

人工智能辅助前列腺癌磁共振影像的精准分级中国专家共识(2022)

中国老年学和老年医学学会 中国医师协会临床精准诊疗专业委员会
中国研究型医院学会医学影像与人工智能专业委员会

通信作者: 卢剑, 北京大学第三医院泌尿外科, 100191, Email: lujian@bjmu.edu.cn; 林健, 北京大学第一医院泌尿外科, 100034, Email: linjianbj@163.com

【摘要】 前列腺癌是老年男性最常见的泌尿生殖系统恶性肿瘤之一, 精准的术前肿瘤分级有助于进一步提高前列腺癌患者的诊疗效果。磁共振影像(MRI)作为前列腺癌术前诊断的主要工具, 已在临床实践中显示出对前列腺癌诊疗的重要价值。但由于我国不同医疗机构在成像设备、检查方法和诊断水平上存在差异, MRI在前列腺癌诊疗中的潜在价值未能得到充分发挥, 因此亟需更为先进的量化分析方法进一步扩大MRI在前列腺癌精准诊疗中的贡献。人工智能(AI)技术应用于前列腺癌MRI分析, 有助于提高其量化分析的质量与评估水平, 在前列腺癌精准分级问题上具有巨大潜力。但是基于影像的AI模型建立过程通常较为复杂, 特别是在其数据集准备、模型构建和模型评估的各个阶段仍需进行统一和规范。本共识旨在推进AI辅助前列腺癌MRI精准分级模型构建全过程的规范化, 从而提高前列腺癌临床精准分级水平。

【关键词】 前列腺癌; 磁共振影像; 人工智能; 分级预测

基金项目: 国家自然科学基金项目(81922040, 61871004); 北京市自然科学基金重点研究专项项目(Z200027); 中国科学院青年创新促进会(2019136)

DOI: 10.3760/cma.j.issn.0254-9026.2022.06.008

Expert consensus on artificial intelligence-assisted precise grading of prostate cancer via magnetic resonance imaging in China (2022)

Chinese Society of Gerontology and Geriatrics, Professional Committee of Clinical Precision Diagnosis and Treatment of Chinese Medical Doctor Association, Medical Imaging and Artificial Intelligence Professional Committee of Chinese Research Hospital Association

Corresponding author: Lu Jian, Department of Urology, Peking University Third Hospital, Beijing 100191, China, Email: lujian@bjmu.edu.cn; Lin Jian, Department of Urology, Peking University First Hospital, Beijing 100034, China, Email: linjianbj@163.com

【Abstract】 Prostate cancer is one of the most common malignant tumors of the genitourinary system in elderly men. Accurate preoperative tumor grading can further improve the diagnosis and treatment outcomes of prostate cancer patients. Magnetic resonance image(MRI), as the main tool for preoperative diagnosis of prostate cancer, has shown an important value for the diagnosis and treatment of prostate cancer in clinical practice. However, due to differences in imaging equipment, examination methods and expertise in diagnosis among different medical institutions in China, the potential value of MRI in the diagnosis and treatment of prostate cancer has not been fully realized. Therefore, more advanced quantitative analysis methods are urgently needed to further expand the contribution of MRI in precise diagnosis and treatment of prostate cancer. The application of artificial intelligence (AI) technology in the analysis of magnetic resonance imaging of prostate cancer can help improve the quality and evaluation capability of quantitative analysis, and has great potential in accurate grading of prostate cancer. However, image-based AI model building is usually complicated, especially in the stages of data set preparation, model construction and model evaluation, which still need to be integrated and standardized. This consensus aims to promote the standardization of the whole process of model construction for AI-assisted precise grading of prostate cancer based on magnetic resonance imaging, so as to improve clinical precision grading of prostate cancer.

【Key words】 Prostate cancer; Magnetic resonance imaging; Artificial intelligence; Grading prediction

Fund Program: Program of the National Natural Science Foundation of China (81922040, 61871004); Key Research Projects of Beijing Municipal Natural Science Foundation (Z200027); Youth

Innovation Promotion Association(2019136)
DOI:10. 3760/cma. j. issn. 0254-9026. 2022. 06. 008

前列腺癌是老年男性最常见的泌尿生殖系统恶性肿瘤之一。2020 年全球范围内男性癌症患者中前列腺癌发病率及死亡率分别位居第 2 和第 5 位^[1]。我国前列腺癌发病率与死亡率均呈现逐年递增态势,且死亡率与发病率比值远高于欧美发达国家^[2-4]。目前,前列腺癌正严重威胁我国乃至世界范围内男性的生命健康。针对前列腺癌患者的精准评估与个性化诊疗有助于改善患者预后,提高全体患者受益^[5]。因此,精准的前列腺癌术前分级有助于进一步提高前列腺癌患者的诊疗效果。

磁共振影像(MRI)作为前列腺癌术前诊断与诊疗方案制定的重要依据之一,充分展现了影像学在描绘肿瘤异质性中所扮演的重要角色^[6]。但是,由于临床采用 MRI 进行前列腺癌诊断中易受到成像设备、检查方法和评估水平的影响^[7],亟需更为先进的量化分析方法,实现基于 MRI 的前列腺癌精准分级预测。

