



(12) 发明专利申请

(10) 申请公布号 CN 120015312 A

(43) 申请公布日 2025. 05. 16

(21) 申请号 202510071546.4

G06N 3/045 (2023.01)

(22) 申请日 2025.01.16

G06N 3/084 (2023.01)

(71) 申请人 北京科技大学

G06N 3/0985 (2023.01)

地址 100083 北京市海淀区学院路30号

G16H 30/20 (2018.01)

申请人 首都医科大学附属北京天坛医院

(72) 发明人 赵小燕 张朝晖 郑绍文 吴先毫

张天尧 杨沛 张锐 孙志延

陶锐 都明辉 袁媛 张宇

(74) 专利代理机构 北京市广友专利事务所有限

责任公司 11237

专利代理师 付忠林 张仲波

(51) Int. Cl.

G16H 50/30 (2018.01)

G01N 21/3586 (2014.01)

G06N 3/0464 (2023.01)

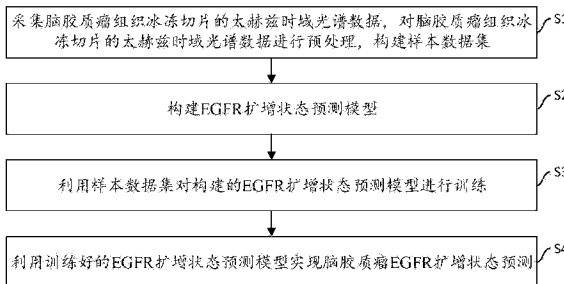
权利要求书2页 说明书9页 附图3页

(54) 发明名称

一种基于太赫兹时域光谱的脑胶质瘤EGFR
扩增状态预测方法

(57) 摘要

本发明公开了一种基于太赫兹时域光谱的脑胶质瘤EGFR扩增状态预测方法,属于机器学习技术领域,所述EGFR扩增状态预测方法包括:采集脑胶质瘤组织冰冻切片的太赫兹时域光谱数据,对脑胶质瘤组织冰冻切片的太赫兹时域光谱数据进行预处理,构建样本数据集;构建EGFR扩增状态预测模型;利用样本数据集对构建的EGFR扩增状态预测模型进行训练;利用训练好的EGFR扩增状态预测模型实现脑胶质瘤EGFR扩增状态预测。本发明利用太赫兹时域光谱结合卷积神经网络实现对脑胶质瘤EGFR扩增状态进行预测,具有无损、快速诊断的优势。有望投入实际应用中,辅助医生在术中对脑胶质瘤EGFR扩增状态进行预测。



1. 一种基于太赫兹时域光谱的脑胶质瘤EGFR扩增状态预测方法,其特征在于,所述基于太赫兹时域光谱的脑胶质瘤EGFR扩增状态预测方法包括:

采集脑胶质瘤组织冰冻切片的太赫兹时域光谱数据,对脑胶质瘤组织冰冻切片的太赫兹时域光谱数据进行预处理,构建样本数据集;

构建EGFR扩增状态预测模型;

利用样本数据集对构建的EGFR扩增状态预测模型进行训练;

利用训练好的EGFR扩增状态预测模型实现脑胶质瘤EGFR扩增状态预测。

2. 如权利要求1所述的基于太赫兹时域光谱的脑胶质瘤EGFR扩增状态预测方法,其特征在于,采集脑胶质瘤组织冰冻切片的太赫兹时域光谱数据,包括:

将手术中切除的脑胶质瘤组织置于冰盒中预冷的磷酸盐缓冲液中;

将脑胶质瘤组织从磷酸盐缓冲液中取出并吸干表面水分后,将其置于盛有冰冻组织包埋剂的冰冻包埋盒中,并将冰冻包埋盒置于 -80°C 冰箱中,待脑胶质瘤组织彻底凝固后利用冰冻组织切片机将脑胶质瘤组织切成100微米厚的切片;

将脑胶质瘤组织的切片贴附于聚乙烯PE片上,作为样本;

使用透射式太赫兹时域光谱系统照射样本,获取其太赫兹时域光谱数据。

3. 如权利要求2所述的基于太赫兹时域光谱的脑胶质瘤EGFR扩增状态预测方法,其特征在于,在使用透射式太赫兹时域光谱系统照射样本时,将透射式太赫兹时域光谱系统的太赫兹光路整体置于充满氮气的有机玻璃箱中,在太赫兹时域光谱数据采集过程中,相对湿度保持在5%以下,温度保持在 $20^{\circ}\text{C} \pm 0.2^{\circ}\text{C}$ 。

4. 如权利要求2所述的基于太赫兹时域光谱的脑胶质瘤EGFR扩增状态预测方法,其特征在于,在使用透射式太赫兹时域光谱系统照射样本时,样本被放置于六孔样品架中,六孔样品架包括六个容纳孔,其中一个容纳孔中放置PE片,作为参考样片,另外五个容纳孔中分别放置一个样本;在采集太赫兹时域光谱数据前,将六孔样品架在充有流动干燥氮气的实验箱里放置预设时长,并保持六孔样品架按照预设速度转动,确保每个样本表面均匀干燥;开始采集太赫兹时域光谱数据时,首先旋转六孔样品架,使参考样片中心在太赫兹光线焦点处,测得参考信号,此后旋转六孔样品架,以保证每一样本依次正好位于太赫兹光线焦点处,以得到五个样本的太赫兹时域光谱数据;其中,在采集太赫兹时域光谱数据时,每个样本采集多次并取多次采集结果的均值作为样本最终的太赫兹时域光谱数据。

5. 如权利要求4所述的基于太赫兹时域光谱的脑胶质瘤EGFR扩增状态预测方法,其特征在于,对脑胶质瘤组织冰冻切片的太赫兹时域光谱数据进行预处理,构建样本数据集,包括:

利用傅里叶变换将所述参考信号和脑胶质瘤组织冰冻切片的太赫兹时域光谱数据分别转换成频域谱,基于所述参考信号的频域谱和脑胶质瘤组织冰冻切片的太赫兹时域光谱数据的频域谱,计算得到脑胶质瘤组织冰冻切片的吸收系数;

