要意义,而对血肿体积的精确测量则为临床决策提供了重要依据。

CT 广泛应用于急诊科颅内出血的诊断。近几十年来,全国每年进行 CT 检查的人数逐年增长。阅读和分析大量的 CT 图像数据不仅耗费医生大量的时间及精力,同时也增加了漏诊和误诊的可能性。目前,急诊颅脑 CT 诊断,尤其是夜班的急诊颅脑 CT 初步诊断多由低年资放射科医生提供,后期再由高年资放射科医生进行二次审核。已有多项研究显示低年资放射科医生提供的初步诊断存在不同程度的漏诊和误诊^[1,2]。

近年来,人工智能(artificial intelligent,AI)特别是深度学习(deep learning)算法作为大数据处理的有效手段,在医疗影像研究中得到了非常广泛的应用。以卷积神经网络为代表的相关深度学习算法在肺结节监测及良恶性判别^[3,4]、肋骨骨折检测^[5]、新型冠状病毒肺炎诊断、病变分割及体积测量^[6,7]、糖尿病视网膜病变分类^[8]、淋巴结组织学切片中转移的检测^[9]及皮肤癌图像的分级^[10]等方面取得了较好的效果,其精确度可与专家评估相媲美。部分学者也尝试将深度学习算法应用于颅内出血的研究^[11-13],但均局限于单一的出血检测或血肿分割,且多数研究建立在较小的训练集样本量基础上。因此,我们开发了一套建立在较大样本量训练集基础上的深度学习算法来检测颅内出血并自动测量血肿体积,以期有助于提高临床诊断准确率及降低误诊漏诊率,同时方便临床医生对血肿进行随访评估,辅助临床决策。

材料与方法

1. 数据集的采集

回顾性搜集 2016 年 1 月—2018 年 12 月期间我院门诊及住院部以及 qure.ai 开源数据集的 9594 例颅脑 CT 平扫图像,其中颅内出血阳性 6643 例,阴性 2951 例,所有患者信息已匿名处理。涉及的 CT 扫描机器主要包括 GE、西门子、东芝、飞利浦、联影等多种机型及多种参数,图像层厚 1.5~7.5 mm 不等,多数为 5 mm。qure.ai 开源数据集所提供的图像数据由三位资深放射科医师的一致诊断为金标准,当三位医师中两位与另外一位的诊断意见不同时以多数者的诊断为金标准;其余图像均由两位资深放射科医生的一致诊断为金标准,当诊断意见不一致时通过协商达成一致。图像排除标准为图像不全、伪影严重及颅脑术后。随机选取 223 例颅内出血阳性患者的数据作为颅内出血类型识别的测试集,剩余 CT 图像数据作为其训练集。

另选取我院门诊及病房 819 例已进行人工手动勾画血肿边界的急性颅内出血 CT 图像数据来研究血肿分割的精准性。排除标准为图像不全、严重伪影、颅脑术后及同一患者复查。随机选取 74 例颅内出血阳性患者的数据作为测试集,剩余 CT 图像数据作为其训练集。人工手动分割采用开源软件 Insight Toolkit SANP(ITK-SNAP) 3.6.0 进行分割,于 CT 平扫横轴面图像沿血肿边界进行勾画,标注后的血肿由软件生成 1 个分割文件,并输出血肿体积。

颅内出血分类识别模型及血肿分割模型的构建由上海联影智能医疗科技有限公司算法部门实施。

2. 模型的建立及算法

首先进行 CT 图像的数据输入,所有 CT 图像均为标准的 DICOM 格式。数据预处理包括图像摆正、去头骨和灰度归一化。在横轴面数据中,基于深度学习对脑中线两个端点的位置进行检测,并对横轴面 CT 图像进行旋转,进而自动摆正脑部位置。然后,基于深度学习自动对脑组织区域进行分割,自动剔除图像中包括颅骨及之外的干扰信息。在将灰度归一化到 [-1, 1]之后,根据五种出血类型及未出血共六个标签,对图像的每一层调用残差网络(ResNet)进行分类。对每一层的分类结果,调用循环神经网络(long short term memory network,LSTM)进行结果修正,