目前,得益于人工智能(AI)技术的发展及其在相关临床应用问题中的成功探索^[8],AI 结合 MRI 可以实现图像中海量信息的提取,进而实现更为完善的肿瘤异质性描述,为基于 MRI 的前列腺癌精准分级提供了新机遇。当前 AI 辅助前列腺癌 MRI 精准分级模型较为多样,在临床应用转化过程中,可能因数据集标准不统一、模型构建过程不明晰和性能评估不全面等因素,导致模型可信度不足,临床应用与推广受限。

基于此,中国老年学与老年医学学会与中国医师协会临床精准诊疗专业委员会、中国研究型医院学会医学影像与人工智能专业委员会共同组织专家编写了 AI 辅助前列腺癌 MRI 精准肿瘤分级预测专家共识。本共识适用于构建 AI 辅助前列腺癌 MRI 分级系统的数据集构建与数据预处理、分级模型构建和分级模型评估等研究目的。旨在推进人工智能辅助前列腺癌术前精准预测,明确 AI 相关模型及系统在该场景下的构建流程,推动临床应用转化。

数据集构建与数据预处理

MRI 数据通常存在来源不一致,成像设备不统一,采集方法多样化,图像质量不标准等问题。如何实现有效且统一的数据集准备是本部分工作

的重点。具体内容包括:数据采集、数据标注和数据预处理。

一、数据采集

1. 伦理审批要求:数据采集应具备伦理委员会的伦理批准,获得参与者或其近亲提供的书面知情同意。由于 AI 辅助前列腺癌 MRI 精准肿瘤分级预测研究涉及人类参与者,因此需遵守 1964 年赫尔辛基宣言及其后来修订版本的基本原则。

2. 成像设备参数及采集时间要求:随着医疗水平的进步与 MRI 成像设备的发展,结合当前医疗中心设备配置现状,推荐采用磁场强度为 3.0T 的 MRI 成像设备进行数据采集^[9-12]。采集过程中所涉及的层厚、层间距、回波时间、重复时间等参数的设置与选择,均应严格按照标准化流程进行设置。另外,MRI 采集应早于穿刺活检,从而避免穿刺手术引发的出血带来的图像信息干扰。

3. 扫描序列要求:根据 PIRADs v2.1^[13]技术规范要求,T2 加权像(T2WI)用于描绘前列腺癌变部位解剖结构,T2WI 图像应获取轴向位和至少一个与之正交方位(矢状位或冠状位)的图像,从而避免因局部肿瘤结构方向性差异所导致的评估受限。部分研究者已经使用 T2WI 进行了 AI 辅助前列腺癌术前精准预测研究^[9-12]。同时,建议计算表面扩散系数(ADC)用于完善肿瘤异质性描述。ADC 图由两个表征不同水平(由 b 值表示)的水分子扩散程度的扩散加权成像(DWI)计算获得。低 b 值推荐采用 $0\text{ s/mm}^2 \leq b \leq 100\text{ s/mm}^2$ ($50\text{ s/mm}^2 \leq b \leq 100\text{ s/mm}^2$ 更优),高 b 值推荐采用 $800\text{ s/mm}^2 \leq b \leq 1\,000\text{ s/mm}^2$,从而避免扩散峰度效应^[14]。此外,若单独采用 DWI 序列用于分析,仅使用高 b 值即 $b \geq 1\,400\text{ s/mm}^2$ 的影像^[14]。部分研究者已经使用 DWI 或 ADC 图进行了基于 MRI 影像的前列腺癌诊断研究^[11-12]。本共识中涉及的预测模型主要为基于平扫序列的模型。当前临床趋势倾向于依靠平扫序列实现前列腺癌诊断相关问题,减少临床诊断对增强序列的依赖,进而提高患者检查收益^[15]。因此,仅基于平扫序列的预测模型如能实现精准预测,将有希望进一步降低甚至取代临床诊断对于增强序列的需求。在提高诊断效率的同时,降低检查成本,减少扫描时间。

4. MRI 图像质量控制:基于不同的研究目的,

由专业从事前列腺癌影像诊断的放射科医师对纳入研究的影像进行图像质量评价。图像质量评价内容包括:图像信噪比,空间分辨率,图像均匀度,伪影等。对于存在伪影、噪点、前列腺部位影像质量过差,图像资料不完整等问题的患者数据,推荐重新扫描以获取满足质控要求的影像。

5. 组织学检查要求:所有纳入研究的患者都有其相对应的组织学检查结果。若患者经历前列腺根治性切除术,则以术后病理为金标准;若患者未经历前列腺根治性切除术,则以活检病理为金标准。临床常用活检方式包括系统性穿刺活检,饱和穿刺活检和靶向穿刺活检等。经活检或根治术后的前列腺组织切片经石蜡包埋与 HE 染色后扫描生成全景数字病理图像。标准化术后病理报告:包含对主要病灶、次要病灶及重要病灶的肿瘤组织分化等级评估及组织学类型等。推荐采用两步检查方案来避免潜在的评估偏差:首先,低年资病理学家阅读前列腺标本和报告,然后,由高级病理学家审核检查。如果有分歧,讨论结果直到达成共识。

