

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2018 年 7 月 26 日 (26.07.2018)



(10) 国际公布号
WO 2018/132997 A1

- (51) 国际专利分类号:
G06F 19/00 (2011.01)
- (21) 国际申请号: PCT/CN2017/071665
- (22) 国际申请日: 2017 年 1 月 19 日 (19.01.2017)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (71) 申请人: 中国科学院深圳先进技术研究院 (SHENZHEN INSTITUTES OF ADVANCED TECHNOLOGY CHINESE ACADEMY OF SCIENCES) [CN/CN]; 中国广东省深圳市南山区西丽大学城学苑大道 1068 号, Guangdong 518055 (CN)。
- (72) 发明人: 李其花 (LI, Qihua); 中国广东省深圳市南山区西丽大学城学苑大道 1068 号, Guangdong

518055 (CN)。孙秋畅 (SUN, Qiuchang); 中国广东省深圳市南山区西丽大学城学苑大道 1068 号, Guangdong 518055 (CN)。李志成 (LI, Zhicheng); 中国广东省深圳市南山区西丽大学城学苑大道 1068 号, Guangdong 518055 (CN)。宋柏霖 (SONG, Bolin); 中国广东省深圳市南山区西丽大学城学苑大道 1068 号, Guangdong 518055 (CN)。王梦巧 (WANG, Mengqiao); 中国广东省深圳市南山区西丽大学城学苑大道 1068 号, Guangdong 518055 (CN)。

- (74) 代理人: 北京三友知识产权代理有限公司 (BEIJING SANYOU INTELLECTUAL PROPERTY AGENCY LTD.); 中国北京市金融街 35 号国际企业大厦 A 座 16 层, Beijing 100033 (CN)。

(54) Title: RADIOMICS-BASED SURVIVAL PREDICTION METHOD AND DEVICE

(54) 发明名称: 一种基于影像组学的生存期预测方法及装置

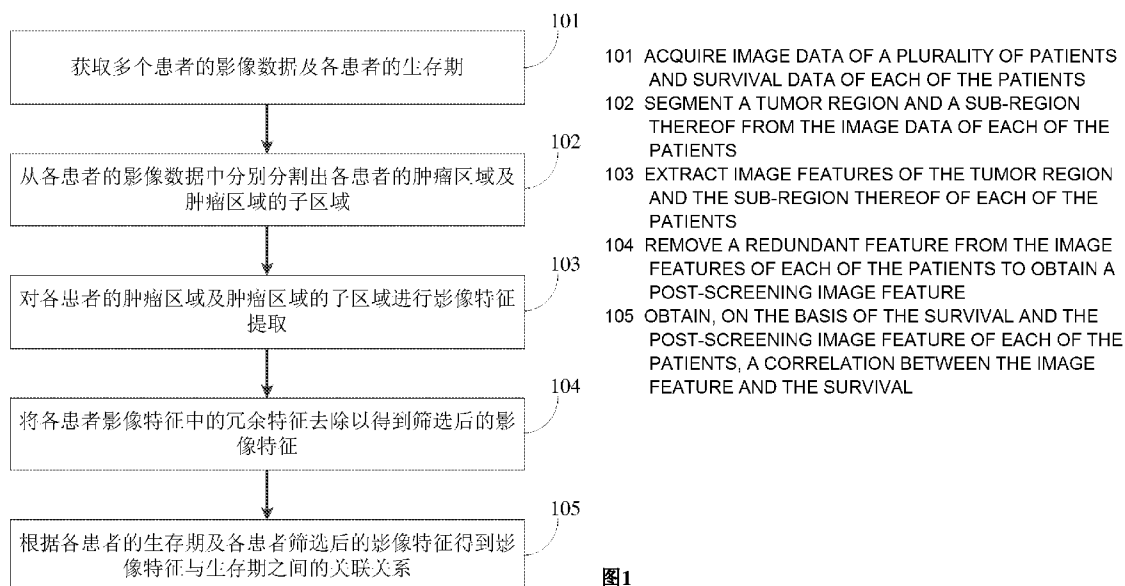


图1

(57) Abstract: The present application provides a radiomics-based survival prediction method and device. The method comprises: acquiring image data of a plurality of patients and survival data of each of the patients; segmenting a tumor region and a sub-region thereof from the image data of each of the patients; extracting image features of the tumor region and the sub-region thereof of each of the patients; removing a redundant feature from the image features of each of the patients to obtain a post-screening image feature of each of the patients; and obtaining, on the basis of the survival and the post-screening image feature of each of the patients, a correlation between the image feature and the survival. The present application segments a tumor region and a sub-region thereof to extract features thereof to obtain a large amount of image features, providing strong support for establishing a relationship between image features and patient survival.

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

(57) 摘要: 本申请提供了一种基于影像组学的生存期预测方法及装置, 其中, 方法包括获取多个患者的影像数据及各患者的生存期; 从各患者的影像数据中分别分割出各患者的肿瘤区域及肿瘤区域的子区域; 对各患者的肿瘤区域及肿瘤区域的子区域进行影像特征提取; 将各患者影像特征中的冗余特征去除以得到各患者筛选后的影像特征; 根据各患者的生存期及各患者筛选后的影像特征得到影像特征与生存期之间的关联关系。本申请通过划分出肿瘤区域及肿瘤区域的子区域, 对肿瘤区域及肿瘤区域的子区域进行特征提取, 从而获得大量的影像特征, 为寻求影像特征和患者生存期的关系提供更多有力支持。

一种基于影像组学的生存期预测方法及装置

技术领域

本申请属于生物医学工程领域，特别涉及一种基于影像组学的生存期预测方法及装置。

5

背景技术

近年来，随着模式识别工具的增加和肿瘤个体化治疗的发展，影像组学应运而生。影像组学从影像、病理、基因等海量数据中利用自动化高通量的数据特征提取算法将影像数据转化为具有高分辨率的可挖掘的影像特征数据。通过这些数据挖掘描述组织特性。有研究报道，影像组学数据可以判断组织特性对治疗的反应，并预测患者的预后。应用影像组学数据对患者或肿瘤表型进行精确区分，可以成为对临床参数的一种有效补充。

10

影像组学是一种可以根据患者医学影像中的肿瘤表型对患者进行分类的方法。它通过对肿瘤图像提取大量先进的影像特征来评估肿瘤表型，然后使用可靠，重现性好的方法论，将这些特征与临床结果进行综合分析作为潜在的预后指标，从而提供一种非侵入性的精准诊疗方法。影像组学生成一个独特的肿瘤数据集，它是一个肿瘤表型的量化，可以提供比目前临床应用的成像指标更高的预测能力。

15

现有技术采用影像组学方法预测患者生存期的方法中分割出了完整的肿瘤区域，根据该完整的肿瘤区域提取了影像特征，建立了完整肿瘤区域影像特征和生存期之间的关联性。

20

发明内容

现有技术中，基于影像组学的分析只分割出了完整的肿瘤区域，没有考虑肿瘤区域的子区域具有不同的病理特征，另只提取了肿瘤区域的影像特征，忽略了肿瘤区域的子区域具有不同的病理的影像特征，从而导致所提取的影像特征不能全面的代表感兴趣区域的特点，极大的限制了基于感兴趣区域所提取的影像特征的数量与质量的问题。