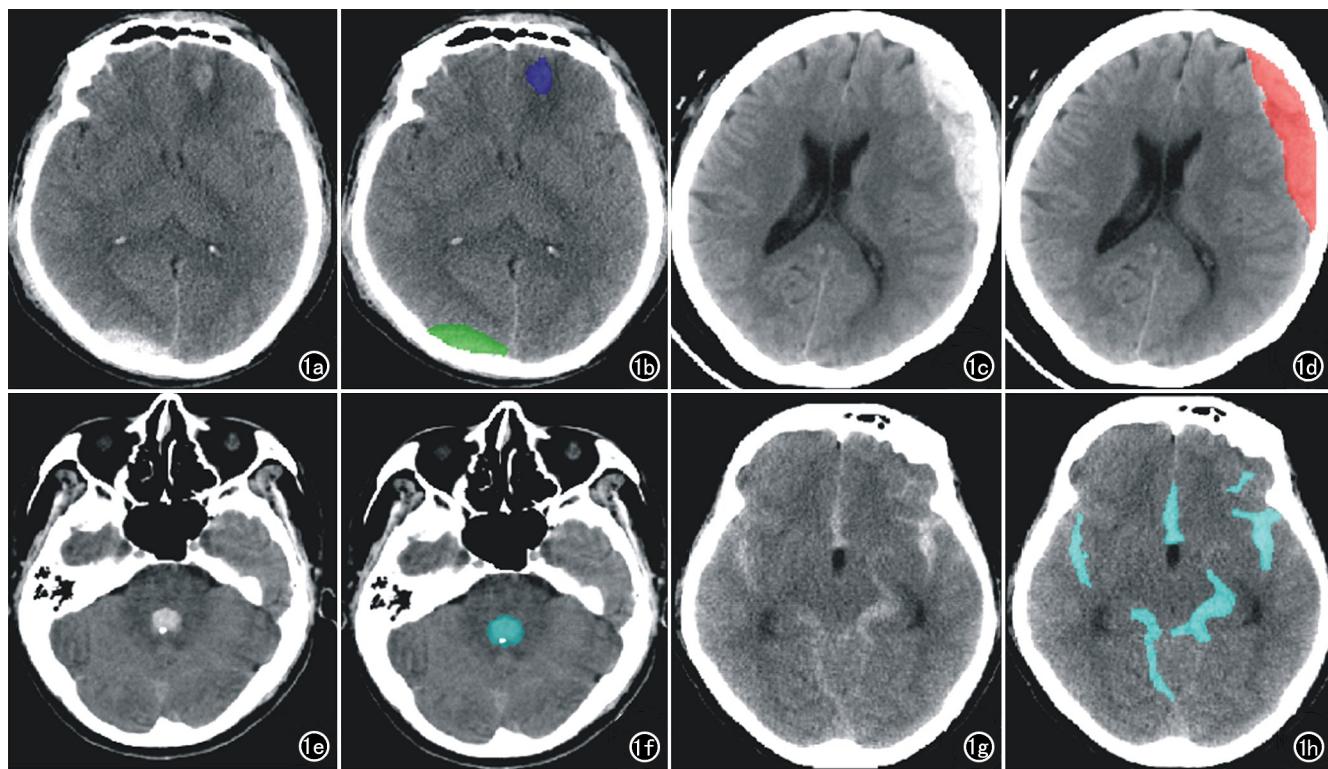


图 1 深度学习算法对五种不同类型颅内出血识别的示意图。a、b)深度学习算法对脑实质出血及硬膜外出血的识别;c、d)深度学习算法对硬膜下出血的识别;e、f)深度学习算法对脑室内出血的识别;g、h)深度学习算法对蛛网膜下腔出血的识别。

最终合并为基于患者个体的分类结果。算法细节参考 2019 年 RSNA 脑出血分类竞赛^[14]。

对于出血灶分割任务,在图像预处理之后,调用 V 型网络(VB-Net)进行模型训练,并通过体素的统计和间距的换算,自动得到每个出血量的统计从而计算血肿体积。同时,采用同样网络对 CT 图像中侧脑室进行分割,并通过与出血灶的匹配自动得到是否存在脑室内出血的结果^[15]。

