

RSNA2019 人工智能影像学

姚义好,李娟,石晶晶,申楠茜,张烁琪,周依然,吴迪,马艳强,朱文珍

【摘要】 随着人工智能(AI)的迅速发展和日益广泛的应用,人们对自动图像处理和分类技术产生了极大的兴趣,本年度 RSNA 年会上关于 AI 的研究报告按照部位进行分类总结。各研究通过利用深度学习分割图像、研究图像特征提取和图像转换,研发机器学习辅助诊断方法,比较不同 AI 模型如何提高成像效率、减少辐射剂量和对比剂剂量等,对临床疾病进行鉴别诊断和准确分级,进而辅助临床个体化治疗方案的制定,且有望大幅降低影像医师的工作负荷。

【关键词】 人工智能; AI; 影像组学; 深度学习; 图像特征; 卷积神经网络

【中图分类号】 TP18;R814.42;R445.2 **【文献标识码】** A

【文章编号】 1000-0313(2020)05-0575-08

DOI:10.13609/j.cnki.1000-0313.2020.05.002

本年度 RSNA 年会上关于人工智能(AI)的研究报告按照部位和疾病类型、检查手段进行分类,各研究通过图像特征提取、集成学习、研发高效机器学习辅助诊断方法,服务于临床疾病的鉴别诊断和准确分级,进而辅助临床个体化治疗方案的制定,有望大幅降低影像医师的工作负荷,同时使患者获得早期诊断和适当的治疗。

中枢神经系统

对于急诊外伤、脑血管事件等, AI 显示出较好的协助诊断功能。采用卷积神经网络(CNN)来检测头颅 CT 平扫中的颅骨骨折,由于头皮血肿是协助诊断骨折的良好指标,将其作为临床证据添加到 CNN 中,可以显著改善平均精度(由 0.7952 升高至 0.8190);在不影响灵敏度的情况下,假阳性率明显下降,可见在采用深度学习模型进行图像分析的基础上加入临床背景可以使诊断更为准确。Jayadeepa 等研究颅内出血(ICH)术前平扫 CT,通过人工智能采用卷积神经网络进行中线移位检测和测量;与放射科医生报告相比,人工智能算法在中线偏移的检测和定量方面显示出良好的效果,其准确度为 91.41%,敏感度为 91.95%,特异度为 90.79%。Balaji 等在头颅 CT 平扫上使用基于 CNN 的 AI 方案来评估 ICH 的准确率,结果发现 28 例 ICH 中 16 例为临床漏诊病例。同样 Jayadeepa 等

通过 AI 模型,对 22 个漏诊的 ICH 病例进行回顾性分析,结果检出 22 个漏诊中的 11 个,敏感度为 59.62%,特异度为 91.07%。因此, AI 可以帮助放射医师诊断 ICH,降低漏诊率。但也有研究显示 AI 技术在有些领域还需要提高评价效能,如 Walsh 等研究前循环大血管闭塞最终梗死体积,就发现与自动 Alberta 卒中项目早期 CT 评分(ASPECTS)相比,医生阅片的 ASPECTS 与前循环出现大血管闭塞的最终梗死体积相关性更好。

对于血管成像和动脉瘤的诊断, AI 也显示出较高的效能。提取动脉瘤 MR 图像,随机分为训练数据集和测试数据集,另随机选择 50 例无动脉瘤的检查数据纳入测试数据集;由两位放射科医师对动脉瘤进行手工标注,建立基于 3D ResNet 架构的深度学习算法,利用训练数据集进行基于像素块的分类,然后进行像素投票算法;测试数据集的敏感度、阳性预测值和特异度分别为 87.1%、95.5%及 92.0%;测试数据集中,算法新诊断了 1 个动脉瘤,19 个未检出的动脉瘤中有 14 个位于颈内动脉,10 个假阳性检测中有 9 个也位于颈内动脉。研究结果显示,深度学习算法检测脑动脉瘤具有较高的诊断效能。与此类似,对 DSA 图像进行分析,为了避免过度拟合,数据按 4:1 的比例分成测试和训练两组,使用基于开源 Tensorflow 框架的商用级机器学习软件(Cognex, ViDi Suite 2.0)进行深度学习;检测并正确定位颅内动脉瘤的敏感度为 79%,特异度为 79%,准确度为 0.75, F1 评分为 0.77, AUC 为 0.85。未检出与已检出动脉瘤大小差异无统计学意义($P=0.19$)。深度学习可在 DSA 图像上辅助检测和定位颅内动脉瘤。Fujita 等对健康志愿者和颅内动脉瘤患者同时进行 TOF-MRA 及 3D 合成 MRI 序列(3D-

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

作者简介: 姚义好(1989—),女,河南安阳人,博士研究生,住院医师,主要从事中枢神经系统影像学诊断和研究工作。

通讯作者: 朱文珍, E-mail: zhuwenzhen8612@163.com

基金项目: 国家自然科学基金(81730049, 81801666, 81873890)

QALAS)扫描,并采用深度学习算法产生深度学习 MRA(DL-MRA)图像,通过对比研究发现 3D 合成 MRI 不仅能获得定量成像及各种对比加权成像,其通过深度学习算法获得的 DL-MRA 能有效显示颅内大动脉,且成像效果与 TOF-MRA 相当,减少 MR 扫描总时间并有助于放射科医生筛选血管病变。

中枢退行性疾病的研究也有可喜进展,采用 AI 方法可从稳定的轻度认知障碍(sMCI)中识别进行性轻度认知障碍(pMCI)。采用迭代注意力聚焦(IAF)神经网络来生成疾病相关的注意图并预测诊断结果。结果表明该方法的准确率达到 81.6%,超过了其他如基于 VBM 的方法(64.3%)和基于像素的深度学习方法(LDMIL,76.9%);同时也获得了较高的敏感度(60.5%),VBM 和 LDMIL 的敏感度为 36.8%和 42.1%,表明该方法能更有效地识别可能的 MCI 转换。除了诊断结果外,聚焦注意力地图还提供了与 MCI 进展相关的特定病理位置。IAF 神经网络可以作为新的计算机辅助诊断痴呆方法,有助于认识和理解 MCI 到 AD 的进展。

