

全自动人工智能身体成分工具的技术完善性：在异质性外部 CT 检查样本中进行评估

摘要

背景：临床可使用的分析影像检查的人工智能（artificial intelligence, AI）工具应该对检查参数的预期变化具有鲁棒性。

目的：评估一组自动化 AI 腹部 CT 身体成分工具在作者所在医院系统外的异质性外部 CT 检查样本上的技术完善性，并探讨工具失效的可能原因。

方法：这项回顾性研究纳入了 8949 例患者（平均年龄 55.5 ± 15.9 岁；4256 名男性，4693 名女性），他们在 777 个不同的外部机构进行了 11,699 次腹部 CT 检查，使用了来自 6 个不同制造商的 82 种不同型号的 CT，随后将图像转移到本地的 PACS 用于临床。使用三个独立的自动化 AI 工具来评估身体成分（骨骼衰减、肌肉量和衰减、内脏脂肪和皮下脂肪量），每项检查评估其中一个轴位序列。技术完善性定义为工具输出值在经验推导的参考范围内。审查失败项目（如工具输出值超出参考范围）以明确可能的原因。

结果：在 11,431/11,699（97.7%）的检查中，这三种工具在技术上都是完善的，在 268/11,699（2.3%）的检查中至少有一种工具不合格。骨骼、肌肉和脂肪工具的技术完善率分别为 97.8%、99.1% 和 98.0%。有一种类型的图像处理错误（等距误差，由于不正确的 DICOM 体素尺寸信息）在三种工具都失败的检查中占了 81/92（88%），每当这种错误出现时，这三种工具均失败了。等距误差是所有工具最常见的失败原因（骨骼为 31.6%，肌肉为 81.0%，脂肪为 62.8%）。79/81（97.5%）的等距误差发生在同一制造商生产的扫描仪中；80/81（98.8%）发生在同一型号的扫描仪中。在骨骼、肌肉和脂肪工具的失败检查中，分别有 59.4%、16.0% 和 34.9% 没有确定失败的原因。

结论：自动化 AI 身体成分工具在异质性外部 CT 检查样本中具有较高的技术完善率，支持工具的普适性和广泛使用的潜力。

临床意义：与技术因素有关的 AI 工具故障的某些原因可以通过适当的采集和重建协议在很大程度上得到预防。