



(12)发明专利申请

(10)申请公布号 CN 110265095 A

(43)申请公布日 2019.09.20

(21)申请号 201910426920.2

G16H 70/60(2018.01)

(22)申请日 2019.05.22

G06K 9/32(2006.01)

(71)申请人 首都医科大学附属北京佑安医院
地址 100071 北京市丰台区右安门外西头
条8号

G06K 9/34(2006.01)

G06K 9/46(2006.01)

(72)发明人 袁春旺 王振常 田捷 魏靖伟
顾东升 赵鹏 何宁 高文峰
杨晓珍 孙玉

(74)专利代理机构 重庆市信立达专利代理事务
所(普通合伙) 50230
代理人 包晓静

(51)Int.Cl.

G16H 10/20(2018.01)

G16H 30/20(2018.01)

G16H 50/30(2018.01)

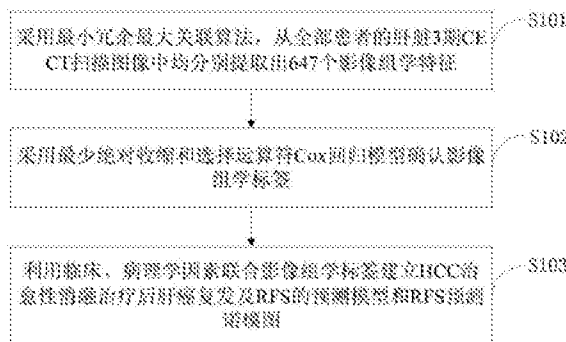
权利要求书2页 说明书11页 附图3页

(54)发明名称

用于HCC复发及RFS的预测模型和诺模图的
构建方法及应用

(57)摘要

本发明属于预测模型和诺模图构建技术领域,公开了一种用于HCC复发及RFS的预测模型和诺模图的构建方法及应用,采用最小冗余最大关联算法(minimum redundancy maximum relevance algorithm,MRMRA),从全部患者的肝脏3期增强CT扫描图像中均分别提取出647个影像组学特征;采用最少绝对收缩和选择运算符(least absolute shrinkage and selection operator,LASSO)-Cox回归模型确认影像组学标签;利用临床、病理学因素联合影像组学标签建立HCC治愈性消融治疗后复发及RFS的预测模型和预测诺模图。本发明通过影像组学方法构建预测模型和预测诺模图,用于对HCC患者在病灶治愈性消融治疗后肿瘤复发及RFS进行预测。



1. 一种用于HCC复发及RFS的预测模型和诺模图的构建方法,其特征在于,所述用于HCC复发及RFS的预测模型和诺模图的构建方法包括:

第一步,采用MRMRA,从肝脏3期CECT扫描图像中均分别提取出647个影像组学特征;

第二步,采用LASSO-Cox回归模型确认影像组学标签;

第三步,利用临床、病理学因素联合影像组学标签建立HCC治愈性消融治疗后HCC复发及RFS的预测模型和预测诺模图。

2. 如权利要求1所述用于HCC复发及RFS的预测模型和诺模图的构建方法,其特征在于,所述步骤一具体包括:

共647个反映机器阅读影像学特征和细微纹理信息的影像组学特征从所分割的感兴趣区中采用软件Matlab 2014a提取;采用非抽样小波变换对原始三维肿瘤层面进行图像滤过处理;从原始图像或分解图像中进行提取的特征可被分为两类:纹理特征和非纹理特征;非纹理特征包括:形状、大小和密度特征;形状和大小特征获取肿瘤的直观特点;密度特征展现病灶的直方图特点;纹理特征基于以下四个纹理矩阵进行提取,即灰度共生矩阵、灰度级长矩阵、灰度级带矩阵和邻域灰度差分矩阵。

3. 如权利要求1所述用于HCC复发及RFS的预测模型和诺模图的构建方法,其特征在于,所述步骤二具体包括:

(1) 临床、病理因素的分析、选择

临床、病理因素采用单因素Cox风险比例回归模型进行分析, p 值 <0.10 的因素入选,入选的因素被整合并进入逐步回归多因素Cox模型;在多因素分析中,变量的 p 值 <0.05 为该因素可能与患者的RFS相关而被纳入模型的建立,否则即被排除;

(2) 影像组学特征的选择

ICC通过test-retest检验法计算并确定某一特征的稳定性;某一特征的属性相关系数 <0.75 将被排除;MRMRA被用于特征选择;选择特征子集,该子集可以取得两组之间差异的最佳特点,最终根据每一特征的MRMRA输出分数,共有20个特征被遴选用于构建模型;irrR软件包被用于计算属性相关系数;mRMReR软件包被用于MRMRA特征选择;

(3) 影像组学标签的建立

对于训练集数据,采用LASSO方法在Cox比例风险模型中对选择的20个特征进行深层次变量选择;最终选择理想的具有最小交叉验证日志部分似然的特征;表明特征和生存风险比例的非零系数被确定为每一个选择特征的权重;每一位患者的影像组学标签通过所选择特征与各自系数的乘积产生;对于CECT图像来说,采用LASSO-Cox模型利用训练集数据建立模型并在验证集对模型预测效能进行验证;影像组学标签的预测价值通过KM曲线评估;入组患者通过影像组学标签的中位值分为高风险组和低风险组;通过对数秩和检验比较KM曲线;采用glmnet R软件包进行LASSO-Cox变量选择和模型建立;采用R软件包对生存曲线进行比较;

最终选择的20个特征被输入LASSO-Cox模型;根据缺一交叉验证法,最终纳入的用于建立影像组学模型的动脉期、门静脉期和延迟期的影像特征;单因素Cox比例风险模型以3期CECT中的动脉期、门静脉期、延迟期的影像组学标签分别建立预测模型,而多因素Cox比例风险模型则整合3期CECT全部影像组学标签建立联合预测模型。