AI 对脑肿瘤的诊断也是关注的热点。Hoebel 等通过对 48 例胶质瘤患者的术前 MRI 的 T_2 -FLAIR 图像进行纹理分析,发现标准化对于 MRI 纹理特征及信号强度的重复性具有显著意义,研究者需要谨慎选择直方图箱,来确保提取的特征具有意义。另外,对比增强 MRI 是脑肿瘤诊断的关键,尽管钆对比剂被广泛使用,但也存在成本和缺陷。分析胶质母细胞瘤患者的术前 MR 影像[包括对比剂注入前后 T_1 、 T_2 、 T_2 -FLAIR 序列及动脉自旋标记、磁敏感加权(SWI)、DWI 序列]。训练深度卷积神经网络,该神经网络基于改进的 U-net 体系结构,具有分解的三维卷积和残差学习能力。该深度学习网络能够生成合成的增强 T_1 加权图像,在定性和定量上都与真实的增强图像相似。对于合成的增强 T_1 加权图像,从全脑来看贡献最大的是非增强 T_1 图像,其次是 SWI;单从肿瘤区域来看贡献最大的是 DWI,其次是 SWI。研究结果表明,深度学习可能有助于减少钆对比剂的使用。基于现在的研究发现异柠檬酸脱氢酶(IDH)突变状态对脑肿瘤的诊断、治疗和预后具有重要意义。一个完全自动化的深度学习网络多数投票算法,仅使用 T_2 图像训练的网络取得了与使用多对比图像训练的网络相当的结果(采用 T_2 -net 的准确率为 96.1%,AUC 为 0.972,TS-net 的准确率为 96.9%,AUC 为 0.981),可用于无创预测 IDH 突变状态。

对于儿童的影像学 AI 研究也出现了有意义的报道。有研究评估纠正足月的早产儿(PNs)的质子磁共振波谱(^1H -MRS)和弥散张量成像(DTI),选择重要

的特征,最终使用前馈神经网络(fNNs)构建了预测因子,能够预测 PN 中的发育不良。对运动发育不良,预测因子具有 100%的准确性;对认知发展不良,获得了 100%的真实阳性率和 83.3%的阳性预测值。因此 fNNs 可用于发现那些需要早期干预的 PN。在肿瘤诊断方面也有了新的进展,评估儿童髓母细胞瘤的 MR 图像,从增强 T_1 加权像中提取放射组学特征,利用最小绝对收缩和选择算子(LASSO)Cox 回归模型识别选择放射学特征(Rad-score),结果显示结合 Rad-score、临床及常规影像特征的一体化模型是预测儿童髓母细胞瘤总体生存期(OS)最准确的模型(C-index: 0.928)。另外,基于 CT 的放射组学特征也可以预测儿童神经母细胞瘤(NBs)中 MYCN 扩增(MNA)的能力。MNA 预测效果最好的是将增强前/后各期图像相结合,检验组 AUC、敏感度、特异度、准确度分别为 93.8%、100%、94.9%、0.98。基于 CT 的放射组学特征还可用于预测儿童恶性神经母细胞瘤(PNTs)的预后,根据临床预后因素将患者分为低、中、高危组,采用 Intelligence Foundry 建立两个回归模型以级联方式应用,model1 用于对高危人群和其余人群进行分类,model2 将其余人群分为中、低风险人群。模型在训练队列和验证队列中均表现出良好的辨别能力,模型 1 的 AUC 为 0.826,模型 2 的 AUC 为 0.811,结果显示基于术前 CT 图像的放射学分析有助于儿童恶性肿瘤的个体化治疗分层。

从数据处理来看,脑组织分割提取是神经影像学分析中图像预处理的第一步,MRI 图像通过 FreeSurfer 进行脑提取和分割作为金标准。Michael 等研究发现改进后的基于深度学习的 U-Net 模型能得到与 FreeSurfer 相当的脑组织提取结果,但每例患者仅需要几秒钟的时间来完成,而不是 FreeSurfer 所需要的数小时。这可以为其他任务释放计算资源,对依赖于这一步骤的神经影像学研究具有重要意义。

头颈部

评估来自头颈部鳞状细胞癌(HNSCC)患者治疗前的颈部增强 CT,分别来源于喉部或下咽(LHP)、嘴唇和口腔(OC)、口咽(OP),提取肿瘤的纹理特征,并与患者的年龄、吸烟状态、饮酒状态和肿瘤 T 分期相结合,构建预测淋巴结状态、淋巴血管侵犯(LVI)和神经周围侵犯(PNI)的模型。采用随机森林(RF)和支持向量机(SVM)两种机器学习方法构建预测模型。OC、LHP、OP 三组肿瘤的纹理特征差异有统计学意义($P < 0.05$)。对基于原发肿瘤位置的纹理数据进行亚分层,与使用组合数据集的模型相比,预测模型的准确性提高了 14%,所以为了获得最佳诊断性能需要根