25

为了解决上述技术问题，本申请的一技术方案为提供一种基于影像组学的生存期预测方法，包括：

获取多个患者的影像数据及各患者的生存期；

从各患者的影像数据中分别分割出各患者的肿瘤区域及肿瘤区域的子区域；
对各患者的肿瘤区域及肿瘤区域的子区域进行影像特征提取；
将各患者影像特征中的冗余特征去除以得到各患者筛选后的影像特征；
根据各患者的生存期及各患者筛选后的影像特征得到影像特征与生存期之间的关联

5 关系。

本申请另一技术方案为提供一种基于影像组学的生存期预测装置，包括：
获取模块，用于获取多个患者的影像数据及各患者的生存期；
分割模块，用于从各患者的影像数据中分别分割出各患者的肿瘤区域及肿瘤区域的子区域；

10 特征提取模块，用于对各患者的肿瘤区域及肿瘤区域的子区域进行影像特征提取；
筛选模块，用于将各患者影像特征中的冗余特征去除以得到各患者筛选后的影像特征；

关联模块，用于根据各患者的生存期及各患者筛选后的影像特征得到影像特征与生存期之间的关联关系。

15 本申请提供的基于影像组学的生存期预测方法及装置能够细化肿瘤区域得到肿瘤区域的子区域，对肿瘤区域及肿瘤区域的子区域进行特征提取，从而获得大量的影像特征，为寻求影像特征和（患者）生存期的关系提供更多有力支持。

附图说明

20 为了更清楚地说明本申请实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图 1 为本申请实施例的基于影像组学的患者生存期预测方法的流程图；

25 图 2 为本申请实施例的基于影像组学的患者生存期预测装置的结构图；

图 3 为本申请实施例的基于影像组学的患者生存期预测装置的结构图；

图 4 为本申请实施例的脑瘤区域的子区域划分的示意图；

图 5 为本申请实施例的电子设备的结构图。

30 具体实施方式

为了使本申请的技术特点及效果更加明显，下面结合附图对本申请的技术方案做进一步说明，本申请也可有其他不同的具体实例来加以说明或实施，任何本领域技术人员在权利要求范围内做的等同变换均属于本申请的保护范畴。

在本说明书的描述中，参考术语“一个实施例”、“一个具体实施例”、“一些实施例”、“例如”、“示例”、“具体示例”或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本申请的至少一个实施例或示例中。在本说明书中，对上述术语的示意性表述不一定指的是相同的实施例或示例。而且，描述的具体特征、结构、材料或者特点可以在任何的一个或多个实施例或示例中以合适的方式结合。各实施例中涉及的步骤顺序用于示意性说明本申请的实施，其中的步骤顺序不作限定，可根据需要作适当调整。

实施例1

如图1所示，图1为本申请实施例的基于影像组学的患者生存期预测方法的流程图。本实施例中，考虑了肿瘤组织空间-时间异质性，提取出的感兴趣区域不局限于肿瘤区域，还提取出了肿瘤区域的子区域，并对肿瘤区域的子区域进行了影像特征提取，使得获得的影像特征更全面，建立的影像特征与生存期之间的关联关系精度更高，能够提高患者生存期预测的准确度。具体的，包括：

步骤101：获取多个患者的影像数据及各患者的生存期。

多个患者的影像数据为患者的影像集合，例如表示为 $V=\{v_i, i=1, \dots, N\}$ ， N 为患者的个数，每个个体数据 v_i 表示一个影像样本。

本步骤中所述的多个患者患有同一种肿瘤疾病，如脑瘤。影像数据为患者进行治疗前的影像图像，包括但不限于包括 PET，CT 及 MRI 等影像。生存期为影像数据产生至患者死亡之间的时间段。实施时，可从医院获取多个患者的影像数据及各患者的生存期。

本步骤中获得的患者的影像数据例如来源于 TCIA（The Cancer Imaging Archive，癌症影像存档）中的 57 例患者影像数据集。本申请对患者的影像数据个数不做限定，一般情况下，个数越多，关联关系确定越准确。

步骤102：从各患者的影像数据中分别分割出各患者的肿瘤区域及肿瘤区域的子区域，肿瘤区域为整个肿瘤区域，肿瘤区域的子区域为整个肿瘤区域内部划分的子区域。以脑瘤为例，脑瘤区域的子区域包括但不限于坏死区、增强区及水肿区，划分结果如图4所示。

肿瘤区域及肿瘤区域的子区域构成感兴趣区域，肿瘤区域除了包括真实肿瘤区域外，还包括疑似肿瘤区域。

病理学已证实：肿瘤水肿区内有异常形态的毛细血管，间质水肿和散在的肿瘤细胞在新生血管或扩张血管的浸润生长；坏死区则是因肿瘤生长过快，养料供应不够，导致肿瘤内部坏死，它可以间接的反应肿瘤的增长速度。将肿瘤区域的子区域作为感兴趣区域可以详细的具有代表性的反应肿瘤所具有的性质特征，为下一步的高通量特征提取提供了更大的提取区域。

步骤 103：对各患者的肿瘤区域及肿瘤区域的子区域进行影像特征提取。

肿瘤区域的子区域有其独特的病理特征，提取这些区域的特征可更全面的反应出肿瘤的的性质，同时，进一步获得大量的影像学特征数据以极大的提高后期预测患者生存期的准确性。

肿瘤区域、肿瘤区域的子区域提取的影像特征可以相同，也可以不同，本申请对此不作具体限定。

步骤 104：将各患者影像特征中的冗余特征去除以得到各患者筛选后的影像特征。

冗余特征包括可以通过其他特征推演出来的特征及与预测无关的特征。

步骤 105：根据各患者的生存期及各患者筛选后的影像特征得到影像特征与生存期之间的关联关系。

本实施例能够细化肿瘤区域得到肿瘤区域的子区域，对肿瘤区域及肿瘤区域的子区域进行影像特征提取，从而获得大量的影像特征，为寻求影像特征和（患者）生存期的关系提供更多有力支持。

一些实施方式中，上述步骤 102 之前还包括：对各患者的影像数据进行预处理。

因患者的影像数据获取参数不同，通过本步骤的处理能够标准化患者的影像数据。具体的，对各患者的影像数据进行预处理包括对各患者的影像数据进行图像配准、图像平滑及数据标准化处理。下面将对预处理过程进行详细的说明：

图像配准：将不同时间，不同成像设备或不同条件（如照度，摄像位置和角度）获取的两幅或多幅影像进行匹配，叠加，从而统一各患者影像数据的坐标系。

具体的配准方法有相对配准和绝对配准。相对配准是指选择多影像中的一张影像作为参考影像，将其它的相关影像与之配准，配准时可任意选择坐标系统。绝对配准是指先定义一个控制网格，所有的影像相对于该控制网络来进行配准，也就是分别完成各影