二、数据标注

针对肿瘤或目标区域的标注可采用自动,半自动,或人工勾画等方式。目前由于用于前列腺癌肿瘤自动识别及分割的软件系统精度和成熟度有限,因此推荐采用医师勾画方式进行基于视觉观察的瘤灶区域勾画。标注医师可选用目前国际广泛采用的 ITKSNAP、3DSlicer 及中国科学院分子影像重点实验室自主研发的 MITK 等医学图像浏览及标注工具对影像中的肿瘤区域进行勾画。

1. 标注序列:标注序列主要分为以下两类标注方式:(1)仅在 T2WI 序列上标注,其余序列通过与 T2WI 配准方式获得相应标注。配准方式包括刚体配准,仿射配准,B 样条配准等方式,从而实现不同序列图像的空间域对齐。(2)分别基于 T2WI 和高 b 值 DWI 序列标注。ADC 图的标注通过直接投射到 DWI 标注结果获得。核心思想为以 T2WI 作为解剖结构参考,辅助 DWI 病变区域的标注,实现基于多序列信息的联合标注。

2. 标注流程:(1)若为回顾性研究,对于接受根治术的患者,医师需结合术后组织病理标本与 MRI 影像,即将术后病理指示的肿瘤区域拓印至 MRI 影像上,生成肿瘤区域掩模。对于未接受根治术的患者,可参考靶向活检结果进行辅助标注。(2)若为前瞻性研究,可遵循前列腺影像报告和数

并尽量保证病灶区域边缘的准确性。标注结果应保存为独立且格式统一的文件。保存格式包括:DICOM、NRRD 和 Nifty 格式等。

3. 标注医师:前列腺癌病灶标注人员应至少由 3 位具备前列腺癌影像诊断经验的医师组成,其中 2 位为标注医师(低年资或中年资医师),对瘤灶区域进行边界勾画。另 1 位为审核医师(高年资医师),针对勾画区域进行复检,并对存在疑义标注样本进行校正或重新勾画。当两位标注医师在决策存在差异时,应遵循审核医师标注结果。对标注医师和审核医师的要求为:(1)熟悉理解前列腺部位解剖及影像的相关知识;(2)熟练掌握标注软件操作流程;(3)经过专门前列腺癌标注培训。

三、数据预处理

目前,由于影像生物标志物标准化计划(IBSI)对于图像预处理的相关研究尚未达成共识,为了防止数字图像预处理阻碍再现性,并确保构建的模型实现最佳的可比性,本共识建议以下预处理方式。

1. 数据去识别化:由于从医院系统中导出的 MRI 影像文件,通常不仅涉及图像信息,还包括患者的姓名、地址、联系方式等私人信息。因此,需首先对数据进行去隐私化处理。推荐以患者的住院号作为数据集的患者唯一标识,使得患者的影像和病理原始检查结果实现一一映射。对于多科室间的数据查询与调配,推荐采用上述患者唯一标识进行关联。

2. 数据一致性处理:为有效解决多中心研究中存在的数据不一致性和偏倚性的问题,需进行数据一致性操作,从而有助于实现跨中心数据建模与验证。具体步骤包括:(1)图像收集过程标准化:各中心按照统一的数据采集标准进行数据收集。(2)基于特征域的一致性处理:可以采用 Z-score 等方法对数据进行标准化。(3)基于图像域的一致性处理:建议基于标注信息,提取包含瘤内及瘤周在内的矩形感兴趣区域(ROI),并将 ROI 统一插值到相同尺寸。

3. 数据增强:为了丰富数据多样性,提高模型泛化能力,以及降低模型过拟合风险,可采取数据增强方法。具体方法包括:图像翻转、图像旋转、图像扭曲、图像仿射变换、图像缩放、图像压缩、图像随机剪切、图像对比度调整、亮度调整和生成对抗网络生成等。

推荐 1:数据采集应获得伦理批准。

推荐 2:本共识中涉及的预测模型主要为基于

平扫序列的模型。

推荐 3: 纳入数据应满足图像质控要求。

推荐 4: 建议对数据进行去识别化、一致性处理、数据增强等预处理操作。

分级模型构建

采用 AI 技术进行前列腺癌分级模型构建的方法较为多样,大部分研究显示出较高的模型准确率及临床应用潜力^[9-12]。但是面对临床应用转化需求,需要对这些方法进一步进行筛选、对比与流程规范化。因此,如何针对前列腺癌分级预测的模型构建问题,选择或设计符合临床应用转化的 AI 方法,采用规范化的建模方式,提高模型在临床应用过程中的可信度是本章节的重点。

一、基于人工定义特征的前列腺癌分级预测模型构建

基于人工定义特征的前列腺癌分级预测模型以具体临床问题为导向,采用与之对应的病理金标准作为预测标签。建模过程通常包括:分类标签确立、肿瘤区域图像特征提取、特征筛选和模型构建四个步骤。该方法主要应用于显著性前列腺癌预测^[16]及高级别前列腺癌预测模型的构建^[17]。