据原发肿瘤部位对患者进行分层。

胸部

基于深度学习(DL)的算法可以协助评估在医院中最常见的影像学检查-胸片(CR)。不同的研究机构分别对 CR 使用 DL (DNetLoc) 算法、自动检测(DLAD)算法等进行处理,结果发现基于深度学习的算法可以在 CR 中准确检测和定位 10 种异常影像学表现(肺不张、钙化、心脏肥大、实变、纤维化、结节、纵隔扩大、胸腔积液、气腹和气胸),特别是气胸的检出 AUC 高达 0.96,总体敏感度和特异度达到 89%(小型气胸敏感度为 82%,大型气胸敏感度为 97%),获得了较高的准确度,可以高精度地有效检测正位胸部 X 线图像中的气胸,用于识别需要优先审视的检查,从而有助于提高工作效率,改善患者结局。对于缺乏经验的年轻医生,在 CR 上确定肺结核的活动性相对比较困难。基于深度学习的自动算法,可以用 CR 将活动性肺结核(TB)与其他肺部异常和正常肺部进行分类。使用卷积神经网络(CNN)对正常与异常 CR 进行分类,然后使用 YOLOv2-densenet 模型进行深度微调。在测试数据集中,该算法在正常与异常 CR 分类中的准确度达到 98.18%。采用该算法对活动性 TB、其他异常 CR 和正常 CR 进行诊断时,准确度为 78.5%,对于每个 CR 该算法的总体准确度为 95.61%;对每个 CR 而言,该算法对活动性结核病的敏感度和特异度分别为 79.73%和 96.96%。使用 CNN 进行的深度学习在分类正常、活动性 TB 与其他异常状态 CR 方面显示出很高的诊断性能,可用于筛查活动性结核,从而提高诊断流程的效能。更有研究专门针对基线和治疗后胸片分别标记为活动性和非活动性结核。相关研究采用这些 CR 训练了一个深层神经网络,以输出有关肺结核活动性的百分比得分。初步结果表明,深层神经网络和放射科医师的诊断 AUC 值分别为 0.80(95%CI:0.71~0.89)和 0.74(95%CI:0.64~0.84)。深层神经网络算法可以在 CR 上较准确地确定结核的活动性,并可与医师判别结果相媲美。在结核病专家有限的情况下,深度神经网络可用于确定结核病的活动性,并监测抗结核治疗的反应。一种新开发的基于深度学习的自动检测算法(DLAD)可检测在 CR 中被放射科医师忽略的肺部恶性肿瘤。相关研究纳入经病理证实的胸片上发生遗漏的 155 例肺恶性肿瘤患者,其中 127 例病变是活动性病灶;当活动阈值为 0.3 时,DLAD 对活动性肺恶性肿瘤的敏感度(54%)高于对非活动性肺恶性肿瘤的敏感度(14%);每例患者的假阳性标记数为 0.24。当活动阈值为 0.15 时,DLAD 对活动性肺恶性肿瘤的敏感度提高到 60%,每例患者的

假阳性标记数为 0.35。DLAD 算法在 CR 上可以检测出 50%以上的肺部恶性肿瘤,而这些恶性肿瘤在 CR 中往往被放射科医生忽略,且假阳性率低。

基于 CT 检测对肺结节进行诊断也是 AI 研究的热点。有研究显示将 AI 算法与医师的诊断效能进行比较,两者结果高度相关($r=0.663, P<0.001$),而 AI 算法的 AUC 高于医师的 AUC($P<0.001$),具有极高的临床应用潜力。Mukherjee 等根据对 CT 扫描直径大于 4mm 的肺结节进行检查,建立 CT 筛查阳性的全国肺部筛查试验(NLST)非侵入性确定肺癌可能性的模型,通过使用一个、两个和三个筛选时间点,以三种方式建立了一个使用 CT 图像进行癌症预测的两阶段机器学习(ML)模型。这三个模型共有的第一个 ML 阶段是训练卷积神经网络(CNN)用于检测结节和预测恶性评分。ML 的第二个阶段是使用 Xgboost 通过第一阶段预测的受试者肺结节的位置和恶性评分来预测癌症发生率。根据一个时间点的 CT 扫描进行预测时 AUC 值为($0.75\pm0.03, 95\%CI$),而根据两次时间点的扫描进行预测时则达到(0.80 ± 0.03),从 NLST 中的所有三个时间点进行扫描预测时可达(0.85 ± 0.02),差异有统计学意义($P<0.01$)。结果发现仅通过筛查肺部 CT 扫描就可以预测肺结节 $>4mm$ 的受试者在以后几年是否会患上癌症,如果将来自多个筛选时间点的 CT 成像数据合并到模型中,则预测性能会更高。另外,Chen 等评估了一种深度学习系统(DLS)的诊断性能,回顾以往胸部 CT 的报告,结果发现在所有的胸部 CT 扫描中,有 0.3%发现了漏诊的肺结节,其中 1/3 具有临床意义。因此,使用 DLS 辅助的自动检测可能会提高放射科医生的工作效率。类似的研究显示,回顾性分析在阴性筛查后又经 NLST 筛查偶然发现肺癌患者的 CT 影像($n=122$)。训练的基于 DenseNet 三维卷积神经网络的计算机辅助设计系统,根据 NLST 标准和肺-RADS 分类分析结果,显示基于深度学习的 CAD 系统对阳性筛查的敏感度为 89%~95%,对漏检肺癌的敏感度为 74%,而假阳性率有限。肺癌在低剂量 CT 筛查中常被遗漏,而基于深度学习的 CAD 系统可能会改善早期诊断。AI 算法还可以帮助进行结节复查决策。卷积神经网络(CNN)模型由 NLST 数据集训练得出,其会生成一个评分,在每例病例中得分最高的结节被用作指标结节。采用阈值为 5.0 的 CNN 评分免去了 38 例良性结节患者的 CT 检查和 3 例 PET-CT 扫描而不遗漏任何癌症,占随访时间内 CT 总数的 41.3%和 PET-CT 的 37.5%;采用 0.56 的阈值,则显示 15 个良性结节不需要任何随访,免去了 23 例 CT 扫描。在 11 例(35.5%)癌症患者中使用 80 分的 CNN 阈值,他们的

检测和干预将会加快 3.4 个月。CNN 生成的模型评分应用于偶然检测到的不确定的小结节,减少了对良性结节的随访扫描,同时潜在地加速了对高分评分癌症结节的监测和治疗。