3. 统计学分析

所有数据采用 SPSS 23.0 和 MedCalc 19.0 统计软件进行处理与分析。算法识别颅内出血类型的效能采用准确率(%)、敏感度、特异度及受试者操作特征(receiver operating characteristic, ROC)曲线分析中曲线下面积(area under curve, AUC)来评估。不同类型的血肿体积先进行 Kolmogorov-Smirnov 正态性检验,符合正态分布的数据以均数±标准差表示,不符合正态分布者以中位数(上、下四分位数)表示。人工手动分割和卷积神经网络算法分割血肿体积之间的一致性采用组内相关系数(intraclass correlation coefficient, ICC)分析并绘制 Bland-Altman 图。以 $P < 0.05$ 为差异有统计学意义。

结 果

1. 深度学习算法对五种类型颅内出血的识别

223 例 CT 图像数据纳入测试集,其中 62.78%(140/223)合并有两种或两种以上类型的颅内出血。测试集中五种类型颅内出血的分布见表 1,占比最高的是脑实质出血(IPH>SAH>IVH>SDH>EDH)。深度学习算法对不同类型颅内出血识别的示意图见图 1。

表 1 测试集不同类型颅内出血的分布(例)

| 总病例数 | IPH | IVH | EDH | SDH | SAH |
|------|-----|-----|-----|-----|-----|
| 223 | 156 | 77 | 60 | 63 | 115 |

深度学习算法对五种类型颅内出血的识别效能见表 2 和图 2,其 AUC 均高于或接近 0.85,特异度均高于 0.95,尤其是对于脑实质出血与脑室内出血,深度学习算法具有良好的识别效能(AUC 为 0.922, 95%CI: 0.879~0.954)。

表 2 深度学习算法对五种类型颅内出血的识别效能

| 类型 | 准确率 | 敏感度 | 特异度 | AUC | 95% 可信区间 |
|-----|--------|--------|--------|-------|-------------|
| IPH | 0.8969 | 0.8590 | 0.9851 | 0.922 | 0.879~0.954 |
| IVH | 0.9462 | 0.8442 | 1 | 0.922 | 0.879~0.954 |
| EDH | 0.9148 | 0.8000 | 0.9571 | 0.879 | 0.828~0.918 |
| SDH | 0.9013 | 0.7143 | 0.9750 | 0.845 | 0.790~0.890 |
| SAH | 0.8565 | 0.7565 | 0.9630 | 0.860 | 0.807~0.902 |

表 3 人工手动分割方法与深度学习算法测量血肿体积的一致性

| 类型 | 病例数 | 体积 (深度学习算法) (mL) | 体积 (人工手动分割) (mL) | ICC | 95% 可信区间 |
|-----|-----|------------------------|------------------------|-------|-------------|
| EDH | 20 | 14.79(5.51~50.92) | 15.31(7.25~54.44) | 0.990 | 0.975~0.996 |
| IPH | 20 | 11.04(5.74~23.87) | 10.75(5.80~23.59) | 1.000 | 0.999~1.000 |
| IVH | 23 | 17.51(10.21~31.66) | 22.13(11.00~33.89) | 0.996 | 0.990~0.998 |
| SDH | 11 | 55.98(29.46~82.35) | 83.13(34.01~111.48) | 0.878 | 0.612~0.966 |

五种类型颅内出血中,算法识别准确率最高的为脑室内出血(IVH>EDH>SDH>IPH>SAH),敏感度最高的为脑实质出血(IPH>IVH>EDH>SAH>SDH)。

2. 人工手动分割方法与深度学习算法测量血肿体积的一致性

人工手动分割方法与深度学习算法对 74 例人工标注的不同类型的颅内出血体积测量结果的一致性见表 3 及图 3。对于四种不同的颅内出血类型,深度学习算法与人工手动分割方法测量的出血体积之间均显示了较好的一致性,其中硬膜下出血、脑实质内出血和脑室内出血体积测量的 ICC 接近于 1,硬膜下出血体积测量的 ICC 接近 0.9。