像的几何校正来实现坐标系的统一。具体实施时，可根据需求选择图像配准方法，本申请对此不作具体限定。

图像平滑：在影像数据获取、形态学处理过程中，可能会产生不平滑的毛刺、锋利的边缘等情况。这通常不是物体原始特征，而是人为造成的，需要去除，图像平滑用于去除各患者影像数据中的毛刺。影像数据平滑的方法有很多种，例如均值滤波器，中值滤波器，这些方法均可通过现有技术实现，此处不再详细叙述。

数据标准化：标准化即将获取的多个患者的影像数据变为统一标准的影像数据。实施时，影像数据标准化的方法有多种，例如最小-最大标准化，Z-score 标准化，小数定标标准化，线性变换等等。这些方法均可通过现有技术实现，下面以线性变换为例进行说明。

设样本 A 为一患者的影像数据，maxA 为样本 A 像素的最大值，A 为样本 A 的任意一原像素值，A' 为与 A 相对的标准化的像素值，M 为所有患者影像数据像素最大值。

其公式为： $A' = (A/\max A) \times (M+100)$ 。

一些实施方式中，上述步骤 103 中提取的影像特征包括但不限于直方图特征数据，形状特征数据及纹理特征数据。

直方图特征用于描述肿瘤区域及各个肿瘤区域的子区域的所有像素点的灰度值，包括均值，中值，最大值，最小值，极差，能量，熵，偏斜度，峰度，标准差，方差，平均绝对差，均方根等等。

形状特征用于描述肿瘤区域及各个肿瘤区域的子区域的三维特征，包括体积，最长径，表面积，硬度，密度，等周商，球形不均衡度，曲率，偏心率，表面积体积比等等。

纹理特征用于描述肿瘤区域及各个肿瘤区域的子区域的纹理特征来量化肿瘤内部的异质性，包括灰度共生矩阵类特征，灰度游程矩阵类特征，灰度尺寸区域矩阵类特征，邻域灰度调差矩阵类特征，小波变换，拉普拉斯变换，高斯变换等等。

直方图特征数据，形状特征数据及纹理特征数据具体包括的特征参见表一。需要说明的是，表一中的所有特征并非都得提取，可根据肿瘤类型进行选取，例如，脑瘤分区后，对脑瘤整个区域及内部子区域分别提取 14 个直方图特征，28 个形状特征，52 个纹理特征。

表一

特征种类	特征名字
------	------

直方图特征		最大值
		中值
		最小值
		均值
		能量
		熵
		方差
		峰度
		均方根
		偏斜度
		平均绝对差
		标准差
形状特征		肿瘤体积
		增强区体积
		非增强区体积
		水肿区体积
		坏死区体积
		核心区体积
		坏死区比例
		水肿区比例
		非增强区比例
		增强区比例
		核心区表面积
		肿瘤表面积
		核心区最长径
		肿瘤最长径
		核心区密度
		肿瘤密度
		核心区偏心率
		肿瘤偏心率
		核心区致密度 1
		肿瘤致密度 1
		核心区致密度 2
		肿瘤致密度 2
		核心区球形不均衡度
		肿瘤球形不均衡度
		核心区曲率
		肿瘤曲率
		核心区表面积体积比
		肿瘤表面积体积比
纹理特征	灰度共生矩阵类特征	对比度
		相关性
		熵差

		熵
		相关信息测度 1
		相关信息测度 2
		平均和
		熵和
		方差和
		方差
		变异数差
		自相关性
		不相关性
		能量
		同质性 1
		同质性 2
		最大概率
		逆差值
		标准化
		逆差标准化
	灰度游程矩阵特征	短游程增强
		长游程增强
		灰度不均匀性
		游程长度不均匀
		游程百分比
		低灰度游程增强
		高灰度游程增强
		短游程低灰度增强
		短游程高灰度增强
		长游程低灰度增强
		长游程高灰度增强
		灰度方差
		游程长度方差
	灰度尺寸区域矩阵特征	小区域增强
		大区域增强
		灰度级不均匀性
		区域面积不均匀性
		区域所占比例
		低灰度级区域增强
		高灰度级区域增强
		小区域低灰度级增强
		小区域高灰度级增强
		大区域低灰度级增强
		大区域高灰度级增强
		灰度级方差
		区域面积方差

	邻域灰度调 差矩阵类特 征	粗糙度
		对比度
		冗余度
		复杂度
		强度

上述步骤 104 可以减轻后期学习中的维数灾难问题，也可以降低后期学习的难度，实施时，可采用现有的方法去除各患者影像特征中的冗余特征，具体的方法包括：

- 1) 广度优先搜索：枚举了所有的特征组合，广度优先遍历特征子空间，即从一个顶点开始，辐射状地优先遍历其周围较广的特征子空间；
- 2) 分支限界搜索：在穷举搜索的基础上加入分支限界。例如：若断定某些分支不可能搜索出比当前找到的最优解更优的解，则可以剪掉这些分支。
- 3) 定向搜索：首先选择 N 个得分最高的特征作为特征子集，将其加入一个限制最大长度的优先队列，每次从队列中取出得分最高的子集，然后穷举向该子集加入 1 个特征后产生的所有特征集，将这些特征集加入队列。特征的得分即特征与生存期的一致性指标（CI: Harrell's Concordance Index）值的大小，CI 值的高低由所选特征和生存期的相关性有关，相关性越高，特征的分数越高。
- 4) 最优优先搜索：与定向搜索类似，唯一的不同点是不限制优先队列的长度。
- 5) 序列前向选择：特征子集 X 从空集开始，每次选择一个特征 x 加入特征子集 X，使得特征函数 J(X) 最优。简单说就是，每次都选择一个使得评价函数的取值达到最优的特征加入，其实就是一种简单的贪心算法。
- 6) 序列后向选择：从特征全集 O 开始，每次从特征集 O 中剔除一个特征 x，使得剔除特征 x 后评价函数值达到最优。
- 7) 增 L 去 R 选择算法：该算法有两种形式：（1）算法从空集开始，每轮先加入 L 个特征，然后从中去除 R 个特征，使得评价函数值最优。（ $L > R$ ）；（2）算法从全集开始，每轮先去除 R 个特征，然后加入 L 个特征，使得评价函数值最优。（ $L < R$ ）。
- 8) 序列浮动选择：序列浮动选择由增 L 去 R 选择算法发展而来，该算法与增 L 去 R 选择算法的不同之处在于：序列浮动选择的 L 与 R 不是固定的，而是“浮动”的，也就是会变化的。序列浮动选择根据搜索方向的不同，有以下两种变种。（1）序列浮动前向选择：空集开始，每轮在未选择的特征中选择一个子集 x，使加入子集 x 后评价函数达到最优，然后在已选择的特征中选择子集 z，使剔除子集 z 后评价函数达到最优。

