



(12) 发明专利

(10) 授权公告号 CN 116563192 B

(45) 授权公告日 2025. 12. 16

(21) 申请号 202210096934.4
 (22) 申请日 2022.01.26
 (65) 同一申请的已公布的文献号
 申请公布号 CN 116563192 A
 (43) 申请公布日 2023.08.08
 (73) 专利权人 香港理工大学
 地址 中国香港九龙
 (72) 发明人 蔡璟 张将 滕信智 张远鹏
 林世杰 周塔 李兵
 (74) 专利代理机构 北京世峰知识产权代理有限公司 11713
 专利代理师 卓霖 许向彤

G06T 7/66 (2017.01)
 G06V 10/25 (2022.01)
 G06V 10/82 (2022.01)
 A61N 5/10 (2006.01)

(56) 对比文件

Sai-Kit Lam, Yuanpeng Zhang等.Multi-Organ Omics-Based Prediction for Adaptive Radiation Therapy Eligibility in Nasopharyngeal Carcinoma Patients Undergoing Concurrent Chemoradiotherapy.frontiers in oncology.2022,全文.

审查员 田秋硕

(51) Int. Cl.
 G06T 7/00 (2017.01)
 G06T 5/90 (2024.01)

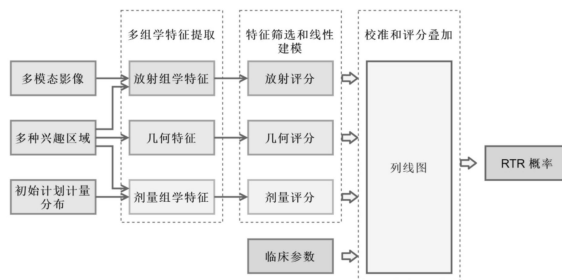
权利要求书4页 说明书10页 附图2页

(54) 发明名称

通过头颈癌患者数据进行多组学处理的方法、相关计算机可读介质及计算设备

(57) 摘要

本发明提供了一种通过头颈癌患者的放疗数据进行多组学处理的方法,其中将多组学应用于所述头颈癌患者的放疗重新计划需求的治疗前评估,包括以下步骤:(1)对头颈癌患者的放疗患者数据的预处理,所述放疗患者数据包括放疗数据和患者的临床特征数据;(2)从经过预处理的放疗患者数据中提取多组学特征;(3)基于预测模型通过拟合所提取的多组学特征来对所述头颈癌患者的RTR需求进行估算或预测。本发明允许针对RTR对有风险的患者进行“治疗前”识别,使得医生可以提前了解这些患者并更好地为他们分配资源。本发明不改变在放疗过程中针对肿瘤的剂量和/或其覆盖范围;患者在放疗过程中仍会经历其解剖结构(器官的体积和位置)的变化,因为这是器官在放疗照射时的自然反应。



1. 一种通过头颈癌患者的放疗患者数据进行多组学处理的方法,其中将多组学应用于所述头颈癌患者的放疗重新计划需求的治疗前评估,包括以下步骤:

(1) 对头颈癌患者的放疗患者数据的预处理以形成所述头颈癌患者的输入图像/剂量数据,确保在放射治疗过程中所述头颈癌患者的输入图像/剂量数据的特征与用于构建放疗重新计划的预测模型的图像/剂量数据的特征一致;

(2) 从经过预处理的放疗患者数据中提取多组学特征;

(3) 基于所述预测模型通过拟合所提取的多组学特征来对所述头颈癌患者的放疗重新计划需求进行估算或预测;

其中,所述多组学特征是通过计算提取的,所述多组学特征的数据包括放射组学特征数据、剂量组学特征数据、几何特征数据;

其中,所述几何特征数据以距离和角度来表示在患者的器官和肿瘤之间的几何关系或位置关系。

2. 根据权利要求1的通过头颈癌患者的放疗数据进行多组学处理的方法,其中所述放疗患者数据包括放疗数据和患者的临床特征数据,并且所述患者的输入图像/剂量数据的特征包括图像分辨率、信号强度可变性和均匀性;

其中,所述放疗数据采用医学数字成像和通信文件的数据格式,并且包括多模态放射图像数据、放疗剂量数据和器官轮廓数据;以及,

所述患者的临床特征数据包括患者人口统计学数据或肿瘤特性数据。

3. 根据权利要求2的通过头颈癌患者的放疗数据进行多组学处理的方法,其中所述多模态放射图像数据包括计算机断层扫描和多参数磁共振图像的数据。

4. 根据权利要求1的通过头颈癌患者的放疗数据进行多组学处理的方法,其中所述对头颈癌患者的放疗患者数据的预处理的步骤包括:将体素尺寸重采样应用于所述放疗患者数据的所有图像、掩码和剂量,以保持1mm x 1mm x 1mm的一致分辨率。

5. 根据权利要求4的通过头颈癌患者的放疗数据进行多组学处理的方法,其中所述对头颈癌患者的放疗患者数据的预处理的步骤包括:通过使用SimpleITK软件的N4偏差场校正图像滤波器功能进行偏差校正,以减少所述放疗患者数据中由不均匀磁场引起的核磁图像的体素强度偏差;以及,通过所述体素强度进行移动平均并重新缩放所述体素强度的标准偏差,来对所述核磁图像进行归一化。

6. 根据权利要求4的通过头颈癌患者的放疗数据进行多组学处理的方法,其中所述对头颈癌患者的放疗患者数据的预处理的步骤还包括:

在提取所述放射组学特征之前,将所有的对比度增强计算机断层扫描图像和核磁图像都重新采样到 $1 \times 1 \times 1\text{-mm}^3$ 的体素尺寸,以减轻在不同患者之间图像采集参数差异的影响;

将所述对比度增强计算机断层扫描图像和核磁图像的灰度级强度离散到固定的桶计数范围中,所述桶计数范围为从50到350,增量间隔为50;

对所有的对比度增强计算机断层扫描图像和核磁图像都在三个高斯半径参数级别下使用高斯拉普拉斯滤波器进行卷积以生成过滤图像,其中突出显示多个尺度的特定纹理放射学特征,所述多个尺度包括从1mm精细、3mm中等到6mm粗糙;

将小波滤波器也应用于所述对比度增强计算机断层扫描图像和核磁图像两者,以产生具有多种分辨率的放射学特征;以及

将对对比度增强计算机断层扫描图像上的原发性肿瘤和颈部淋巴结病变进行重新分割,以将亨氏单位限制在从-150到180的范围内,以消除所述头颈癌患者的感兴趣体积内的非软组织成分。

7. 根据权利要求1的通过头颈癌患者的放疗数据进行多组学处理的方法,其中所述预测模型是基于多组学特征数据开发的,所述多组学特征数据用于表征特定患者及其所接受的特定治疗的信息,所述多组学特征数据包含通过使用多模式患者的治疗前图像和临床参数而从多个器官结构提取的多个基于多组学的预测因子。

8. 根据权利要求7的通过头颈癌患者的放疗数据进行多组学处理的方法,其中所述预测模型是通过以下步骤开发的:将患者队列经由20次迭代以7比3的比例分为训练集和测试集数据集;将监督特征选择算法仅应用于每次迭代的训练数据集,以保持残余特征的临床相关性;然后,应用无监督特征选择算法以去除高度冗余的特征,从而得到减少的特征集;其中,所述预测模型是通过使用岭回归算法经由在所述训练集之中的10倍交叉验证而进行的。

