



(12) 发明专利

(10) 授权公告号 CN 108335756 B

(45) 授权公告日 2022.05.24

(21) 申请号 201810047223.1

(22) 申请日 2018.01.18

(65) 同一申请的已公布的文献号
申请公布号 CN 108335756 A

(43) 申请公布日 2018.07.27

(73) 专利权人 中山大学
地址 510275 广东省广州市新港西路135号
中山大学

(72) 发明人 陈明远 蔡宏民 刘友平 陈佳洲
邹雄 游瑞

(74) 专利代理机构 广州三环专利商标代理有限公司 44202
专利代理师 宋静娜 郝传鑫

(56) 对比文件

CN 106997389 A, 2017.08.01

CN 103690240 A, 2014.04.02

CN 1313891 A, 2001.09.19

CN 105144178 A, 2015.12.09

CN 107346372 A, 2017.11.14

US 9095266 B1, 2015.08.04

CN 107491630 A, 2017.12.19

审查员 魏仁然

(51) Int. Cl.

G16H 50/70 (2018.01)

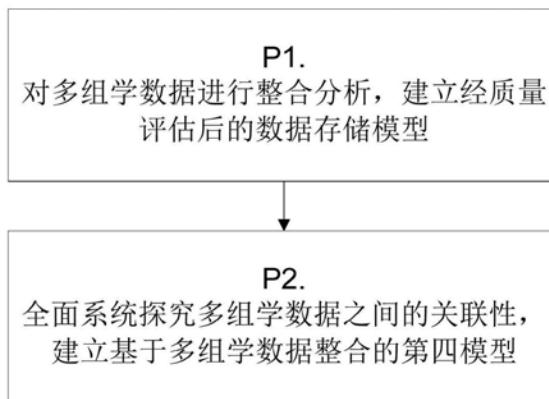
权利要求书2页 说明书11页 附图10页

(54) 发明名称

鼻咽癌数据库及基于所述数据库的综合诊疗决策方法

(57) 摘要

本发明提供了一种鼻咽癌数据库及基于所述数据库的综合诊疗决策方法,所述鼻咽癌数据库包括临床大数据库和基因数据库;所述临床大数据库中记录有由临床文字数字信息形成的第一数据以及根据临床图像信息提取并识别出的第二数据;鼻咽癌数据库中的第一数据、第二数据和/或基因数据都伴随其时间属性;所述第一数据、第二数据和基因数据均为单组学数据,所述鼻咽癌数据库构成能够存储包含第一数据、第二数据和基因数据的多组学数据库;所述鼻咽癌数据库中的数据用于支持诊疗决策服务器进行基于第一模型、第二模型、第三模型、第四模型和/或第五模型的诊疗决策。本发明能够为用户提供基于人工智能的精准的智能诊断决策服务,具备广阔的使用前景。



1. 一种综合诊疗决策方法,其特征在于,所述诊疗决策方法使用鼻咽癌数据库,所述鼻咽癌数据库作为综合诊疗决策方法中的数据来源,包括临床大数据库和基因数据库;所述临床大数据库中记录有由临床文字数字信息形成的第一数据以及根据临床图像信息提取并识别出的第二数据;所述鼻咽癌数据库中的第一数据、第二数据和/或基因数据都伴随其时间属性;所述第一数据、第二数据和基因数据均为单组学数据,所述鼻咽癌数据库构成能够存储包含第一数据、第二数据和基因数据的多组学数据库,所述方法包括:

根据所述鼻咽癌数据库中的数据构建诊疗决策模型,所述诊疗决策模型包括第一模型、第二模型、第三模型、第四模型和/或第五模型;所述第一模型以所述第一数据为输入,所述第二模型以所述第二数据为输入,所述第三模型以所述基因数据为输入,所述第四模型和所述第五模型以所述鼻咽癌数据库中多组学数据为输入;所述第四模型为基于所述多组学数据库中多组学数据的整合模型;

对多组学数据进行整合分析,建立经质量评估后的数据存储模型;

在所述数据存储模型框架下获取多组学数据之间的关联性,建立基于多组学数据整合的第四模型,所述第四模型中包括公共发现模型,所述公共发现模型基于稀疏表达理论构建;

选择诊疗决策模型;

根据所述诊疗决策模型为用户提供诊疗决策服务。

2. 根据权利要求1所述的一种综合诊疗决策方法,其特征在于:所述第四模型中包括公共发现模型,所述公共发现模型基于稀疏表达理论构建。

3. 根据权利要求2所述的一种综合诊疗决策方法,其特征在于:所述第四模型中包括公共发现模型,所述公共发现模型的构建方法包括:

在张量表达和矩阵表达的框架下,提出面向张量描述的特征模型;

建立基于先验知识约束的张量配对能量方程;

分析特征张量描述方法、先验知识约束设计、张量配对方法与局部公共子结构的内在关系;

针对高噪音污染的数据,测试模型的鲁棒性;

建立约束问题分解机制,得到公共发现模型。

4. 根据权利要求3所述的一种综合诊疗决策方法,其特征在于,所述公共发现模型的构建方法包括:

建立面向多组学数据的矩阵分解模型,将多维数据分解到同一个基空间中,分析系数矩阵之间的互动关系,得到多组学数据之间的关联关系;

建立面向多组学数据的张量分解模型,通过采用三个隐式变量基空间的方法,将多组学数据分解到三个基空间中,提取出独立于多组学数据的中心核张量,研究核张量间的公共关系;

构建基于先验知识约束的张量配对模型;

建立公共模块发现模型。

5. 根据权利要求4所述的一种综合诊疗决策方法,其特征在于:所述公共发现模型以整合模型优化函数为核心,用于在多组学数据中挑选出对鼻咽癌患者生存预后具备指向性的指标。

6. 根据权利要求5所述的一种综合诊疗决策方法,其特征在于:在获取对鼻咽癌患者生存预后具备指向性的指标之后,获取所述指标对应的权重。

鼻咽癌数据库及基于所述数据库的综合诊疗决策方法

技术领域

[0001] 本发明涉及医疗领域,尤其涉及一种鼻咽癌数据库及基于所述数据库的综合诊疗决策方法。

背景技术

[0002] 人工智能是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门技术科学。人工智能是计算机科学的一个分支,它企图了解智能的实质,并生产出一种新的能以人类智能相似的方式做出反应的智能机器,该领域的研究包括机器人、语言识别、图像识别、自然语言处理和专家系统等。人工智能医疗是其在医疗领域的扩展应用。目前人工智能医疗主要包括虚拟助理、医疗大数据、医学影像、智能语音、健身生物技术、健康生活方式管理、医疗搜索、癌症早筛、人工智能芯片等方面。现有出现医疗相关智能系统包括IBM公司Watson机器人、中山眼科中心“CC—Cruiser先天性白内障人工智能平台”等。

[0003] 但是,现有的人工智能医疗系统未有如鼻咽癌的这类专科医疗领域的智能系统,以Watson机器人为例,其功能相对齐全,可覆盖多种常见肿瘤,但对目前一些我国常见的特色的癌症专科领域深入不足。而除去Watson机器人之外,国内一些职能平台等系统功能较为单一,主要针对某一特定小范围疾病或检查项目,远不能达到系统化辅助诊疗的效果。

[0004] 鼻咽癌是广东省地方特色癌种,复发和转移是主要死因,精准预测复发转移并指导个体化治疗是提高疗效的关键,而如何基于智能化的数据分析并由此进行精准的决策是目前亟待解决的问题。

发明内容

[0005] 为了解决上述问题,本发明提供一种鼻咽癌数据库及基于所述数据库的综合诊疗决策方法。

[0006] 本发明是以如下技术方案实现的:

[0007] 一种鼻咽癌数据库,包括临床大数据库和基因数据库;

[0008] 所述临床大数据库中记录有由临床文字数字信息形成的第一数据以及根据临床图像信息提取并识别出的第二数据;鼻咽癌数据库中的第一数据、第二数据和/或基因数据都伴随其时间属性;

[0009] 所述第一数据、第二数据和基因数据均为单组学数据,所述鼻咽癌数据库构成能够存储包含第一数据、第二数据和基因数据的多组学数据库;

[0010] 所述鼻咽癌数据库中的数据用于支持诊疗决策服务器进行基于第一模型、第二模型、第三模型、第四模型和/或第五模型的诊疗决策。

[0011] 一种综合诊疗决策方法,所述诊疗决策方法使用上述鼻咽癌数据库,包括:

[0012] 根据鼻咽癌数据库中的数据构建诊疗决策模型,所述诊疗决策模型包括第一模型、第二模型、第三模型、第四模型和/或第五模型;所述第一模型以第一数据为输入,所述

第二模型以第二数据为输入,所述第三模型以基因数据为输入,所述第四模型和所述第五模型以鼻咽癌数据库中多组学数据为输入;

[0013] 选择诊疗决策模型;

[0014] 根据所述诊疗决策模型为用户提供诊疗决策服务。

[0015] 进一步地,所述第四模型的构建方法包括:

[0016] 对多组学数据进行整合分析,建立经质量评估后的数据存储模型;

[0017] 在所述数据存储模型框架下获取多组学数据之间的关联性,建立基于多组学数据整合的第四模型。

[0018] 进一步地,所述第四模型中包括公共发现模型,所述公共发现模型基于稀疏表达理论构建。

[0019] 进一步地,所述第四模型中包括公共发现模型,所述公共发现模型的构建方法包括:

[0020] 在张量表达和矩阵表达的框架下,提出面向张量描述的特征模型;

[0021] 建立基于先验知识约束的张量配对能量方程;

[0022] 分析特征张量描述方法、先验知识约束设计、张量配对方法与局部公共子结构的内在关系;

[0023] 针对高噪音污染的数据,测试模型的鲁棒性;

[0024] 建立约束问题分解机制,得到公共发现模型。

[0025] 进一步地,所述公共发现模型的构建方法包括:

[0026] 建立面向多组学数据的矩阵分解模型,将多维数据分解到同一个基空间中,分析系数矩阵之间的互动关系,得到多组学数据之间的关联关系;

[0027] 建立面向多组学数据的张量分解模型,通过采用三个隐式变量基空间的方法,将多组学数据分解到三个基空间中,提取出独立于多组学数据的中心核张量,研究核张量间的公共关系;

[0028] 构建基于先验知识约束的张量配对模型;

[0029] 建立公共模块发现模型。

[0030] 进一步地,所述公共发现模型以整合模型优化函数为核心,用于在多组学数据中挑选出对鼻咽癌患者生存预后具备指向性的指标。

[0031] 进一步地,在获取对鼻咽癌患者生存预后具备指向性的指标之后,获取所述指标对应的权重。

[0032] 本发明的有益效果是:

[0033] 本发明提供一种鼻咽癌数据库及基于所述数据库的综合诊疗决策方法,能够为用户提供基于人工智能的精准的智能诊断决策服务,具备广阔的使用前景。

附图说明

[0034] 图1是本发明实施例提供的鼻咽癌数据库示意图;

[0035] 图2是本发明实施例提供的第一模型的形成方法流程图;

[0036] 图3是本发明实施例提供的第二模型的构建方法流程图;

[0037] 图4是本发明实施例提供的第三模型的构建方法流程图;

- [0038] 图5是本发明实施例提供的第四模型的构建方法流程图；
- [0039] 图6是本发明实施例提供的对多维数据进行整合分析的方法流程图；
- [0040] 图7是本发明实施例提供的第五模型的构建方法流程图；
- [0041] 图8是本发明实施例提供的云系统示意图；
- [0042] 图9是本发明实施例提供的分析决策服务器示意图；
- [0043] 图10是本发明实施例提供的诊疗决策终端示意图；
- [0044] 图11是本发明实施例提供的一种在线辅助诊疗系统中使用方法流程图。

具体实施方式

[0045] 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作进一步地详细描述。

[0046] 本发明实施例提供一种鼻咽癌人工智能辅助诊疗决策系统，为构建所述诊疗决策系统，本发明实施例首先构建鼻咽癌数据库。具体地，本发明实施例中鼻咽癌数据库的组成结构的设计是基于下述研究目的的考量：

[0047] 深入研究宏观多模态图像、临床病理表型和微观基因型之间的多尺度异构关联关系，以期建立临床表型和微观基因型之间的联动关系，从而完成对鼻咽癌从宏观到微观各个状态的多尺度、多模态、大数据描述，建立可以精确、全面预测鼻咽癌发生远处转移及复发的预测模型。

[0048] 收集鼻咽癌患者临床资料、影像资料和基因数据，建立大型的多源异构多组学数据库，最终开发鼻咽癌大数据云端诊疗分析系统，并用于辅助指导精准治疗。

[0049] 有鉴于此，本发明实施例中提供的鼻咽癌数据库如图1所示，包括临床大数据库和基因数据库。

[0050] 具体地，所述临床大数据库中记录有由临床文字数字信息形成的第一数据以及根据临床图像信息提取并识别出的第二数据。其中，第一数据可以为文字和/或数字形式，所述第二数据可以为图片和/或视频形式，并以数字、二进制大对象、链接或多媒体的形式进行存储，所述第一数据和第二数据可以作为一种多维可分析数据的数据源，供后续的诊疗决策使用。

[0051] 具体地，临床大数据库可以包括文字数字数据库，所述文字数字数据库用于存储第一数据，其对应的临床文字数字信息包括但不限于检查检验报告、治疗数据以及治疗情况数据。具体地，所述检查检验报告包括病历数据，检验指标，病理、超声、影像和/或核医学等检查检验报告；所述治疗数据包括但不限于放疗、化疗、手术、靶向、生物治疗和/或中医治疗，所述治疗情况数据包括但不限于随访预后数据以及治疗效果、并发症数据。

[0052] 具体地，临床大数据库还可以包括图像数据库，所述图像数据库存储第二数据，其对应的临床图像信息包括但不限于影像、放疗靶区、超声、核医学、病理等各类临床图像。

[0053] 需要强调的是，鼻咽癌数据库中的第一数据、第二数据和/或基因组学数据都伴随其时间属性一并记录，所述时间属性在各种诊疗决策中均具备一定作用，通过将各指标时间属性纳入诊疗决策的考量，能够为患者生存预后提供依时的决策结果。

[0054] 在鼻咽癌数据库的数据支持下，所述诊疗决策系统可以基于多种诊疗决策模型对鼻咽癌进行分析、预测和决策指导。

[0055] 在一个可行的实施例中,所述诊疗决策系统可以包括第一诊疗模块,所述第一诊疗模块可以基于第一模型来提供鼻咽癌分析、预测和决策指导。所述第一模型以第一数据为数据源,所述第一模型的形成方法如图2所示,包括:

[0056] S1.对第一数据进行分析,将第一数据中各因素与患者的预后及治疗效果进行关联,去除相关性较强,共线性较高的因素。

[0057] S2.筛选出对患者预后具有确定显著影响的特征因素,建立第一模型。

[0058] S3.对第一模型进行验证分析。

[0059] 所述第一模型可以以第一数据或第一数据中的某个或某几个字段为输入,以患者的预后指标为输出,所述预后指标包括但不限于:n年生存率,以及肿瘤复发、转移和/或并发症;预后指标用于指示患者的最终结局以及鼻咽癌病情的发展趋势。进一步地,在所述第一模型中可以在第一数据中获取对患者预后具有指向性的指标及所述指标对应的权重。

[0060] 进一步地,随着临床大数据库数据量的不断扩张,以及各地的医疗共享数据的发展,还可以将来源于自身或者不同数据中心或者医疗系统的数据引入S2中得到的第一模型进行验证分析,并校正相关参数。在不断扩大临床大数据库的内容的同时,还可以实时地校正第一模型的相关参数,使得第一模型不断完善以达到最优,从而为鼻咽癌分析、预测和决策指导提供更有价值的参考。

[0061] 所述第一模型为以第一数据为研究对象建立的模型,在临床工作之中第一数据的获取途径较为简单,在数据获取方面具有一定优势,因此,第一模型在临床工作中具备较好的应用前景。

[0062] 在另一个可行的实施例中,所述诊疗决策系统还可以包括第二诊疗模块,所述第二诊疗模块可以基于第二模型来提供鼻咽癌分析、预测和决策指导。所述第二模型以第二数据为数据源。所述第二模型的构建方法如图3所示,包括:

[0063] S10.对图像数据进行预处理。

[0064] S20.对图像进行分割及三维重建。

[0065] S30.对图像进行高通量定量分析,构建基于图像的第二模型。

[0066] 所述第二模型可以以第二数据或第二数据中的某个或某几个部分为输入,以患者的预后指标为输出,所述预后指标包括但不限于:n年生存率,以及肿瘤复发、转移和/或并发症;预后指标用于指示患者的最终结局以及鼻咽癌病情的发展趋势。进一步地,在所述第二模型中可以在第二数据中获取对患者预后具有指向性的指标及所述指标对应的权重。

[0067] 具体地,所述图像可优选为鼻咽癌CT/MRI图像,对图像进行分割及三维重建后,即可进行数字化多参数CT/MRI等图像的高通量定量分析。具体分析内容包括但不限于特征参数数据提取、特征参数数据相关性分析、聚类及统计学特征探索,从大批量的复发转移鼻咽癌的多模态医学影像中寻找高灵敏度的转移复发影像参数,结合患者复发、转移等预后情况,构建基于鼻咽癌CT/MRI等图像的第二模型。

[0068] 进一步地,随着临床大数据库数据量的不断扩张,以及各地的医疗共享数据的发展,还可以将来源于自身或者不同数据中心或者医疗系统的数据引入S30中得到的第二模型进行验证分析,并校正相关参数。进一步地,在不断扩大临床大数据库的内容的同时,还可以实时地校正第二模型的相关参数,使得模型不断完善以达到最优。

[0069] 所述第二模型为以第二数据为研究对象的模型,随着医疗条件的改善和影像技术

的普及,第二数据在临床工作之中的作用日益突出,第二数据对于医疗诊断的指向性也逐步增强,因此第二模型在临床工作中同样具备较好的应用前景。

[0070] 进一步地,基因数据库用于存储基因数据,相较于第一数据和第二数据,基因数据的获取难度更高,分析难度更大,但是其对于鼻咽癌的预后分析的作用也是非常突出的。基因数据库的建立能够为下述工作创造条件:微观层面分析复发转移鼻咽癌的全基因组测序数据,探究肿瘤发生、发展、凋亡过程中的基因型变异,以期发现潜在转移复发驱动基因及分子标志物,从而为临床精准医疗研究提供重要基础。

[0071] 可见,本发明实施例中所述鼻咽癌数据库为整合了临床大数据库和基因数据库的多源异构多组学数据库,其存储的数据构成了多维可分析数据。

[0072] 在另一个可行的实施例中,所述诊疗决策系统还可以包括第三诊疗模块,所述第三诊疗模块可以基于第三模型来提供鼻咽癌分析、预测和决策指导。所述第三模型以基因数据为数据源。所述第三模型的构建方法如图4所示,包括:

[0073] S100.对配对的鼻咽癌原发灶、转移灶、复发灶进行全基因组测序。

[0074] S200.对测序后的数据进行生物信息学分析。

[0075] 对测序后的数据进行全面生物信息学分析,于鼻咽癌转移复发样本中发现高度相关基因靶点、差异分子和分子标志物,找到有意义的转移、复发驱动基因及突变信息和转移、复发相关通路。

[0076] S300.根据分析结果构建基于鼻咽癌基因数据的第三模型。

[0077] 对分析得到的转移驱动基因及相关突变信息,采用体内体外实验,进行功能验证,构建基于鼻咽癌基因数据的第三模型。所述第三模型可以全部或部分基因数据为输入,以患者的预后指标为输出,所述预后指标包括但不限于:n年生存率,以及肿瘤复发、转移和/或并发症;预后指标用于指示患者的最终结局以及鼻咽癌病情的发展趋势。进一步地,在所述第三模型中可以在基因数据中获取对患者预后具有指向性的指标及所述指标对应的权重。