利用Savitzky-Golay滤波器对所述吸收系数进行平滑处理;其中,Savitzky-Golay滤波器的窗口大小设置为15,多项式阶数设置为3;

将平滑处理后的吸收系数转化成GASF图像和GADF图像;

将生成的GASF图像和GADF图像分别转换为灰度图;

利用脑胶质瘤组织冰冻切片对应的GASF图像和GADF图像的灰度图,以及脑胶质瘤组织

冰冻切片对应的标记结果,构建样本数据集;其中,脑胶质瘤组织冰冻切片对应的标记结果为对应的脑胶质瘤组织冰冻切片是否存在EGFR扩增。

6.如权利要求5所述的基于太赫兹时域光谱的脑胶质瘤EGFR扩增状态预测方法,其特征在于,在利用傅里叶变换将所述参考信号和脑胶质瘤组织冰冻切片的太赫兹时域光谱数据分别转换成频域谱时,针对待转换的数据,以其第一次回波出现前的11ps处作为截取点,保留截取点前面的数据,对剩余数据进行补零操作至2048点后执行基2快速傅里叶变换,将数据变换至频域。

7.如权利要求5所述的基于太赫兹时域光谱的脑胶质瘤EGFR扩增状态预测方法,其特征在于,所述EGFR扩增状态预测模型包括两个结构相同的卷积神经网络,用于对输入图像进行特征提取,其中一个卷积神经网络的输入为GASF图像的灰度图,另外一个卷积神经网络的输入为GADF图像的灰度图;两个卷积神经网络提取的特征被输入中端融合层,中端融合层将两个卷积神经网络提取的特征相互沿一个维度拼接成一个一维特征,得到融合特征;中端融合层后连接有三个全连接层,由最后的全连接层输出预测结果,也即是否存在EGFR扩增。

8.如权利要求7所述的基于太赫兹时域光谱的脑胶质瘤EGFR扩增状态预测方法,其特征在于,卷积神经网络由两个卷积核大小为 5×5 的卷积层和两个 2×2 大小的最大池化层构成;三个全连接层的神经元个数依次为512、256、256。

9.如权利要求8所述的基于太赫兹时域光谱的脑胶质瘤EGFR扩增状态预测方法,其特征在于,卷积神经网络和全连接层后面添加了RELU函数,并且二分类输出采用Sigmoid函数,选用Adam优化器编译模型,并使用二元交叉熵作为损失函数来计算模型优化过程中的损失值。

10.如权利要求9所述的基于太赫兹时域光谱的脑胶质瘤EGFR扩增状态预测方法,其特征在于,在对构建的EGFR扩增状态预测模型进行训练时,采用五折交叉验证训练模型,训练轮次设置为100;

使用受试者工作特征曲线下面积值作为模型评估的性能指标。

一种基于太赫兹时域光谱的脑胶质瘤EGFR扩增状态预测方法

技术领域

[0001] 本发明涉及机器学习技术领域,特别是指一种基于太赫兹时域光谱的脑胶质瘤EGFR扩增状态预测方法。

背景技术

[0002] 脑胶质瘤是成人最常见的原发于中枢神经系统的恶性肿瘤,其中胶质母细胞瘤恶性程度最高,危害性最大。表皮生长因子受体EGFR由一个酪氨酸激酶受体家族组成,直接调节胶质瘤血管生成,而血管生成在胶质瘤的生长、侵袭和复发中起着重要作用。随着越来越多的研究表明,EGFR扩增和突变已被确定为多种癌症的驱动事件,特别是非小细胞肺癌,乳腺癌和胶质母细胞瘤。

[0003] 目前,主要采用分子生物学技术对EGFR扩增状态进行检测。采取的方法主要有荧光原位杂交、显色原位杂交、实时定量PCR和下一代测序。这些方法可以准确诊断EGFR扩增状态,但获取结果的等待时间较长。

发明内容

[0004] 本发明提供了一种基于太赫兹时域光谱的脑胶质瘤EGFR扩增状态预测方法,以解决目前EGFR扩增状态检测时间较长的技术问题。

[0005] 为解决上述技术问题,本发明提供了如下技术方案:

[0006] 一方面,本发明提供了一种基于太赫兹时域光谱的脑胶质瘤EGFR扩增状态预测方法,所述基于太赫兹时域光谱的脑胶质瘤EGFR扩增状态预测方法包括:

[0007] 采集脑胶质瘤组织冰冻切片的太赫兹时域光谱数据,对脑胶质瘤组织冰冻切片的太赫兹时域光谱数据进行预处理,构建样本数据集;

[0008] 构建EGFR扩增状态预测模型;

[0009] 利用样本数据集对构建的EGFR扩增状态预测模型进行训练;

[0010] 利用训练好的EGFR扩增状态预测模型实现脑胶质瘤EGFR扩增状态预测。

[0011] 进一步地,采集脑胶质瘤组织冰冻切片的太赫兹时域光谱数据,包括:

[0012] 将手术中切除的脑胶质瘤组织置于冰盒中预冷的磷酸盐缓冲液中;

[0013] 将脑胶质瘤组织从磷酸盐缓冲液中取出并吸干表面水分后,将其置于盛有冰冻组织包埋剂的冰冻包埋盒中,并将冰冻包埋盒置于-80℃冰箱中,待脑胶质瘤组织彻底凝固后利用冰冻组织切片机将脑胶质瘤组织切成100微米厚的切片;

[0014] 将脑胶质瘤组织的切片贴附于聚乙烯PE片上,作为样本;

[0015] 使用透射式太赫兹时域光谱系统照射样本,获取其太赫兹时域光谱数据。

[0016] 进一步地,在使用透射式太赫兹时域光谱系统照射样本时,将透射式太赫兹时域光谱系统的太赫兹光路整体置于充满氮气的有机玻璃箱中,在太赫兹时域光谱数据采集过程中,相对湿度保持在5%以下,温度保持在 $20^{\circ}\text{C} \pm 0.2^{\circ}\text{C}$ 。