4. 如权利要求1所述用于HCC复发及RFS的预测模型和诺模图的构建方法,其特征在于,

所述步骤三具体包括：

(1) 将临床、病理因素融入上述4个影像组学预测模型,门静脉期影像组学联合预测模型在验证组显示出最佳预测能力；

(2) 诺模图建立:因整合有临床、病理因素的门静脉期影像组学联合模型具有最佳预测能力,因此基于该模型建立诺模图;同时对于训练组和验证组均绘制诺模图校正曲线。

5. 一种由权利要求1所述用于HCC复发及RFS的预测模型和诺模图的构建方法构建的用于HCC复发及RFS的预测模型和诺模图。

6. 一种如权利要求5所述用于预测HCC复发及RFS的模型和诺模图在HCC局部治愈性消融治疗后进行肿瘤复发及RFS预测的应用。

用于HCC复发及RFS的预测模型和诺模图的构建方法及应用

技术领域

[0001] 本发明属于预测模型和诺模图构建技术领域,尤其涉及一种用于HCC复发及RFS的预测模型和诺模图的构建方法及应用。具体为一种用于肝细胞肝癌 (hepatocellular carcinoma,HCC) 复发及无复发生存 (recurrence free survival,RFS) 的预测模型和诺模图的构建方法及应用。

背景技术

[0002] 目前,最接近的现有技术:

[0003] 尚无与现有技术类似的预测模型,目前有根据患者甲胎蛋白 (alpha feto protein,AFP) 结果对HCC复发进行预测,但一旦AFP升高,可能肿瘤已经比较明显;此外尚有相当一部分患者 (20%左右) 为AFP阴性HCC,即便出现肿瘤复发,AFP仍为阴性,因此对阴性HCC患者监测AFP无法预测HCC复发。亦有根据HCC的病理学分化程度预测复发的方法,但敏感性及特异性均较低,因临床发现部分低分化HCC患者在病灶切除/消融后3年无复发,而相当比例的中高分化HCC患者在病灶切除/消融后半年内即发生肿瘤复发,此为HCC生物学特性及肿瘤所致,因活检仅穿刺极少部分HCC组织,不能代表整个HCC的生物学行为。

[0004] 也有根据患者的意大利肝癌协作组 (cancer of the liver Italian program, CLIP) 评分和肝功能Child-pugh评分及联合AFP进行复发预测的模型,同样灵敏度及特异度明显不足,文献具体如下:

[0005] (1) Nakagawa S, Hayashi H, Nitta H, et al. Scoring system based on tumor markers and Child-Pugh classification for HCC patients who underwent liver resection. *Anticancer Res.* 2015; 35 (4) :2157-2163.

[0006] (2) Zhao WH, Ma ZM, Zhou XR, Feng YZ, Fang BS. Prediction of recurrence and prognosis in patients with hepatocellular carcinoma after resection by use of CLIP score. *World J Gastroenterol.* 2002; 8 (2) :237-242.

[0007] (3) Tateishi R, Shiina S, Yoshida H, et al. Prediction of recurrence of hepatocellular carcinoma after curative ablation using three tumor markers. *Hepatology.* 2006; 44 (6) :1518-1527.

[0008] 综上所述,现有技术存在的问题是:

[0009] (1) 普适性差;

[0010] (2) 敏感性低;

[0011] (3) 特异性低;

[0012] (4) 准确率低;

[0013] 解决上述技术问题的难度:

[0014] 通过影像组学的研究方法,利用人工智能技术,对患者术前肝脏CECT中整个肿瘤的扫描数据 (DICOM格式) 进行挖掘、提取和处理分析,在上百个影像组学特征中提取与HCC复发密切相关的特征,再结合患者的临床及实验室、病理结果建立复发预测模型及诺模图,

可明显提高预测准确性。

[0015] 解决上述技术问题的意义在于：

[0016] 在较准确预测HCC消融术后复发几率的前提下，可指导临床医师调整治疗策略（预测为高复发率的患者术后辅助靶向药或中医中药治疗预防复发）及随访强度（视所预测复发率的高低，调强或调低随访强度，如高复发概率者，每月进行1次CECT或增强MRI（CEMRI）、抽血查肿瘤标志物检查等以早期发现复发病灶，反之，如低复发概率者，每3-6个月进行上述复发即可）。

发明内容

[0017] 针对现有技术存在的问题，本发明提供了一种用于HCC复发及RFS的预测模型和诺模图的构建方法及应用。

[0018] 本发明是这样实现的，一种用于HCC复发及RFS的预测模型和诺模图的构建方法包括：

[0019] 第一步，采用MRMRA（最小冗余最大关联算法minimum redundancy maximum relevance algorithm,MRMRA），从肝脏3期CECT扫描图像中均分别提取出647个影像组学特征；

[0020] 第二步，采用最少绝对收缩和选择运算符（least absolute shrinkage and selection operator,LASSO）-Cox回归模型确认影像组学标签；

[0021] 第三步，利用临床、病理学因素联合影像组学标签建立HCC治愈性消融治疗后HCC复发及RFS的预测模型和预测诺模图。

[0022] 进一步，所述步骤一具体包括：

[0023] 共647个反映机器阅读影像学特征和细微纹理信息的影像组学特征从所分割的兴趣区中采用软件Matlab 2014a提取；采用非抽样小波变换对原始三维肿瘤层面进行图像滤过处理；从原始图像或分解图像中进行提取的特征可被分为两类：纹理特征和非纹理特征；非纹理特征包括：形状、大小和密度特征；形状和大小特征获取肿瘤的直观特点；密度特征展现病灶的直方图特点；纹理特征基于以下四个纹理矩阵进行提取，即灰度共生矩阵（gray level co-occurrence matrix,GLCM）、灰度级长矩阵（gray level run-length matrix,GLRLM）、灰度级带矩阵（gray level size zone matrix,GLSZM）和邻域灰度差分矩阵（neighborhood gray-tone difference matrix,NGTDM）。