9. 根据权利要求7的通过头颈癌患者的放疗数据进行多组学处理的方法,其中通过所述预测模型,对所述头颈癌患者的放疗重新计划需求的治疗前评估的最终预测概率包括:放射评分、剂量评分、几何评分和/或预测临床参数,其中所述放射评分是放射组学特征及其系数的线性组合,所述剂量评分是剂量组学特征及其系数的线性组合,所述几何评分是几何特征及其系数的线性组合。

10. 根据权利要求1的通过头颈癌患者的放疗数据进行多组学处理的方法,其中所述放射组学特征数据用于表征在治疗扰动时的内在组织生物学反应,并且所述放射组学特征数据是在患者的器官轮廓内从多模态图像中提取的。

11. 根据权利要求1的通过头颈癌患者的放疗数据进行多组学处理的方法,其中所述剂量组学特征数据通过提供来自常规剂量体积直方图和器官轮廓内放疗剂量分布的局部空间模式的剂量信息来表征个体放疗计划的积极性,所述剂量组学特征是从个体患者的放疗计划中提取的。

12. 根据权利要求1的通过头颈癌患者的放疗数据进行多组学处理的方法,其中所述放射组学特征数据、剂量组学特征数据和几何特征数据分别从三维医学影像、剂量分布图、和距离分布图中提取。

13. 根据权利要求12的通过头颈癌患者的放疗数据进行多组学处理的方法,其中所述放射组学特征是包括形状特征、一阶特征和高阶特征,

其中,所述形状特征描述勾画区域的大小以及形状,包括包含三个维度的长度、体积、圆度的16个特征;

所述一阶特征表示在勾画区域内的像素值本身的统计分布,包括包含像素平均值、最大值、最小值的19个特征;以及

所述高阶特征表示像素排布的异质性,与不同位置像素值的差有直接的关系,所述高阶特征分为5大类,包括基于GLCM、GLRLM、GLSZM、GLDM、NGTDM的5个描述矩阵的总共75个特征。

14. 根据权利要求13的通过头颈癌患者的放疗数据进行多组学处理的方法,其中对于医学影像,在原始图像和 $\text{Sigma} = 1\text{mm}, 3\text{mm}, 6\text{mm}$ 的高斯-拉普拉斯滤波图像以及包含Coil11

HHH, HHL, HLH, LHH, HLL, LHL, LLH, LLL的8种小波滤波图像提取一阶和高阶特征;并且,分别用50、100、150、200、250、300、350个等级对图像灰度进行离散化;以及

对于剂量分布和距离分布,只在非离散化的原始图像上计算所述放射组学特征。

15. 根据权利要求1的通过头颈癌患者的放疗数据进行多组学处理的方法,其中所述剂量组学特征包括用于描述加权重心的缩放不变矩、三维梯度、直方图点位以对体积分布进行降维;

所述加权重心的缩放不变矩和三维梯度的数学定义分别是:

(1) 缩放不变矩:三维体积分布在 x, y, z 三个方向上加权平均值,用于描述三维体积分布在特定区域的包括重心和方向的关键信息;

所述缩放不变矩通过平移不变矩的归一化得到:

$$\eta_{pqr} = \frac{\mu_{pqr}}{\frac{p+q+r}{3} + 1} \mu_{000}$$

其中 p, q, r 分别为用户指定的在 x, y, z 方向上的指数,

平移不变矩 μ_{pqr} 的数学表达式为:

$$\mu_{pqr} = \sum_x \sum_y \sum_z (x - \bar{x})^p (y - \bar{y})^q (z - \bar{z})^r D(x, y, z) I(x, y, z);$$

(2) 三维梯度:该类特征计算了三维体积分布在 x, y, z 三个方向上的梯度,在 x 方向的剂量梯度计算公式为:

$$\text{梯度}_x = \frac{\sum_{x,y,z} D(x+1,y,z)I(x+1,y,z) - D(x-1,y,z)I(x-1,y,z)}{2 \sum_{x,y,z} I(x,y,z)}$$

其中 $D(x, y, z)$ 代表在体素位置 (x, y, z) 的剂量值, $I(x, y, z)$ 描述了该体素是否位于兴趣区域内。

16. 根据权利要求1的通过头颈癌患者的放疗数据进行多组学处理的方法,其中所述几何特征数据的距离分布图描述任意位置到靶区表面的最近距离,靶区外距离为正,靶区内距离为负;所述距离分布图的计算过程通过SimpleITK软件的SignedMaurerDistanceMapFilter实现。

17. 根据权利要求16的通过头颈癌患者的放疗数据进行多组学处理的方法,其中所述几何特征数据还包括角度特征,所述角度特征用于描述肿瘤器官到靶区的角度分布,其中,所述角度特征包括角度直方图的点位,并且描述在所有角度下靶区的平行投影与肿瘤器官的重叠体积。

18. 根据权利要求17的通过头颈癌患者的放疗数据进行多组学处理的方法,其中所述角度直方图由修改后的拉东变换产生的正弦图积分得到:肿瘤器官在不同角度位于靶区背部前方部分的密度投影的积分,修改过的拉东变换反映了在靶区和肿瘤器官之间的前后方位。

19. 根据权利要求1的通过头颈癌患者的放疗数据进行多组学处理的方法,其中所述预测模型是通过使用所述头颈癌患者的多组学特征和临床参数进行预训练和开发的,其中通过整合所述头颈癌患者的患者特异性放射组学特征、剂量组学特征和几何特征以及临床数据,对所述头颈癌患者的放疗重新计划的需求进行综合预测或治疗前评估。

20. 根据权利要求19的通过头颈癌患者的放疗数据进行多组学处理的方法,其中所述

预测模型通过整合来自所述头颈癌患者的个体放射影像、放疗治疗计划、在目标器官之间的几何关系、以及相关临床属性的患者特异性预测因子来进行基于多组学的预测,以对所述头颈癌患者的放疗重新计划的需求进行个性化的预测或治疗前评估。

21. 根据权利要求19的通过头颈癌患者的放疗数据进行多组学处理的方法,其中所述预测模型的输入包括来自于放疗计划数据的多模态影像、多种兴趣区域和初始计划剂量分布,所述预测模型的输入还包括临床预测参数,所述预测模型的输出是RTR预测概率。

22. 根据权利要求21的通过头颈癌患者的放疗数据进行多组学处理的方法,其中所述预测模型的中间步骤包括多组学特征提取、特征筛选和线性建模、以及校准和评分叠加,其中,多组学特征提取包括预处理和特征计算,放射组学特征由多组学影像和多种兴趣区域计算得到,剂量组学特征由初始计划剂量分布和多种兴趣区域计算得到,几何特征仅由多种兴趣区域计算得到,所述预测模型对于放射组学特征、剂量组学特征和几何特征进行特征筛选和线性建模,以分别得到放射评分、剂量评分和几何评分,并且对所述放射评分、剂量评分和几何评分以及临床参数进行校准并叠加得到最终的放疗重新计划概率,其中,所述校准和叠加的过程通过列线图实现,所述列线图中的每条标尺代表对所述放射评分、剂量评分和几何评分中的每个评分和临床参数的校准模型,校准之后的评分相加并再次校准从而得到所述最终的放疗重新计划概率。

