

人工智能在肝脏和胰腺影像学中的应用概述

陈国旗，彭奇，刘金润，李金蔚

中南大学数学与统计学院

2025 年 12 月 17 日



中南大学
CENTRAL SOUTH UNIVERSITY

- ① 课题背景
- ② 研究内容
- ③ 扩展
- ④ 参考文献

① 课题背景

② 研究内容

③ 扩展

④ 参考文献

AI 在医学领域的应用

- 人工智能 (AI) 是迄今为止最具前景的研究领域之一。基于 AI 算法的医学应用包括药物开发、健康监测、疾病诊断和个性化医疗治疗。在 AI 的广泛定义下, 存在着显著不同的模型、范例和实现。

机器学习 (ML)

机器学习 (ML) 其系统设计基于模仿人类决策过程并依赖于统计模型来辨别或预测特征的算法。这种方法的主要局限性之一是需要进行特征提取, 其中特征需要事先定义。

方法

深度学习 (DL)

定义一个物体的显著特征有时会很困难。为了克服这一限制，深度学习 (DL) 方法可以从数据中学习，无需事先定义这些特征。DL 网络架构基于对人工神经网络 (ANN) 进行建模，以执行一个特定的任务。ANN 模拟生物神经系统，由多个人工神经元组成。卷积神经网络 (CNN) 是 ANN 的一个子组，其输入由图像组成，特别适用于图像识别任务，因此，CNN 广泛应用于放射学中的各种任务，可以分为四个主要类别。

展示内容

有几篇论文描述了在肝脏和胰腺成像中解决不同问题的 AI 方法。这些问题可以总结为四个不同的类别：分割、量化、表征和图像质量改进。在本文中，我们报道这四个主要类别中的最新进展。

- 分割 (segmentation)：省略时间，优于手动方法。
- 量化 (quantification)：例如可以为肝肿瘤负担提供快速有效的纵向跟踪。
- 表征 (characterization)：借助与放射组学和大数据的关联，还可以提出诊断建议。
- 图像质量改进 (image quality improvement)：缩短扫描时间，提高图像质量。

① 课题背景

② 研究内容

③ 扩展

④ 参考文献

分割的标准

本组任务为医学影像分割算法，故以四个主题中的分割为主。

- 在分割中，AI 算法被用来执行给定器官的半自动或自动分割。
- 黄金标准通常是一个手动分割的器官数据库，称为地面真相 (ground truth)，以比较 AI 分割的结果。
- 比较通常通过使用基于重叠、基于大小、基于边界距离或基于边界重叠的方法来进行。
- 结果上最常用的是相似性系数 (Dice-Sørensen 系数, DSC)，其数值描述了两个组之间的相似性 (地面真相与算法得到的分割)。

DSC 系数

- DSC 数值表示了两个分割之间的空间重叠或共有体素的百分比 ($DSC = 1$, 完全重叠; $DSC < 1$ 且 > 0 , 部分重叠; $DSC = 0$, 没有重叠)
- 在比较对象上, 可用于比较手动与手动、手动与自动以及自动与自动的分割。
- Zou 等人将 DSC 定义为 kappa 统计的一种特殊情况, kappa 统计是一种广泛使用且可靠的一致性指数。正如 Zijdenbos 等人建议的那样, $DSC > 0.7$ 时表示有很好的重叠。

肝脏图像分割的表现

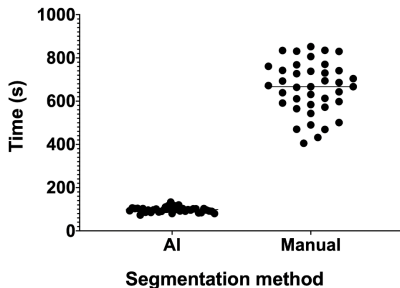
- AI 算法在肝脏分割方面表现出色，训练案例数为几十个时，Dice-Sørensen 系数 (DSC) 可超过 0.9。
- 使用更少的病例进行训练和测试会导致 DSC 分数较低。通过增加用于训练的检查数量或在算法本身中引入一些优化，可以获得增强的算法性能。
- DSC 在不同的肝脏状况之间没有显著差异，因此开发的算法对可能的肝脏形态变化具有鲁棒性。
- AI 算法也可以训练用于检测和分割肝脏病变。在这种情况下，与整个肝脏分割相比，DSC 通常较低。
- Qayyum 等人提出了一种混合的 3D 残差网络结构，具有挤压和激励机制。该方法应用于肝脏肿瘤分割挑战。