[0017] 进一步地,在使用透射式太赫兹时域光谱系统照射样本时,样本被放置于六孔样

品架中,六孔样品架包括六个容纳孔,其中一个容纳孔中放置PE片,作为参考样片,另外五个容纳孔中分别放置一个样本;在采集太赫兹时域光谱数据前,将六孔样品架在充有流动干燥氮气的实验箱里放置预设时长,并保持六孔样品架按照预设速度转动,确保每个样本表面均匀干燥;开始采集太赫兹时域光谱数据时,首先旋转六孔样品架,使参考样片中心在太赫兹光线焦点处,测得参考信号,此后旋转六孔样品架,以保证每一样本依次正好位于太赫兹光线焦点处,以得到五个样本的太赫兹时域光谱数据;其中,在采集太赫兹时域光谱数据时,每个样本采集多次并取多次采集结果的均值作为样本最终的太赫兹时域光谱数据。

[0018] 进一步地,对脑胶质瘤组织冰冻切片的太赫兹时域光谱数据进行预处理,构建样本数据集,包括:

[0019] 利用傅里叶变换将所述参考信号和脑胶质瘤组织冰冻切片的太赫兹时域光谱数据分别转换成频域谱,基于所述参考信号的频域谱和脑胶质瘤组织冰冻切片的太赫兹时域光谱数据的频域谱,计算得到脑胶质瘤组织冰冻切片的吸收系数;

[0020] 利用Savitzky-Golay滤波器对所述吸收系数进行平滑处理;其中,Savitzky-Golay滤波器的窗口大小设置为15,多项式阶数设置为3;

[0021] 将平滑处理后的吸收系数转化成GASF图像和GADF图像;

[0022] 将生成的GASF图像和GADF图像分别转换为灰度图;

[0023] 利用脑胶质瘤组织冰冻切片对应的GASF图像和GADF图像的灰度图,以及脑胶质瘤组织冰冻切片对应的标记结果,构建样本数据集;其中,脑胶质瘤组织冰冻切片对应的标记结果为对应的脑胶质瘤组织冰冻切片是否存在EGFR扩增。

[0024] 进一步地,在利用傅里叶变换将所述参考信号和脑胶质瘤组织冰冻切片的太赫兹时域光谱数据分别转换成频域谱时,针对待转换的数据,以其第一次回波出现前的11ps处作为截取点,保留截取点前面的数据,对剩余数据进行补零操作至2048点后执行基2快速傅里叶变换,将数据变换至频域。

[0025] 进一步地,所述EGFR扩增状态预测模型包括两个结构相同的卷积神经网络,用于对输入图像进行特征提取,其中一个卷积神经网络的输入为GASF图像的灰度图,另外一个卷积神经网络的输入为GADF图像的灰度图;两个卷积神经网络提取的特征被输入中端融合层,中端融合层将两个卷积神经网络提取的特征相互沿一个维度拼接成一个一维特征,得到融合特征;中端融合层后连接有三个全连接层,由最后的全连接层输出预测结果,也即是否存在EGFR扩增。

[0026] 进一步地,卷积神经网络由两个卷积核大小为 5×5 的卷积层和两个 2×2 大小的最大池化层构成;三个全连接层的神经元个数依次为512、256、256。

[0027] 进一步地,卷积神经网络和全连接层后面添加了RELU函数,并且二分类输出采用Sigmoid函数,选用Adam优化器编译模型,并使用二元交叉熵作为损失函数来计算模型优化过程中的损失值。

[0028] 进一步地,在对构建的EGFR扩增状态预测模型进行训练时,采用五折交叉验证训练模型,训练轮次设置为100;

[0029] 使用受试者工作特征曲线下面积值作为模型评估的性能指标。

[0030] 再一方面,本发明还提供了一种电子设备,其包括处理器和存储器;其中,存储器中存储有至少一条指令,所述指令由处理器加载并执行,以实现上述方法。

[0031] 又一方面,本发明还提供了一种计算机可读存储介质,所述存储介质中存储有至少一条指令,所述指令由处理器加载并执行,以实现上述方法。

[0032] 本发明提供的技术方案带来的有益效果至少包括:

[0033] 1、本发明利用太赫兹时域光谱结合卷积神经网络实现对脑胶质瘤EGFR扩增状态进行预测,具有无损、快速诊断的优势。

[0034] 2、本发明首次将吸收系数转化为GASF和GADF的二维图像做数据集来预测脑胶质瘤EGFR扩增状态,可有效提高EGFR扩增状态预测准确率。

[0035] 3、相比于常规的机器学习方法,本发明具有较高的分类精度和泛化能力,有望投入实际应用中,辅助医生在术中对脑胶质瘤EGFR扩增状态进行预测。

附图说明

[0036] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例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发明的一些实施例,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。

[0037] 图1是本发明实施例提供的基于太赫兹时域光谱的脑胶质瘤EGFR扩增状态预测方法的执行流程示意图;

[0038] 图2是本发明实施例提供的太赫兹时域光谱仪光路示意图;

[0039] 图3是本发明实施例提供的时域谱曲线图;

[0040] 图4是本发明实施例提供的样品数据处理示意图;

[0041] 图5是本发明实施例提供的EGFR扩增状态预测模型结构图;

[0042] 图6是本发明实施例提供的电子设备的系统框图。

具体实施方式

[0043] 为使本发明的目的、技术方案和优点更加清楚,下面将结合附图对本发明实施方式作进一步地详细描述。

[0044] 首先,需要说明的是,在本发明实施例中,“示例地”、“例如”等词用于表示作例子、例证或说明。本发明中被描述为“示例”的任何实施例或设计方案不应被解释为比其它实施例或设计方案更优选或更具优势。确切而言,使用“示例地”一词旨在以具体方式呈现概念。此外,在本发明实施例中,“和/或”所表达的含义可以是两者都有,或者可以是两者任选其一。

