





天津医科大学总医院

用于心肌灌注显像的人工智能辅助诊断系统的建立与验证

 作者：张如意，王澎，王深，李宁，郑薇，谭建

 作者电子信箱：1663918881@qq.com

 作者电话：15139093625

 通讯作者：孟召伟，贾强

 通讯作者电子信箱：zmeng@tmu.edu.cn
jiaqiang4321@163.com

 通讯作者电话：186 2203 5159
137 5272 8158

背景

Background

冠状动脉疾病 (CAD) 是世界范围内发病和死亡率较高的主要原因之一。目前用于CAD诊断的影像学方法有心电图、有创冠状动脉造影 (ICA) 或无创CT冠状动脉造影 (CTCA)、心肌灌注成像 (MPI)、超声等。这些方法均得到了较广泛的应用。其中，基于单光子发射计算机断层扫描仪 (SPECT) 的MPI是一种较为成熟的无创检测方法，可直接反映心脏对示踪剂的摄取程度，从而较准确的评估缺血、瘢痕、左室容积、射血分数等。随着过去几十年的发展，MPI已经从一种检测冠心病的高精度诊断试验 (据报道其平均灵敏度和特异性分别为90%和75%) 发展为危险度分层的必要工具。

传统上，医学影像的诊断过程较为耗时。同时，患者对于高效的医疗服务需求越来越高。近年来，AI与医学影像的结合一直是一个热门话题，目前有较多的研究证明其可行性和实用性 (如AI与超声、计算机断层扫描、磁共振成像或组织病理学等结合应用)。然而，人工智能在核医学成像，尤其是MPI中的应用较少。

目的

Objectives

本研究的主要目的是探讨人工智能在核医学MPI中应用的可行性，建立并验证一种用于鉴别冠心病心肌缺血的人工智能 (artificial intelligence, AI) 辅助诊断系统。

研究人群

本研究回顾性地选择了来自4个医疗中心的599例接受门控MPI检查方案的患者。由三个轴位图像组成：分别为短轴(13, 267幅)、水平长轴(11, 465幅)和垂直长轴(11, 676幅)。然后将患者分为两个子集，包括训练集(占70%)和验证集(占30%)。利用训练集对神经网络进行训练和开发，利用验证集对神经网络的预测能力进行测试。

数据标注与模型构建

图像按照ATA于2002年提出的标准化心肌分割和心脏断层成像命名法进行标记。因此，在标记过程中总共使用了17个标签(从基底部到心尖部)(图1)。如果发现病变存在于心尖部，我们在所有三个轴(即SA、HLA和VLA)上用17来标记该区域。在本研究中，我们选择了有监督学习方法-区域深度学习技术，这是一种基于ROI(感兴趣区域)的传统神经网络，名为YOLO (you only look once, version 3)来完成训练。YOLO算法由四个主要阶段组成：标记图像的预处理、利用深度卷积网络进行特征提取(训练)、带置信度的病灶检测(计算)、最后利用全连接神经网络(FC-NNs, 输出)进行病灶分类。机器学习网络使用Python软件(版本3.7.3)实现。

数据分析

为了测试神经网络的训练性能，计算了准确率、召回率和平均准确率等指标。为检验其临床有效性，对验证集进行ROC分析。我们还将人工智能的预测精度与其他阅片者(初学者、经验不足和经验丰富的诊断者)的预测精度进行了比较。通过计算Cohen's Kappa系数，对人工智能和临床诊断进行了一致性分析。