23. 根据权利要求22的通过头颈癌患者的放疗数据进行多组学处理的方法,其中所述特征筛选和线性建模包括对于放射组学模型、剂量组学模型和几何模型各生成一个方程,所述方程分别代表所述模型中所有选定预测变量的线性组合:放射组学模型的放射评分、剂量组学模型的剂量评分和几何模型模型的几何评分,分别包括:

$$\text{放射评分} = \text{截距}r + R_1(r_1) + R_2(r_2) + R_3(r_3) + \dots + R_i(r_i)$$

$$\text{剂量评分} = \text{截距}d + D_1(d_1) + D_2(d_2) + D_3(d_3) + \dots + D_j(d_j)$$

$$\text{几何评分} = \text{截距}g + G_1(g_1) + G_2(g_2) + G_3(g_3) + \dots + G_k(g_k)$$

其中,截距r、截距d、截距g是从所述预测模型针对放射组学、剂量组学和几何模型通过使用Ridge算法而分别推导出来的,R,D,G和r,d,g分别是放射组学、剂量组学和几何模型中的预测变量及其系数,i,j,k分别表示放射学、剂量学和几何特征的最终数量。

24. 一种计算机可读存储介质,所述计算机可读存储介质上存储有程序,其特征在于,所述程序被执行时实现如权利要求1至23中任一项所述的方法的步骤。

25. 一种计算设备,其包括处理器和可读存储介质,所述可读存储介质上存储有程序,其特征在于,所述程序被所述处理器执行时实现如权利要求1至23中任一项所述的方法的步骤。

通过头颈癌患者数据进行多组学处理的方法、相关计算机可读介质及计算设备

技术领域

[0001] 本发明涉及一种通过头颈癌患者的放疗患者数据进行多组学处理的方法。

背景技术

[0002] 头颈癌(Head-and-Neck Cancer, HNC)是存在于所述的上呼吸消化道之中的常见肿瘤,是世界范围内的第五大最为常见的恶性肿瘤。调强放疗(intensity modulated radiotherapy, IMRT)是目前在香港治疗HNC患者的标准治疗方法。在对癌症患者进行放射治疗之前,放疗(RT)治疗计划是根据各个患者或肿瘤器官的解剖结构而预先量身定制的。但是,在整个放疗过程中,接受放疗的患者器官的体积和位置会发生显著变化。因此,这种变化可能会削弱对肿瘤的辐射剂量覆盖范围和/或导致周围健康正常组织的剂量过大。

[0003] 目前普遍采用放疗重新计划(RadioTherapy Re-planning, RTR)来修改原放疗计划,以补偿这些变化,以保持满意的治疗率。例如,在调强放疗过程中发生的解剖学改变可以通过利用重复计算机断层扫描(CT)的适应性放射治疗(ART)进行补偿,在治疗过程中进行扫描以进行重新计划。放疗重新计划对HNC患者的临床和剂量学益处已被广泛报道,并且在当今的临床环境中,放疗重新计划在全球的使用已变得司空见惯。然而,放疗重新计划治疗程序的当前实施效率非常低,因为它通常以即时方式完成。此外,整个放疗重新计划的监测和风险评估程序是资源密集型和耗时的,因为它需要来自图像重新扫描(放射治疗师)、器官重新分割(医学物理学家和肿瘤学家)、治疗重新计划(医学物理学家)到新计划批准(肿瘤学家)等的多个医学科室的努力,以进行基于患者的需求评估、图像重新扫描、器官重新分割以及治疗重新计划等。特别是,每个患者的放疗重新计划需求筛查目前只能在治疗开始后进行,这需要对每个HNC患者进行密集的中期监测,而全球肿瘤科室还没有通用的筛查方法。考虑到上述因素,当前的放疗重新计划实践效率非常低,并且给肿瘤中心带来了巨大的临床负担。

[0004] 例如,Ting-ting Yu等人在2019年10月16日发表的论文“使用基于MRI的放射组学对晚期鼻咽癌患者进行适应性放射治疗资格的预处理预测”(“Pretreatment Prediction of Adaptive Radiation Therapy Eligibility Using MRI-Based Radiomics for Advanced Nasopharyngeal Carcinoma Patients”),《肿瘤学前沿(Frontiers in oncology)》,9,1050. 参见<https://doi.org/10.3389/fonc.2019.01050>,成功证明了基于磁共振成像(MRI)的放射组学(radiomics)在对于鼻咽癌(NPC)患者的适应性放射治疗(ART)资格的预处理之前的识别方面的有前景的能力。特别是,与其他研究模型相比,所述论文中具有6个选定放射组学特征的联合T1-T2模型看起来是更可取的预测系统。这将使放射肿瘤学家能够更有效、更准确地根据个体患者开出ART处方,以实现真正的鼻咽癌患者个性化放射治疗,同时精简临床环境中的资源管理。在未来的工作中,需要使用更大的患者样本进行多机构前瞻性研究,以提高该论文中的模型的临床疗效。但是,所述论文仅涉及通过使用头颈癌患者的仅MR图像而只来自原发肿瘤的放射组学特征。该论文中没有生成列线

图,因此没有对各个患者进行风险概率评估。

[0005] 因此,发明人提出了将放射组学应用于对头颈癌(Head-and-Neck Cancer,HNC)患者的放疗过程,以对HNC患者的放疗重新计划需求的治疗前评估。

发明内容

[0006] 本发明涉及一种通过头颈癌患者的放疗数据进行多组学处理的方法,对头颈癌患者的放疗重新计划(RTR)需求提供治疗前评估,其中将多组学应用于所述HNC患者的RTR需求的治疗前评估。

[0007] 本发明的技术主要包括三个方面:1)放疗患者数据的预处理,2)多组学特征的提取,3)RTR需求的估算。

[0008] 因此,本发明的通过头颈癌患者的放疗数据进行多组学处理的方法,其中将多组学应用于所述头颈癌患者的放疗重新计划需求的治疗前评估,包括以下步骤:(1)对头颈癌患者的放疗患者数据的预处理,以确保在放射治疗过程中所述头颈癌患者的输入图像/剂量数据的特征与用于构建放疗重新计划的预测模型的图像/剂量数据的特征一致;(2)从经过预处理的放疗患者数据中提取多组学特征;和(3)基于所述预测模型通过拟合所提取的多组学特征来对所述头颈癌患者的放疗重新计划需求进行估算或预测。

[0009] 本发明旨在允许对HNC患者的RTR需求进行“治疗前评估”,即,如果该特定患者根据最初的治疗计划进行RT治疗,则该患者是否将有更多的RTR需求机会。

[0010] 一旦确定了有风险的患者,则由医生决定调整(以及如何调整)最初的治疗计划、或者提前更好地分配医院资源(例如,多学科人力、时间、机器可用性等)、或采取其他补救措施。本发明不改变针对肿瘤的剂量和/或其覆盖范围。值得注意的是,患者在放疗过程中仍会经历其解剖结构(器官的体积和位置)的变化,因为这是器官在放疗照射时的自然反应。本发明并没有阻止这种情况的发生。

[0011] 本发明允许针对RTR对有风险的患者进行“治疗前”识别,使得医生可以提前了解这些患者并更好地为他们分配资源。否则的话,传统的在整个RT过程中对每一位患者进行定期密切监测无疑会给临床带来巨大的负担,而且效率极低。

附图说明

[0012] 图1是根据本发明的预测模型的示范性方框图。

[0013] 图2是根据本发明的预测模型对HNC患者的RTR需求的预测列线图。

[0014] 图3(a)和图3(b)是根据本发明的预测模型对HNC患者的高低剂量缩放不变矩的比较。

[0015] 图4是根据本发明的预测模型对HNC患者的从任意位置到放疗治疗的靶区表面的最近距离的距离分布图和角度直方图。

具体实施方式

[0016] 下面参照附图,进一步详细描述本发明的技术方案和具体实施例。

[0017] 多组学(multi-omics),也称为综合组学(integrative omics)、“panomics”或“pan-omics”,是一种生物学分析方法,其中数据集是多个“组”(omes),例如基因组、蛋白质