(2) 序列浮动后向选择：与 SFFS 类似，不同之处在于 SFBS 是从全集开始，每轮先剔除特征，然后加入特征。

9) 决策树：在训练样本集上运行 C4.5 或其他决策树生成算法，待决策树充分生长后，再在树上运行剪枝算法。则最终决策树各分支处的特征就是选出来的特征子集了。

5 决策树方法一般使用信息增益作为评价函数。

10) 遗传算法：首先随机产生一批特征子集，并用评价函数给这些特征子集评分，然后通过交叉、突变等操作繁殖出下一代的特征子集，并且评分越高的特征子集被选中参加繁殖的概率越高。这样经过 N 代的繁殖和优胜劣汰后，种群中就可能产生了评价函数值最高的特征子集。

10 11) 聚类算法：基于 K 均值聚类的特征选择方法，其基本思想就是对每一个特征子集利用 K 均值聚类算法确定其最佳分类数，从而删除掉相关性较大的特征之一。本发明采用的是一致性聚类方法：计算在多次运行的情况下两个子样本聚集在一起的频率，并根据一致率的结果做出视觉评价：比较稳定性和确定最优聚类数(K)。基本假设是，如果存在最优聚类数 K，K 所对应的子样本稳定性会达到最佳。一致性聚类是用 (0-1) 来描述稳定。我们用层次聚类算法与基于皮尔森相关的不同性进行测量，经过 2000 次重采样迭代，为了确定最佳聚类数，我们首先计算不同聚类数的累积分布函数。最佳聚类数即为累积分布函数中函数收敛时所对应的数值。

15

一些实施方式中，上述步骤 102 从各患者的影像数据中分别分割出各患者的肿瘤区域及肿瘤区域的子区域的过程包括：

20 从预先分割好的影像数据中提取影像特征集；将该提取出的影像特征集训练分类器，通过该分类器来分割（预处理后的）各患者的影像数据以得到肿瘤区域及肿瘤区域的子区域。

实施时，预先分割好的影像数据可由有经验的专家按照肿瘤区域及肿瘤区域的子区域的解剖结构将患者的影像数据划分成肿瘤区域和肿瘤区域的子区域。以脑瘤为例，分割好的脑部影像包括水肿区域、增强区域、坏死区域及非增强区域。

25

提取的影像特征集包括但不限于直方图特征，形状特征，纹理特征。直方图特征，形状特征及纹理特征具体包括的内容如表一。影像特征集具体包括的特征由预先分割好的影像数据各区域确定。对各分割出的子区域分别提取特征，例如提取的特征包括 23 个低阶特征：14 个灰度值特征（每个像素点和与它相邻的 6 个像素点的灰度值及以这 7 个

像素点为中心用 3×3 模块提取的均值灰度值)；6 个一阶纹理特征（均值，方差，偏斜度，峰值，能量，熵）；3 个位置特征（X,Y,Z 三个方向）。

本实施方式中所述的学习方法包括但不限于支持向量机，随机森林，卷积神经网络等等，本申请对学习方法的限定。

- 5 具体实施时，上述步骤 102 还可采用手动分割方式实现肿瘤区域及肿瘤区域的子区域的分割。例如，让有经验的专家按照肿瘤区域及肿瘤区域的子区域的解剖结构将患者的影像划分为肿瘤区域及肿瘤区域的子区域。亦或采样人工和计算机处理结合的方式进行分割处理。

- 10 一实施方式中，上述步骤 102 之后还包括：针对每一患者，对该患者的部分肿瘤区域的子区域进行合并处理以得到该患者的肿瘤区域的合并子区域。

- 15 实施时，肿瘤区域的子区域合并方式有很多种，例如随机合并、具有相似病理意义的肿瘤区域的子区域合并，本申请对具体的合并方法不作限定。以脑部胶质母细胞瘤为例，该胶质母细胞瘤包括坏死区、增强区、非增强区及水肿区，因坏死区、增强区和非增强区具有相似病理意义（致命性），所以可以将坏死区、增强区和非增强区进行合并。

上述步骤 103 对各患者的肿瘤区域及肿瘤区域的子区域进行影像特征提取进一步为：对各患者的肿瘤区域、肿瘤区域的子区域及肿瘤区域的合并子区域进行影像特征提取。

- 20 本实施方式更适用于实际需求，提取肿瘤区域的合并子区域的特征可以更全面的、多角度的反应出肿瘤的性质。同时这一步获得的大量影像学特征可以极大的提高后期预测患者生存期的准确性。

- 25 一些实施方式中，上述步骤 105 得到的影像特征与生存期之间的关联关系可以表示为：生存期为筛选后的多个影像特征或全部筛选后的影像特征进行线性组合的结果。假设每个影像特征代表一个因子，则生存期就是一个多因子方程，该方程的构建流程如下：

步骤 201：随机确认一个生存期的方程，该方程可以是单因子或多因子，线性或非线性，本申请对此不作限定。例如，方程可表示为：

$$Y = W * X + W_1 * X_1 + \dots + W_i * X_i + \dots + W_n * X_n + z,$$

其中, $X = [x_1, x_2, \dots, x_{m_0}]^T$, $X1 = [x_{1,1}, x_{1,2}, \dots, x_{1,m_1}]^T$, $Xi = [x_{i,1}, x_{i,2}, \dots, x_{i,m_i}]^T$, $Xn = [x_{n,1}, x_{n,2}, \dots, x_{n,m_n}]^T$;

Y 表示生存期; X 为肿瘤区域筛选出的影像特征; m_0 为肿瘤区域筛选出的影像特征个数; W 为 X 的系数, 为未知量; X1、Xi 及 Xn 为肿瘤区域的子区域筛选出的影像特征; m_1 、 m_i 及 m_n 为肿瘤区域的子区域筛选出的影像特征个数; W1、Wi 及 Wn 分别为 X1、Xi 及 Xn 的系数, 为未知量; z 表示常数, 为未知量。W、W1、Wi 及 Wn 为常数系数。

具体构建方程时, 可选用多参数 COX 逻辑回归模型构建一个线性的多因子方程, 并用梯度下降的算法寻求上述方程的最优参数组合, 即计算得到未知量的值。

步骤 202: 分别将各患者筛选后的影像特征值及生存期代入步骤 201 中的方程, 建模得到影像特征的系数 $W = \{Wi, i = 1, \dots, m\}$ 及常数 z, 将得到的影像特征的系数及常数 z 代入至步骤 201 中得到构建好的生存期的方程。

进一步的, 构建好的生存期的方程还可由时间-生存概率曲线表示, 该时间-生存概率曲线的生成过程包括:

1) 对于步骤 101 中多个患者, 分别通过构建好的方程计算方程值。例如, 步骤 101 中包括 59 个病例样本, 即包括 59 个患者, 则需要将这 59 个病例筛选后的影像特征代入构建好的方程中, 共得到 59 个方程值。

2) 提取计算出的方程值中的中间值, 将该中间值为界, 计算出的方程值若大于或等于中间值则赋为 1, 计算出的方程值若小于中间值则赋为 0。

同时, 记录患者的生存状态, 若最后随访时患者已死亡, 则该患者生存状态赋值为 1, 若最后随访时患者活着或者失访, 则患者的生存状态赋值为 0。

继续步骤 1) 中的例子, 假设中间值对应第 3 个患者, 则步骤 2) 得到的结果如表二,

表二:

患者编号	根据方程值的赋值		患者生存状态 (1: 表示死亡; 0 表示删失)	患者总的生存期 (天)
1	1		1	1000
2		0	0	500
4	1		1	

...	...	...	...	...
59		0	0	2000

3) 将各患者方程值的赋值及其对应患者的生存期、患者生存状态作为输入，通过SPSS 软件绘制得到时间-生存概率的曲线。

实施例 2

5 基于同一发明构思，本申请实施例中还提供了一种基于影像组学的患者生存期预测装置，如下面的实施例所述。由于该装置解决问题的原理与基于影像组学的患者生存期预测方法相似，因此该装置的实施可以参见方法的实施，重复之处不再赘述。

如图 2 所示，图 2 为本申请实施例的基于影像组学的患者生存期预测装置的结构图。该装置可以通过逻辑电路实现运行于智能终端，例如手机、平板电脑、计算机等设
10 备中，或者以功能模块的方式由软件实现各部件的功能，运行于所述智能终端上。具体的，该装置包括：

获取模块 201，用于获取多个患者的影像数据及各患者的生存期。

分割模块 202，用于从各患者的影像数据中分别分割出各患者的肿瘤区域及肿瘤区域
15 的子区域。详细的说，分割模块 203 具体用于从预先分割好的影像数据中提取影像特征；将该提取出的影像特征作为训练集，通过学习方法来分割预处理后的各患者的影像数据以得到肿瘤区域及肿瘤区域的子区域。

特征提取模块 203，用于对各患者的肿瘤区域及肿瘤区域的子区域进行影像特征提取。详细的说，影像特征包括直方图特征数据，形状特征数据及纹理特征数据。

筛选模块 204，用于将各患者影像特征中的冗余特征去除以得到各患者筛选后的影
20 像特征。

关联模块 205，用于根据各患者的生存期及各患者筛选后的影像特征得到影像特征与生存期之间的关联关系。

本申请提供的基于影像组学的生存期预测装置能够细化肿瘤区域得到肿瘤区域的子
25 区域，对肿瘤区域及肿瘤区域的子区域进行影像特征提取，从而获得大量的影像特征，为寻求影像特征和（患者）生存期的关系提供更多有力支持。

一些实施方式中，如图 3 所示，基于影像组学的患者生存期预测装置还包括预处理
模块 206，用于对各患者的影像数据进行预处理。预处理模块 206 具体用于对各患者的影像数据进行图像配准、图像平滑及数据标准化处理。

一些实施方式中，复请参阅图 3，基于影像组学的患者生存期预测装置还包括合并模块 207，用于针对每一患者，对该患者的部分肿瘤区域的子区域进行合并处理以得到该患者的肿瘤区域的合并子区域。

特征提取模块 203 进一步用于对各患者的肿瘤区域、肿瘤区域的子区域及肿瘤区域的合并子区域进行影像特征提取。

一些实施方式中，关联模块 205 得到的影像特征与生存期之间的关联关系的表达式为：

$$Y = W * X + W_1 * X_1 + \dots + W_i * X_i + \dots + W_n * X_n + z,$$

$$\text{其中, } X = [x_1, x_2, \dots, x_{m_0}]^T, \quad X_1 = [x_{1,1}, x_{1,2}, \dots, x_{1,m_1}]^T, \quad X_i = [x_{i,1}, x_{i,2}, \dots, x_{i,m_i}]^T,$$

$$X_n = [x_{n,1}, x_{n,2}, \dots, x_{n,m_n}]^T;$$

Y 表示生存期， X 表示肿瘤区域筛选出的影像特征， m_0 表示肿瘤区域筛选出的影像特征个数， W 表示 X 的系数， X_1 、 X_i 及 X_n 表示肿瘤区域的子区域筛选出的影像特征， m_1 、 m_i 及 m_n 表示肿瘤区域的子区域筛选出的影像特征个数， W_1 、 W_i 及 W_n 分别表示 X_1 、 X_i 及 X_n 的系数， z 表示常数。

实施例 3

本申请还提供一种电子设备，该电子设备可以是台式计算机、平板电脑及移动终端等，本实施例不限于此。在本实施例中，该电子设备可以参照实施例 1 的方法的实施及实施例 2 所述的装置，其内容被合并于此，重复之处不再赘述。

如图 5 所示，图 5 为本申请实施例的电子设备 500 的系统构成示意图。该电子设备 500 可以包括中央处理器 100 和包括计算机可读指令的存储器 140；存储器 140 耦合到中央处理器 100。

计算机可读指令在被执行时使所述处理器执行以下操作：获取多个患者的影像数据及各患者的生存期；从各患者的影像数据中分别分割出各患者的肿瘤区域及肿瘤区域的子区域；对各患者的肿瘤区域及肿瘤区域的子区域进行影像特征提取；将各患者影像特征中的冗余特征去除以得到各患者筛选后的影像特征；根据各患者的生存期及各患者筛选后的影像特征得到影像特征与生存期之间的关联关系。

其中，所述影像特征数据包括直方图特征数据，形状特征数据及纹理特征数据。

从各患者的影像数据中分别分割出各患者的肿瘤区域及肿瘤区域的子区域之前还包括：对各患者的影像数据进行预处理，其中，预处理包括图像配准、图像平滑及数据标准化处理。

从各患者的影像数据中分别分割出各患者的肿瘤区域及肿瘤区域的子区域之后还包括：针对每一患者，对该患者的部分肿瘤区域的子区域进行合并处理以得到该患者的肿瘤区域的合并子区域；对各患者的肿瘤区域及肿瘤区域的子区域进行影像特征提取进一步为对各患者的肿瘤区域、肿瘤区域的子区域及肿瘤区域的合并子区域进行影像特征提取。

在另一个实施方式中，基于影像组学的患者生存期预测装置可以与中央处理器 100 分开配置，例如可以将基于影像组学的患者生存期预测装置配置为与中央处理器 100 连接的芯片，通过中央处理器的控制来实现基于影像组学的患者生存期预测装置的功能。

如图 5 所示，该电子设备 500 还可以包括：通信模块 110、输入单元 120、音频处理单元 130、显示器 160、电源 170。值得注意的是，电子设备 500 也并不是必须要包括图 5 中所示的所有部件；此外，电子设备 500 还可以包括图 5 中没有示出的部件，可以参考现有技术。

如图 5 所示，中央处理器 500 有时也称为控制器或操作控件，可以包括微处理器或其他处理器装置和/或逻辑装置，该中央处理器 100 接收输入并控制电子设备 500 的各个部件的操作。

其中，存储器 140，例如可以是缓存器、闪存、硬驱、可移动介质、易失性存储器、非易失性存储器或其它合适装置中的一种或更多种。可储存上述与失败有关的信息，此外还可存储执行有关信息的程序。并且中央处理器 100 可执行该存储器 140 存储的该程序，以实现信息存储或处理等。其他部件的功能与现有类似，此处不再赘述。电子设备 500 的各部件可以通过专用硬件、固件、软件或其结合来实现，而不偏离本发明的范围。