1

2

3

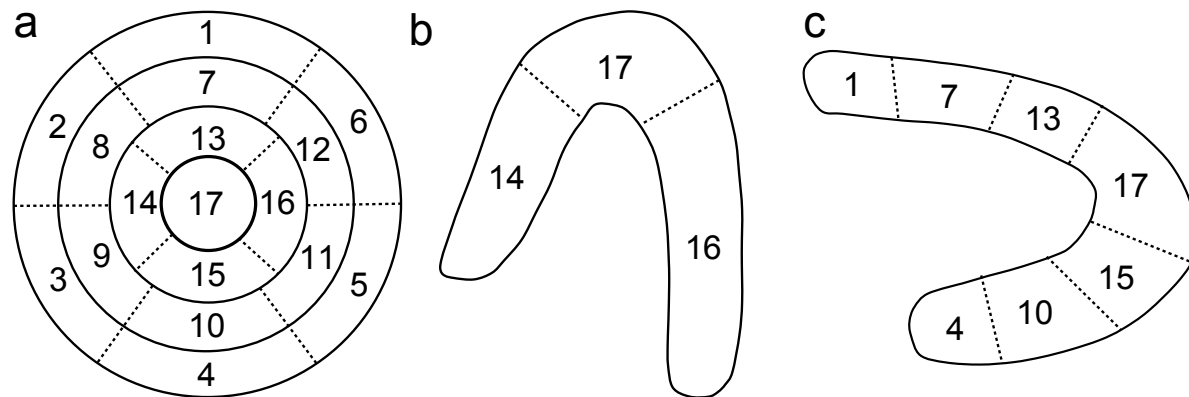


图1 心脏断层显像的标准化心肌分割和命名。a, 短轴; b, 水平长轴; c, 垂直长轴; 1、前壁基底部; 2、前间壁基底部; 3、下间壁基底部; 4, 后壁; 5, 下侧壁基底部; 6, 前侧壁基底部; 7, 前壁中间部; 8, 前间壁中间部; 9, 下间壁中间部; 10, 下壁中间部; 11, 下侧壁中间部; 12, 前侧壁中间部; 13 前壁心尖部; 14, 间壁心尖部; 15, 下壁心尖部; 16, 侧壁心尖部; 17, 心尖

表1 短轴训练准确度

	Accuracy (%)	Recall (%)	Average precision (%)
前壁 (1区)	78.78	98.01	96.88
前间壁 (2区)	77.67	96.26	94.64
下间壁 (3区)	81.37	95.31	93.16
后壁 (4区)	84.73	98.76	98.15
下侧壁 (5区)	74.37	96.10	93.78
前侧壁 (6区)	74.73	94.04	90.04

在SA中，准确率为74.37%–84.73%，召回率为94.04%–98.76%，平均准确率为90.04%–98.15%(表1)。后壁的正确率、召回率和平均准确率均高于其他段(分别为84.73%、98.76%和98.15%)。在HLA中，正确率为66.20%–79.32%，召回率为88.50%–92.77%，平均准确率为81.90%–90.34%(表2)。在14、17、16段中，第17段正确率最高(79.32%)，召回率最高(92.77%)，平均准确率最高(90.34%)，而第14段正确率最低(66.20%，召回率为88.50%，平均准确率为81.90%)。VLA的正确率在88.02%–94.64%之间，召回率在76.96%–91.8%之间，平均准确率在80.17%–93.37%之间(表3)。17段的正确率最高(94.64%)，而1段的召回率和平均准确率分别为91.80%和93.37%。第15段的准确率最小(87.91%)，第13段的召回率和平均准确率分别最小(76.96%和80.17%)。

表2：水平长轴训练准确度

	Accuracy (%)	Recall (%)	Average precision (%)
间壁 (14区)	66.20	88.50	81.90
心尖 (17区)	79.32	92.77	90.34
侧壁 (16区)	70.83	88.85	87.02

表3：垂直长轴训练准确度

	Accuracy (%)	Recall (%)	Average precision (%)
前壁底部 (1区)	90.32	91.80	93.37
前壁中部 (7区)	89.36	85.14	86.47
前壁心尖部 (13区)	88.02	76.96	80.17
心尖部 (17区)	94.64	84.13	88.51
下壁心尖部 (15区)	87.91	81.92	85.39
下壁中部 (10区)	90.91	86.33	88.30
下壁底部 (4区)	92.69	87.57	88.71