[0045] 第一实施例

[0046] 本实施例提供了一种基于太赫兹时域光谱的脑胶质瘤EGFR扩增状态预测方法,该方法可以由电子设备实现,该方法的执行流程如图1所示,包括以下步骤:

[0047] S1,采集脑胶质瘤组织冰冻切片的太赫兹时域光谱数据,对脑胶质瘤组织冰冻切片的太赫兹时域光谱数据进行预处理,构建样本数据集;

[0048] 具体地,在本实施例中,上述S1的实现过程如下:

[0049] S11,将手术切除的肿瘤及瘤周组织立刻置于冰盒中预冷的磷酸盐缓冲液中,并带回实验室进行冰冻切片处理;

[0050] S12,将标本从磷酸盐缓冲液中取出并吸干表面水分后,将其置于盛有冰冻组织包埋剂的冰冻包埋盒中,并将其置于-80℃冰箱中,待其彻底凝固后利用冰冻组织切片机将组织切成100微米厚的薄片,然后用定制厚度为2mm的聚乙烯PE片作为载体,将冰冻薄片均匀贴附于PE片上,用于太赫兹测量实验。其中,所有切片完成后立即置于-80℃低温保存,直至太赫兹测量实验开始。

[0051] S13,使用透射式太赫兹时域光谱系统(Terahertz time-domain spectroscopy, THz-TDS)照射切片获取样品信号;

[0052] 其中,需要说明的是,透射式太赫兹时域光谱系统的光路如图2所示,其太赫兹波的产生和探测使用了基于低温砷化镓的光电导天线装置,飞秒激光器(Toptica Photonics Femtofiber Smart 780,中心波长780nm,脉冲宽度<90fs)输出的红外光经分束后分别作为泵浦光和探测光照射在光电导天线中心,强度均为15mW,泵浦光路中增加了步进电机延迟线用于实现太赫兹脉冲的波形扫描。太赫兹脉冲通过四个离轴抛物面镜实现波束转换,在样品区实现聚焦,束腰直径为5mm。时域信号的扫描步长为33.3fs,共采集1024点,在太赫兹波传播路径中存在生物样品时,选用可用频带为0.2-1.4THz。

[0053] 此外,还需要说明的是,太赫兹波传播路径中的水蒸气对信号具有很强的吸收,为排除湿度变化引起的表征影响,将太赫兹光路整体置于充满氮气的有机玻璃箱中,测试过程中相对湿度保持在5%以下,温度保持在20℃±0.2℃。为提高测试效率并使同一对照组内的测试环境相同,本实验采用自制的六孔样品架,分别放置1个参考样片(纯PE片)和5个实验样片(带有组织样片的PE片)。实验时把样片依次安装放置到多孔旋转样品架上,测量前为了除去从冰冻环境转移到室温环境后样片表面产生的冷凝水,需要在充有流动干燥氮气的实验箱里放置一段时间,过程中保持样品架缓慢转动,确保每个样品表面均匀干燥;开始测量时首先旋转调整样品位置使作为参考信号的纯PE片中心在THz光线焦点处,测得参考信号,此后样品架每次旋转60度以保证后续被测样品正好位于聚焦光处,这样依次得到后续5组样品信号,每次测量数据采集两次取平均,以减小误差。

[0054] S14,利用傅里叶变换将参考信号和脑胶质瘤组织冰冻切片的太赫兹时域光谱数据分别转换成频域谱,基于参考信号的频域谱和脑胶质瘤组织冰冻切片的太赫兹时域光谱数据的频域谱,计算得到脑胶质瘤组织冰冻切片的吸收系数;

[0055] 其中,需要说明的是,把采集的原始数据绘制时域谱,会得到一条有回波的曲线,此时直接进行傅里叶变换将会出现由样品内部多重反射引起的量器具效应,为了消除频谱信号中的振荡,如图3所示,本实施例在第一次回波出现前的某一点(11ps处)作为截取点,保留截取点前面的数据,对剩余时域数据进行补零操作至2048(2¹¹)点后执行基2快速傅里叶变换至频域。如图4所示,参考信号频谱记为 $E_{ref}(\omega)$,样品信号记为 $E_{sam}(\omega)$,并利用两者的比值求得样品内部的传递函数 $T(\omega)$ 。由于前面对数据进行了截取,预处理之后只保留了主峰信号,计算太赫兹参数时就不用考虑多重反射,于是再通过计算即可得到样品折射率 $n_s(\omega)$ 、消光系数 $\kappa_s(\omega)$ 、和吸收系数 $\alpha_s(\omega)$,具体计算公式如下:

$$n_s(\omega) = 1 - \angle T(\omega) \cdot \frac{c}{\omega d}$$

[0056]

$$\kappa_s(\omega) = -\frac{c}{\omega d} \cdot \ln \left\{ |T(\omega)| \cdot \frac{[1 + n_s(\omega)] \cdot [n_{PE} + n_s(\omega)]}{2n_s(\omega) \cdot (1 + n_{PE})} \right\}$$

[0057] 样品的吸收系数可由消光系数计算：

$$[0058] \quad \alpha_s(\omega) = \frac{2\omega\kappa_s(\omega)}{c} = -\frac{2}{d} \cdot \ln \left\{ |T(\omega)| \cdot \frac{[1+n_s(\omega)] \cdot [n_{PE} + n_s(\omega)]}{2n_s(\omega) \cdot (1+n_{PE})} \right\}$$

[0059] 其中,d为样品厚度,T(ω)为样品传递函数,c为真空中光速,n_{PE}为PE的折射率,实验测得n_{PE}=1.50。

[0060] S15,利用Savitzky-Golay滤波器对吸收系数进行平滑处理;其中,Savitzky-Golay滤波器的窗口大小设置为15,多项式阶数设置为3;