本申请实施例还提供一种计算机可读程序，其中当在电子设备中执行所述程序时，所述程序使得计算机在所述电子设备中执行如上面实施例 1 所述的基于影像组学的患者生存期预测方法。

本申请实施例还提供一种存储有计算机可读程序的存储介质，其中所述计算机可读程序使得计算机在电子设备中执行上面实施例 1 所述的基于影像组学的患者生存期预测方法。

本领域内的技术人员应明白，本申请的实施例可提供为方法、系统、或计算机程序产品。因此，本申请可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质（包括但不限于磁盘存储器、CD-ROM、光学存储器等）上实施的计算机程序产品的形式。

本申请是参照根据本申请实施例的方法、设备（系统）、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上，使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

以上所述仅用于说明本申请的技术方案，任何本领域普通技术人员均可在不违背本申请的精神及范畴下，对上述实施例进行修饰与改变。因此，本申请的权利保护范围应视权利要求范围为准。

权利要求书

1. 一种基于影像组学的患者生存期预测方法，其特征在于，包括：

获取多个患者的影像数据及各患者的生存期；

从各患者的影像数据中分别分割出各患者的肿瘤区域及肿瘤区域的子区域；

5 对各患者的肿瘤区域及肿瘤区域的子区域进行影像特征提取；

将各患者影像特征中的冗余特征去除以得到各患者筛选后的影像特征；

根据各患者的生存期及各患者筛选后的影像特征得到影像特征与生存期之间的关联关系。

2. 如权利要求 1 所述的基于影像组学的患者生存期预测方法，其特征在于，从各患
10 者的影像数据中分别分割出各患者的肿瘤区域及肿瘤区域的子区域之前还包括：

对各患者的影像数据进行预处理，其中，预处理包括图像配准、图像平滑及数据标准化处理。

3. 如权利要求 1 所述的基于影像组学的患者生存期预测方法，其特征在于，所述影像特征包括直方图特征数据，形状特征数据及纹理特征数据。

4. 如权利要求 1 所述的基于影像组学的患者生存期预测方法，其特征在于，从各患
15 者的影像数据中分别分割出各患者的肿瘤区域及肿瘤区域的子区域之后还包括：

针对每一患者，对该患者的部分肿瘤区域的子区域进行合并处理以得到该患者的肿瘤区域的合并子区域；

20 对各患者的肿瘤区域及肿瘤区域的子区域进行影像特征提取进一步为对各患者的肿瘤区域、肿瘤区域的子区域及肿瘤区域的合并子区域进行影像特征提取。

5. 如权利要求 1 所述的基于影像组学的患者生存期预测方法，其特征在于，影像特征与生存期之间的关联关系的表达式为：

$$Y = W * X + W1 * X1 + \dots + Wi * Xi + \dots + Wn * Xn + z,$$

$$\text{其中, } X = [x_1, x_2, \dots, x_{m_0}]^T, X1 = [x_{1,1}, x_{1,2}, \dots, x_{1,m_1}]^T, Xi = [x_{i,1}, x_{i,2}, \dots, x_{i,m_i}]^T,$$

$$25 \quad Xn = [x_{n,1}, x_{n,2}, \dots, x_{n,m_n}]^T;$$

Y 表示生存期，X 表示肿瘤区域筛选出的影像特征， m_0 表示肿瘤区域筛选出的影像特征个数，W 表示 X 的系数，X1、Xi 及 Xn 表示肿瘤区域的子区域筛选出的影像特征， m_1 、 m_i 及 m_n 表示肿瘤区域的子区域筛选出的影像特征个数，W1、Wi 及 Wn 分别表示 X1、Xi 及 Xn 的系数，z 表示常数。

6. 一种基于影像组学的患者生存期预测装置，其特征在于，包括：

获取模块，用于获取多个患者的影像数据及各患者的生存期；

分割模块，用于从各患者的影像数据中分别分割出各患者的肿瘤区域及肿瘤区域的子区域；

5 特征提取模块，用于对各患者的肿瘤区域及肿瘤区域的子区域进行影像特征提取；

筛选模块，用于将各患者影像特征中的冗余特征去除以得到各患者筛选后的影像特征；

关联模块，用于根据各患者的生存期及各患者筛选后的影像特征得到影像特征与生存期之间的关联关系。

10 7. 如权利要求6所述的基于影像组学的患者生存期预测装置，其特征在于，还包括预处理模块，用于对各患者的影像数据进行预处理，其中，预处理包括图像配准、图像平滑及数据标准化处理。

8. 如权利要求6所述的基于影像组学的患者生存期预测装置，其特征在于，所述影像特征包括直方图特征数据，形状特征数据及纹理特征数据。

15 9. 如权利要求6所述的基于影像组学的患者生存期预测装置，其特征在于，还包括合并模块，用于针对每一患者，对该患者的部分肿瘤区域的子区域进行合并处理以得到该患者的肿瘤区域的合并子区域；

所述特征提取模块进一步用于对各患者的肿瘤区域、肿瘤区域的子区域及肿瘤区域的合并子区域进行影像特征提取。

20 10. 如权利要求6所述的基于影像组学的患者生存期预测装置，其特征在于，所述关联模块得到的影像特征与生存期之间的关联关系的表达式为：

$$Y = W * X + W1 * X1 + \dots + Wi * Xi + \dots + Wn * Xn + z,$$

$$\text{其中, } X = [x_1, x_2, \dots, x_{m_0}]^T, \quad X1 = [x_{1,1}, x_{1,2}, \dots, x_{1,m_1}]^T, \quad Xi = [x_{i,1}, x_{i,2}, \dots, x_{i,m_i}]^T, \\ Xn = [x_{n,1}, x_{n,2}, \dots, x_{n,m_n}]^T;$$

25 Y表示生存期，X表示肿瘤区域筛选出的影像特征， m_0 表示肿瘤区域筛选出的影像特征个数，W表示X的系数，X1、Xi及Xn表示肿瘤区域的子区域筛选出的影像特征， m_1 、 m_i 及 m_n 表示肿瘤区域的子区域筛选出的影像特征个数，W1、Wi及Wn分别表示X1、Xi及Xn的系数，z表示常数。