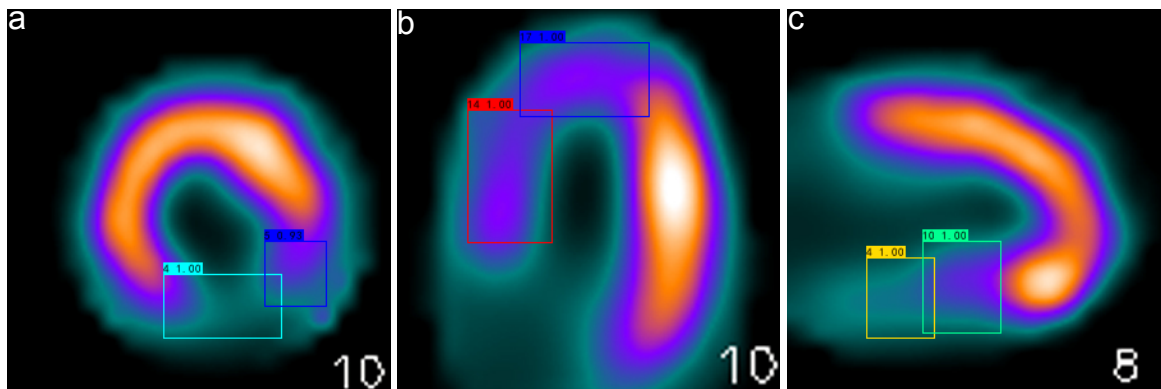


图2 YOLO v3网络预测病灶。A, 短轴;b、水平长轴;c, 垂直长轴

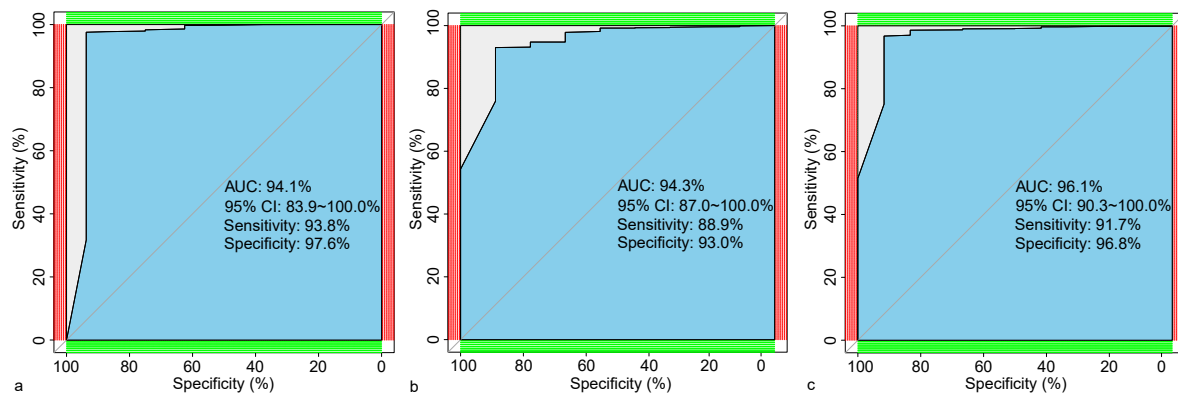


图3 不同轴的ROC分析。a, 短轴; b, 水平长轴; c, 垂直长轴



图2是AI预测效果图，所有病变的节段都得到了较准确的识别。在验证集的ROC分析中，SA的敏感性、特异性和AUC分别为93.8%、97.6%和94.1%(图3a)。HLA的敏感性为88.9%，特异性为93.0%，AUC为94.3%(图3b)。VLA的敏感性为91.7%，特异性为96.8%，AUC为96.1%(图3c)。

如图4所示，在与AI的敏感性和特异性比较时，所有诊断者在随机验证集上的差异均具有统计学意义（所有P值 <0.05 ）。

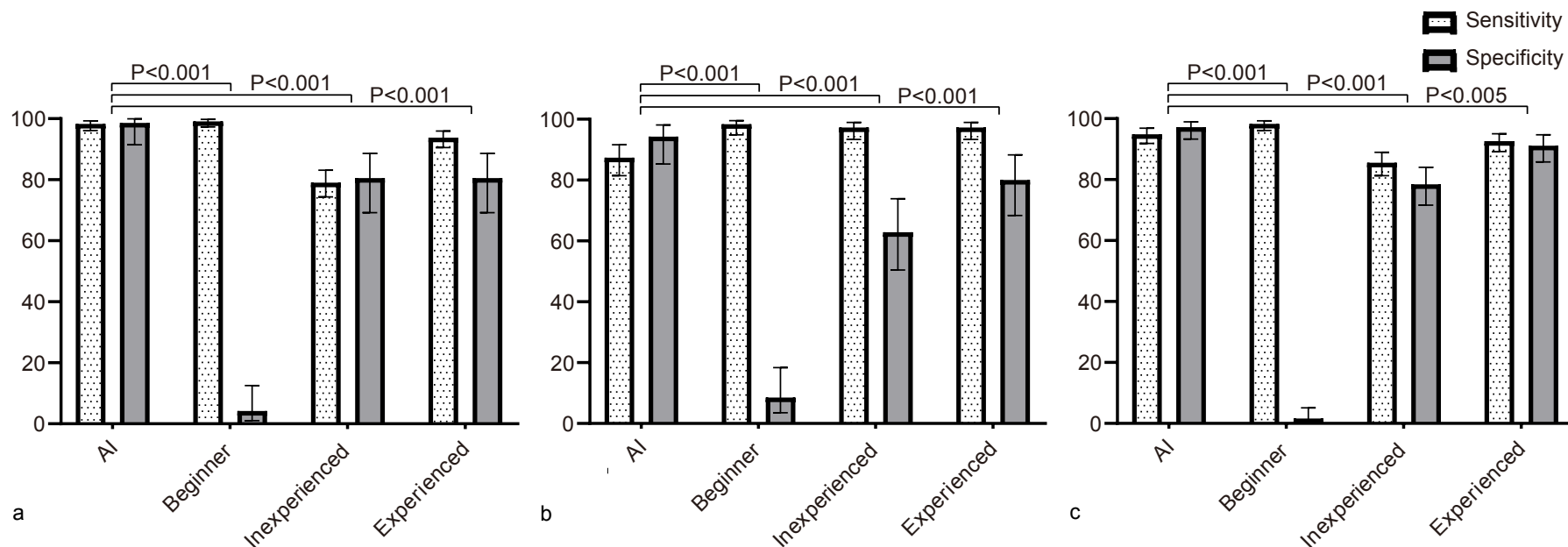


图4 不同诊断者敏感性和特异性的比较。 a, 短轴; b, 水平长轴; c, 垂直长轴

表4 AI、初学者、经验不足、以及经验丰富诊断者在验证机中的诊断灵敏度对比 †, P值<0.05; ††, P值 <0.001; (), 95% 置信区间; SA, 短轴; HLA, 水平长轴; VLA, 垂直长轴