[0061] S16,将平滑处理后的吸收系数转化成格拉姆角和场(GASF)图像和格拉姆角差场(GADF)的二维图像,具体如下:

$$[0062] \quad GASF = \begin{bmatrix} \cos(\phi_1 + \phi_1) & \cdots & \cos(\phi_1 + \phi_n) \\ \cos(\phi_2 + \phi_1) & \cdots & \cos(\phi_2 + \phi_n) \\ \vdots & \ddots & \vdots \\ \cos(\phi_n + \phi_1) & \cdots & \cos(\phi_n + \phi_n) \end{bmatrix}$$

$$[0063] \quad GADF = \begin{bmatrix} \sin(\phi_1 + \phi_1) & \cdots & \sin(\phi_1 + \phi_n) \\ \sin(\phi_2 + \phi_1) & \cdots & \sin(\phi_2 + \phi_n) \\ \vdots & \ddots & \vdots \\ \sin(\phi_n + \phi_1) & \cdots & \sin(\phi_n + \phi_n) \end{bmatrix}$$

[0064] 其中,GASF为角度和的余弦,GADF为角度差的余弦,n为频率点数,φ_n为第n个频率点处的角度值。

[0065] S17,将GASF图像和GADF图像分别转换成分辨率大小为256×256的灰度图,所以单个图像的大小为256×256×1(像素×像素×通道);

[0066] S18,利用脑胶质瘤组织冰冻切片对应的GASF图像和GADF图像的灰度图,以及脑胶质瘤组织冰冻切片对应的标记结果,构建样本数据集;其中,冰冻切片对应的标记结果为对应的脑胶质瘤组织冰冻切片是否存在EGFR扩增。

[0067] S2,构建EGFR扩增状态预测模型;

[0068] 其中,本实施例搭建的EGFR扩增状态预测模型结构如图5所示,该网络主要由两个结构相同的卷积神经网络在特征提取后进行融合和全连接。其中卷积神经网络模型中主要由两个卷积核大小为5×5的卷积层和两个2×2大小的最大池化层构成,中端融合则是将提取出来的特征相互沿一个维度拼接成一个一维特征,模型的最后则是构建了三个全连接层全连接层的神经元个数依次为512、256、256。此外,在卷积层和全连接层后面添加了RELU函数,并且二分类输出采用Sigmoid函数,选用Adam优化器编译模型,并使用二元交叉熵作为损失函数来作为模型优化过程中的损失值,采用五折交叉验证训练模型,训练轮次设置为100。

[0069] S3,利用样本数据集对构建的EGFR扩增状态预测模型进行训练;

[0070] 其中,本实施例所采集的样本数量共有440例,其中120例为EGFR扩增,剩下的320例为EGFR非扩增。通过Savitzky-Golay滤波器平滑处理以及利用GASF和GADF生成对应的二维图像后,共有440组图片数据,每组里面有对应的GASF和GADF图像,将这些数据按4:1的比

例划分为训练集和测试集,其中训练集采用五折交叉验证对模型进行训练和验证。

[0071] 进一步地,本实施例使用受试者工作特征曲线下面积 (Area Under the Receiver Operating Characteristic Curve, ROC AUC) 值作为模型评估的主要性能指标,ROC曲线是一种用于表示分类模型性能的图形工具。它绘制了不同阈值设置下的真阳性率 (True Positive Rate, TPR) 与假阳性率 (False Positive Rate, FPR) 的关系图。AUC代表ROC曲线下的面积,用于衡量分类器的性能。AUC值越接近1表示分类器性能越好,而AUC值接近0表示性能较差。

[0072] 经实验,本实施例的模型最终在测试集上的预测AUC值为94.74%,由此证明本实施例的模型具有较高的分类精度,其有望投入实际应用中,辅助医生在术中对脑胶质瘤EGFR扩增状态进行预测。

[0073] S4,利用训练好的EGFR扩增状态预测模型实现脑胶质瘤EGFR扩增状态预测。

[0074] 综上,本实施例提供了一种基于太赫兹时域光谱的脑胶质瘤EGFR扩增状态预测方法,利用太赫兹时域光谱结合卷积神经网络实现对脑胶质瘤EGFR扩增状态进行预测,具有无损、快速诊断的优势。且相比于常规的机器学习方法,本方法首次将吸收系数转化为GASF和GADF的二维图像做数据集来预测脑胶质瘤EGFR扩增状态,可有效提高EGFR扩增状态预测准确率。具有较高的分类精度和泛化能力,可用于辅助医生在术中对脑胶质瘤EGFR扩增状态进行预测。

[0075] 第二实施例

[0076] 本实施例提供一种电子设备,如图6所示,所述电子设备包括:处理器和存储器;其中,所述处理器与所述存储器可以通过通信总线连接;所述存储器中存储有至少一条指令,所述指令由所述处理器加载并执行,以实现上述第一实施例的方法。此外,所述电子设备还可以包括收发器,所述处理器与所述收发器可以通过通信总线连接,所述收发器用于与其它设备进行通信。

[0077] 下面,结合图6对该电子设备的各个构成部件进行具体的介绍:

[0078] 其中,处理器是电子设备的控制中心,所述电子设备可以包括多个处理器,这些处理器中的每一个可以是一个单核处理器(single-CPU),也可以是一个多核处理器(multi-CPU)。这里的处理器可以是一个处理器,也可以是多个处理元件的统称。例如,处理器是一个或多个中央处理器(central processing unit,CPU),也可以是其他通用处理器、特定集成电路(application specific integrated circuit,ASIC),或者是被配置成实施本发明实施例的一个或多个集成电路,例如:一个或多个微处理器(digital signal processor,DSP),或,一个或者多个现场可编程门阵列(field programmable gate array,FPGA),或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者也可以是任何常规的处理器等。处理器可以通过运行或执行存储在存储器内的软件程序,以及调用存储在存储器内的数据,执行所述电子设备的各种功能。