1. 分类标签确立:基于人工定义特征的前列腺癌分级预测模型的构建主要集中应用于显著性前列腺癌预测和高级别前列腺癌预测两个临床问题。在不考虑上下调风险情况下^[18],通常选择活检 Gleason 评分作为肿瘤侵袭性表征,按照国际泌尿病理学会(ISUP)拟定标准中的要求^[19],将患者划分为显著性前列腺癌(阳性组)与非显著性前列腺癌(阴性组),或高级别前列腺癌(阳性组)与低级别前列腺癌(阴性组)。分类标签中阳性组定义为 1,阴性组定义为 0,其在训练过程中被用于与模型的预测结果进行对比,并通过计算分类误差损失的方式引导模型训练。

2. 肿瘤区域图像特征提取:基于肿瘤勾画区域提取基于影像的预定义量化图像特征,原则上应尽可能纳入全面描述肿瘤异质性的特征,涵盖但不局限于下述三类图像特征^[20],即:低维特征(肿瘤病灶区强度、形状):一维统计特征又称作强度特征,与 ROI 内像素灰度值的分布有关,主要包含反映其灰度分布直方图整体信息的特征,比如:灰度均值、灰度标准差、峰度、偏度等。形状特征主要反映肿瘤的三维尺寸和形状,用以表征病变组织的尺寸信息、体积比例和生长形状等。高维特征(肿瘤

纹理):纹理特征属于二阶统计学特征,常见的纹理特征可以通过灰度共生矩阵等方式获得。该类特征考虑二维空间中像素之间的分布关系,是量化肿瘤异质性的重要特征之一。复杂特征(滤波特征):基于空间变换的特征,如:将原图进行 Gabor 变换、小波变换等变换后,分析其在频率或尺度空间的纹理。经过不同方式滤波后的图像能反映出与原图不同的纹理特性和像素分布特性,从而更好地表现图像中存在的深层特征。由于前列腺区域 MRI 横断位成像通常层间分辨率有限,优先推荐计算二维的肿瘤特征,通过对多层图像对应的特征位点进行均值处理的方式^[17]整合多层影像特征。基于人工定义的三维影像特征在部分研究中展现了更有效的疾病关联性^[21],但在前列腺癌分级预测问题上有待进一步考证。

3. 特征筛选:特征筛选对冗余度高或共线性强的特征集进行选择或变换,从而获得更优的特征组合,进而提高分级模型构建水平。特征筛选的方法包括但不限于:方差选择法、相关系数法、卡方(χ^2)检验、U 检验法、皮尔森检验法、递归消除特征法、套索回归(LASSO)和基于树模型的方法等。其中,采用表征显著性的 P 值作为评估特征重要性的手段时,可在显著性阈值($P < 0.05$)基础上结合建模需求进行适当调整,使得模型能够尽可能纳入充分且必要的特征集合用于预测模型的构建。

4. 模型构建:目前,众多机器学习的模型可被用于建立基于人工定义特征的前列腺癌显著性预测模型。常用的模型包括但不限于:随机森林(RF)、支持向量机(SVM)、人工神经网络(ANN)、聚类分析(Clustering)及一系列基于此类基础算法模型的优化算法模型等^[20]。模型训练过程中通常推荐采用交叉验证方式挑选最优参数,包括多折交叉验证、留一交叉验证等,从而消除由于数据随机划分带来的选择偏倚性,提高模型拟合效果和鲁棒性。模型若在训练过程中存在较严重的过拟合或欠拟合,可返回特征筛选步骤对特征筛选方法和特征筛选约束条件做优化调整。

二、基于深度学习的前列腺癌分级预测模型构建

基于深度学习的前列腺癌分级预测模型同样首先需要确立与具体临床问题对应的病理评估结果作为金标准。该方法的建模及预测过程通常对单层影像进行预测,进而通过投票策略获得患者的评估结果^[11]。该类方法除应用于显著性前列腺癌

预测及高级别前列腺癌预测外^[11],也可被拓展为分级预测,但由于前列腺横断位成像过程中,层间分辨率有限,通常该类模型用于分级预测模型准确率较为有限。部分研究者尝试通过学习专家经验,从多层影像中搜索关键信息构建决策路径的方式,生成预测结果,获得了成功的临床应用,展现了临床应用转化潜力^[9-10]。具体步骤包括:分级标签确立、输入数据结构确立、卷积神经网络结构选取和模型构建。

1. 分级标签确立:基于深度学习的前列腺癌分级预测模型根据不同的监督标签表征内容可以分别实现:显著性前列腺癌预测、高级别前列腺癌预测和分级预测。除用于显著性前列腺癌与高级别前列腺癌标签构建外,可将 Gleason 评分按照 ISUP 国际标准^[19]划分为五级变量,并将其定义为分级模型监督标签。针对分级标签选择,由于前列腺癌活检和术后病理评估存在偏差^[18],应采用术后病理评估结果作为金标准。