[0024] 进一步，所述步骤二具体包括：

[0025] （1）临床、病理因素的分析、选择

[0026] 临床、病理因素采用单因素Cox风险比例回归模型进行分析， p 值 <0.10 的因素入选，入选的因素被整合并进入逐步回归多因素Cox模型；在多因素分析中，变量的 p 值 <0.05 被认定为该因素可能与患者的RFS相关而被纳入模型的建立，否则即被排除；

[0027] （2）影像组学特征的选择

[0028] ICC通过test-retest检验法计算并确定某一特征的稳定性；某一特征的属性相关系数 <0.75 将被排除；为了减少计算和建模的冗余和不必要的复杂性，MRMRA被用于特征选择；该算法的目的是选择特征子集，该子集可以取得两组之间差异的最佳特点，其中考虑一项限制因素，即这些特征完全不同但与所关注的临床结局又有极其微弱的相关性；而且

MRMRA被证实可使所提取的特征更加稳定,尤其对于影像组学研究来说更是如此;最终根据每一特征的MRMRA输出分数,共有20个特征被遴选出来用于构建模型;“irr”R软件包被用于计算属性相关系数;“mRMRe”R软件包被用于MRMRA特征选择。

[0029] (3) 影像组学标签的建立

[0030] 对于训练集数据,采用LASSO方法在Cox比例风险模型中对选择的20个特征进行深层次变量选择;最终选择理想的具有最小交叉验证日志部分似然的特征;表明特征和生存风险比例的非零系数被确定为每一个选择特征的权重;每一位患者的影像组学标签通过所选择特征与各自系数的乘积产生;对于CECT图像来说,采用LASSO-Cox模型利用训练集数据建立模型并在验证集对模型预测效能进行验证;影像组学标签的预测价值通过KM曲线评估;入组患者通过影像组学标签的中位值分为高风险组和低风险组;通过对数秩和检验比较KM曲线;采用glmnet R软件包进行LASSO-Cox变量选择和模型建立;采用R软件包对生存曲线进行比较;

[0031] 最终选择的20个特征被输入LASSO-Cox模型;根据缺一交叉验证法,最终纳入的用于建立影像组学模型的动脉期、门静脉期和延迟期的影像特征;单因素Cox比例风险模型以3期CECT中的动脉期、门静脉期、延迟期的影像组学标签分别建立预测模型,而多因素Cox比例风险模型则整合3期CECT全部影像组学标签建立联合预测模型。

[0032] 进一步,所述步骤三具体包括:

[0033] (1) 将临床、病理因素融入上述4个影像组学预测模型,门静脉期影像组学联合预测模型在验证组显示出最佳预测能力;

[0034] (2) 诺模图建立:因整合有临床、病理因素的门静脉期影像组学联合模型具有最佳预测能力,因此基于该模型建立诺模图;同时对于训练组和验证组均绘制诺模图校正曲线。

[0035] 本发明的另一目的在于提供一种由所述构建方法构建的用于HCC复发及RFS的预测模型和诺模图。

[0036] 本发明的另一目的在于提供一种所述用于预测HCC复发及RFS的模型和诺模图在HCC局部治愈性消融治疗后进行肿瘤复发及RFS预测的应用。

[0037] 综上所述,本发明的优点及积极效果为:

[0038] 本发明提供的用于肝细胞肝癌 (hepatocellular carcinoma, HCC) 复发及无复发生存 (recurrence free survival, RFS) 的预测模型和诺模图的构建方法及应用,采用最小冗余最大关联算法 (minimum redundancy maximum relevance algorithm, MRMRA), 从全部患者的肝脏3期增强CT扫描图像中均分别提取出647个影像组学特征;采用最少绝对收缩和选择运算符 (least absolute shrinkage and selection operator, LASSO)-Cox回归模型确认影像组学标签;利用临床、病理学因素联合影像组学标签建立HCC治愈性消融治疗后复发及RFS的预测模型和预测诺模图。本发明通过影像组学方法构建预测模型和预测诺模图,用于对HCC患者在病灶治愈性消融治疗后肿瘤复发及RFS进行预测。

[0039] 本发明通过影像组学 (radiomics) 方法构建预测模型和预测诺模图,用于HCC患者在病灶局部治愈性消融治疗后对肿瘤复发及RFS进行预测。

[0040] 本发明由所述用于HCC复发及RFS的预测模型和诺模图的构建方法构建的用于HCC复发及RFS的预测模型和诺模图,是可准确预测HCC治愈性消融后复发概率的工具。

[0041] 本发明用于HCC复发及RFS的预测模型和诺模图对于HCC在病灶局部治愈性消融治

疗后HCC复发及无复发生存预测中的应用。利用该诺模图即可根据某一患者的影像组学数值,临床及病理特征直接读取其1/2/3年复发概率。

附图说明

[0042] 图1是本发明实施例提供的用于HCC复发及RFS的预测模型和诺模图的构建方法流程图。

[0043] 图2是本发明实施例提供的用于HCC复发及RFS的预测模型和诺模图的技术路线图。

[0044] 图3是本发明建立预测模型和诺模图的具体实施步骤示意图。

[0045] 图4是本发明实施例提供的临床、病理特征和门静脉期影像组学标签整合优化后的预测诺模图。

[0046] 图5是本发明实施例提供的训练组联合模型预测诺模图校正曲线图。

[0047] 图6是本发明实施例提供的验证组联合模型预测诺模图校正曲线图。

具体实施方式

[0048] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合实施范例,对本发明进行进一步详细说明。应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。

[0049] 下面结合附图对本发明的应用原理作详细的描述。

[0050] 如图1所示,本发明实施范例提供的用于HCC复发及RFS的预测模型和诺模图的构建方法包括以下步骤:

[0051] S101:采用MRMRA,从全部患者的肝脏3期CECT扫描图像中均分别提取出647个影像组学特征;

[0052] S102:采用LASSO-Cox回归模型确认影像组学标签;