[0078] 进一步地,随着基因数据库数据量的不断扩张,以及全国各地的医疗共享数据的发展,还可以将来源于不同数据中心或者医疗系统的数据引入S300中得到的第三模型进行验证分析,并校正相关参数。进一步地,在不断扩大基因数据库的内容的同时,还可以实时地校正第三模型的相关参数,使得模型不断完善以达到最优。

[0079] 所述第三模型为以基因数据为研究对象建立的模型,随着基因技术以及相关研究的普及,基因数据在临床工作之中的作用日益突出,基因数据对于医疗诊断的指向性也逐步增强,因此第三模型在临床工作中同样具备较好的应用前景。

[0080] 第一模型、第二模型以及第三模型的数据来源各不相同,而基于多种数据来源进行对于鼻咽癌的诊疗决策可以将各种数据源纳入考量,从而为患者提供更加全面的诊疗指导。基于这种认知,在另一个可行的实施例中,所述诊疗决策系统还可以包括第四诊疗模块,所述第四诊疗模块可以基于第四模型来提供鼻咽癌分析、预测和决策指导。所述第四模型能够以鼻咽癌数据库中第一数据、第二数据以及第三数据中的多种数据为研究对象,为用户进行诊疗决策指导。其中第四模型的构建方法如图5所示,包括:

[0081] P1.对多组学数据进行整合分析,建立经质量评估后的数据存储模型。

[0082] 具体地,所述多组学数据可以为第一数据、第二数据和基因数据中的多种,其中第

一数据、第二数据和基因数据都是组学数据。

[0083] P2.全面系统探究多组学数据之间的关联性,建立基于多组学数据整合的第四模型。

[0084] 本发明实施例中第四模型中包括公共发现模型,具体可以基于稀疏表达理论构建所述公共发现模型。

[0085] 在另一种可行的实施方式中,在P2执行过程中,可以在张量表达和矩阵表达的框架下,提出面向张量描述的特征模型;建立基于先验知识约束的张量配对能量方程;分析特征张量描述方法、先验知识约束设计、张量配对方法与局部公共子结构的内在关系;针对高噪音污染的数据,测试模型的鲁棒性;针对海量数据,建立高效的约束问题分解机制,以期快速优化求解,从而最终得到公共发现模型。

[0086] 具体地,本发明实施例提出了对多组学数据进行整合分析的方法,如图6所示,包括:

[0087] 首先,建立面向多组学数据的矩阵分解模型,将多维数据分解到同一个基空间中,分析系数矩阵之间的互动关系,从而发现多组学数据之间的关联关系。

[0088] 具体地,建立研究基因型与表型之间关联规则,发现重要基因靶点与图像表型、肿瘤分期分级等因素之间的动态关联。

[0089] 其次,建立面向多组学数据的张量分解模型,通过采用三个隐式变量基空间的方法,将多组学数据分解到三个基空间中,提取出独立于多组学数据的中心核张量,研究核张量间的公共关系,以期发现其公共关联。

[0090] 具体地,研究在不同约束下,隐式变量基空间的分解特性,设计面向多约束中心核张量的关联函数。

[0091] 再次,建立基于张量描述的张量配对模型,探讨单组学数据的张量特征描述方法,研究张量特征维数和定量描述对张量配对的影响;设计特征张量的配对方法,构建基于先验知识约束的张量配对模型。

[0092] 最后,针对上述分析结果建立公共模块发现模型。

[0093] 基于上述公共发现模型的研究方法,本发明实施例提供了独创性整合模型优化函数,所述整合模型优化函数为公共发现模型的核心内容,用于在多组学数据中挑选出对鼻咽癌患者预后具备指向性的指标:

$$[0094] \quad \min_X \sum_{i,j} H_{i,j} X_{i,j} + \lambda_1 \text{tr}(X^T L_{M_1} X) + \lambda_2 \text{tr}(X L_{M_2} X^T) + \beta \|X\|_0$$

[0095] Subject to: $X_{i,j} \in \{0,1\}$

[0096] $X I_1 \leq b_1$

[0097] $X^T I_2 \leq b_2$

[0098] 其中H表示组学数据 S_1 和 S_2 之间的相似度,X表示分配矩阵, L_{M_1} 表示数据 S_1 内部特征关联关系, L_{M_2} 表示数据 S_2 内部特征关联关系,公式第一项表示分配矩阵X尽量满足相似度矩阵H的分布,第二项和第三项分别表示融入了数据 S_1 和数据 S_2 的先验知识,提高准确率,第四项表示分配矩阵具有稀疏性,使得可以更好的解释数据 S_1 和数据 S_2 之间的关联关系。