2. 输入数据结构确立:模型输入推荐使用 MRI T2WI,用于软组织结构像信息全貌获取。在此基础上,可拓展多序列 MRI 数据,包括但不限于 DWI 和 ADC 等影像数据,用于从不同观测角度完善肿瘤异质性描述。多序列数据以并行输入方式输入网络。对由于扫描参数差异导致的维度不对应问题,应采用计算机辅助插值方法进行数据维度统一。针对仅分析单个肿瘤区域的模型构建,按照肿瘤边界区域,提取包含肿瘤区域信息的外接矩形。该过程中可适度放宽外接矩形与边界像素距离,用于保全肿瘤及癌旁信息。最后,将其等比例缩放至符合预定网络输入维度的尺寸,并对无像素值的区域通过赋予零值的方式进行补全。

3. 卷积神经网络结构选取:基于深度学习的前列腺癌分级预测模型以卷积神经网络模型作为基础展开。卷积神经网络的结构及性能决定了模型对于原始影像进行编码并提取深层次特征的质量,并进一步决定了通过肿瘤图像量化肿瘤异质性的全面性^[9-10]。目前卷积神经网络结构应用于前列腺癌肿瘤分级预测问题的典型结构包括:U-Net、ResNet、DenseNet、InceptionV3 等。网络结构选取包括但不限于上述典型网络结构,并可通过微调与修改实现自定义设计,提高模型整体性能。

4. 模型构建:模型构建的模式通常可被划分

为:端到端模式、基于投票的弱监督学习模式和基于深度学习特征提取的机器学习模式。(1)端到端模式,该模式直接通过输入患者的影像数据,通过模型计算,无需分级处理直接获得患者前列腺癌分级预测结果。技术方法包括:将多层影像打包为三维数据结构形式通过三维卷积神经网络进行计算方式^[22]、通过深度强化学习对多层影像关键信息进行联立推理获得整体预测结果的方式及采用多示例学习通过动态分配不同层面的预测结果的权重获得整体预测结果的方式^[9-10]。(2)基于投票的弱监督学习模式,该模式将单一患者的三维影像拆解为多个单层影像进行分析,最终的患者预测结果通常选用多个单层影像预测结果的均值^[11]。训练过程中采用弱监督方式,即将患者的评估金标准平等赋予影像的每一层用作监督信息。为突出样本间差异性,也可在模型训练过程中对样本的权重做适量调整,用以获得更好的模型预测效能^[9]。(3)基于深度学习特征提取的机器学习模式。该模型训练与预测模式为:首先利用基于二维自然图像预训练模型的网络对原始的每一层图像进行信息编码与特征提取,进而利用机器学习模型将多层影像的特征通过均值方式整合,最后建模训练获得患者预测结果。模型的训练阶段包括:训练策略选择、损失函数设计、优化策略选择和超参数选取。训练策略选择:强监督训练,通常应用于端到端模式和基于深度学习特征提取的机器学习模式;弱监督训练,通常应用于基于投票的弱监督学习模式。损失函数设计:损失函数根据预测任务类型进行选择,通常为分类任务中常用的交叉熵损失函数。优化策略选择:通常可选随机梯度下降法、携带动量的梯度下降法和 Adam 优化等。超参数选取:模型训练过程中涉及初始学习率、学习率下降率、类别权重和样本权重等超参数的设置。通常以训练过程中内部验证集的模型性能作为评判条件。当建模过程中用于观测验证效果的数据集的准确率保持较为稳定的统计结果时,确立模型的超参数组合。

推荐 5:基于人工定义特征的前列腺癌分级预测模型应以高质量、可重复、一致性高的影像专家勾画为基础。

推荐 6:基于深度学习的前列腺癌分级预测模型应以大数据为支撑。

推荐 7:分级预测模型分类标签的确立应以临床问题为导向。显著性前列腺癌预测和高级别前

列腺癌预测模型推荐选择活检 Gleason 评分确立分类标签。精准分级预测模型推荐选择术后 Gleason 评分确立五分类标签。

推荐 8:精准预测模型及指标的构建与高质量的特征提取手段密切相关。

推荐 9:模型构建的过程应符合具体临床诊疗逻辑和临床需求。

分级模型评估

由于模型针对的临床问题具有多样性,因此需以临床问题为导向,且有针对性地进行规范化的模型性能评估,从而增加临床应用可信度。评估内容包括但不局限于:评估方式选择、评估指标选择、特征与建模可重复性评估和预测独立性评估。

一、评估方式选择

模型评估需结合回顾性数据同时具备:训练集评估、内部验证集评估、同中心外部测试集评估和跨中心外部测试集评估的四类评估结果。其中,训练集评估结果和内部验证集评估结果用于描述模型在建模过程中模型过拟合的风险程度。同中心外部测试用于评估在排除数据分布偏倚性和中心干扰条件下的模型性能。

此外,可补充前瞻性数据,构建前瞻性模型评估实验,进一步提高模型临床应用的可信度。

二、评估指标选择

1. 显著性前列腺癌/高级别前列腺癌预测模型评估指标:显著性前列腺癌预测和高级别前列腺癌预测均为二分类预测任务类型。因此,所产生的最终预测结果需结合二分类模型性能评估指标实现规范化的评估。主要评价指标包括:计算模型分类正确率(ACC)、受试工作特征曲线下面积(AUC)、阳性预测值(PPV)和阴性预测值(NPV)等。受试工作特征曲线(ROC)、决策曲线(DCA)和校准曲线用于评估预测模型临床应用可靠性。