关于肺癌的诊断和疗效监测 AI 也表现出其优势。由基于相关合并的层次聚类算法(CHCA)有效提取的 EGFR 突变的非小细胞肺癌(NSCLC)的放射学表型,可以帮助识别可能受益于靶向 EGFR 抑制剂治疗的 NSCLC 患者。另有研究显示,分析每例 NSCLC 铂类化疗患者治疗前 CT 图像的放射学特征,影响 PFS 的独立因素为放射性组学特征、PS 和 N 期,均被选择到 nomogram 中,nomogram 预测 PFS 的 c 指数为 0.721 (95%CI:0.713~0.729),高于基于临床病理的模型(c 指数:0.641,95%CI:0.631~0.651),可方便地用于对铂类化疗的Ⅳ期非小细胞肺癌患者进行 PFS 的个体化预测。也有相关研究进行 NSCLC 的淋巴结转移 LNM 的预测,从增强 CT 中提取放射学特征,降维采用 Mann-Whitney U 检验和单变量方差分析,采用最小绝对收缩选择算子(LASSO)算法进行径向组学特征选择。建立了三个模型(临床模型、放射学模型和联合模型)来预测早期 NSCLC 的 LNM,结果显示联合模型优于单纯应用放射学和临床特征建立的模型,约 20% 的术前 IA 期 NSCLC 患者可能存在 LNM,放射学临床模型有预测 LNM 的潜力,可以帮助改善治疗方案。

基于 AI 算法的 XGBoost 模型可能对预测局部晚期食管癌(LAEC)术前化疗的效果有帮助,有助于改善 LAEC 患者术前管理的临床决策。

相关研究对多创伤患者中的急性呼吸窘迫综合征(ARDS)患者的胸部 CT 图像进行了基于深度学习的肺分割,利用每个肺的空间视觉词的直方图训练一个支持向量机(SVM)分类器来预测 ARDS,并将该算法与常用的预后评估评分(ISS)和胸损伤评分(AIS)进行比较。基于机器学习的 ARDS 风险评分的 AUC 为 0.78 (ISS:0.66;AIS:0.68)。以机器学习为基础的多创伤患者肺的放射学特征能够预测 ARDS,其水平高于一般临床评分。

对于各种 CTA 的诊断和重建,AI 也显示了极高的临床应用效能。Noothout 等评估了 CTA 中用于自动主动脉根部标记定位的深度学习性能,结果发现,对于左半月瓣、右半月瓣、右半月瓣铰链点及右冠状动脉口、左冠状动脉口的定位,自动标记定位与观察者人工定位差异无统计学意义,而平均每个 CTA 分析只需要 0.3 秒,大大缩短了定位时间。肺动脉造影(CT-PA)用于检测肺动脉栓塞时会出现假阳性(FP)结果,大多数 FP 是由肺树外结构和/或非管状结构引起,因

此,动脉树的分割和检测结节结构的算法的集成可能是进一步减少 FP 的措施。人工智能算法还可以通过 CTPA 的放射组学特征对急性肺栓塞(APE)患者进行危险分级,即区分高危组与非高危组。有研究应用深度学习方法计算 APE 的 CTPA 血栓体积,同时由放射科医师采用梗阻评分评估 APE 患者的凝血负荷。DL 模型的平均测量时间为(12.9±3.8)s,而第二年住院医师的平均测量时间为(10±4)min。当模型临界值设为 0.1 时,敏感度和特异度最高,分别为 94.6%和 76.5%,AUC 值为 0.926(95%CI:0.884~0.968),两次测量的一致性为 100%。DL 模型具有高度的敏感性和重复性,血栓体积与梗阻评分高度相关。采用 DL 模式检测 APE 可大大提高诊断效率,减轻放射科医生的工作量。另有研究比较了 DLIR 算法与 CTPA 标准图像重建算法的图像质量,分别采用 FBP、ASIRv50 和 3 个 DLIR 原型[低(L)、中(M)和高(H)]进行回顾性重建,定量测量主要 PA 的噪声(标准差)、信噪比(SNR)和对比噪声比(CNR)。两位放射科医生分别对 ASIRv50 和 DLIR-M 的主观图像噪声、噪声纹理、伪影和诊断质量进行了 1~5 级的评分。除 ASIRv50 与 DLIR-L 的 CNR 比较差异无统计学意义($P=0.175$)外,其余各项比较差异均有统计学意义。与 FBP 和 ASIRv 相比,DLIR 降低了图像噪声,提高了 CNR 和 SNR。与 ASIRv50 相比,DLIR 中等强度显示降低了图像噪声,改善了图像纹理,在诊断质量或伪影的主观评估方面差异无统计学意义。

AI 技术在乳腺疾病的诊断中得到了较广泛的应用。使用基于深度学习的软件 Lunit INSIGHT 对乳腺 X 线检查图像进行分析,显示其对乳腺癌有较高的诊断效能,敏感度为 90.2%,特异度为 90.9%,准确率为 90.2%,AUC 值为 0.960。另有研究显示,经过 AI 处理后的乳腺癌检测的 AUC 都有所提高:数字乳房 X 光检查(DM/DBT)分别为 0.781 和 0.848;合成乳房 X 光检查(SM/DBT)分别为 0.812 和 0.846。相应的 AI 处理后两种检查模式的敏感性、特异性也有所提高,读取时间更短。从技术角度来看,Duan 等采用卷积神经网络构建了一套用于增强数字乳腺断层摄影(CEDBT)散射校正的算法,能提高图像质量。有研究采用卷积神经网络(CNN)进行乳腺良、恶性疾病分类,利用全二维乳腺 X 线片和全容积 DBT 相结合的方法(首先使用 MMG 进行训练,然后使用 DBT 进行微调)来提高模型的性能,AUC 为 0.93,高于单独使用 2D 或 3D 乳腺摄片,可以提高乳腺癌的诊断效率。采用 DIB-MMG 对其进行分析,在 47 例漏诊病例中检测出 32 例,在 61 例间变性癌症中检测出 30 例,在 17 例隐匿性癌症中检测出 7 例,总的 AUC 为 0.738,DIB-