组、转录组、表观基因组、代谢组和微生物组,即,元基因组和/或元转录组,取决于它的测序方式);换句话说,使用多种组学技术以协调一致的方式研究生命。通过结合这些“组”,研究人员可以分析复杂的生物大数据,以发现生物实体之间的新关联,确定相关生物标志物并构建疾病和生理学的精细标志物。在这样做的过程中,多组学整合了不同的组学数据,以找到一致匹配的基因-表型-环境型关系或关联。OmicTools (组工具) 服务列出了超过99个与多组学特征数据分析相关的软件,以及超过99个关于该主题的数据库。

[0018] 如上所述,本发明的技术主要包括三个方面:1) 放疗患者数据的预处理,2) 多组学特征的提取,3) RTR需求的估算。

[0019] 具体地说,首先,本发明的技术包括对放疗患者数据的预处理,其中RTR需求预测涉及的放疗患者数据包括RT数据和患者的临床特征数据。所述RT数据采用医学数字成像和通信 (DICOM) 文件的格式,包括多模态放射图像(即计算机断层扫描 (CT) 和多参数磁共振 (MRI) 图像) 数据、RT剂量数据和器官轮廓数据。患者的临床特征数据包括患者人口统计学、肿瘤特性等数据。

[0020] DICOM表示“医学中的数字成像和通信”,是医学成像中处理、存储、打印和传输信息的国际标准——用于管理和传输医学图像和相关数据。它包括文件格式定义和网络通信协议。DICOM是由美国国家电气制造商协会 (NEMA) 和美国放射学院 (ACR) 联合开发的,以允许成像设备之间以及与其他设备之间的互操作性。DICOM标准负责管理图像格式以及传输在许多与医疗保健相关的成像“方式”(例如磁共振,核医学,计算机断层扫描和超声) 期间生成的医学图像信息所需的各种网络协议。DICOM既是一种通信协议,也是一种文件格式,这意味着它可以将超声和MRI图像等医疗信息以及患者的信息存储在一个文件中。该格式确保所有数据保持在一起,并提供了在支持DICOM格式的设备之间传输信息的能力。例如,扫描一次CT,可以获得几十到上百张的CT图像,一张CT图像对应一个.dcm文件,文件名为序列号。DICOM标准中存储的内容除了患者的CT值,还包含了CT机器的信息、层厚、时间戳、患者基本信息等等,这些文件都可以在CT文件中查到。DICOM在全球范围内被用于存储、交换和传输医学图像,使来自不同制造商的医学成像设备集成成为可能,将放疗患者数据和相关图像以标准格式交换和存储,便于患者获得更准确的治疗。

[0021] 对放疗患者数据执行预处理,以确保在放射治疗过程中患者的输入图像/剂量数据的特征(例如,图像分辨率、信号强度可变性和均匀性) 与用于构建本发明的RTR预测模型的图像/剂量数据的特征一致。具体而言,将体素 (voxel) 尺寸重采样应用于所有图像、掩码和剂量,以保持1mm x 1mm x 1mm的一致分辨率。使用SimpleITK(一种专门处理医学影像的软件)的“N4Bias Field Correction Image Filter”(N4偏差场校正图像滤波器) 功能进行偏差校正,以减少由不均匀磁场引起的MR图像的体素强度偏差。通过移动平均体素强度并重新缩放其标准偏差,来对MR图像进行归一化。体素是体积元素 (Volume Pixel) 的简称,包含体素的立体可以通过立体渲染或者提取给定阈值轮廓的多边形等值面表现出来。“体素”是数字数据在三维空间分割上的最小单位,用于三维成像、科学数据与医学影像等领域。概念上类似于用在二维计算机图像的影像数据上的二维空间的最小单位“像素”。

[0022] 除了体素尺寸重采样之外,所述预处理步骤还包括:在提取放射组学特征之前,所有CECT(对比度增强计算机断层扫描) 和MR(核磁) 图像都被重新采样到 $1 \times 1 \times 1\text{-mm}^3$ 的体素尺寸,以减轻在不同患者之间图像采集参数差异的影响。其中,灰度级的量化被应用于归一化

两种图像类型中的图像信号强度。图像的灰度级强度被离散到一个固定的桶(bin)计数范围中,所述范围从50到350,增量间隔为50。此外,所有图像都在三个高斯半径参数级别下使用高斯拉普拉斯(LoG)滤波器进行卷积以生成过滤图像,其中突出显示从精细(1mm)、中等(3mm)到粗糙(6mm)的多个尺度的特定纹理放射学特征。小波滤波器(HHH、HLL、LHL、LLH、LHH、HLH、HHL、LLL,H表示高,L表示低)也应用于CECT和MR图像两者,以产生具有多种分辨率的放射学特征。对CECT图像上的原发肿瘤和颈部淋巴结病变进行重新分割,以将亨氏单位(HU,Hounsfield Unit)限制在[-150,180]的范围内,以消除所研究的VOI(感兴趣体积)内的非软组织成分,例如气腔和骨骼结构。特别是对于MR图像,图像像素值的不均匀性校正正是通过使用SimpleITK(v1.2.4)中的“N4偏置场校正图像滤波器”中提供的N4B偏置校正来实现的。

[0023] 本发明的技术是通过RTR预测模型来执行的。本发明的RTR预测模型是基于多组学特征数据开发的,所述多组学特征数据用于表征患者特定和治疗特定的信息。它包含通过使用多模式患者的治疗前图像(CECT、MR图像)和临床参数而从多个器官结构(包括同侧和对侧腮腺、原发肿瘤、颈部淋巴结病变、脊髓、脑干、淋巴结病变的计划性-标靶-体积,等等)提取的多个基于多组学的预测因子(例如,放射组学数据、剂量组学数据、几何数据)。

[0024] 在模型开发过程中,患者队列经由20次迭代以7比3的比例分为训练集和测试集数据集。监督特征选择算法仅应用于每次迭代的训练数据集,以保持残余特征的临床相关性。随后应用无监督特征选择算法以去除高度冗余的特征,从而得到减少的特征集。RTR预测模型的开发是通过使用岭回归(Ridge regression)算法经由在所述训练集之中的10倍交叉验证而进行的,以降低模型过度拟合的风险。

[0025] 图1中示出了根据本发明的预测模型的示范性方框图。在图1中,预测模型的输入是指定的多模态影像(Multi-modality images)、多种兴趣区域(Multiple ROIs)、初始计划剂量分布(Initial planning dose)、和临床预测参数(Clinical parameters)。最终输出则为RTR预测概率。前三类输入均来自于放疗计划数据,在输入预测模型之前需要经过数据清洗,包括数据筛选和格式转换,以满足下列条件:

[0026] 1.选择的影像数据模态是否正确,是否为增强图像,拍摄时间是否为放疗前,兴趣区域的勾画和计划剂量分布是否属于初始计划。

[0027] 2.所有的图像和结构均在在同一个坐标系下,从而确保每个兴趣区域可以对应到影像和剂量分布的正确位置。

[0028] 3.将所有影像、兴趣区域、剂量分布数据从DICOM转换为Simple ITK图像数据,其中转换之后的兴趣区域为三维二元掩膜图像(3D binary mask image)。