讨 论

本研究结果显示,我们开发的基于卷积神经网络的深度学习算法对不同类型颅内出血具有良好的识别效能(AUC 大于或接近 0.85,特异度均高于 0.95),且对除蛛网膜下腔出血以外的其他类型的颅内出血血肿

分割具有较高的一致性(ICC 高于或接近 0.9)。Ye 等^[13]及 Kuo 等^[12]采用不同的深度学习算法对颅内出血亚型进行检测,前者对不同亚型颅内出血识别的 AUC 达到 0.8 以上,后者更是达到专家水平,但前者的训练集仅包含 2836 例 CT 图像,后者的结果则建立在单中心小样本训练集上,测试集中颅内出血阳性仅 25 例。Chilamkurthy 等^[16]的研究纳入了印度 20 余家医疗中心的 30 余万例 CT 图像数据,对各种亚型颅内出血识别的 AUC 接近或高于 0.9,但该研究未涉及对血肿体积的测量。我们开发的模型纳入了较大样本量的 CT 数据,对不同类型颅内出血的识别效能高于或接近既往研究结果,同时可以自动测量血肿体积,更适合于临床应用与推广。

硬膜下出血和蛛网膜下腔出血是颅内出血中最容易漏诊或误诊的类型^[17],给深度学习算法的识别带来了很大挑战。Ye 等^[13]开发的深度学习算法对蛛网膜下腔出血识别的敏感度仅为 0.69,本模型对蛛网膜下腔出血和硬膜下出血识别的敏感度分别为 0.7565 和 0.7143。62.78%(140/223)的病例合并有两种或两种