MMG 有望用于筛查性乳房 X 线摄影中乳腺癌的早期检测。另一项研究显示,筛查性乳房 X 线检查由 AI 进行分析,根据不同的召回率做出召回复查决定,其敏感度分别为 99.3%、87.7% 和 76.1%,召回率分别为 50%、10% 和 4%。召回率为 50% 时,达到了极高灵敏度,这意味着不必复查的那部分人几乎可以确定是正常的,因此 AI 可以帮助放射科医生减少工作量,在乳腺癌筛查方面具有巨大的潜力。采用基于卷积神经网络(CNN)的算法分析乳腺钼靶 X 线图像可以鉴别不典型导管增生(ADH)与导管原位癌(DCIS),AUC 为 0.90(95%CI±0.04),诊断准确度为 80.7%,敏感度为 63.9%,特异度为 93.7%,具有很高的特异度,可以用来预测能安全观察而无需手术的单纯 ADH 患者。DIB-MMG 在乳腺 X 线摄影中也可以预测肿瘤侵袭性,对乳腺导管原位癌(DCIS)、微浸润导管癌(DCIS-MI)与浸润性导管癌(IDC)进行鉴别诊断。研究结果显示 DCIS-MI 与 DCIS 的鉴别较与 IDC 的鉴别困难,DIB-MMG-TI 用于区分 IDC 与其他类别是可行的。来自 DBT 的合成乳房 X 线摄影的放射组学特征在区分乳腺癌 TN 与非 TN 分子亚型方面表现出了很高的性能,并可能影响治疗方向,其 AUC 为 0.838, TN 的敏感度和特异度分别为 83.3% 和 79.7%。与此类似,机器学习算法在 luminal A/luminal B、HER2 富集、三阴性分子亚型鉴别诊断中也显示出较高的效能。基于 CNN 的算法可以预测乳腺癌的风险,并且可以根据已知的化疗方案进行调整,从而有望监测治疗反应。

乳腺 MRI 的日渐成熟为 AI 辅助乳腺疾病诊断增加了新的研究方向。从技术上看,Wang 等在乳腺 MRI 中自动进行纤维腺体组织(FGT)分割时发现,模糊边缘注意生成对抗网络(FEA-GAN)在骰子相似系数(DSC)、Jaccard 索引(JI)和 Hausdorff 距离(HD)等方面与深度学习 U-Net 相比,其改进均具有统计学意义($P<0.0001$)。评估乳腺动态对比增强 DCE-MRI 图像的时间和空间分辨(4D)放射组学,以体素方式提取 BI-RADS 衍生的药代动力学增强特征,通过主成分分析(PCA)和人工神经网络(ANN)进行进一步分析,鉴别乳腺病变的良恶性,结果显示 AUC 为 0.836(95%CI:0.799~0.868),显示出较高的诊断效能。另有研究显示,基于乳腺的 DCE-MRI 成像,使用 Res-Net 深度学习的 ROI+放射组学模型的诊断准确度为 86%,较单独的 ROI 和放射组学模型高,且在肿瘤附近加入少量的肿瘤周组织可以获得更高的准确度(达到 90% 左右)。癌型 DX 评分(ODX)对雌激素受体(ER)阳性乳腺癌患者的风险分层起着关键作用,只有高危患者才能从辅助化疗中获益。采用多参数 MRI 放射组学鉴别 ER 阳性乳腺癌患者的高与低(exp1)、

高与中低(exp2)、低与中高(exp3)的 ODX 风险,结果显示在 exp1 中, $T_2 + DCE$ 的 AUC 最高(0.83, 95%CI:0.75~0.91),对 exp2 也是如此(0.78, 95%CI:0.69~0.85);而对于 exp3 来说, T_2 的 AUC 最高(0.74, 95%CI:0.65~0.83)。这些结果显示了 $T_2 + DCE$ 对从低危或中危/低危患者中分层出高危患者的重要性。相比之下, T_2 能够独立从中高危患者中预测出低风险患者。与之前报道的只使用 DCE 结果相比,增加 T_2 特征可以使 AUC 的风险预测性能提高 11%~15%。为预测腋窝淋巴结转移情况(阳性/阴性),提取乳腺癌 DCE-MRI 早期和晚期的放射学特征,仅使用 DCE-MRI 放射组学特征、DCE-MRI 放射组学与临床病理相结合都能进行有效预测,而后者效能更高,AUC 值高达 0.912(95%CI:0.819~0.979),敏感度、特异度和准确度分别为 0.91、0.88 和 0.71。