[0029] 本发明的预测模型的建立包含多个中间步骤,包括多组学特征提取(Multi-omics feature extraction)、特征筛选和线性建模(Feature selection and linear modeling)、校准和评分叠加(Calibration and score combination)。

[0030] 多组学特征提取包括预处理和特征计算。放射组学特征由多组学影像和多种兴趣区域计算得到,剂量组学特征由初始计划剂量分布和多种兴趣区域计算得到,几何特征仅由多种兴趣区域计算得到。模型训练过程中用到了完整的多组学特征,但是对于模型应用只需要计算最终筛选得到的多组学特征。本发明人对于每一类特征进行特征筛选和线性建模,分别得到放射评分、剂量评分、和几何评分。这三个线性模型可以由特征系数(feature

coefficients) 和截距 (intercept) 来描述, 用于模型部署 (deployment)。最终本发明人对这三个评分和临床参数进行校准并叠加得到最终的RTR概率。校准和叠加的过程通过列线图 (nomogram) 实现。列线图的每条标尺代表了每个评分和临床参数的校准模型, 校准之后的评分相加并再次校准则得到RTR概率。

[0031] 本发明的预测模型在开发过程中的关键步骤包括患者队列划分和特征选择。首先, 在模型开发过程中, 患者队列通过20次迭代以7比3的比例分为训练集数据集和测试集数据集。然后, 将监督特征选择算法仅应用于每次迭代的训练集数据集, 以保持剩余特征的临床相关性。在这个特征选择过程中, 将统计单变量分析用于评估每个提取的多组学特征的临床关联和感兴趣的结果, 即, 训练集数据集之中的患者是否具有由临床肿瘤学家确定的RTR风险。随后应用无监督特征选择算法 (例如, Pearson的相关性测试算法) 来去除高度冗余的特征, 从而减少特征集。因此, 训练集数据集的主要组成包括多组学特征 (从多模态医学图像中提取的放射特征、来自多个ROI的几何特征和来自初始剂量的剂量多组学特征) 和患者的临床参数, 均以数值表示。

[0032] 因此, 通过本发明的RTR预测模型, RTR需求的最终预测概率是: Radiomic (放射组学) 特征及其系数的线性组合 (Rad-Score, 放射评分), 和/或Dosiomic (剂量组学) 特征及其系数的线性组合 (Dose-Score, 剂量评分), 和/或几何特征及其系数 (Geo-Score, 几何评分) 的线性组合, 和/或预测临床参数, 如图2所示。

[0033] 在一个实施例中, 本发明的RTR预测模型的开发是使用Ridge算法 (一种使用线性函数解决偏差-方差权衡的典型统计方法) 进行的, 其中经由在训练集之中的10倍交叉验证来使用剩余特征以减轻所述预测模型的过拟合的风险。在这个过程中, 首先通过使用训练集数据集分别开发放射组学模型、剂量组学模型和几何模型, 这个过程可以称为建模。

[0034] 所述建模过程的关键是, 为三个模型中的每一个模型生成一个方程, 所述方程代表所述模型中所有选定预测变量的线性组合, 分别称为: 放射组学 (Radiomic) 模型的放射评分 (Rad-score)、剂量组学 (Dosiomics) 模型的剂量评分 (Dose-score) 和几何模型 (Geometric) 模型的几何评分 (Geo-score), 所述方程如下所示:

[0035] 放射评分 = 截距 $r + R_1(r_1) + R_2(r_2) + R_3(r_3) + \dots + R_i(r_i)$

[0036] 剂量评分 = 截距 $d + D_1(d_1) + D_2(d_2) + D_3(d_3) + \dots + D_j(d_j)$

[0037] 几何评分 = 截距 $g + G_1(g_1) + G_2(g_2) + G_3(g_3) + \dots + G_k(g_k)$

[0038] 其中, 截距是Intercept的译文, 截距 r 、截距 d 、截距 g 是从本发明开发的RTR预测模型针对放射组学、剂量组学和几何模型通过使用Ridge算法而分别推导出来的。 (R, D, G) 和 (r, d, g) 分别是放射组学、剂量组学和几何模型中的预测变量及其系数 (权重)。 (i, j, k) 分别表示放射学、剂量学和几何特征的数量。

[0039] 所述各个截距的示例显示在图2的列线图中。

[0040] 作为结果, 计算了训练集数据集之中每个患者的放射评分、剂量评分和几何评分 (即, 作为三个变量)。这三个变量与临床预测变量一起用作输入, 通过使用在R Studio软件 (即R语言集成开发环境) 中的rms包中的逻辑回归建模 (lrm) 函数, 来开发用于RTR需求的预测列线图, 如图2所示。其中, 所述临床预测变量显示了统计显著的关联, 例如 $p < 0.05$ 。在列线图构建期间, 基于训练集数据集校准了每个预测变量 (剂量评分、几何评分、放射评分、性别和肿瘤阶段) 的权重, 所述权重有助于最终RTR需求。在图2的示范性列线图中, 各个变

量的每一行的长度是不同的,这代表了它们各自对于RTR需求的权重。为了从实际的角度进一步解释这个列线图,例如,如果一个具有第IV期肿瘤(参见列线图中的~2点位)的新的男性患者(参考列线图中的~2点位)来到诊所,则对于该特定患者,剂量评分为0.4分(参考列线图中~3点位)、几何评分为0.62分(参考列线图中~11点位)、并且放射评分为0.8(参考列线图中~70点位),总的评分将等于 $2+2+3+11+70=88$ 点位,这将得出该特定患者的估计RTR需求为0.9,如图2的最后一行RTR需求中的虚线所示。

[0041] 预测的RTR风险概率值越高,特定患者的RTR需求就越高。列线图模型使用所有多组学特征和临床参数进行预训练和开发。一旦获得了预期患者模型指定的所有特征,就可以使用本发明的列线图模型预测该患者的RTR。本发明的工作(以及预测列线图)仍在继续,并将使用更大的数据队列进一步优化模型开发。

[0042] 在第二方面,本发明的方案包括多组学特征的提取:从预处理的DICOM数据中计算出三种类型的多组学特征。所述三种类型的多组学特征包括放射组学特征、剂量组学特征和几何特征,它们分别如下所述:

[0043] -放射组学特征:放射组学能够表征在治疗扰动时的内在组织生物学反应,本发明人不断努力将放射组学应用于基于肿瘤缩小来预测患者特异性治疗反应,为应用放射组学来预测HNC患者的患者特异性RTR需求奠定了良好的基础,因为RTR的实施可以由多个器官(例如原发性肿瘤、转移性颈部淋巴结和腮腺)的缩小来触发。在本发明中,放射组学特征是在器官轮廓内从多模态图像中提取的,放射组学特征的计算是通过使用公开可用的Python包-PyRadiomics来执行的,遵循图像生物标记标准化倡议(IBSI)指南。Python是一种计算机程序设计语言,PyRadiomics是一种开源的python软件包。

[0044] -剂量组学特征:剂量组学通过提供来自常规剂量体积直方图和器官轮廓内放疗剂量分布的局部空间模式的剂量信息来表征个体放疗计划的积极性。剂量组学特征是从个体患者的放疗计划中提取的。

[0045] 此外,类似放射组学的剂量组学信息也已被广泛研究用于各种癌症的预后和毒性预测,显示出了在预测在治疗扰动时的组织反应方面的巨大潜力。在本发明中,类似放射组学的剂量组学信息也被计算。剂量组学特征的计算由发明人自己开发的算法进行。