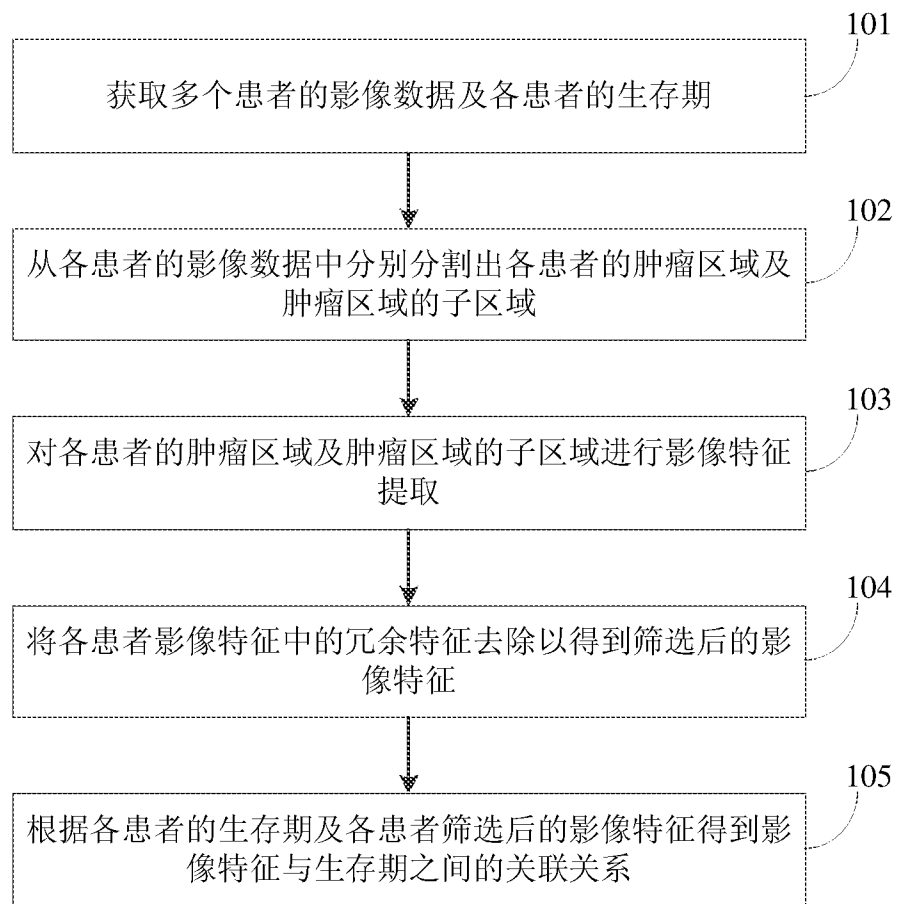


图1

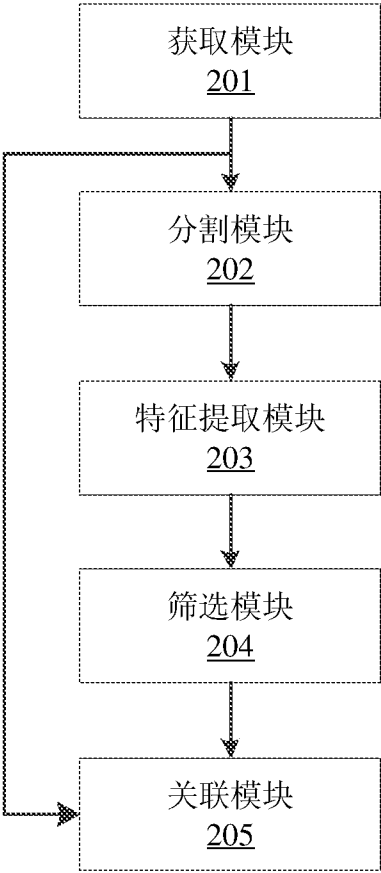


图2

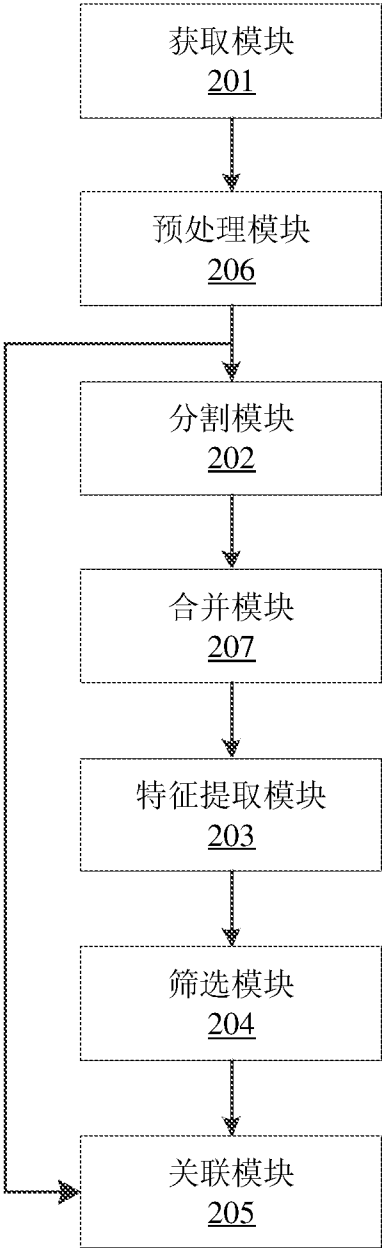


图3

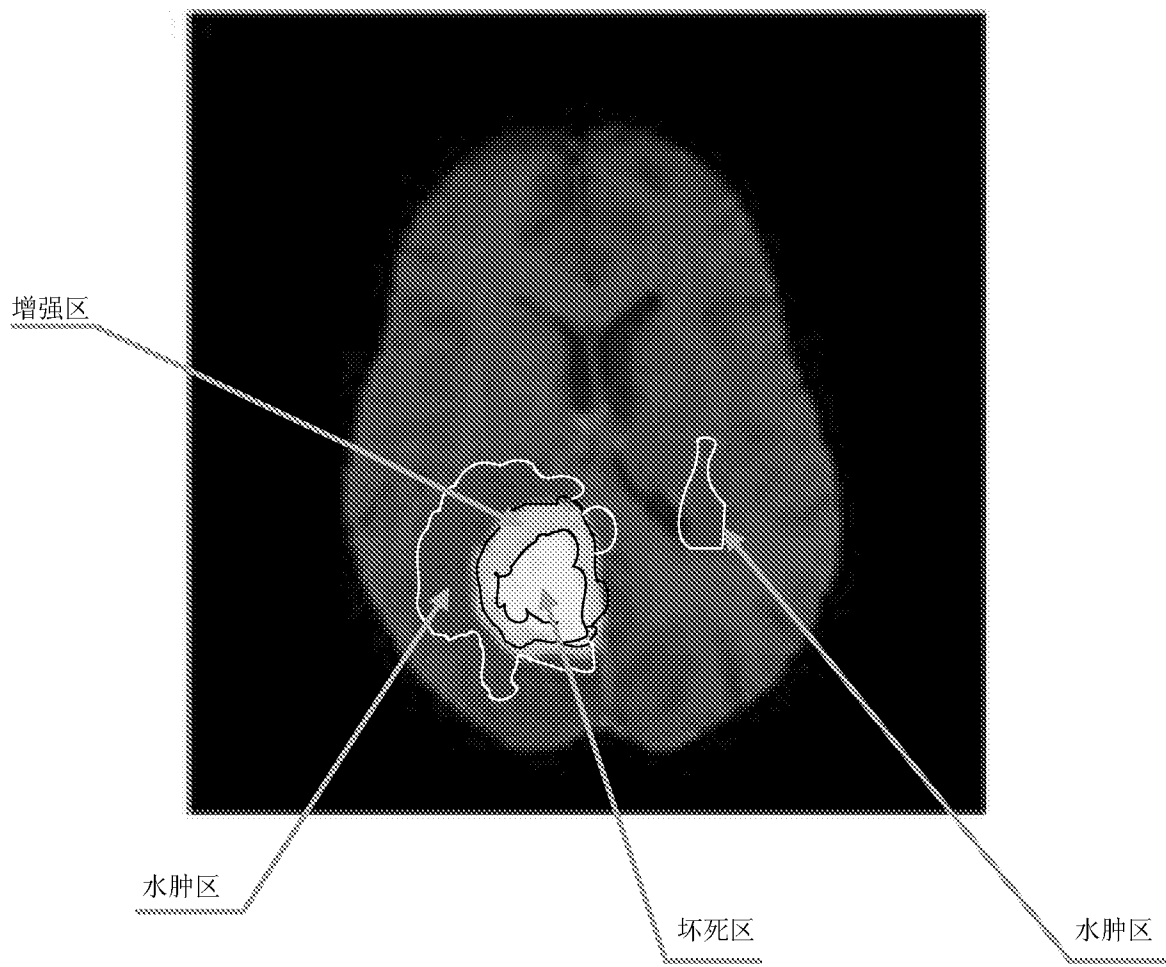


图4

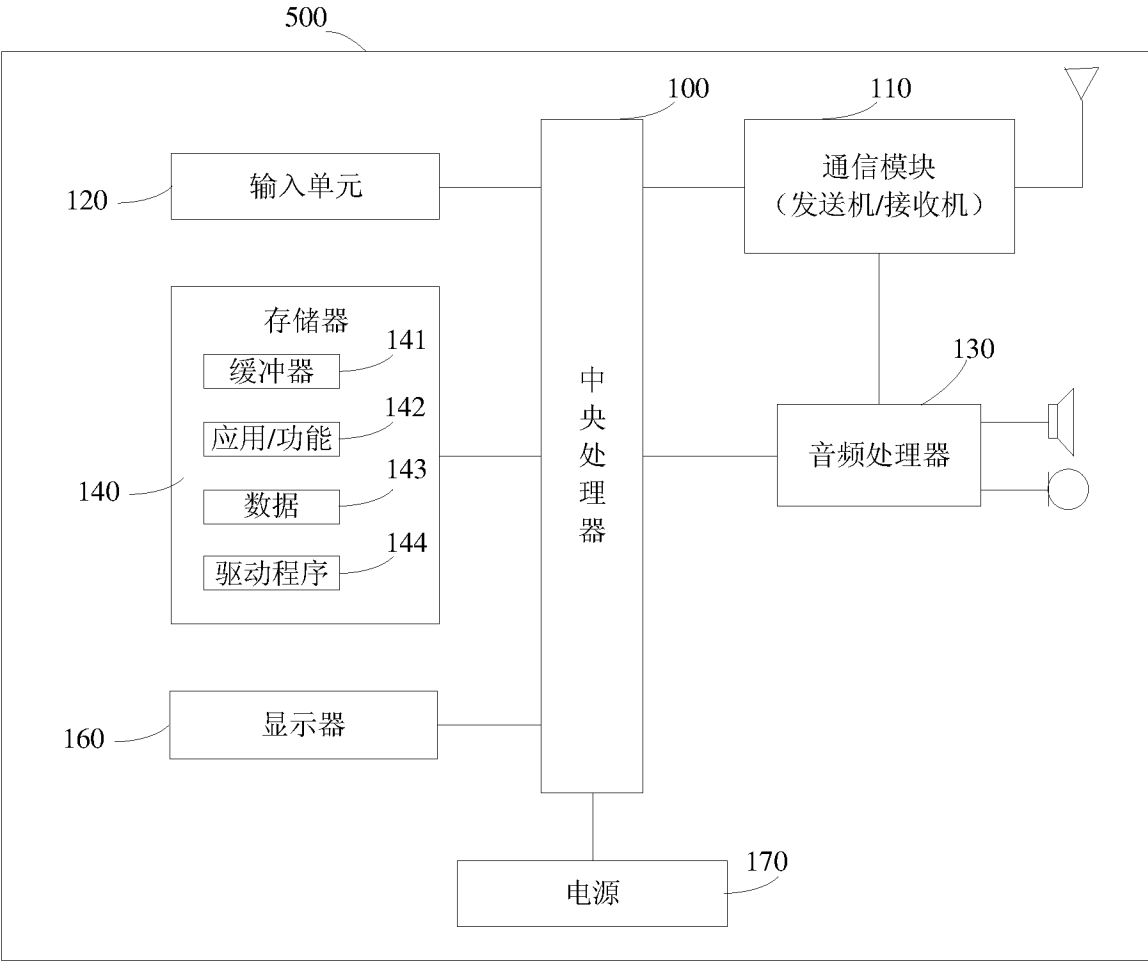


图5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/071665

A. CLASSIFICATION OF SUBJECT MATTER

G06F 19/00 (2011.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F 19/-; G06T 7/-

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CPRSABS, DWPI, CNKI: 医学, 影像, 特征, 提取, 分割, 生存期, medical, image, characteristic, extraction, divide, lifetime

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 105653858 A (INSTITUTE OF AUTOMATION, CHINESE ACADEMY OF SCIENCES), 08 June 2016 (08.06.2016), claims 1-10	1-10
A	CN 105005714 A (INSTITUTE OF AUTOMATION, CHINESE ACADEMY OF SCIENCES), 28 October 2015 (28.10.2015), entire document	1-10
A	CN 104252570 A (SHANGHAI UNITED IMAGING HEALTHCARE CO., LTD.), 31 December 2014 (31.12.2014), entire document	1-10

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
“A” document defining the general state of the art which is not considered to be of particular relevance	“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
“E” earlier application or patent but published on or after the international filing date	“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	“&” document member of the same patent family