[0053] S103:利用临床、病理学因素联合影像组学标签建立HCC治愈性消融治疗后HCC复发及RFS的预测模型和RFS预测诺模图。

[0054] 下面结合具体实施例对本发明的应用原理作进一步描述。

[0055] 如图1-3所示,本发明实施范例提供的用于HCC复发及RFS的预测模型和诺模图的构建方法包括以下步骤:

[0056] 步骤一,采用MRMRA,从全部患者的肝脏3期CECT扫描图像中均分别提取出647个影像组学特征;

[0057] 步骤二,采用LASSO-Cox回归模型确认影像组学标签;

[0058] 步骤三,利用临床、病理学因素联合影像组学标签建立HCC治愈性消融治疗后HCC复发及RFS的预测模型和RFS预测诺模图。

[0059] 进一步,所述步骤一具体包括:

[0060] 共647个反映机器阅读影像学特征和细微纹理信息的影像组学特征从所分割的感兴趣区(ROI)中采用软件Matlab 2014a(MathWorks,Natick,MA,USA)提取。采用非抽样小波变换(可将原始图像分解为8个组分)对原始三维HCC层面进行图像滤过处理。从原始图像或分解图像中进行提取的特征可被分为两类:纹理特征和非纹理特征。非纹理特征包括:形

状、大小和密度特征。形状和大小特征获取HCC的直观特点。密度特征展现病灶的直方图特点。纹理特征基于以下四个纹理矩阵进行提取,即GLCM、GLRLM、GLSZM、NGTDM。全部特征的类型和命名见表1。

[0061] 表1.影像组学模型中CECT各期所选特征的详细信息

[0062]

特征名称	公式	内容
Coif3_glcmaximum_probability	maximum_probability= $\max\{C(i, j)\}$	The maximum probability of the Gray-Level Co-Occurrence Matrix of CT image transformed by wavelet filter XLHL
Coif1_glrml_RLGL	LRGLLE= $\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[\frac{R(i, j) \theta_j^2}{i^2} \right]$	Long Run Low Gray Level Emphasis of Gray-Level Run-Length matrix of the CT image transformed by wavelet filter XLLL
Coif3_fos_mean	Entropy= the mean intensity value	Mean intensity value of the CT image transformed by wavelet filter XLHL
Coif7_glszm_LGLZE	LGLZE= $\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[\frac{Z(i, j)}{i^2} \right]$	Low Gray Level Zone Emphasis of Gray-level size zone matrix of the CT image transformed by wavelet filter XHHL
Coif3_glcmaxcluster_shade	cluster_shade= $\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} [i + j - \mu_x(i) - \mu_y(j)]^2 C$	Cluster shade of the Gray-Level Co-Occurrence Matrix of CT image

[0063]

	Coif3_glem_entropy	Entropy= $-\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} C(i, j) \log_2[C(i, j)]$	Entropy of the Gray-Level Co-Occurrence Matrix of CT image transformed by wavelet filter XLHL
	Coif1_glszm_LZLGE	LZLGE= $\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[\frac{Z(i, j)j^2}{i^2} \right]$	Large Zone Low Gray-Level Emphasis of Gray-level size zone matrix of the CT image transformed by wavelet filter XLLL
门静脉期	Coif4_glszm_LGLZE	LGLZE= $\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[\frac{Z(i, j)}{i^2} \right]$	Low Gray Level Zone Emphasis of Gray-level size zone matrix of the CT image transformed by wavelet filter XLHH
	Coif3_glem_maximum_probability	maximum_probability= $\max\{C(i, j)\}$	Maximum probability of the Gray-Level Co-Occurrence Matrix of CT image transformed by wavelet filter XLHL
	Ori_fos_maximumum	fos_maximum= The maximum intensity value of X	Maximum intensity value of the original image
	Coif4_glszm_SZLGE	SZLGE= $\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[\frac{Z(i, j)}{i^2 j^2} \right]$	Small Zone Low Gray Level Emphasis of Gray-level size zone matrix of the CT image transformed by wavelet filter XLHH
	Coif1_glrhm_LGLRE	LGLRE= $\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[\frac{RG(i, j) \theta }{i^2} \right]$	Low Gray Level Run Emphasis of Gray-Level Run-Length matrix of the CT image transformed by wavelet filter XLLL
延迟期	ori_Max3D	Max3D= maximum three-dimensional tumor diameter	Maximum three-dimensional tumor diameter of the original CT image
	Coif7_fos_mean	fos_mean= $\frac{1}{N} \sum_{i=1}^{N_x} X(i)$	Mean intensity of grey histogram of CT image transformed by wavelet filter XHHH
	Coif4_glem_correlation	Correlation= $\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} ijC(i, j)$	Correlation of the Gray-Level Co-Occurrence Matrix of CT image transformed by wavelet filter XLHL
	Coif1_glrhm_LRHGLE	LRHGLE= $\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} (R(i, j) \theta)^2 j^2$	Long Run High Gray Level Emphasis of Gray-Level Run-Length matrix of the CT image transformed by wavelet filter XLLL

[0064]

<p>inverse_variance Coif2_glem_in verse_variance</p>	$\text{inverse_variance} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{C(i, j)}{1 + \left(\frac{ i-j ^2}{N^2}\right)}$	<p>Inverse variance of the Gray-Level Co-Occurrence Matrix of CT image transformed by wavelet filter XLHL</p>
<p>Coif3_glszm_S ZHGE</p>	$\text{SZHGE} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \left[\frac{Z(i, j) i^2}{j^2} \right]$	<p>Small Zone High Gray-Level Emphasis of Gray-level size zone matrix of the CT image transformed by wavelet filter XLHL</p>
<p>Coif5_glem_co variance</p>	<p>Covariance= $\sum_{i=2}^{2N_g} (i - SE)^2 P_{xy}(i)$</p>	<p>Sum of covariance of the Gray-Level Co-Occurrence Matrix of CT image transformed by wavelet filter XLHL</p>
<p>Coif8_glszm_S ZE</p>	$\text{SZE} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \left[\frac{Z(i, j) i^2}{j^2} \right]$	<p>Small Zone Emphasis of Gray-level size zone matrix of the CT image transformed by wavelet filter XLHL</p>

[0065] X is the intensity value of the original image.