[0046] -几何特征:在患者的器官/肿瘤之间的几何关系可能会影响治疗时的解剖学变化程度,从而导致对RTR的需求。在本发明中,患者的几何器官以距离和角度来表示。几何特征的计算由发明人自己开发的算法进行。

[0047] 本发明的算法首先包括多组学特征设计。本发明人设计了多种创新的放疗数据多组学特征来全面有效的描述放疗计划信息,从而能够在建模过程中挖掘出更多和更加有针对性的预测因子。具体来讲,本发明所采用的多组学特征包括:在放疗前的多种影像模态和区域的影像组学特征(Radiomic features,即放射组学特征),在靶区和目标器官轮廓之间的几何特征(Geometric features),和基于三维计划放射剂量分布的剂量组学特征(Dosiomic features)。其中几何特征包括了距离特征(Distance features)和角度特征(Angle features)。距离和角度的结合才能够最准确的描述靶区和目标器官之间的位置关系。

[0048] 放射组学特征、剂量组学特征和几何特征分别从三维医学影像、剂量分布图、和距离分布图中提取。由于这三种分布图均为体积分布,因此这三类特征共享同一套特征定义,

包括了传统的影像组学特征定义和其他的对于体积分布的空间描述 (Spatial description)。传统的影像组学特征的提取参照“图像生物标志物标准化倡议” (Image biomarker standardisation initiative, IBSI) 的标准, 所述的“图像生物标志物标准化倡议”由[1]Zwanenburg A, Leger S, Vallières M, Löck S. 在“放射学” (Radiology) 杂志上在线公开于2020年3月10日: 191145. doi:10.1148/radiol.2020191145。所述三类特征可以由开源Python软件PyRadiomics计算得到, 主要分为形状特征、一阶特征、和高阶特征 (纹理特征)。一阶和高阶特征可以由多种滤波 (Filter) 之后的图像中计算, 如下所述。

[0049] a. 形状特征: 该类特征主要描述勾画区域的大小以及形状, 与图像本身无关。

[0050] 包括三个维度的长度, 体积, 圆度等16个特征。

[0051] b. 一阶特征: 该类特征概括了在勾画区域内的像素值本身的统计分布, 与像素点所在位置无关, 不涉及不同像素点之间的差。包括像素平均值, 最大值, 最小值等19个特征。

[0052] c. 高阶特征: 该类特征描述了像素排布的异质性, 与不同位置像素值的差有直接的关系。高阶特征又可以分为5大类, 包括基于GLCM, GLRLM, GLSZM, GLDM, NGTDM这5个描述矩阵的总共75个特征。

[0053] 对于医学影像, 本发明人在原始图像和 $\sigma = 1\text{mm}, 3\text{mm}, 6\text{mm}$ 的高斯拉普拉斯 (Laplacian-of-Gaussian) 滤波图像以及8种小波滤波图像 (Coil1 HHH, HHL, HLH, LHH, HLL, LHL, LLH, LLL) 提取一阶和高阶特征。同时, 本发明人分别用50, 100, 150, 200, 250, 300, 350个等级对图像灰度进行离散化。对于剂量分布和距离分布, 我们只在非离散化的原始图像上计算传统影像组学特征。

[0054] 对于剂量组学特征, 基于以上传统影像组学特征定义, 本发明人借鉴了历史研究中对于三维剂量空间分布的描述, 包括描述加权重心的缩放不变矩 (Scale-invariant moments)、三维梯度 (3D gradients)、直方图点位 (Histogram points) 以对体积分布进行降维。

[0055] 所述加权重心的缩放不变矩的概念可以参见 Buettner F, Miah AB, Gulliford SL, 等人的“改进头颈部放疗对的治疗指数的新方法: 来自PARSPORT随机III期试验的数据分析” (Novel approaches to improve the therapeutic index of head and neck radiotherapy: An analysis of data from the PARSPORT randomised phase III trial), *Radiother Oncol.* 2012; 103 (1): 82-87. doi:10.1016/j.radonc.2012.02.006, 所述三维梯度的概念可以参见 Gabryś HS, Buettner F, Sterzing F, Hauswald H, Bangert M. 的“用于口腔干燥症的正常组织并发症概率建模的使用放射组学和剂量组学的机器学习方法的设计和选择” (Design and Selection of Machine Learning Methods Using Radiomics and Dosiomics for Normal Tissue Complication Probability Modeling of Xerostomia), *Front Oncol.* 2018; 8. doi:10.3389/fonc.2018.00035)。

[0056] 所述加权重心的缩放不变矩和三维梯度的数学定义为:

[0057] a. 缩放不变矩: 三维体积分布在 x, y, z 三个方向上加权平均值, 描述的是三维体积分布在特定区域的重心, 方向等关键信息。其中, 平移不变矩的数学表达式为:

$$[0058] \quad \mu_{pqr} = \sum_x \sum_y \sum_z (x - \bar{x})^p (y - \bar{y})^q (z - \bar{z})^r D(x, y, z) I(x, y, z)$$

[0059] 缩放不变矩可以通过平移不变矩的归一化得到:

$$[0060] \quad \eta_{pqr} = \frac{\mu_{pqr}}{\frac{\mu_{000}^3}{p+q+r+1}}$$

[0061] 高低剂量的缩放不变矩的比较由图3(a)和图3(b)所示。本发明人从x、y、z三个方向计算得到三个三维梯度值,采用0,1,2,3阶进行缩放不变矩的计算。由于全0阶的结果恒为1,故不纳入多组学特征组,最终得到63个缩放不变矩特征。

[0062] 图3(a)和图3(b)是高低剂量缩放不变矩的比较。图3(a)每一行的三张图片为某一病人的剂量分布图在x-y平面的不同切面。可以看出高 η_{110} 的病人的剂量分布中心偏向右上角(第一行)。图3(b)每一行的三张图片为某一病人的剂量分布图在x-z平面的不同切面。高 η_{003} 的病人(第一行)的剂量分布的重心明显偏向正上方。

[0063] b. 三维梯度:该类特征计算了三维体积分布在x、y、z三个方向上的梯度,在x维度的剂量梯度计算公式为:

$$[0064] \quad \text{梯度}x = \frac{\sum_{x,y,z} D(x+1,y,z)I(x+1,y,z) - D(x-1,y,z)I(x-1,y,z)}{2 \sum_{x,y,z} I(x,y,z)}$$

[0065] 其中D(x,y,z)代表在体素位置(x,y,z)的剂量值,I(x,y,z)描述了该体素是否位于感兴趣区域内。

[0066] 本发明的预测模型中的几何特征包括距离分布图和角度特征。所述距离分布图描述了任意位置到靶区表面的最近距离,靶区外距离为正,靶区内距离为负。距离分布图的计算过程通过SimpleITK的SignedMaurerDistanceMapFilter实现,其示例由图4(b)所示。可以看到距离主肿瘤区域(primary tumor)表面越远的点有大的距离值,主靶区内部的点均为负的距离值。除了距离特征,几何特征还包括角度特征,用来描述危及器官到靶区的角度分布。在本应用中角度特征包括角度直方图的点位(Angle histogram points)。角度直方图描述了在所有角度下靶区的平行投影与危及器官的重叠体积。角度直方图可以由修改后的拉东变换(radon transform)产生的正弦图(sinogram)积分得到:危及器官在不同角度位于靶区背部前方部分的密度投影的积分。图4(c)展示了传统的左腮腺(Parotid_L)的正弦图和主肿瘤区域正弦图的轮廓,修改过的拉东变换能够反映出靶区和危及器官之间的前后方位,如图4(d)所示,只有在左腮腺位于主肿瘤区域之前正弦图才有信号。