[0099] 其中 L_M 为归一化拉普拉斯矩阵,如下表示

$$[0100] \quad L_{M_k} = D_{M_k}^{-\frac{1}{2}} (D_{M_k} - W_{M_k}) D_{M_k}^{-\frac{1}{2}}$$

[0101] 根据拉普拉斯矩阵性质,可以得到以下公式

$$tr(X^T L_{M_1} X)$$

$$= \sum_{k=1}^{n_2} X_{.k}^T L_{M_1} X_{.k}$$

$$[0102] \quad = \sum_{k=1}^{n_2} \sum_{i,j} \left(\frac{X_{ik}}{\sqrt{(D_{M_1})_{ii}}} - \frac{X_{jk}}{\sqrt{(D_{M_1})_{jj}}} \right)^2 (W_{M_1})_{ij}$$

[0103] 最终通过梯度下降法求解上述最优化函数,获得数据整合结果,如下算法1所示

[0104] 算法1

[0105] 输入: $M_1, M_2, W_{M_1}, W_{M_2}$

[0106] 1:初始化 $\lambda_1, \lambda_2, \beta, X^{(0)}, \gamma^{(1)}, M, \rho > 1, \delta \in (0, 1), \epsilon > 0, k=1$

[0107] 2:归一化 M_1, M_2

[0108] 3:利用梯度下降法求解以下公式更新分配矩阵 $X^{(k)}$

$$\min_X G(X) = tr(HX^T) + \lambda_1 tr(X^T L_{M_1} X) + \lambda_2 tr(X L_{M_2} X^T) + \beta tr(UX^T)$$

$$[0109] \quad + \frac{1}{2M} \sum_{k=1}^N \{ [\max(Mg_k(X) - \gamma_k, 0)]^2 - \gamma_k^2 \}$$

[0110] 4:设置 $M = \rho M$, 如果 $\frac{\|g(X^{(k)})\|}{\|g(X^{(k-1)})\|} \geq \delta$

[0111] 5:更新 $\gamma^{(k+1)} = \min(\gamma^{(k)} - Mg(X^{(k)}), 0)$

[0112] 6:设置 $k = k+1$

[0113] 7:重复步骤3,4,5,6直到 $\|g(X^{(k)})\| < \epsilon$

[0114] 输出: X

[0115] 进一步地,在基于所述公共发现模型获取预后具备指向性指标的基础之上,为了提升综合诊疗效果,所述第四模型中还能够将各种指向性指标在预后判断中的权重纳入考量。基于第四模型获取权重的方法有很多,本发明实施例概举三个可行的实施方式。

[0116] 在一个可行的实施方式中,可以将第四模型的输出简单被表示为 $P = K_1 A_1 + K_2 A_2 + K_3 A_3 + \dots + K_m A_m$ 。P为在 $A_1 + A_2 + A_3 + \dots + A_m$ 条件下,生存5年/5年内无疾病进展等各类生存可能的发生概率。而根据鼻咽癌数据库中的数据获取 $A_1 \dots A_m$ 的值,即可得到各种指向性指标在预后判断中的权重。

[0117] 在另一种可行的实施方式中,可以将第四模型的输出P表示为多个指标综合作用的结果,而所述综合作用包括各个预后具备指向性指标同时作用、分阶段作用以及混合作用。在这种表示方法中通过分析鼻咽癌数据库中的数据获取到各种指向性指标在预后判断中的权重。

[0118] 在另一种可行的实施方式中,可以基于第一模型、第二模型和/或第三模型的研究结果,获得各种指向性指标在预后判断中的权重。在这种实施方式中,第四模型的构建包括

下述步骤：

[0119] 根据公共发现模型挑选出对于预后具有指向性的指标；

[0120] 判断所述指标属于第一数据、第二数据还是基因数据；

[0121] 若所述指标属于第一数据，则根据第一模型获取所述指标的权重；

[0122] 若所述指标属于第二数据，则根据第二模型获取所述指标的权重；

[0123] 若所述指标属于基因数据，则根据第三模型获取所述指标的权重；

[0124] 待获取全部具有指向性的指标的权重后，根据指标及其权重的加权结果进行诊疗决策指导。

[0125] 第四模型为诊疗决策系统实现全面化的诊疗决策创造了条件，显然，在诊疗决策系统的实际使用过程中，需要不断获取医生和患者的反馈，结合不断更新的数据以及文献，不断修正公共发现模型、第一模型、第二模型以及第三模型，从而为人工智能辅助诊疗提供更多助力。

[0126] 为了基于鼻咽癌数据库得到更为全面的诊疗结果，在另一个可行的实施例中，所述诊疗决策系统还可以包括第五诊疗模块，所述第五诊疗模块可以基于第五模型来提供鼻咽癌分析、预测和决策指导。所述第五模型同样能够以鼻咽癌数据库中多组学数据为数据源，不同于第四模型的构建方法，第五模型基于神经网络为用户提供诊疗决策。其中第五模型的构建方法如图7所示，包括：