[0079] 在具体的实现中,作为一种实施例,所述处理器可以包括一个或多个CPU,例如图6中所示出的CPU0和CPU1,当然,此仅为示例性说明。

[0080] 所述存储器用于存储执行本发明方案的软件程序,并由所述处理器来控制执行,具体实现方式可以参考上述方法实施例,此处不再赘述。

[0081] 可选地,所述存储器可以是只读存储器(read-only memory,ROM)或可存储静态信

息和指令的其他类型的静态存储设备,随机存取存储器(random access memory, RAM)或者可存储信息和指令的其他类型的动态存储设备,也可以是电可擦可编程只读存储器(electrically erasable programmable read-only memory, EEPROM)、只读光盘(compact disc read-only memory, CD-ROM)或其他光盘存储、光碟存储(包括压缩光碟、激光碟、光碟、数字通用光碟、蓝光光碟等)、磁盘存储介质或者其他磁存储设备、或者能够用于携带或存储具有指令或数据结构形式的期望的程序代码并能够由计算机存取的任何其他介质,但不限于此。存储器可以和处理器集成在一起,也可以独立存在,并通过所述电子设备的接口电路(图6中未示出)与处理器耦合,本发明实施例对此不作具体限定。

[0082] 所述收发器可以包括接收器和发送器(图6中未单独示出)。其中,所述接收器用于实现接收功能,所述发送器用于实现发送功能。所述收发器可以和所述处理器集成在一起,也可以独立存在,并通过所述电子设备的接口电路(图6中未示出)与所述处理器耦合,本发明实施例对此不作具体限定。

[0083] 此外,需要说明的是,图6中示出的所述电子设备的结构并不构成对该设备的限定,实际的设备可以包括比图示更多或更少的部件,或者组合某些部件,或者不同的部件布置。此外,该电子设备在执行上述第一实施例的方法时所实现的技术效果可以参考上述第一实施例所述的技术效果,故,此处不再赘述。

[0084] 第三实施例

[0085] 本实施例提供一种计算机可读存储介质,该存储介质中存储有至少一条指令,所述指令由处理器加载并执行,以实现上述第一实施例的方法。其中,该计算机可读存储介质可以是ROM、随机存取存储器、CD-ROM、磁带、软盘和光数据存储设备等。其内存储的指令可由终端中的处理器加载并执行上述方法。

[0086] 此外,需要说明的是,本发明可提供为方法、装置或计算机程序产品。因此,本发明实施例可采用完全或部分硬件实施例、完全或部分软件实施例或结合软件和硬件方面的实施例的形式。而且,当使用软件实现时,本发明实施例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上实施的计算机程序产品的形式。所述计算机程序产品包括一个或多个计算机指令或计算机程序。在计算机上加载或执行所述计算机指令或计算机程序时,全部或部分地产生按照本发明实施例所述的流程或功能。所述计算机可以为通用计算机、专用计算机、计算机网络、或者其他可编程装置。所述计算机指令可以存储在计算机可读存储介质中,或者从一个计算机可读存储介质向另一个计算机可读存储介质传输,例如,所述计算机指令可以从一个网站站点、计算机、服务器或数据中心通过有线(例如红外、无线、微波等)方式向另一个网站站点、计算机、服务器或数据中心进行传输。所述计算机可读存储介质可以是计算机能够存取的任何可用介质或者是包含一个或多个可用介质集合的服务器、数据中心等数据存储设备。所述可用介质可以是磁性介质(例如,软盘、硬盘、磁带)、光介质(例如,DVD)、或者半导体介质。半导体介质可以是固态硬盘。

[0087] 本发明实施例是参照根据本发明实施例的方法、终端设备(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、嵌入式处理机或其他可编程数据处理终端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终端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。

[0088] 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终端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终端设备上,使得在计算机或其他可编程终端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终端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。

[0089] 还需要说明的是,在本文中,诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来,而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者终端设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者终端设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括所述要素的过程、方法、物品或者终端设备中还存在另外的相同要素。此外,术语“和/或”,仅仅是一种描述关联对象的关联关系,表示可以存在三种关系,例如,A和/或B,可以表示:单独存在A,同时存在A和B,单独存在B这三种情况,其中A,B可以是单数或者复数。另外,本文中字符“/”,一般表示前后关联对象是一种“或”的关系,但也可能表示的是一种“和/或”的关系,具体可参考前后文进行理解。“至少一个”是指一个或者多个,“多个”是指两个或两个以上。“以下至少一项(个)”或类似表达,是指的这些项中的任意组合,包括单项(个)或复数项(个)的任意组合。例如,a,b或c中的至少一项(个),可以表示:a,b,c,a-b,a-c,b-c,或a-b-c,其中,a,b,c可以是单个,也可以是多个。

[0090] 此外,可以理解的是,在本发明的各种实施例中,上述各过程的序号的大小并不意味着执行顺序的先后,各过程的执行顺序应以其功能和内在逻辑确定,而不应对本发明实施例的实施过程构成任何限定。

[0091] 本领域普通技术人员可以意识到,结合本文中所公开的实施例描述的各示例的单元及算法步骤,能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行,取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本发明的范围。

[0092] 在本发明所提供的几个实施例中,应该理解到,所揭露的设备、装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,功能模块/单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个设备,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口,装置或单元的间接耦合或通信连接,可以是电性,机械或其它的形式。作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分

或者全部单元来实现本实施例方案的目的。另外,在本发明各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。

[0093] 所述方法如果以软件功能单元的形式实现并作为独立的产品销售或使用时,可以存储在一个计算机可读取存储介质中。基于这样的理解,本发明的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)执行本发明各个实施例所述方法的全部或部分步骤。而前述的存储介质包括U盘、移动硬盘、只读存储器(read-only memory,ROM)、随机存取存储器(random access memory,RAM)、磁碟或光盘等各种可以存储程序代码的介质。

[0094] 最后需要说明的是,以上所述仅是本发明的优选实施方式,应当指出,尽管已描述了本发明的优选实施例,但对于本技术领域普通技术人员来说,一旦得知了本发明的基本创造性概念,在不脱离本发明所述原理的前提下,还可以做出若干改进和润饰,这些改进和润饰也应视为本发明的保护范围。所以,所附权利要求意欲解释为包括优选实施例以及落入本发明实施例范围的所有变更和修改。