	AI (%)	Beginner (%)	Inexperienced (%)	Experienced (%)
SA	98.30 (96.15~99.31)	99.15 (97.33~99.78)	79.04 (74.34~83.09) ††	93.77 (90.58~95.96) †
HLA	87.36 (81.44~91.66)	98.35 (94.87~99.57) ††	97.25 (93.36~98.98) ††	97.25 (93.36~98.98) ††
VLA	94.89 (91.89~96.85)	98.30 (96.14~99.31) †	85.51 (81.30~88.93) ††	92.61 (89.23~95.03)

表4显示了随机验证集中AI与初学者、无经验和经验丰富的阅片者之间的敏感度比较。初学者在SA (99.15%, 95% CI: 97.33-99.78%)、HLA (98.35%, 95% CI: 94.87-99.57%)和VL (98.30%, 95% CI: 96.14-99.31%)中敏感性最高。而在SA中, 初学者与AI之间的差异无统计学意义。经验丰富的阅片者在HLA中的敏感性高于AI, 差异有统计学意义(P值<0.001), 但在SA和VLA中均低于AI (SA, P值<0.05, VL, P值<0.05)。经验不足的阅片者在SA和VLA中敏感性最低, 与AI相比差异有统计学意义(P值<0.001)。然而, 在HLA, 具有与经验丰富的诊断者相同的敏感性。

表5 AI、初学者、经验不足、以及经验丰富诊断者在验证集中的诊断特异度对比 †, P值 <0.05; ††, P值 <0.001; (), 95% 置信区间; SA, 短轴; HLA, 水平长轴; VLA, 垂直长轴

	AI (%)	Beginner (%)	Inexperienced (%)	Experienced (%)
SA	98.61 (91.46~99.93)	4.17 (1.08~12.50) ††	80.56 (69.20~88.59) ††	80.56 (69.20~88.59) ††
HLA	94.29 (85.27~98.15)	8.57 (3.53~18.35) ††	62.86 (50.43~73.86) ††	80.00 (68.39~88.26) †
VLA	97.24 (93.33~98.98)	1.66 (0.43~5.16) ††	78.45 (71.61~84.06) ††	91.16 (85.80~94.70) †

如表5所示，初学者在三个轴上的特异性最小(SA为4.17%，HLA为8.57%，VLA为1.66%)，与AI相比差异均有统计学意义(P值<0.001)。AI在SA (98.61%，95%CI为91.46–99.93%)、HLA (94.29%，95%CI为85.27–98.15%)和VLA (97.24%，95%CI为93.33–98.98%)中的特异性最大。虽然无经验和有经验的阅片者在SA中的特异性相同，但可以注意到，有经验的阅片者在HLA和VLA的特异性要大于经验不足阅片者。此外，经验丰富的阅片者与经验不足的阅片者的特异性差异均有统计学意义(P值<0.05)。

表6 在随机验证集中的敏感性和特异性的一致性分析（诊断者vs. 金标准）†, P值<0.05; ††, P值<0.001; SA, 短轴; HLA, 水平长轴; VLA, 垂直长轴

	Cohen's Kappa			
	AI	Beginner	Inexperienced	Experienced
SA	0.943 ††	0.052 †	0.447 ††	0.712 ††
HLA	0.754 ††	0.095 †	0.662 ††	0.804 ††
VLA	0.905 ††	0.013	0.630 ††	0.827 ††

一致性检验表明, AI在三个轴位上与金标准的一致性最好 (Cohen's Kappa 系数 :SA 为 0.943, HLA为0.754, VLA为0.905, P值<0.001, 表6), 同样, 初学者与金标准的一致性最小 (Cohen's Kappa 系数 :SA 为 0.052, HLA 为 0.095, VLA为0.013; SA和HLA P值<0.05, VLA P值 > 0.05)。经验丰富阅片者的Cohen's Kappa系数小于AI, 但大于经验不足阅片者 (SA 为0.712比0.447, HLA为0.804比0.662, VLA为0.827比0.630, P值<0.001)。



研究结论

我们研究证明人工智能系统在MPI大型多中心人群中具有良好的预测准确性、一致性和临床实用性，因此可以用于辅助临床诊断。



感谢您的观看！