[0127] P10，提取不同的组学数据中对于第五模型的生成有价值的特征。

[0128] 具体地的提取方法可以有多种，也可以参考第四模型中的公共发现模型进行特征提取。

[0129] P20，对不同组学数据中相关联的表征特征进行高维抽象得到高维特征。

[0130] P30，完成对冗余噪音特征的过滤。

[0131] 在上述三个步骤中，可以对第一数据和基因数据采用卷积神经网络提取底层到高层的特征表达；对第二数据采用集成方法挖掘有用的特征，并通过传统的逻辑回归结合不同的范数约束过滤无关的干扰特征。

[0132] 通过对不同组学数据中相关联的表征特征进行高维抽象得到高维特征，并完成对冗余噪音特征的过滤；从而为P40创造条件。

[0133] P40，通过全连接层将所有的高维特征进行连接，从而生成第五模型。

[0134] 具体地，可以通过全连接层将所有的高维特征进行连接，生成诊疗决策模型，当用于训练模型的新收集的数据量达到规定量级时，重新启动模型训练，优化模型参数，使得模型不断自我学习，自我调整，自我完善，从而形成第五模型。

[0135] 具体地，本发明实施例第五模型可通过建立卷积神经网络以及卷积神经网络的学习实现。

[0136] 在第五模型的构建过程中，对于鼻咽癌数据库中的数据通过一个全连接层进行集成，通过误差反向传播算法结合随机梯度下降来调整参数，并利用dropout, leaky relu激活函数以及范数约束来限制参数的调整范围，保证特征的稀疏性，避免过拟合的情况，从而进一步保证第五模型的效能。通过第五模型在训练集以及验证集的预测效能来调整模型最后全连接层的数目以及相关参数，以期为医师提供有价值的诊断意见。

[0137] 第四模型和第五模型为以多组学数据为研究对象的模型，可以全面使用鼻咽癌数

数据库的内容以期得到较为全面的决策结果,第一模型、第二模型和第三模型为以单组学数据为研究对象的模型,能够在某一种组学数据的基础上进行诊疗,在一种优选的实施方式中,第一模型、第二模型、第三模型、第四模型和第五模型可以并行存在或者择优存在一种或多种,从而从多个角度为用户提供更优质的诊疗决策服务。

[0138] 在上述内容的基础上,本发明实施例进一步提供所述一种鼻咽癌人工智能辅助诊疗决策系统的逻辑架构,鼻咽癌人工智能辅助诊疗决策系统具体可以为一种在线诊疗系统,具体地,所述在线诊疗系统可以为一种云系统,作为所述在线诊疗系统的数据支持者(鼻咽癌数据库)也可以布设在云端,显然,使用云存储更有利于数据整合以及诊疗分析。

[0139] 所述诊疗决策系统如图8所示,可以包括设置在云端的诊疗决策服务器和与所述诊疗决策服务器通信连接的诊疗决策终端,所述诊疗决策系统具体可以使用B-S架构或者C-S架构,所述诊疗决策终端响应于用户指令,向所述诊疗决策服务器集群发出对应于所述用户指令的数据请求;所述诊疗决策服务器响应于所述数据请求,生成数据响应并将所述数据响应传输至所述诊疗决策终端,以便于所述诊疗决策终端向用户展示对应于所述用户指令的结果。

[0140] 进一步地,为了全面为用户提供诊疗服务,所述诊疗决策服务器具体可以为服务器集群(诊疗决策服务器集群),所述诊疗决策服务器集群包括用户交互服务器、分析决策服务器以及数据服务器,所述用户交互服务器用于与诊疗决策终端进行数据交互并进行用户管理,所述分析决策服务器用于为用户提供诊疗决策服务,所述数据服务器用于进行数据处理,其中设置有鼻咽癌数据库进行交互。所述分析决策服务器如图9所示,优选包括第一诊疗模块、第二诊疗模块、第三诊疗模块、第四诊疗模块和第五诊疗模块,所述第一诊疗模块、第二诊疗模块、第三诊疗模块、第四诊疗模块和第五诊疗模块的功能如前所述,此处不再赘言。

[0141] 进一步地,在所述诊疗决策终端,提供有用户注册、数据管理、查询、注释、比较、检索、展示等功能。具体地,所述诊疗决策终端如图10所示包括:

[0142] 用户注册模块,用于向所述诊疗决策服务器集群进行注册,以便于用户使用诊疗决策服务器集群提供的各种服务。

[0143] 数据管理模块,用于管理用户的个人数据。

[0144] 查询模块,用于用户向诊疗决策服务器集群查询目标数据。

[0145] 注释模块,用于用户在显示界面进行标记。

[0146] 比较模块,用于对于类似数据进行比较,并标识出比较结果以便于用户快速发现类似数据之间的差异。

[0147] 检索模块,用于用户快速搜索目标数据。

[0148] 展示模块,用于进行数据可视化。

[0149] 所述诊疗决策终端可以为装载有与诊疗决策服务器交互的软件的任意智能设备,比如PC、手机或平板等设备。

[0150] 在所述诊疗决策终端,还包括界面切换模块,所述界面切换模块用于根据用户登录身份的不同为用户提供不同的显示界面。具体地,根据使用用户的身份不同,界面切换模块还额外提供不同的功能。若用户身份为患者用户,则向用户提供访问数据、查看诊断结果和治疗方案等功能;若用户身份为医生用户,则用户提供访问数据、作出诊断决策以及预后