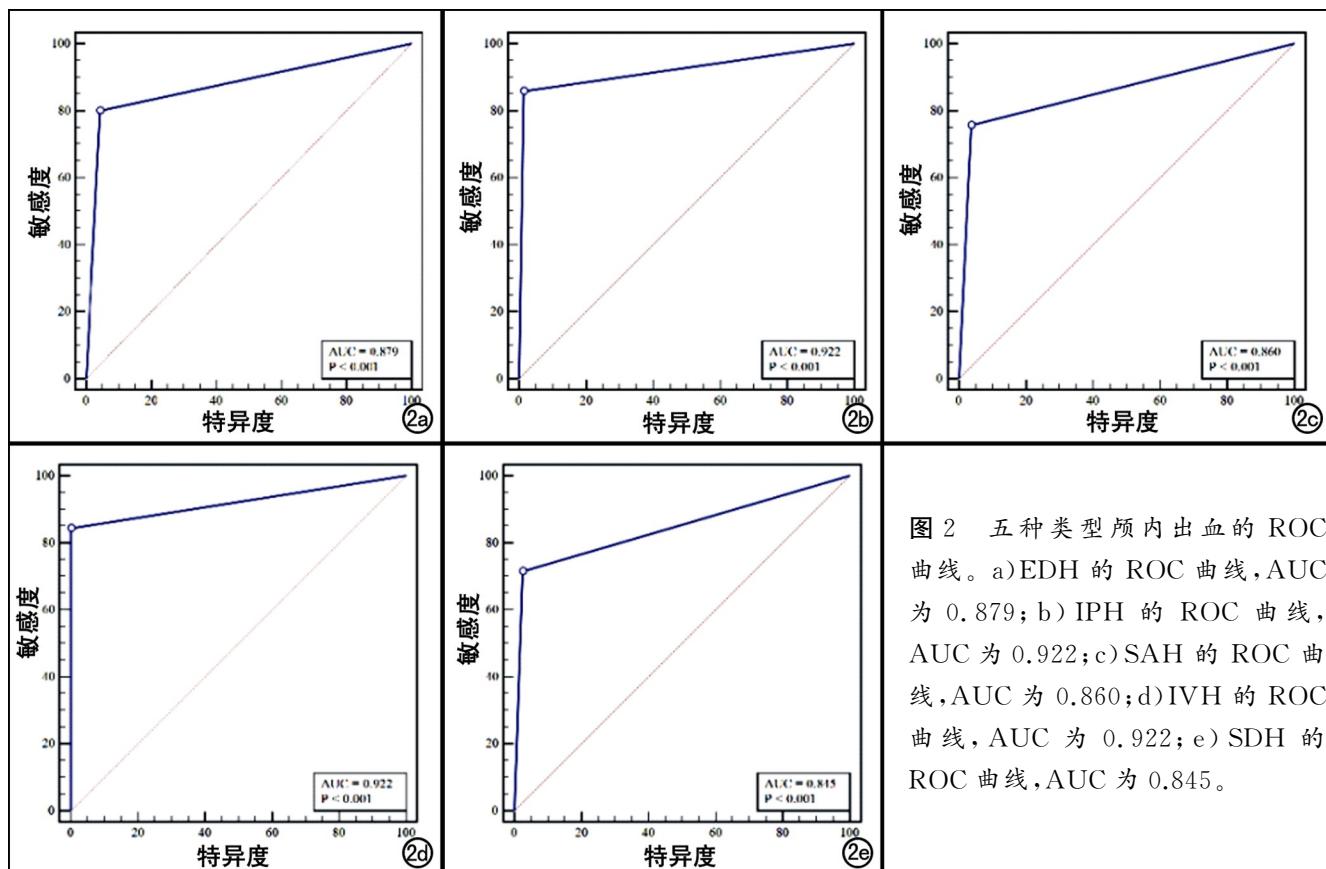


图 2 五种类型颅内出血的 ROC 曲线。a) EDH 的 ROC 曲线, AUC 为 0.879; b) IPH 的 ROC 曲线, AUC 为 0.922; c) SAH 的 ROC 曲线, AUC 为 0.860; d) IVH 的 ROC 曲线, AUC 为 0.922; e) SDH 的 ROC 曲线, AUC 为 0.845。

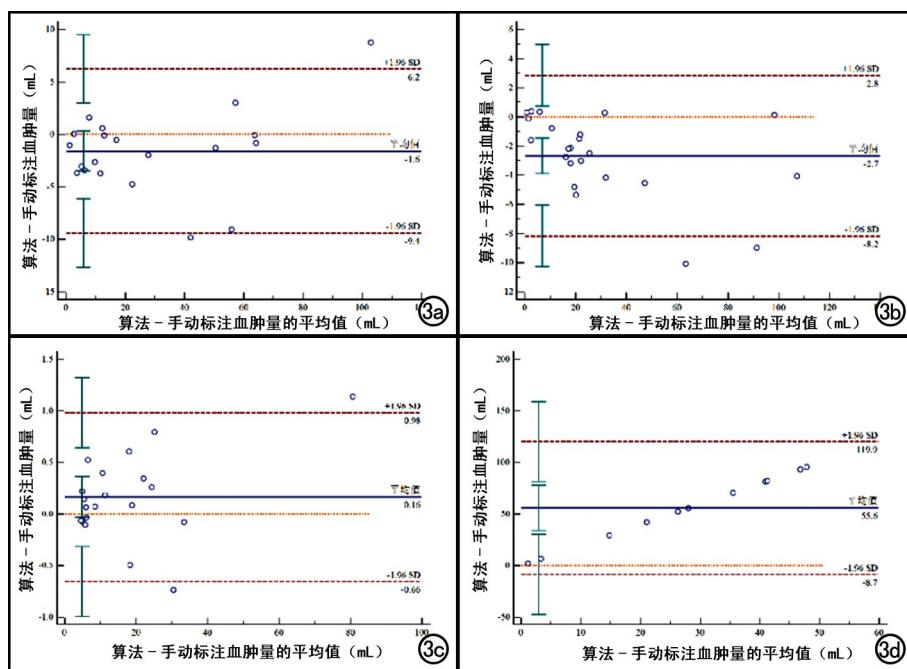


图 3 人工手动分割方法与深度学习算法测量血肿体积的一致性 Bland-Altman 图。a) 硬膜外出血(EDH): 95%一致性界限(95%LoA)范围为 $-9.41\sim6.22\text{mL}$, 10%(2/20)的点在95%LoA以外; b) 脑室内出血(IVH): 95%LoA范围为 $-8.18\sim2.84\text{mL}$, 8.7%(2/23)的点在95%LoA以外; c) 脑实质出血(IPH): 95%LoA范围为 $-0.66\sim0.98\text{mL}$, 10%(2/20)的点在95%LoA以外; d) 硬膜下出血(SDH): 95%LoA范围为 $8.71\sim119.94\text{mL}$, 0%(0/11)的点在95%LoA以外。