[0067] 图4(a)表示主肿瘤区域和左腮腺(primary tumor-GTVnp, and Parotid_L),图4(b)表示主肿瘤区域和左腮腺距离分布图,图4(c)VS传统的左腮腺(Parotid_L)的正弦图和主肿瘤区域正弦图的轮廓,图4(d)VS修改过的拉东变换反映出靶区和危及器官之间的前后方位,只有在左腮腺位于主肿瘤区域之前正弦图才有信号。

[0068] 在第三方面,本发明的技术包括对RTR需求的估算或预测。基于本发明的预测模型,通过拟合上述提取的特征来估计个体患者的RTR需求,本发明的模型是通过使用发明人的训练数据集之中的HNC患者的多组学特征和临床参数进行预训练和开发的。

[0069] 在一个实施例中,用于开发RTR预测模型的训练数据来自一组鼻咽癌患者。所述鼻咽癌患者包括:(1)被诊断为活检证实的原发性NPC的患者,他们在就诊时不存在远处转移和其他类型的共存肿瘤;(2)接受治愈性同步放化疗(CCRT)或CCRT加辅助化疗(AC)的患者;以及(3)接受了螺旋断层放疗的患者。所述患者鼻咽癌不包括:(1)在CCRT治疗前接受诱导

化疗的患者；(2) 接受单独放疗而没有同步化疗的患者；(3) 未接受造影剂注射以获得计划性对比度增强CT (CECT) 图像或计划性对比度增强T1-w (CET1-w) MR图像的患者；或者(4) 没有完整的临床/图像数据集的患者。

[0070] 如上所述,本发明涉及多组学在高危HNC患者的分层以进行放疗重新计划中的应用以及多组学特征计算方法。其中,用于剂量组学和几何特征提取的内部算法的开发和验证是本发明在研发过程中最耗时的部分。

[0071] 综上所述,本发明包括至少以下贡献特征:(1) 针对HNC RTR需求的预处理人工智能(AI) 评估方法,这是本领域中首次用于相同的目的。(2) 基于多组学的个性化预测,本发明首次整合了HNC患者的患者特异性放射组学、剂量组学、几何学和临床数据,用于综合评估RTR需求。(3) 本发明解决了目前临床RTR实践效率高的问题。目前RTR的实施效率极低,患者需求筛查只能在治疗开始后进行治疗,需要对每一位个体患者进行密集的治疗中期监测。整个监测和风险评估程序非常耗时,需要多个医学科室的努力。一旦患者被确定为RTR候选者,就需要对该特定患者实施即时的RTR,这需要来自图像重新扫描(放射治疗师)、器官重新分割(医学物理学家和肿瘤学家)、治疗重新计划(医学物理学家)到新计划批准(肿瘤学家)的多个医学科室的努力。本发明允许对每位患者的RTR需求进行治疗前评估,大大减轻了密集的中期治疗评估带来的临床负担,能够简化专门用于高危患者的医疗资源和人力的分配,并有可能在临床RTR交付中实现更高的程序效率。

[0072] 另外,本发明解决了临床RTR需求筛选缺乏个性化策略的问题。大多数现有研究的潜在RTR触发因素与一般标准有关,例如肿瘤分期、患者治疗前体重指数(BMI)、或由腮腺接受的治疗中期辐射剂量阈值。然而,在相同治疗下,患者之间的治疗反应存在差异;具有相同肿瘤分期或治疗前BMI或在治疗期间其腮腺接受超过辐射剂量的患者可能表现出异质性治疗反应,导致患者在整个RT过程中的解剖学结构发生不同程度的变化。因此,这种建议的RTR触发因素不是患者特异性的,也不是有效的,尤其是在个性化医疗时代。本发明的基于多组学的预测允许整合来自个体放射影像、RT治疗计划、在特定器官之间的几何关系、以及相关临床属性的患者特异性预测因子,以对HNC中的RTR需求进行个性化评估。

[0073] 本发明能够获得至少以下优点:允许进行治疗前风险评估,大大减轻临床负担;该模型是根据治疗前患者信息开发的;实现个性化风险评估,本发明的模型是基于用于表征患者和治疗特定信息的多组学特征数据开发的;具有成本效益,本发明是基于计算机的方法,不需要精心设计的工作站或其他硬件,同时它可以在治疗开始之前提供基于多组学的患者特定的RTR需求预测。

[0074] 本发明的测试数据是可用的,本发明的当前模型是基于小样本量开发的,可以在其中加入或包含患者的临床参数,也就是说,将来可以完成具有更大的队列并包含相关的临床属性(例如治疗前体重、肿瘤分期等)的模型训练。通过增加样本量和包含患者的临床数据来重新训练模型,这将增强本发明模型的可预测性和统计推断。本发明可以用于针对RTR需求的治疗前HNC患者风险评估,以及根据预测的RTR需求优化初始治疗方案。

[0075] 本发明仅包括在放疗开始之前基于“最初的治疗前放疗计划”计算HNC患者的放疗剂量和/或覆盖范围和/或放疗剂量的2D/3D分布图或轮廓曲线,而不是计算所述放疗剂量和/或覆盖范围在放疗期间的实时变化。本发明的输入都是在RT开始之前提供的,例如治疗前医学图像和治疗前RT计划,基于这些图像或计划来估计RTR需求。这允许进行“治疗前评

估”。因此,通过使用本发明,没有对肿瘤进行任何治疗中的评估/计算剂量/覆盖范围/分布图/轮廓曲线。

[0076] 本发明通过使用用于患有晚期鼻咽癌的患者多模态图像(计算机断层扫描(CT)和多参数磁共振(MR)图像)而利用了来自多个器官(双侧腮腺、原发性肿瘤、淋巴结病变等)的多组学特征。本发明构建了集成各种预测因子的预测列线图,以评估各个患者的风险概率。

[0077] 虽然本发明已结合有限的实施方案进行了详细说明,但应当理解,本发明不限于这些公开的实施方案。本领域的普通技术人员可设想符合本发明的精神和范围的其它实施方式,包括部件的数量变化、更改、替代或等效布置,这些实施方式均落入本发明的范围。