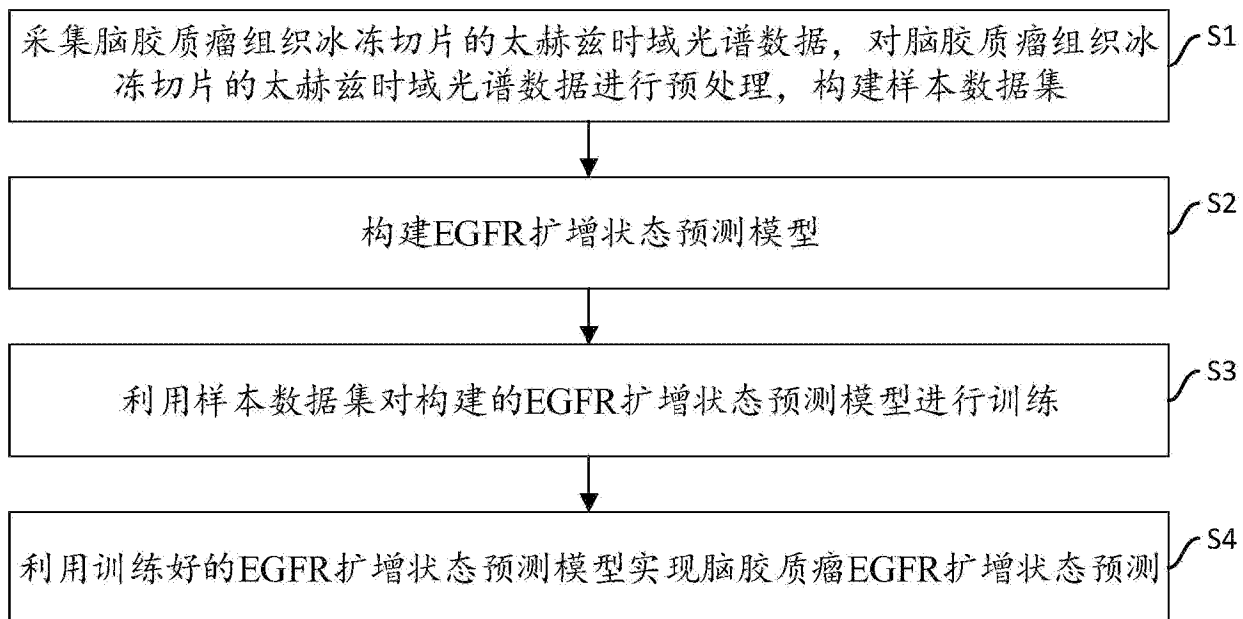


图1

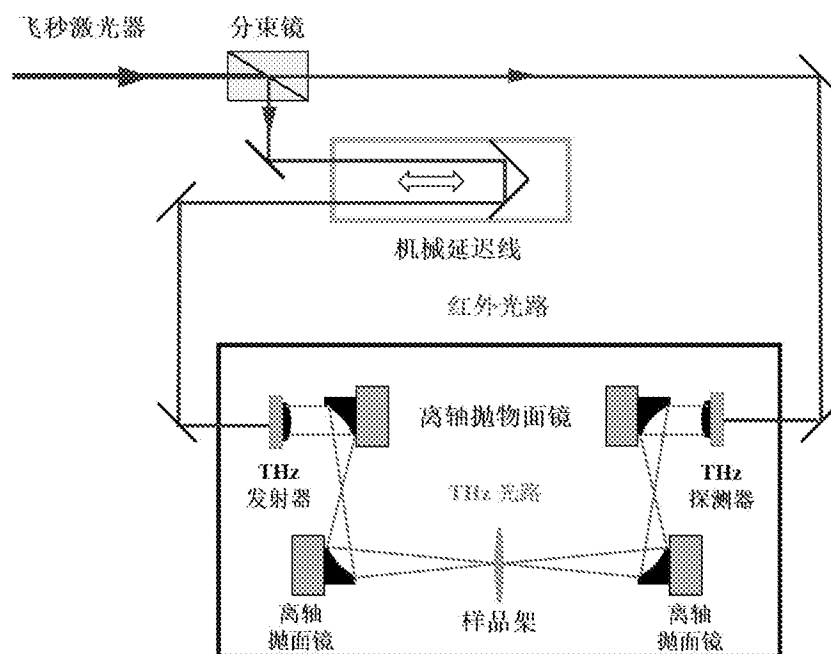


图2

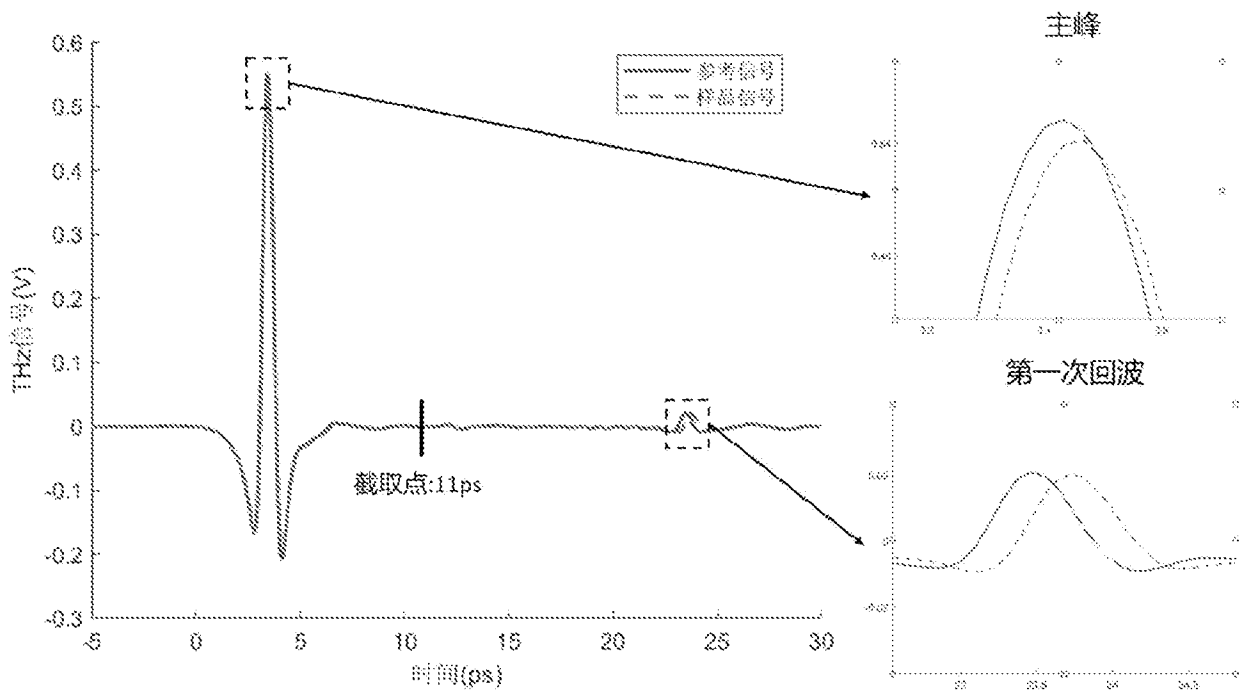


图3