[0066] XLLL, XLLH, XLHL, XLHH, XHLL, XHLH, XHHL, and XHHH are the intensity value of the transformation images from the original image by eight three-dimensional wavelet filters. L: low-pass filter; H: high-pass filter. For example, XLHL represent the intensity value resulting from directional filtering of X with a low-pass filter along the x-direction, a high pass filter along the y-direction and a low-pass filter along the z-direction.

[0067] \bar{X} is median intensity value of X.

[0068] R(i, j | θ) is the value of row i and column j in the Gray-Level Run-Length Matrix for a direction θ .

[0069] C(i, j) is the value of row i and column j in the Gray-Level Co-Occurrence Matrix.

[0070] Z(i, j) is the value of row i and column j in the Gray-Level Size Zone Matrix.

[0071] N_g is the number of discrete intensity values in the image.

[0072] N_r is the number of different run lengths.

[0073] N_z is the size of the largest homogeneous region.

[0074] N_p is the number of voxels in the image.

[0075] $\mu_x(i)$ is the mean of row i.

[0076] $\mu_y(j)$ is the mean of column j.

[0077] $\sigma_x(i)$ is the standard deviation of row i.

[0078] $\sigma_y(j)$ is the standard deviation of column j.

[0079] 进一步, 所述步骤二具体包括:

[0080] (1) 临床、病理因素的分析、选择

[0081] 临床、病理因素采用单因素Cox风险比例回归模型进行分析， p 值 <0.10 的因素入选，入选的因素被整合并进入逐步回归多因素Cox模型。在多因素分析中，变量的 p 值 <0.05 被认定为该因素可能与患者的RFS相关而被纳入模型的建立，否则即被排除。

[0082] (2) 影像组学特征的选择

[0083] 组内属性相关系数 (intra-class correlation coefficient, ICC) 通过test-retest检验法计算并确定某一特征的稳定性。某一特征的属性相关系数 <0.75 将被排除。为了减少计算和建模的冗余和不必要的复杂性，MRMRA被用于特征选择。该算法的目的是选择特征子集，该子集可以取得两组之间差异的最佳特点，这其中考虑一项限制因素，即这些特征完全不同但与所关注的临床结局又有极其微弱的相关性。而且MRMRA被证实可使所提取的特征更加稳定，尤其对于影像组学研究来说更是如此。最终根据每一特征的MRMRA输出分数，共有20个具有潜能特征被遴选出用于构建模型。“irr”R软件包被用于计算属性相关系数。“mRMRe”R软件包被用于MRMRA特征选择。

[0084] (3) 影像组学标签的建立

[0085] 对于训练集数据，采用深入的LASSO方法在Cox比例风险模型中对选择的20个特征进行深层次变量选择。最终选择理想的具有最小交叉验证日志部分似然的特征。表明特征和生存风险比例的非零系数被确定为每一个选择特征的权重。每一位患者的影像组学标签通过所选择特征与各自系数的乘积产生。对于CECT图像来说，采用LASSO-Cox模型利用训练集数据建立模型并在验证集对模型预测效能进行验证。影像组学标签的预测价值通过KM曲线 (Kaplan-Meier Curve, 卡普兰-迈耶曲线) 评估。入组患者通过影像组学标签的中位值分为高风险组和低风险组。通过对数秩和检验 (log-rank test) 比较KM曲线。采用glmnet R软件包进行LASSO-Cox变量选择和模型建立。采用R软件包对生存曲线进行比较。

[0086] 最终选择的20个特征被输入LASSO-Cox模型。根据缺一交叉验证法 (leave-one-out cross-validation)，最终纳入的用于建立影像组学模型的动脉期、门静脉期和延迟期的影像特征。所选择的征象见表2。3期CECT影像组学标签所的计算公式见表3。单因素Cox比例风险模型以3期CECT中的动脉期、门静脉期、延迟期的影像组学标签分别建立预测模型，而多因素Cox比例风险模型则整合3期CECT全部影像组学标签建立联合预测模型。

[0087] 表2. CECT各期特征选择

[0088]

动脉期(n=5)	门静脉期(n=5)	延迟期(n=10)
Coif3_glcm_maximum_probability	Coif3_glcm_entropy	Coif4_glszm_SZLGE
Coif1_glrIm_LRLGLE	Coif1_glszm_LZLGE	Coif1_glrIm_LGLRE
Coif3_fos_mean	Coif4_glszm_LGLZE	ori_Max3D
Coif7_glszm_LGLZE	Coif3_glcm_maximum_probability	Coif7_fos_mean
Coif3_glcm_cluster_shade	Ori_fos_maximum	Coif4_glcm_correlation
		Coif1_glrIm_LRHGLE
		Coif2_glcm_inverse_variance
		Coif3_glszm_SZHGE
		Coif5_glcm_covariance
		Coif8_glszm_SZE

[0089] 注:采用缺一交叉验证法,通过LASSO模型对CECT 3期扫描图像进行特征选择

[0090] 表3.CECT各期影像组学标签计算公式

[0091]