2. 分级预测模型评估指标:前列腺癌分级预测为多分类预测任务类型,所产生的最终预测结果需结合多分类模型性能评估指标实现规范的模型性能评估,所述指标应包括但不限于:多分类全局准确率、单一类别内查全率(Recall)、单一类别内查全率(Precision)和单一类别内 F1 分数(F1-score)等。

3. 特征与建模可重复性评估:基于人工定义特征的预测模型,在特征筛选步骤易受特征筛选方法的影响,导致筛选后特征类型存在偏差问题。因

此,针对此类方法,可在同一数据集上使用不同的特征筛选方法进行特征筛选,并分别评估不同筛选方法下的特征筛选结果是否具备可重复性。特征的可重复性直接决定模型预测性能的稳定性。AI 模型在构建过程中易受数据分布影响,导致在模型训练过程中存在数据选择偏倚性风险。因此,在多中心模型性能评估条件下,可通过调换不同中心的数据分别进行建模与验证,评估模型在建模过程中是否会受训练数据选择的影响,实现预测模型的建模可重复性评估。

4. 预测独立性评估:为了将模型转化为方便临床使用的定制化及可视化的评估模型,可通过逻辑回归模型纳入临床信息和分级预测指标,采用优势比(OR)评估多个预测指标的重要性及显著性,提供预测模型在回顾性数据集上的性能评估结果,进而供临床医师参考。

推荐 10:以临床问题为导向,且有针对性地进行规范化的模型性能评估,从而增加临床应用可信度。

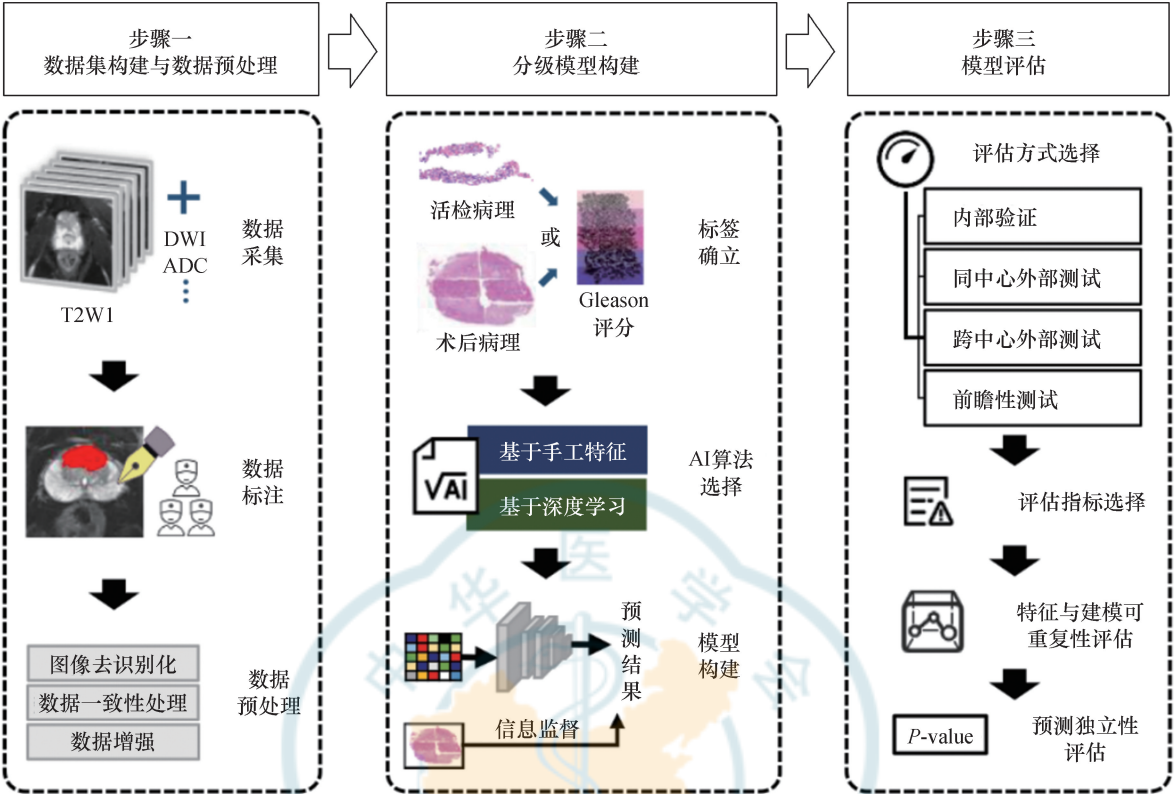
推荐 11:模型评估推荐具备:训练集评估、内部验证集评估、同中心外部测试集评估和跨中心外部测试集评估的四类评估结果。此外,可补充前瞻性数据,进一步提高模型临床应用的可信度。

总结

本共识结合目前前列腺癌临床及 AI 辅助前列腺癌诊疗决策的研究现状,建立了对 AI 辅助前列腺癌 MRI 分级预测模型构建的共识,包括:数据集构建与数据预处理、分级模型构建和分级模型评估等具体操作步骤的规范。希望本共识能够帮助广大泌尿外科和影像科医师快速明晰 AI 辅助诊断模型的构建过程,提高 AI 在前列腺癌分级预测中的应用质量和水平,加快相关研究的临床应用转化。