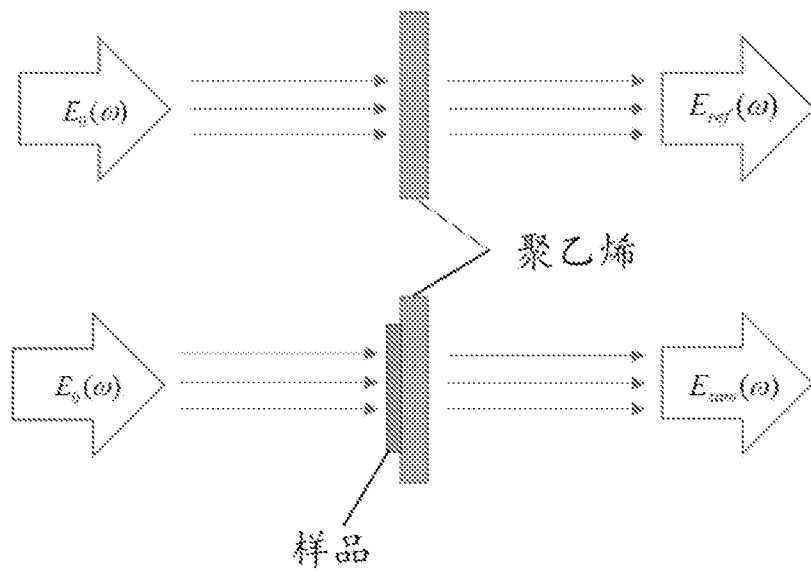


图4

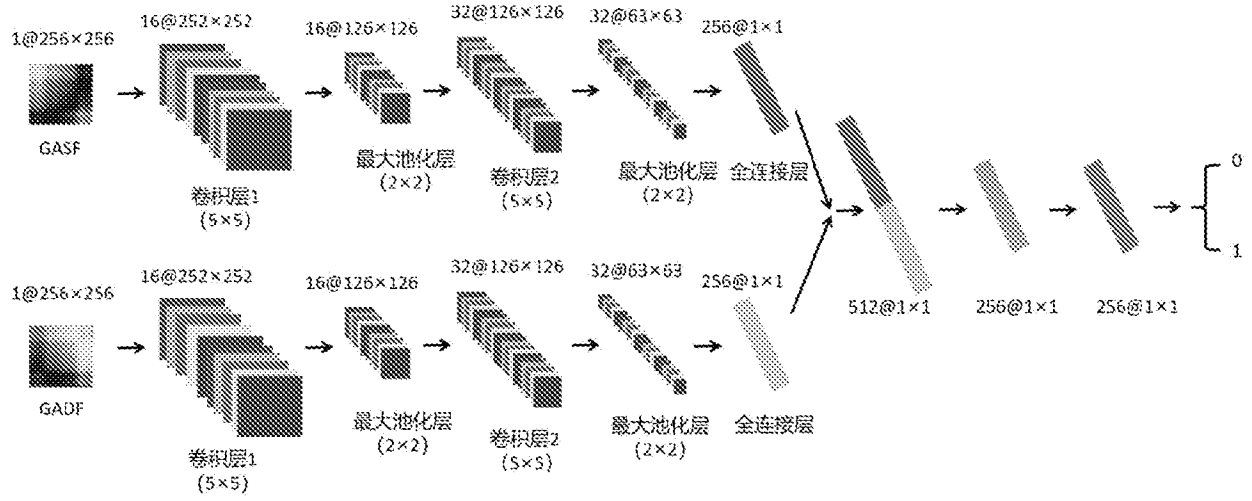


图5

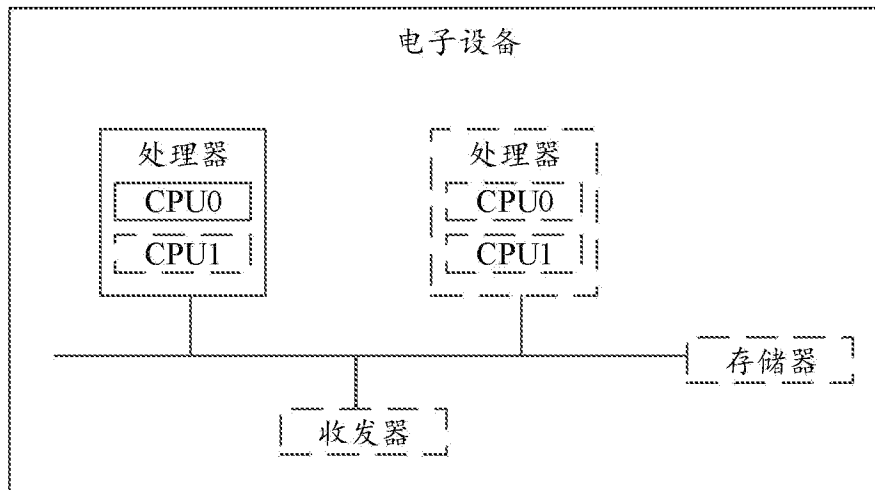


图6