腹部

AI 在肝脏疾病的相关研究应用广泛。基于深度学习的模型可帮助放射科医生在增强 CT 中准确检测肝脏恶性病灶,检测灵敏度为 89%,假阳性(FP)率为 2.54%。类似的研究显示,进行增强双能 CT(DECT)后,借助机器学习(ML)模型,影像组学可以准确地区分良性与恶性肝病灶,具有高特异度(97%)和阳性预测值(94%)(AUC=0.82, $P<0.0001$),其中仅 1/35 的良性(充血性血管瘤)病变被分类为恶性病变(假阳性)。肝静脉期和增强前的融合 MRI 放射学特征可能用于区分 cHCC-CC(肝细胞-胆管混合癌)与 HCC、CC,将有助于改善术前影像学诊断和治疗计划。在活体供体肝移植术 LDLT 中,通过术前评估肝脏和节段容积来确保适当的移植植物大小的安全性是成功预后的主要指标。使用 3 种不同方法进行术前 LDLT 评估:A,完全手动;B,手动切除的 AI(使用 AI 自动分割肝解剖,放射科医生手动切除);C,全自动。结果显示 B 和 C 效果最佳,且 B 显示的总体差异最小。AI 系统与专家干预相结合,有望在肝移植计划中提供更好的结果。对于进行肝大部分切除术后的肝硬化患者,基于放射组学的术前增强 MR 图像模型可用于预测术后患者的肝衰竭,通过使用最小绝对收缩和选择算子方法,并采用多元逻辑回归模型建立影像组学模型,发现 15 分钟的吲哚菁绿清除率(ICG-R15)在预测肝功能衰竭方面表现出最高的性能(AUC=0.894, 95%CI:0.823~0.964)。所有肝脏 MRI 图像均由受过训练的临床分析人员手动描绘 ROI,以量化肝脏 T_1 、 T_2^* 和 PDF,提供有关纤维炎症、铁和肝脂肪的信息来表征肝脏状态。使用这些手动生成的分割蒙版,训练了基于 U-Net 的深度学习方法来自动识别并排除

较差的模型拟合和伪影区域,模拟手动执行的 ROI 分析。将手动 ROI 与自动 ROI 进行比较,结果显示基于机器学习的自动化处理,与手动处理产生的结果非常接近,有可能极大地提高评估定量结果的效率,并提高标准临床工作流程中定量 MRI 分析的可行性。对 T_1 WI 的同相位图像使用纹理分析和机器学习量化处理,用于评估慢性肝病患者的肝纤维化水平,其准确度与 MR 弹性成像(MRE)相似(AUC 分别为 0.82 和 0.92, $P=0.4066$),可以在无法获得 MRE 的情况下用作评估肝纤维化水平的替代方法。Smith 等通过人工智能算法定量分析 CT 图像中的肝脏表面结节,得到评分(LSN),用来评估肝脏纤维化及肝硬化的分期;结果显示该评分对早期肝纤维化、进展期肝纤维化、肝硬化诊断的 AUC 分别为 0.88~0.90、0.89~0.93 和 0.90~0.96,具有较高的准确性,并可以预测具有临床意义的门脉高压及其他包括肝脏失代偿及死亡等相关预后事件。同样,基于机器学习的超声造影特征提取也可以帮助影像学医师提高对局灶性结节性增生(FNH)与非典型肝细胞癌(aHCC)的鉴别诊断。

在肾脏肿物的研究中 AI 也显示出其优势。在放射组学中使用机器学习结合肾脏肿瘤的三维形态和纹理特征,有助于影像学诊断中肾脏肿块良恶性的鉴别。单独使用形状指标 AUC 为 0.65,添加纹理指标后 AUC 为 0.69,形状指标是重要变量,其中矢状凸包周长率(CHP)是表现最好的形状指标。基于多期增强 CT 纹理特征的机器学习可用于鉴别透明细胞肾细胞癌(cc-RCC)与癌细胞肿瘤(嫌色性肾细胞癌、嗜酸细胞瘤),肾脏造影(NG)期 TA 的鉴别能力最高,AUC 为 0.822,敏感度和特异度分别为 71.3%和 81.4%,达到中等精度。类似的人工智能算法也可用于乳头状细胞癌(prcc)、透明细胞癌(crcc)、嗜酸细胞瘤(oc)的鉴别诊断,检测 crcc、oc、prcc 的敏感度分别为 0.76、0.68、0.71,特异度分别为 0.81、0.93、0.87,准确度分别为 0.70、0.58、0.76。增强 CT 纹理特征的机器学习可以改善实体肾肿物的分类。对肿瘤病灶进行自由区域分割,得出肿瘤长度和血管肿瘤负荷(VTB)的变化,然后使用 TexRAD 软件进行 CT 纹理过滤,结果显示抗血管生成治疗后 CT 图像上 CT 纹理的定量变化可以预测无进展生存期(PFS),且精细纹理变化、肿瘤长度和 VTB 是 PFS 的独立预测因素。基于常规 MR 影像的深度学习,区分肾良性肿瘤与肾细胞癌的准确性较高,达到了 87.2%,AUC 为 0.939,有助于指导临床治疗决策的制定。另有研究显示基于 T_2 WI 的纹理分析能有效鉴别低脂 AML 与常见肾细胞癌。基于全相位 CT 和全序列 MR 图像的机器学习模型可以鉴别低级别与高级别 ccRCC,AUC 分别为 0.76 和 0.77,两者

之间差异无统计学意义,可用于术前预测。

Chen 等通过对包括 100 例胰腺肿瘤及 100 例对照病例的对比增强 CT 静脉期图像进行纹理分析及机器学习,建立了一个用以识别胰腺肿瘤的机器学习模型,其基于患者的诊断准确度为 95.12%,基于模块的诊断准确度为 93.43%。

对于胃癌,应用机器学习方法,放射学特征显示出良好的预测早期复发的性能,AUC 为 0.799(95%CI: 0.741~0.857),此外临床 N 期、临床 T 期和癌胚抗原水平是早期复发的独立预测因子,结合所有这些预测因素的列线图显示出更强大的预测能力,AUC 为 0.842(95%CI: 0.791~0.894),可用于术前预测晚期胃癌(AGC)患者术后的早期复发,是指导个性化治疗的潜在工具。通过使用 AI 算法从 CT 结肠造影 CTC 数据集中自动检测息肉,该算法通过 3D 卷积神经网络检测锯齿状息肉的对比涂层现象,有很高的敏感性;91%的息肉可见对比涂层,平均每个息肉的检测敏感度为 93%±7%,平均每例患者的敏感度为 94%±6%。准确的术前肿瘤分期评估对直肠癌的治疗和预后至关重要。相关研究从高分辨率 T_2 加权成像(HR- T_2 WI)和扩散加权成像(DWI)通过人工神经网络(ANN)构建了放射学评估模型,并显示出良好的术前 T 分期表现;结合了放射学特征和临床独立的危险因素,放射学评估模型显示出更好的性能,可用于提高其 T 分期的准确性。影像组学特征是评估局部进展期直肠癌 LARC 患者无病生存期 DFS 的独立生物学指标,结合影像组学较临床病理或仅 Rad-score 诺模图更能提高个体化 DFS 的评估效果(C 指数为 0.715, 95%CI: 0.67~0.79)。