影像组学标签	
动脉期	Signature= $2.750826249 - 0.022446033 * N_{25_Coif3_glcm_maximum_probability} - 0.054947769 * N_{25_Coif1_glrIm_LRLGLE} + 0.008390395 * N_{25_Coif3_fos_mean} + 0.152563563 * N_{25_Coif7_glzm_LGLZE} - 0.034839997 * N_{25_Coif3_glcm_cluster_shade}$
门静脉期	Signature= $-2.633335158 + 0.062617223 * N_{25_Coif3_glcm_entropy} - 0.025120597 * N_{25_Coif1_glzm_LZLGE} + 0.034852387 * N_{25_Coif4_glzm_LGLZE} - 0.061244532 * N_{25_Coif3_glcm_maximum_probability} + 0.055584297 * N_{25_ori_fos_maximum}$
延迟期	Signature= $3.410207937 + 0.207406911 * N_{25_Coif4_glzm_SZLGE} - 0.082837573 * N_{25_Coif1_glrIm_LGLRE} + 0.060390091 * N_{25_ori_Max3D} + 0.035731692 * N_{25_Coif7_fos_mean} - 0.124767859 * N_{25_Coif2_glcm_inverse_variance}$

[0092]

$$\begin{aligned}
 &N25_Coif4_glm_correlation+ \\
 &0.047483351*N25_Coif1_glm_LRHGLE+0.00022*N25_Coif2_glm_inverse_variance+ \\
 &0.067083626*N25_Coif3_glm_SZHGE+0.071125387*N25_Coif5_glm_covariance- \\
 &0.045437444*N25_Coif8_glm_SIZE
 \end{aligned}$$

[0093] 进一步,所述步骤三具体包括:

[0094] (1) 将临床、病理因素融入上述4个影像组学预测模型时,门静脉期影像组学联合预测模型在验证组显示出最佳预测能力,该模型较单纯的临床模型预测能力也有明显改善(ANOVA, $p < 0.0001$) (表4)。

[0095] (2) 诺模图建立:因整合有临床、病理因素的门静脉期影像组学联合模型具有最佳预测能力,因此基于该模型建立诺模图(图4)。同时对于训练组和验证组均绘制联合诺模图校正曲线(图5,图6),其中y轴代表实际RFS比例;x轴代表预测RFS的可能性;对角虚线代表理想预测诺模图的完美预测曲线)。Hosmer-Leme拟合优度指标评价显示该诺模图在训练组($p = 0.791$)和验证组($p = 0.471$)表现出良好的一致性,无显著的统计学差异。

[0096] 表4. 各模型RFS预测效能比较

模型	训练组 (n=129)	验证组 (n=55)
	C-index (95%CI)	C-index (95%CI)
单纯临床模型		
临床、病理特征	0.649 (0.592-0.706)	0.556 (0.471-0.641)
单纯影像组学模型		
动脉期	0.767 (0.702-0.832)	0.694 (0.623-0.832)
门静脉期	0.757 (0.692-0.821)	0.736 (0.632-0.841)
延迟期	0.789 (0.723-0.853)	0.686 (0.582-0.791)
3期联合	0.791 (0.726-0.856)	0.690 (0.586-0.795)
影像组学与临床联合模型		
动脉期+临床病理特征	0.797 (0.732-0.862)	0.732 (0.628-0.837)
门静脉期+临床病理特征	0.792 (0.727-0.857)	0.755 (0.651-0.860)
延迟期+临床病理特征	0.806 (0.741-0.871)	0.728 (0.624-0.834)
增强3期+临床病理特征	0.809 (0.744-0.874)	0.724 (0.620-0.829)

[0099] 单纯临床模型预测效能很低,单纯影像组学模型预测模型较高,联合临床、病理因

素的门静脉期影像组学联合预测模型预测效能最佳,具体见表4。

[0100] 以上所述仅为本发明的较佳实施范例而已,并不用以限制本发明,凡在本发明的精神和原则之内所作的任何修改、等同替换和改进等,均应包含在本发明的保护范围之内。