胰腺图像分割的表现

- 相对于肝脏来说稍微更具挑战性，这可能是由于胰腺的复杂形态和该器官的更高个体变异性造成的。
- Bagheri 等人调查了可能影响胰腺深度学习分割成功率的技术和临床因素。他们确定了五个参数，所有这些参数都与脂肪有关。
- 胰腺分割的 Dice-Sørensen 系数 (DSC) 比肝脏的要低，
- 在一个包含 102 个 MR 数据库的完全自动化的多器官分割研究中 (66 个用于训练, 16 个用于验证, 20 个用于测试), 肝脏的 Dice-Sørensen 系数 (DSC) 达到了 0.96, 胰腺为 0.88。类似的方法也应用于一个 CT 数据集 (验证使用了 66 个样本, 测试使用了 16 个样本), 肝脏和胰腺的 DSC 分别为 0.95 和 0.79

图像分割的时间

- 一旦训练完成，AI 算法在所需时间方面可以超越手动分割。例如：一个完整的肝脏可以在几秒钟内提取出来，每个 CT 切片的时间最长为 0.04 秒，DSC 为 0.95。



① 课题背景

② 研究内容

③ 扩展

④ 参考文献

量化

关于肝脏，有几篇论文探讨了在各种成像模式下量化肝脏脂肪变性和纤维化。

- Treacher 等人采用了 CNN 方法，将灰度弹性成像图像纹理与剪切波速度 (SWV) 相关联，但并未发现显著的统计关联。
- Schawkat 等人采用了机器学习方法，根据 T1 加权成像衍生的纹理参数将患者分为两类，即低级别和高级别纤维化。以组织病理学为金标准，机器学习方法的正确评估百分比为 85.7%，曲线下面积 (AUC) 为 0.82，而磁共振弹性成像 (MRE) 的 AUC 为 0.92。
- Cao 等人在非酒精性脂肪性肝病 (NAFLD) 中使用 DL 定量分析应用于 2D 超声成像。在区分中度和重度 NAFLD 方面非常有前景 (AUC=0.958)。

特征描述与诊断

- AI 特征描述算法使用 CT 成像作为输入进行开发。
- Cao 等人提出了一种多相卷积神经网络，Shi 等人发现，三相 DCE-CT 协议（动脉期、门静脉期和延迟期），结合卷积密集网络，其 AUC 为 0.920，用于区分 HCC 与其他肝脏病变，而四相协议模型的 AUC 为 0.925 ($p = 0.765$)。
- 一篇最近发表的论文提出了一种基于 CT 成像对癌症和非癌症患者进行分类的基于 CNN 的分析方法。这种方法显示出非常有前景，具有高灵敏度和特异度。尽管 CNN 模型有个别疏漏，但比之人工更有效。在基于 EUS 成像的 IPMN 恶性肿瘤诊断方面也报告了类似的结果。
- DL 方法显示出 95.7% 的灵敏度、92.6% 的特异度和 94.0% 的正确分类率。总之，如果正确开发和训练，AI 算法可能在特定任务中胜过人类。

图像质量改进

该应用主要基于 DL 方法:

- 通过深度学习 (DL) 重建的 CT 图像比标准的迭代重建 (IR) 算法在模拟体和真实腹部 CT 扫描上更少噪音, 具有更高的空间分辨率和更好的检测能力。
- 对不同的血管和腹部器官 (包括肝脏) 进行了重建的评估, 发现 DL 重建的对比噪声比 (CNR)、信噪比 (SNR) 和锐度明显更高。
- DL 重建应用于 MR 成像。Hermann 等人提出的一项研究中, DL 重建被用于加速上腹部 T2 加权 (T2W) 成像的获取。相比传统的单次触发 T2W (1:30 分钟) 和非笛卡尔呼吸触发 T2W (4:00 分钟), DL 采集只需要一个呼吸暂停 (16 秒)。而在噪声、锐度和伪影与非笛卡尔 T2W 之间在统计学上没有显著差异。
- DL 应用于 MR 可减少采集时间并提高图像质量。

① 课题背景

② 研究内容

③ 扩展

④ 参考文献

参考

- [1] Cardobi N, Dal Palù A, Pedrini F, Beleù A, Nocini R, De Robertis R, Ruzzenente A, Salvia R, Montemuzzi S, D'Onofrio M. An Overview of Artificial Intelligence Applications in Liver and Pancreatic Imaging. *Cancers (Basel)*. 2021 Apr 30;13(9):2162. doi: 10.3390/cancers13092162. PMID: 33946223; PMCID: PMC8124771.