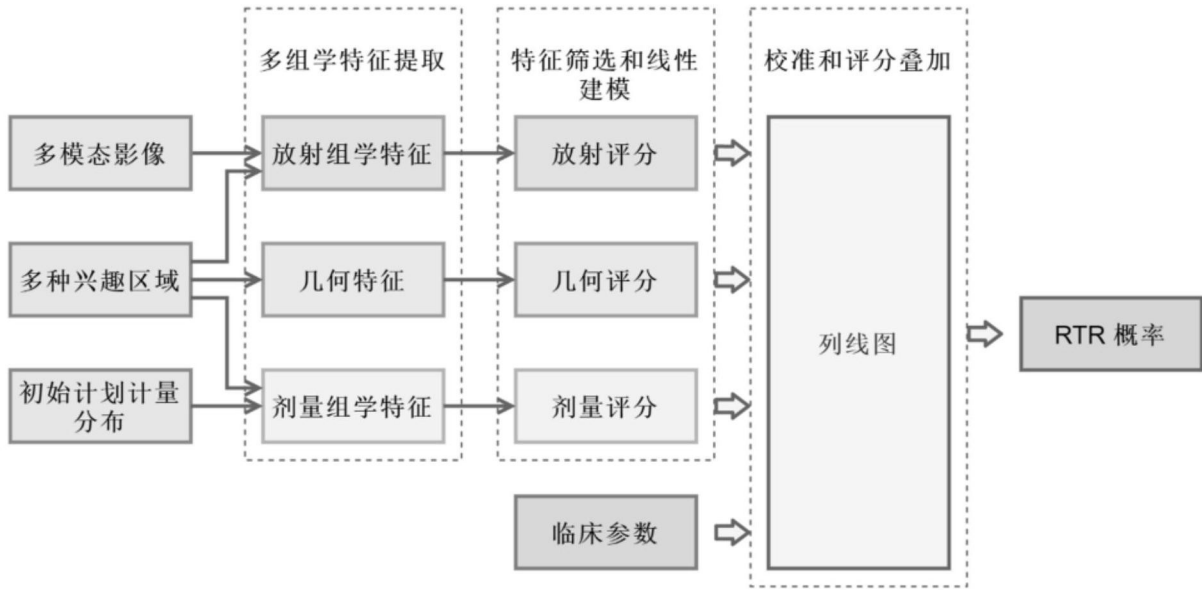


图1

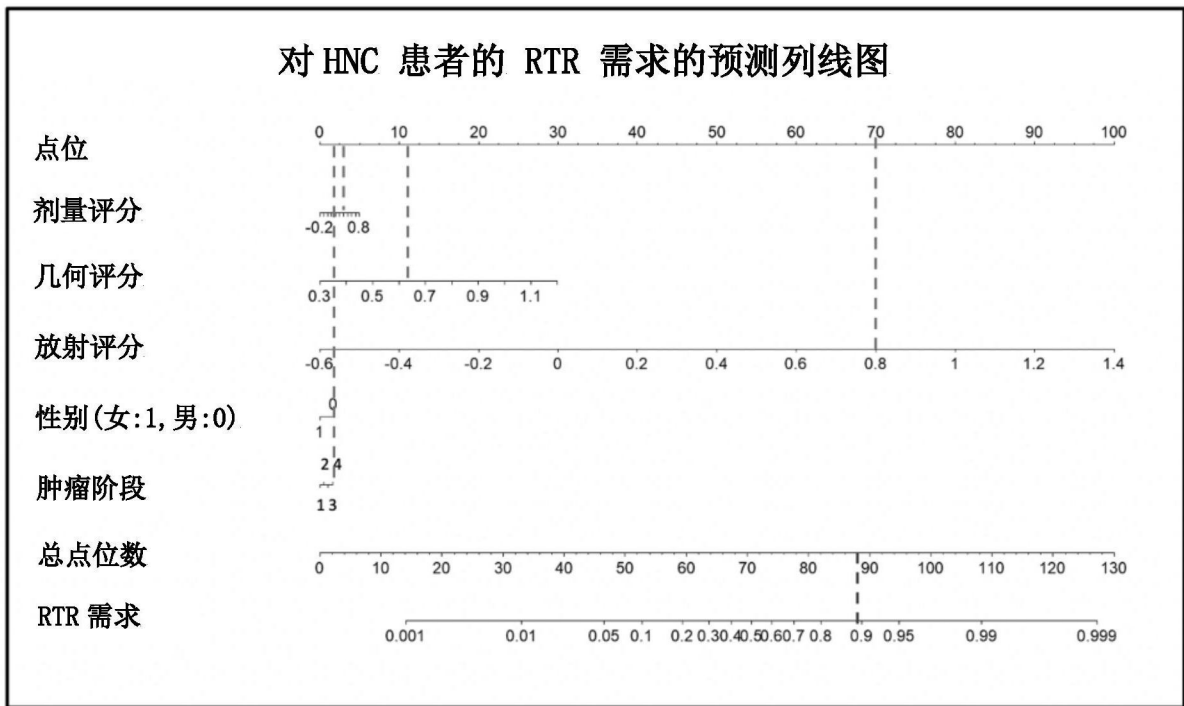


图2

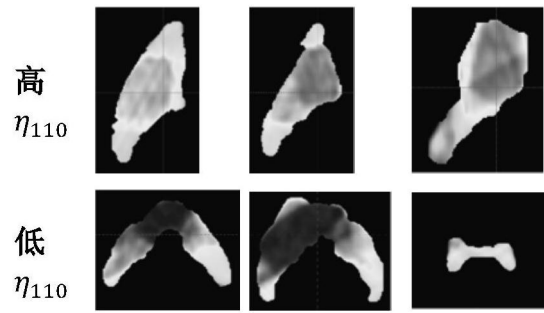


图3(a)

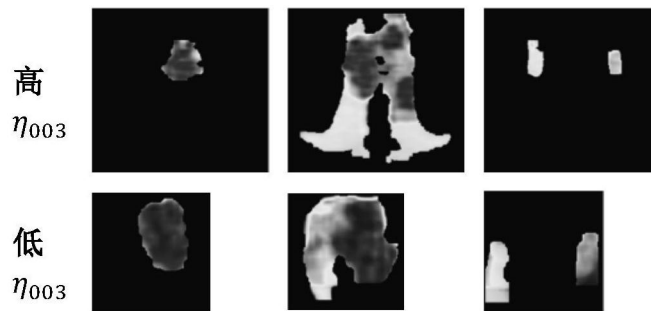


图3(b)

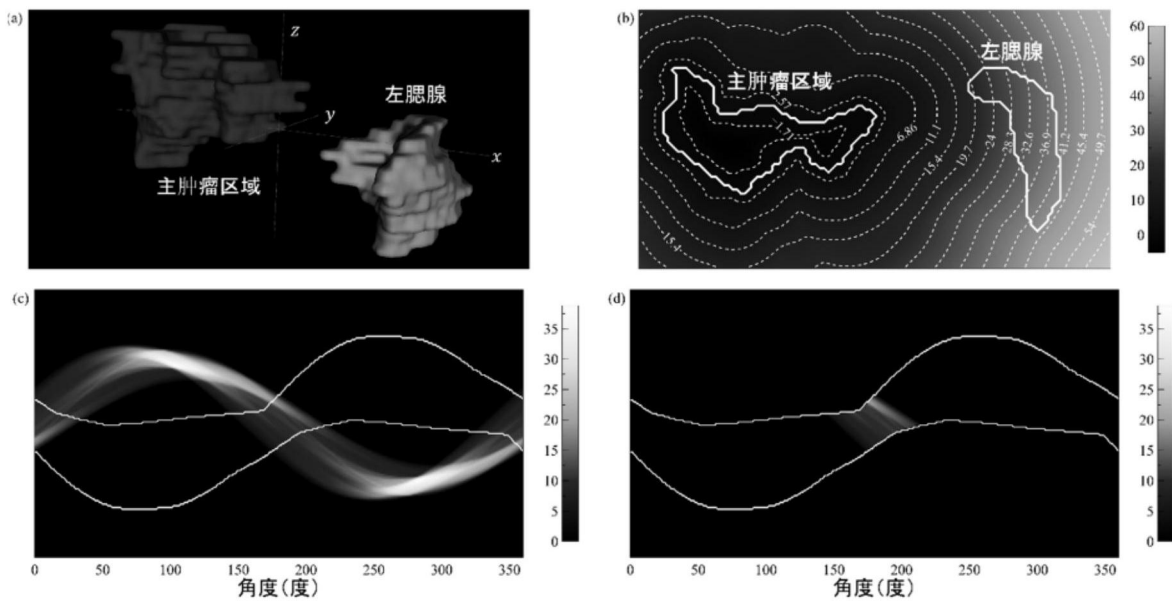


图4