执笔:刘振宇(中国科学院自动化研究所)、刘建刚(北京航空航天大学)、邵立智(中国科学院自动化研究所)、赵立涛(北京航空航天大学)

专家组名单(按照姓氏拼音排序):陈俊星(中山大学第一医院泌尿外科)、何为(北京大学第三医院放射科)、黄毅(北京大学第三医院泌尿外科)、林健(北京大学第一医院泌尿外科)、刘明(北京医院泌尿外科)、刘树声(河北省廊坊市人民医院泌尿外科)、卢剑(北京大学第三医院泌尿外科)、陆敏(北京大学医学部病理学系)、刘维林(中国老年学和老年医学学会)、马林(中国人民解放军总医院第一医学中心)、马潞林(北京大学第三医院泌尿外科)、苗文隆(河北北方学院泌尿外科)、穆涛涛(中国人民解放军总医院第三医学中心放射科)、平浩(首都医科大学附属同仁医院泌尿外科)、田捷



注:T2WI为T2加权像;DWI为扩散加权成像;ADC为表面扩散系数;AI为人工智能。此流程图适用于基于前列腺癌磁共振影像的AI分级模型的构建

图1 AI辅助前列腺癌磁共振影像精准分级模型构建评价流程

(北京航空航天大学)、王磊(首都医科大学附属友谊医院泌尿外科)、王良(首都医科大学附属友谊医院放射科)、王建业(北京医院泌尿外科)、肖雨(北京协和医院病理科)、颜野(北京大学第三医院泌尿外科)、曾晓勇(华中科技大学同济医学院附属同济医院泌尿外科)、叶雄俊(中国医学科学院肿瘤医院泌尿外科)、赵勇(山东省立医院泌尿外科)、周晓峰(中日友好医院泌尿外科)、祖雄兵(中南大学湘雅医院泌尿外科)、周利群(北京大学第一医院泌尿外科)、张旭(中国人民解放军总医院泌尿外科)

利益冲突 所有作者均声明不存在利益冲突

作者贡献声明 刘振宇、刘建刚、林健、卢剑: 酝酿和设计、实施研究; 邵立智、赵立涛: 起草文章; 陈俊星、黄毅、刘明、刘树声、马林、苗文隆、穆学涛、平浩、王良、曾晓勇、叶雄俊、赵勇、周晓峰、祖雄兵: 实施研究、技术或材料支持; 何为、陆敏、王磊、肖雨、颜野: 分析/解释数据; 马路林、刘维林: 支持性贡献; 王建业、田捷、周利群、张旭: 对文章的知识性内容作批评性审阅

参考文献

- [1] Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA Cancer J Clin, 2021, 71(3): 209-249. DOI: 10.3322/caac.21660.
- [2] Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015 [J]. CA Cancer J Clin, 2016, 66(2): 115-132. DOI: 10.3322/caac.21338.
- [3] Siegel RL, Miller KD, Fuchs HE, et al. Cancer statistics, 2021 [J]. CA Cancer J Clin, 2021, 71(1): 7-33. DOI: 10.3322/caac.21654.
- [4] Cao W, Chen HD, Yu YW, et al. Changing profiles of cancer burden worldwide and in China: a secondary analysis of the global cancer statistics 2020 [J]. Chin Med J (Engl), 2021, 134(7): 783-791. DOI: 10.1097/CM9.0000000000001474.
- [5] Mohler JL, Antonarakis ES, Armstrong AJ, et al. Prostate cancer, version 2. 2019, NCCN Clinical Practice Guidelines in Oncology [J]. J Natl Compr Canc Netw, 2019, 17(5): 479-505. DOI: 10.6004/jncn.2019.002.
- [6] Ahmed HU, El-Shater Bosaily A, Brown LC, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study [J]. Lancet, 2017, 389(10071): 815-822. DOI: 10.1016/S0140-6736(16)32401-1.
- [7] Rosenkrantz AB, Ginocchio LA, Cornfeld D, et al. Interobserver reproducibility of the PI-RADS version 2 lexicon: a multicenter study of six experienced prostate radiologists [J]. Radiology, 2016, 280(3): 793-804. DOI: 10.1148/radiol.2016152542.
- [8] Hosny A, Parmar C, Quackenbush J, et al. Artificial intelligence in radiology [J]. Nat Rev Cancer, 2018, 18(8): 500-510. DOI: 10.1038/s41568-018-0016-5.
- [9] Shao L, Liu Z, Yan Y, et al. Patient-level prediction of multi-classification task at prostate MRI based on