在生殖系统疾病中 AI 亦显示出其临床意义。Cheng 等研究发现基于网络的多功能辅助系统可以早期检测妇科肿瘤的远距离复发,对治疗策略的制定有很大帮助。Luo 等对进行子宫动脉栓塞治疗的肌瘤患者,通过训练残余卷积神经网络模型分析术前 MRI 图像,可以预测肌瘤体积的减少和临床结果,测试准确度分别为 78.1%和 84.7%,AUC 分别为 0.865 和 0.916。这将有助于临床医生确定哪些患者会是该疗法的最大受益者。Moeskops 等研究基于深度学习的方法从全腹部 CT 扫描中自动分割皮下脂肪、脏器脂肪和腰大肌是可行的,手动测量体积与自动测量体积之间的相关性系数分别为 0.996、0.997 和 0.941,平均约 15 秒内完成全扫描的分割,从而有助于心血管疾病和癌症的个性化风险评估。

骨骼肌肉系统

深度学习使用多尺度 U-Net 作为检测器,诊断效

能非常出色,检测骨折的特异度、敏感度和 AUC 分别为 87%、86%和 92%,还检测出 39 个漏诊的骨折。因此在传统 CR 图像上引入新型的深度学习驱动系统,可以协助放射科医生进行肋骨骨折的检测,并减少漏诊。同样,Nicolaes 等和 Cheng 等基于对椎体 CT 和骨盆 X 线片进行深度学习,诊断骨折也有较高的效能。

Lang 等对骨折患者的 CT 图像进行分析,使用 ResNet50 对脊柱病变进行良、恶性鉴别诊断,具有较高的诊断准确率,在每例患者的诊断中,敏感度为 0.95,特异度为 0.80,准确度为 0.88。但是当对整个脊柱进行评估时,自动检测异常的准确率较低,还需要进一步改进。Chianca 等回顾性研究脊柱病变患者的平扫 MRI,基于 HCAD 软件纹理分析(TA)可将 89%的病变正确地分类为良、恶性病变,而使用 Pyradiomics 软件的支持向量机的分类准确率可以达到 87%。对于良性、原发性恶性和转移性病变的分类,两种 TA 软件的 RF 模型都准确地对 70%的病变进行了分类。基于 MRI 检查的 ML 算法在脊柱病变分类中具有较高的准确度,不同软件进行的特征提取显示了一致的结果。

在膝关节和肩关节的研究中,AI 技术也有了新的进展。有学者开发了一种基于 CNN 的 U-Net 的改良版深度学习系统,以从非脂肪抑制(FS)图像创建合成人工智能的脂肪抑制 MR(AFMRI)图像;原始序列与 AFMRI 序列之间 CNR 测量的相关系数非常好(均 >0.95)。AFMRI 在检测内外半月板撕裂方面具有出色的灵敏度,对软骨和骨髓异常的评估具有中等特异性。尽管 AFMRI 会降低图像质量,但该技术对于诊断是可行的。深度学习合成 AFMRI 为膝关节的快速成像提供了一种新技术。另有研究显示,自动膝关节软骨分割法取得了与放射科医生相似的性能,可用于评估膝关节骨关节炎。一位有 18 年工作经验的放射科医生根据 3DUTE-Cones 序列的 UTE adiab 1ρ 加权 MR 图像手动分割膝软骨,然后将 UTE adiab $T1\rho$ 图像和相应掩码的 2D 切片输入到深度学习网络中,采用 U-Net 结构分割膝软骨。结果显示手动与自动分割所得的平均 T_1 、 $adiabT1\rho$ 、 T_2^* 值相关性较高,分别为 0.95、0.88 和 0.86。深度卷积神经网络(DCNN)在 X 线图片上可自动识别肩关节成形术(SA)模型,并区分 TSA 与 RTSA,以及 5 种特定的 TSA 模型,为自动关节置换图谱的快速模型识别奠定了基础。

对于儿童生长发育和骨折的诊断评估,AI 也显示出其优势。采用粗-精级联神经网络对小儿腿部的股骨和胫骨进行分割和测量,训练一个粗糙的 CNN 模型对被试的腿进行左右两边的分类,然后对每一边进

行精细的腿部分割。研究结果表明,粗-精级联深度学习方法可以实现儿童腿部的精确分割,分割精度为 0.90,其准确度与专科训练的儿科放射医师相当,且自动测量每个受试者所用的时间不到 1s。然而,在骨龄评估(BAA)中 AI 的评估效能还有待提高。Greulich&Pyle(GP)图谱是 BAA 的最常见标准,比较 GP-BAA 与基于三模型 MobileNetV2 卷积神经网络深度学习的 BAA (DL-BAA),分析儿童手 X 线片,DL-BAA 与 GP-BAA 的平均绝对误差(MAE)差异有统计学意义,实际年龄高估了 5.2~8.3 个月。使用经过适当设计和训练的卷积神经网络(CNN)对儿童骨折 X 线片进行评估,表现最好的 CNN 达到 94.35%的准确度,AUC 为 0.97。相似的研究结果显示,深度学习可对儿童肘关节髁上骨折做出与放射医师准确度相当的诊断。