以上类型的颅内出血可能是本模型敏感度低的原因之一。对模型中不同类型颅内出血假阴性病例进一步分析,发现28例蛛网膜下腔出血假阴性病例中35.71%(10/28)为脑挫裂伤合并少许蛛网膜下腔出血,21.43%(6/28)为脑室积血合并少许蛛网膜下腔积血;18例硬膜下出血假阴性病例中44.44%(8/18)被误诊为硬膜外出血(同时也增加了硬膜外出血的假阳性率),16.67%(3/18)为脑挫裂伤合并少许硬膜下出血;22例脑实质出血假阴性病例全部合并有两种或两种以上出血类型,其中63.64%(14/22)为脑挫裂伤合并少许脑实质出血,18.18%(4/22)为脑室周围脑实质出血破入脑室被误诊为原发性脑室内出血。深度学习算法对脑挫裂伤合并少量脑实质出血或蛛网膜下腔出血的识别以及对硬膜外出血与硬膜下出血的鉴别有待进一步提升。

目前比较公认的出血体积测量的金标准是人工手动勾画分割测量,但需要耗费大量的时间及精力,无法在临床大规模推广应用。对于脑实质出血,临床普遍采用多田公式($V=a\times b\times c\times 1/2$, 其中V代表血肿体积,a、b分别是CT片中血肿最大层面的最长径、最宽径,c是CT片中出现血肿的层面数)来计算血肿体

积^[18]。多项研究显示多田公式得到的血肿体积存在较大误差^[19,20],特别是在血肿体积较小或形态不规则的情况下^[21],而且测量的主观性较强,可重复性不高。而其它类型的颅内出血则没有公认的测量标准,部分研究中硬膜下出血和硬膜外出血延用了多田公式或调整的多田公式计算体积^[22-24],但由于多田公式的计算建立在血肿为椭球形的基础上^[18],而硬膜下出血及硬膜外出血的形态多为新月形及梭形,这种测量方式得到的血肿体积与实际出血量之间必然存在误差。我们开发的深度学习算法自动计算的出血体积与人工手动分割得到的血肿体积之间有很好的一致性,尤其是对于脑实质内出血(ICC为1),与既往研究结果一致或高于既往研究结果^[25];而且深度学习算法测量1例出血体积耗时约10秒,明显快于多田公式测量或人工手动勾画分割。鉴于蛛网膜下腔出血体积测量的临床意义较低,未将其纳入本研究中。

本研究存在一定局限性:首先,本研究为回顾性研究,可能存在选择偏倚;其次,虽然纳入了较大样本量的CT数据作为训练集,但测试集样本量较少,结果可能存在误差,后续仍需加大样本量进行验证。

综上所述,基于卷积神经网络的深度学习算法可以较准确地识别各种类型的颅内出血,并精确测量出血体积,有望应用于临床帮助减轻放射科医生的工作负担并降低其漏诊误诊率,为辅助临床决策提供依据。

参考文献:

- Lal NR, Murray UM, Eldevik OP, et al. Clinical consequences of misinterpretations of neuroradiologic CT scans by on-call radiology residents[J]. AJNR Am J Neuroradiol, 2000, 21(1): 124-129.
- Erly WK, Berger WG, Krupinski E, et al. Radiology resident evaluation of head CT scan orders in the emergency department[J]. AJNR Am J Neuroradiol, 2002, 23(1): 103-107.
- Gruetzmacher R, Gupta A, Paradice D. 3D deep learning for detecting pulmonary nodules in CT scans[J]. J Am Med Inform Assoc, 2018, 25(10): 1301-1310.
- 戴正行,胡春洪,王希明,等.基于 DenseNet 网络深度学习法 CT 图像人工智能分析技术判断肺结节良恶性[J].放射学实践, 2020, 35(4): 484-488.
- Zhou QQ, Wang J, Tang W, et al. Automatic detection and classifi-