“O” document referring to an oral disclosure, use, exhibition or other means	
“P” document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 11 October 2017	Date of mailing of the international search report 25 October 2017
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer MING, Mei Telephone No. (86-10) 62411851

International application No.
PCT/CN2017/071665

Form PCT/ISA/210 (patent family annex) (July 2009)

国际检索报告

国际申请号

PCT/CN2017/071665

A. 主题的分类

G06F 19/00(2011.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F19/-;G06T7/-

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CPRSABS, DWPI, CNKI:医学, 影像, 特征, 提取, 分割, 生存期, medical, image, characteristic, extraction, divide, lifetime

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	CN 105653858 A (中国科学院自动化研究所) 2016年 6月 8日 (2016 - 06 - 08) 权利要求1-10	1-10
A	CN 105005714 A (中国科学院自动化研究所) 2015年 10月 28日 (2015 - 10 - 28) 全文	1-10
A	CN 104252570 A (上海联影医疗科技有限公司) 2014年 12月 31日 (2014 - 12 - 31) 全文	1-10

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2017年 10月 11日

国际检索报告邮寄日期

2017年 10月 25日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局(ISA/CN)
中国北京市海淀区蓟门桥西土城路6号 100088

受权官员

明媚

传真号 (86-10)62019451

电话号码 (86-10)62411851

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/071665

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	105653858	A	2016年 6月 8日	无	
CN	105005714	A	2015年 10月 28日	无	
CN	104252570	A	2014年 12月 31日	无	