骨密度的评估通常依赖于双能 X 线吸收骨密度测量法(DEXA)。Savage 等提出一种基于小波特征、AdaBoost 和局部几何约束的人工智能算法,发现该人工智能模型能成功地通过 CT 值测定骨密度,并与 DEXA 有良好的相关性($P<0.001$),可以成功用于 CT 筛查骨质疏松症。与此类似,Kamel 等研究发现深度卷积神经网络(DCNNs)也可用于骨密度测量,为 CR 预测骨质疏松提供了一种客观的方法,也为此类疾病的筛查提供了可能。

在软骨变性的早期仅存在组成上的微小变化,如胶原蛋白(CO)和蛋白聚糖(PG)含量的变化。采用定量 MRI(qMRI)进行软骨标本成像,运用傅里叶变换-红外光谱法测定 CO 和 PG 组分;建立人工神经网络(ANN)和多元线性模型(MLM),并通过遗漏交叉验证进行训练,实现了 CO 和 PG 组分的预测;模型与光谱测量之间存在很强的相关性(Pearson 相关系数 $r>0.7$),ANN 的表现优于 MLM。机器学习方法基于多参数 qMRI 预测局部 CO 和 PG 含量,具有较高的准确度,对诊断(早期)退变和监测治疗效果具有潜在意义。

影像技术

很多研究人员在 AI 总体的技术改进和应用方面做了许多工作。通过深度学习构建能用于 CT/能谱 CT 降低噪声的算法,Missert 等发现与传统的卷积神经网络模型相比,针对个体特定调整的深度学习神经网络模型能够降低噪声,获得更好的 CT 图像。Erath 等提出了一种 mbDSE 三射线评估方法,并将其应用于锥形束 CT 图像中,该方法不需要基于设备参数和散射参数构建模型,其在去除图像散射伪影、提高图像质量方面有一定的应用前景。Kim 等研究发现对于 CT 图

像,深度学习算法重建的图像质量优于滤波反投影算法,与 ADMIRE 重建算法相当,但深度学习重建算法更简单。

关于 CT 图像深度学习重建方法(DLIR)的探讨, Timothy 等研究发现胸部 CT 扫描在所有层厚和剂量下,DLIR 的性能均优于 FBP 和 ASiR-V。与此类似, Hata 等也发现与 ASiR-V 相比,DLIR 降低了胸部 CT 的噪声和条纹伪影,提高了客观参数和主观图像质量。Li 等对结石患者的 CT 图像进行分析,发现辐射剂量减少 60% 的 ASiR-V 重建图像和辐射剂量减少 90% 的 DL 图像不会影响尿路结石的评估。Zhao 等研究表明 NNC 模型能够将薄层 mDCT 转换成较参考标准全剂量 CT 图像质量高的 VHDCT,而辐射剂量减少了 90%。Brady 等对儿童患者的 CT 图像不同重建算法进行比较,发现采用基于深度学习的 DL-CT 重建(AiCE)技术对目标信号的检测和噪声幅度有了实质性改善,具有显著降低辐射剂量的潜力。基于深度学习的去噪(DLD)提高了超低剂量 CT(ULDCT)上混合迭代重建(IR)和基于模型的迭代重建(MBIR)图像

的质量,其中 MBIR 较混合 IR 更有优势。因此,DLD 和 MBIR 均可通过提高超低剂量胸部 CT 的图像质量来为临床实践做贡献。

关于生成对抗网络(cGAN)有不少研究报道。Xiao 等对多模态 CT 成像(包括平扫 CT 和增强 CT)进行分析,结果发现通过生成对抗网络的迁移学习,与直接训练 CT 图像相比,图像的峰信噪比和结构相似度都有显著改善;因此,迁移生成对抗网络可以有效提高多模态 CT 图像的分辨率和图像质量。Tang 等通过训练生成对抗网络学习来进行正常与异常 CR 的识别,该方法获得的 AUC 为 0.841,而没有使用对抗性学习的 AUC 则为 0.627。Uemura 等基于条件生成对抗网络开发了一种称为 pix2surv 的图像新型生存分析模型,通过分析胸部 CT 图像,将 pix2surv 模型特发性肺纤维化(IPF)患者的存活预测与已建立的临床预后生物标记物[性别、年龄和生理学(GAP)指数]进行比较,结果发现,pix2surv 模型在预测 IPF 患者整体生存方面的性能较 GAP 指数有显著提高。

《请您诊断》栏目征文启事

《请您诊断》是本刊 2007 年新开辟的栏目,该栏目以临床上少见或容易误诊的病例为素材,杂志在刊载答案的同时配发专家点评,以帮助影像医生更好地理解相关影像知识,提高诊断水平。栏目开办 13 年来受到广大读者欢迎。《请您诊断》栏目荣获第八届湖北精品医学期刊“特色栏目奖”。

本栏目欢迎广大读者踊跃投稿,并积极参与《请您诊断》有奖活动,稿件一经采用稿酬从优。

《请您诊断》来稿格式要求:①来稿分两部分刊出,第一部分为病例资料和图片;第二部分为全文,即病例完整资料(包括病例资料、影像学表现、图片及详细图片说明、讨论等);②来稿应提供详细的病例资料,包括病史、体检资料、影像学检查及实验室检查资料;③来稿应提供具有典型性、代表性的图片,包括横向图片(X 线、CT 或 MRI 等不同检查方法得到的影像资料,或某一检查方法的详细图片,如 CT 平扫和增强扫描图片)和纵向图片(同一患者在治疗前后的动态影像资料,最好附上病理图片),每帧图片均需详细的图片说明,包括扫描参数、序列、征象等,病变部位请用箭头标明。

具体格式要求请参见本刊(一个完整病例的第一部分请参见本刊正文首页,第二部分请参见 2 个月后的杂志最后一页,如第一部分问题在 1 期杂志正文首页,第二部分答案则在 3 期杂志正文末页)

栏目主持:石鹤 联系电话:027-69378385 15926283035