- end-to-end framework learning from diagnostic logic of radiologists[J]. IEEE Trans Biomed Eng, 2021, 68 (12): 3690-3700. DOI: 10.1109/TBME.2021.3082176.
- [10] Shao L, Yan Y, Liu Z, et al. Radiologist-like artificial intelligence for grade group prediction of radical prostatectomy for reducing upgrading and downgrading from biopsy[J]. Theranostics, 2020, 10 (22): 10200-10212. DOI: 10.7150/thno.48706.
- [11] Schelb P, Kohl S, Radtke JP, et al. Classification of cancer at prostate MRI: deep learning versus clinical PI-RADS assessment[J]. Radiology, 2019, 293 (3): 607-617. DOI: 10.1148/radiol.2019190938.
- [12] Dinh AH, Melodelima C, Souchon R, et al. Characterization of prostate cancer with Gleason score of at least 7 by using quantitative multiparametric MR Imaging: validation of a computer-aided diagnosis system in patients referred for prostate biopsy[J]. Radiology, 2018, 287 (2): 525-533. DOI: 10.1148/radiol.2017171265.
- [13] Turkbey B, Rosenkrantz AB, Haider MA, et al. Prostate imaging reporting and Data System version 2.1: 2019 update of Prostate Imaging Reporting and Data System version 2[J]. Eur Urol, 2019, 76 (3): 340-351. DOI: 10.1016/j.eururo.2019.02.033.
- [14] 王良. 前列腺影像报告和数据分析系统 (PI-RADS V2.1) 解读[J]. 中华放射学杂志, 2020, 54 (4): 273-278. DOI: 10.3760/cma.j.cn112149-20190429-00382.
- [15] Kang Z, Min X, Weinreb J, et al. Abbreviated biparametric versus standard multiparametric MRI for diagnosis of prostate cancer: a systematic review and meta-analysis[J]. Am J Roentgenol, 2019, 212 (2): 357-365. DOI: 10.2214/AJR.18.20103.
- [16] Bonekamp D, Kohl S, Wiesenfarth M, et al. Radiomic machine learning for characterization of prostate lesions with MRI: comparison to ADC values[J]. Radiology, 2018, 289 (1): 128-137. DOI: 10.1148/radiol.2018173064.
- [17] Hectors SJ, Cherny M, Yadav KK, et al. Radiomics features measured with multiparametric magnetic resonance imaging predict prostate cancer aggressiveness[J]. J Urol, 2019, 202 (3): 498-505. DOI: 10.1097/JU.000000000000272.
- [18] Epstein JI, Feng Z, Trock BJ, et al. Upgrading and downgrading of prostate cancer from biopsy to radical prostatectomy: incidence and predictive factors using the modified Gleason grading system and factoring in tertiary grades[J]. Eur Urol, 2012, 61 (5): 1019-1024. DOI: 10.1016/j.eururo.2012.01.050.
- [19] Epstein JI, Egevad L, Amin MB, et al. The 2014 International Society of Urological Pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma: definition of grading patterns and proposal for a new grading system[J]. Am J Surg Pathol, 2016, 40: 244-252. DOI: 10.1097/PAS.0000000000000530.
- [20] Liu Z, Wang S, Dong D, et al. The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges [J]. Theranostics, 2019, 9 (5): 1303-1322. DOI: 10.7150/thno.30309.
- [21] Arefan D, Chai R, Sun M, et al. Machine learning prediction of axillary lymph node metastasis in breast cancer: 2D versus 3D radiomic features[J]. Med Phys, 2020, 47 (12): 6334-6342. DOI: 10.1002/mp.14538.
- [22] Aldoj N, Lukas S, Dewey M, et al. Semi-automatic classification of prostate cancer on multi-parametric MR imaging using a multi-channel 3D convolutional neural network[J]. Eur Radiol, 2020, 30 (2): 1243-1253. DOI: 10.1007/s00330-019-06417-z.

(收稿日期: 2021-10-28)

(本文编辑: 于普林)

· 征订启事 ·

欢迎订阅《中国神经免疫学和神经病学杂志》

《中国神经免疫学和神经病学杂志》是与神经免疫学相关的神经病学学术性期刊, 由国家卫健委主管, 北京医院、中国免疫学会神经免疫分会主办, 国内外公开发行人。本刊以从事神经免疫学、神经病学临床及科研人员等为主要读者对象, 报道神经疾病领域尤其是神经免疫学领域的科研成果。本刊属中国科技论文统计源期刊、中国科技核心期刊、中国科技引文数据库来源期刊及中国学术期刊综合评价数据库来源期刊等。

本刊主要栏目包括论著、述评、综述、进展、专论、讲座、临床病理讨论、疑难病例报告与讨论、论著摘要、研究简报、短篇报道等。既报道与神经免疫学相关疾病的临床和基础研究, 又报道一般神经病学方面的研究。不仅适合神经科临床医生、神经免疫学工作者、高等医药院校师生阅读, 而且亦适合基层相关专业临床医生阅读。

本刊为双月刊, 逢单月 15 日出版。每册定价 25.00 元, 全年 6 期共 150.00 元。国内外公开发行人, 国内邮发代号 82-718, 国外发行代号 BM1302。欢迎广大读者到当地邮局订购。亦可联系编辑部订购。

投稿方式: 请登录本刊官方网站: zsmc.cbpt.cnki.net 进行投稿

编辑部地址: 100010 北京东城区大佛寺东街 6 号北京医院《中国神经免疫学和神经病学杂志》编辑部

联系电话: 010-65242081, 010-64012981-8110

Email: zgsm@263.net, zgshenmian@163.com