- cation of rib fractures on thoracic CT using convolutional neural Network: accuracy and feasibility[J]. Korean J Radiol, 2020, 21(7):869-879.
- [6] Wang M, Xia C, Huang L, et al. Deep learning-based triage and analysis of lesion burden for COVID-19: a retrospective study with external validation[J]. Lancet Digit Health, 2020, 2(10): e506-e515.
- [7] 应伟峰,陈穹,王钢,等. AI与不同年资放射医师对新冠肺炎CT诊断效能比较[J].放射学实践,2020,35(10):1231-1234.
- [8] Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs[J]. JAMA, 2016, 316(22):2402-2410.
- [9] Ehteshami BB, Veta M, Johannes VDP, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer[J]. JAMA, 2017, 318(22): 2199-2210.
- [10] Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks[J]. Nature, 2017, 542(7639):115-118.
- [11] Lee H, Yune S, Mansouri M, et al. An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets[J]. Nat Biomed Eng, 2019, 3(3):173-182.
- [12] Kuo W, Hne C, Mukherjee P, et al. Expert-level detection of acute intracranial hemorrhage on head computed tomography using deep learning[J]. Proc Natl Acad Sci USA, 2019, 116(45): 22737-22745.
- [13] Ye H, Gao F, Yin Y, et al. Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network[J]. Eur Radiol, 2019, 29(11):6191-6201.
- [14] Seutao. RSNA2019_Intracranial-Hemorrhage-Detection: code for 1st place solution in intracranial hemorrhage detection challenge @ RSNA2019[Z]. 2019.
- [15] Han M, Yao G, Zhang W. Segmentation of CT thoracic organs by multi-resolution VB-nets[Z]. 2019.
- [16] Chilamkurthy S, Ghosh R, Tanamala S, et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study[J]. Lancet, 2018, 392(10162):2388-2396.
- [17] Strub WM, Leach JL, Tomsick T, et al. Overnight preliminary head CT interpretations provided by residents: locations of misidentified intracranial hemorrhage[J]. AJNR Am J Neuroradiol, 2007, 28(9):1679-1682.
- [18] Kothari RU, Brott T, Broderick JP, et al. The ABCs of measuring intracerebral hemorrhage volumes[J]. Stroke, 1996, 27(8):1304-1305.
- [19] 徐兴华,陈晓雷,张军,等.多田公式计算脑内血肿体积的准确性和可靠性[J].中国神经精神疾病杂志,2015,41(2):87-91.
- [20] 白莉,徐刚,郎嘉兴,等.颅内血肿CT容积测量与多田法测量的比较[J].临床放射学杂志,2010,29(10):1415-1417.
- [21] Huttner HB, Steiner T, Hartmann M, et al. Comparison of ABC/2 estimation technique to computer-assisted planimetric analysis in warfarin-related intracerebral parenchymal hemorrhage[J]. Stroke, 2006, 37(2):404-408.
- [22] Gebel JM, Sila CA, Sloan MA, et al. Comparison of the ABC/2 estimation technique to computer-assisted volumetric analysis of intraparenchymal and subdural hematomas complicating the GUSTO-1 trial[J]. Stroke, 1998, 29(9):1799-1801.
- [23] Dubey A, Pillai SV, Kolluri SV. Does volume of extradural hematoma influence management strategy and outcome? [J]. Neurol India, 2004, 52(4):443-445.
- [24] Ma H, Guo ZN, Sun X, et al. Hematoma volume is a predictive factor of disturbed autoregulation after spontaneous intracerebral hemorrhage[J]. J Neurol Sci, 2017, 382(15):96-100.
- [25] 常健博,姜桑种,陈显金,等.基于卷积神经网络的自发性脑出血血肿分割方法的一致性评价[J].中国现代神经疾病杂志,2020, 20(7):585-590.

(收稿日期:2020-12-07 修回日期:2020-12-22)