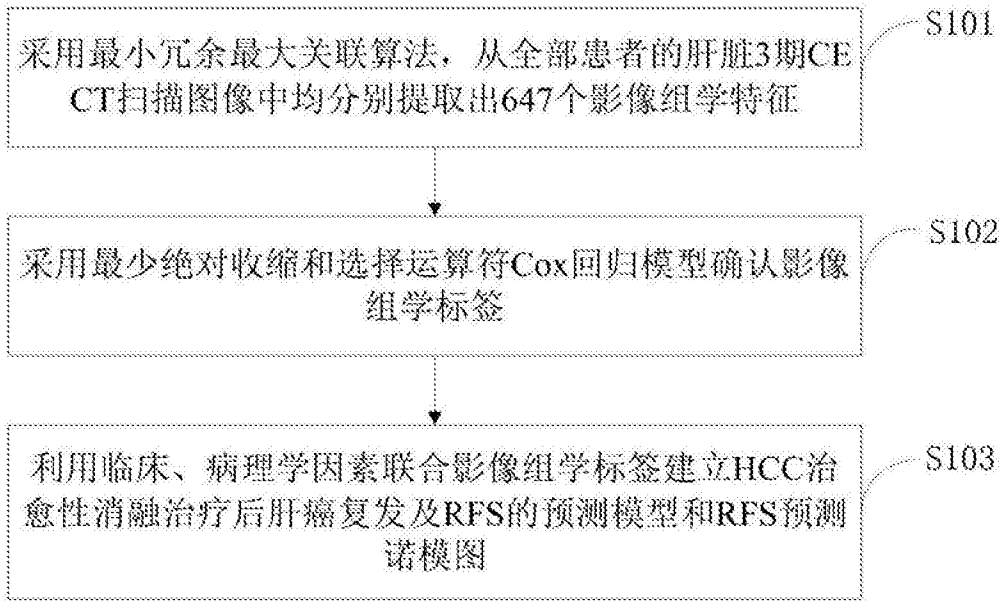


图1

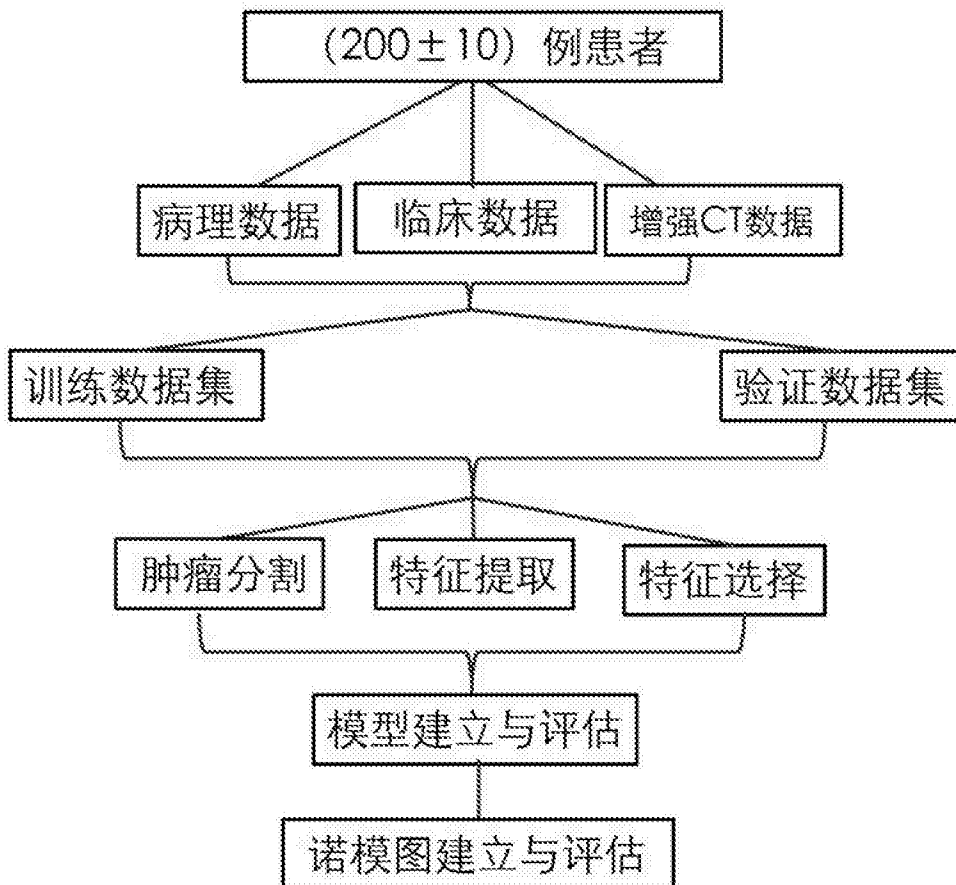


图2

影像组学方法建立诺模图步骤示意图

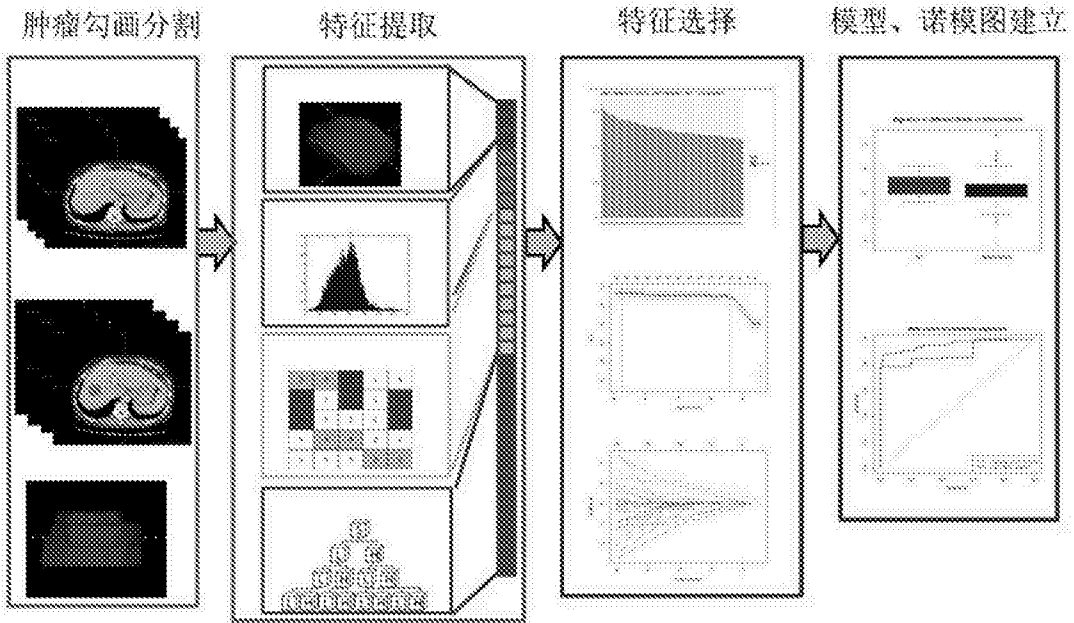


图3

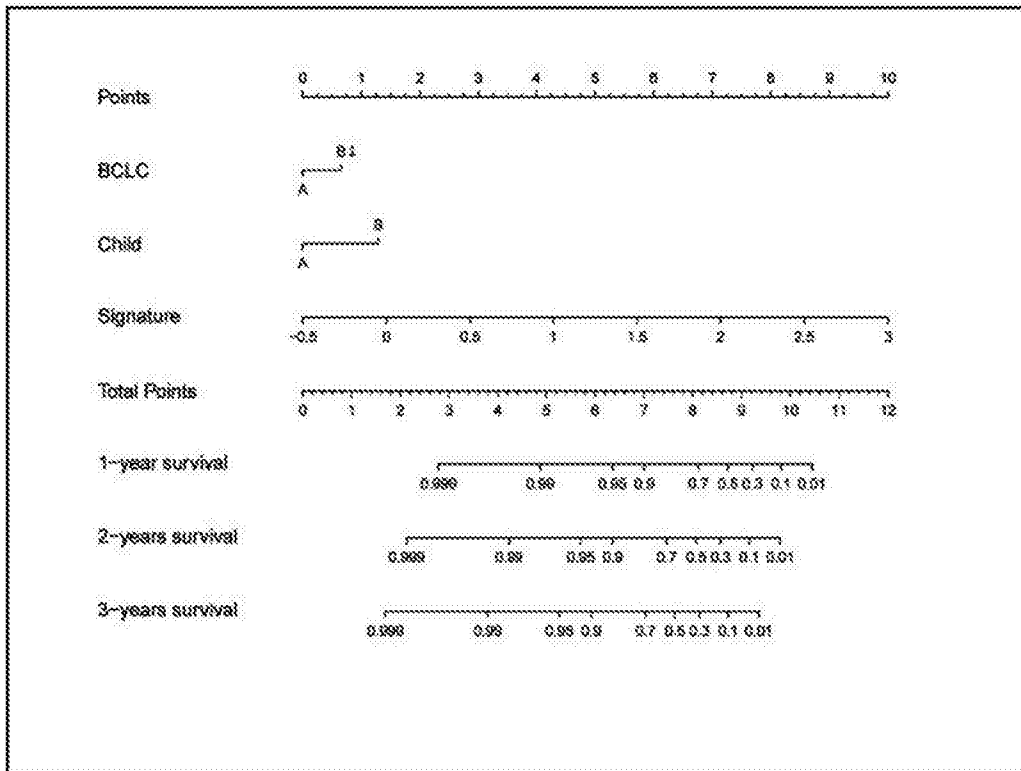


图4

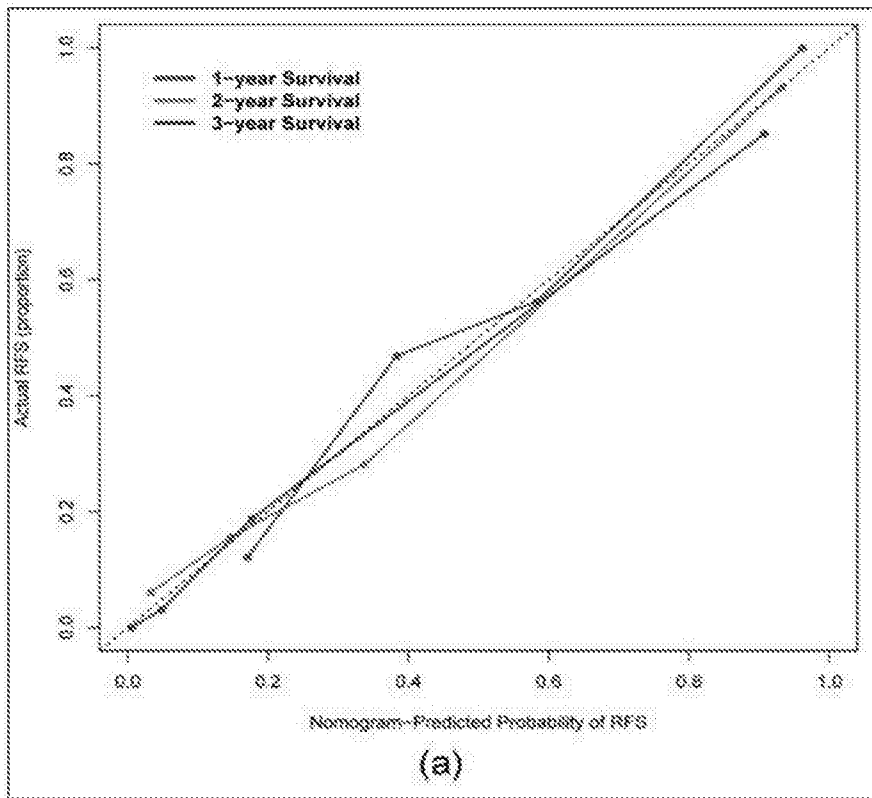


图5

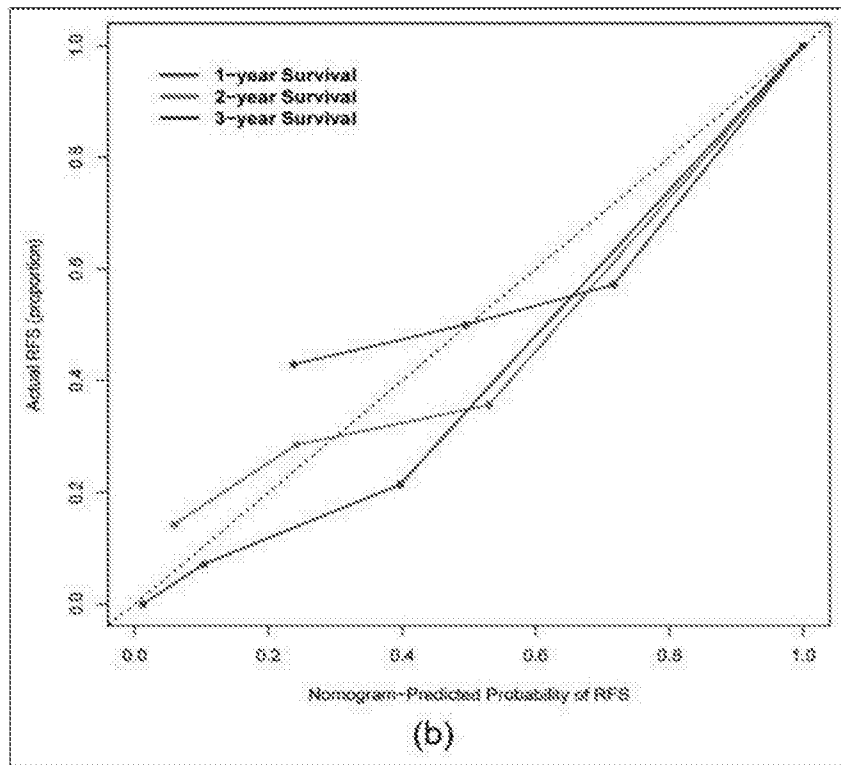


图6