评估等功能。

[0151] 具体地,在B-S架构中,可以通过在网站前端采用简便友好的操作界面,将云端分析的诊疗结果可视化。所述可视化具体地可以由诊疗决策终端的可视化模块实现,所述可视化模块包括数据预警单元、图表联动单元和对比拆分单元。不仅方便临床专业医生操作,而且还提供了清晰明了的医学图像显示界面和病例资料显示界面,以期临床专业医生可以更好地作出评估和诊断。

[0152] 所述在线诊疗系统不仅为临床专业医生提供鼻咽癌复发转移分期分级初步预测结果,还提供了管理、注释、比较、检索和展示PB以上量级组学数据与临床信息功能。所述在线诊疗系统能够为基因靶点、图像肿瘤区域、精准个体化医疗提供方向指导,为医生和患者提供疾病诊断可视化展示平台,为肿瘤研究治疗提供一个可靠有效的模范体系,也为其他肿瘤研究提供可参考的蓝本。

[0153] 具体地,在上述内容的基础上,本发明实施例提供一种在线诊疗系统使用方法作为诊疗示例,如图11所示,包括:

[0154] S101.获取临床文字数字信息、临床图像信息和/或基因数据,并结合数据的时间属性,将其纳入鼻咽癌数据库。

[0155] 具体地,数据库表结构可以如图表1所示:

[0156] 表1

	A001	A002	A003	A003	A999	D	S
编号	性别	年龄	项目 A	项目 A 时 间	鼻咽肿瘤影 像特征参数	生存状态	生存时 间
[0157] 00001	男	55	0	67
00002	女	54	1	76
00003	男	65	1	86
00004	男	45	0	56
[0158]
99999	女	51	0	59

[0159] S102.选择能够用于决策的诊疗决策模型,并使用其进行生存预后、并发症、分析。

[0160] 具体地,所述用于决策的诊疗决策模型可以为第一模型、第二模型、第三模型、第四模型和第五模型中的一种或多种。

[0161] 本步骤中,用于决策的模型可以为第四模型,第四模型的输出为一种对预后有指

向性的指标的加权结果,可暂简单被表示为 $P=K_1A_1+K_2A_2+K_3A_3+\dots+K_mA_m$ 。P为在 $A_1+A_2+A_3+\dots+A_m$ 条件下,生存5年/5年内无疾病进展等各类生存可能的发生概率。P越大,证实患者在 $A_1+A_2+A_3+\dots+A_m$ 条件下获得更久。可建立列线图模型等用于预测 $A_1+A_2+A_3+\dots+A_m$ 生存多少年的概率等生存数据。

[0162] 而其中,根据不同项目 A_i 系数 K_i 的大小,可对 A_i 进行方向性调整,以达到最佳的生存获益。

[0163] S103.根据已有模型,结合当前病例,获取诊疗决策参考方案,并获取随访数据。

[0164] S104.随着新的病例的加入,不断学习并优化已有的诊疗决策模型。

[0165] 在本发明实施例的具体实施过程中,能够得到如下成果:

[0166] (1) 患者方面:

[0167] 患者登录后填写自身基本信息,输入相关检查、检验结果,可获取相关的疾病诊断、预后估计、治疗期间注意事项等相关内容。并随着治疗进展,根据其不同输入,获得不同建议。

[0168] (2) 研究者方面

[0169] 依据已有数据库,建立治疗建议、预后等相关模型。并根据患者输入信息,形成新的大数据,并对模型相关参数进行校正,获得最能反应真实情况的模型。

[0170] 整合临床资料和基因数据两方面的数据资料,以期发现宏观表型——微观基因型多水平下高度关联的标记物,从而建立基因靶点到图像表型的联系,进而建立基于大数据的人工智能精准医疗体系。

[0171] 此外,本发明实施例中的下述实施步骤在相关领域中均是首次提出:

[0172] (1) 构建具备时间属性的多组学数据库,建立多组学依时预后及并发症等预测模型。

[0173] (2) 从批量复发转移鼻咽癌的多模态医学影像中寻找高灵敏度的转移复发影像标志;

[0174] (3) 应用基因组学生物分析技术于鼻咽癌转移复发样本中发现高度相关基因靶点、差异分子和分子标志物;

[0175] (4) 整合临床文字数字信息、临床图像信息和基因数据等多水平异构数据,以期发现宏观表型——观基因型多水平下高度关联的标记物,从而建立基因靶点到图像表型的关联;

[0176] (5) 建成鼻咽癌大规模不同层次组学数据的数据库,开发出了集管理、注释、比较、检索、展示组学数据和复发转移分期及预测预后于一体的在线诊疗系统。

[0177] 以上所揭露的仅为本发明较佳实施例而已,当然不能以此来限定本发明之权利范围,因此依本发明权利要求所作的等同变化,仍属本发明所涵盖的范围。

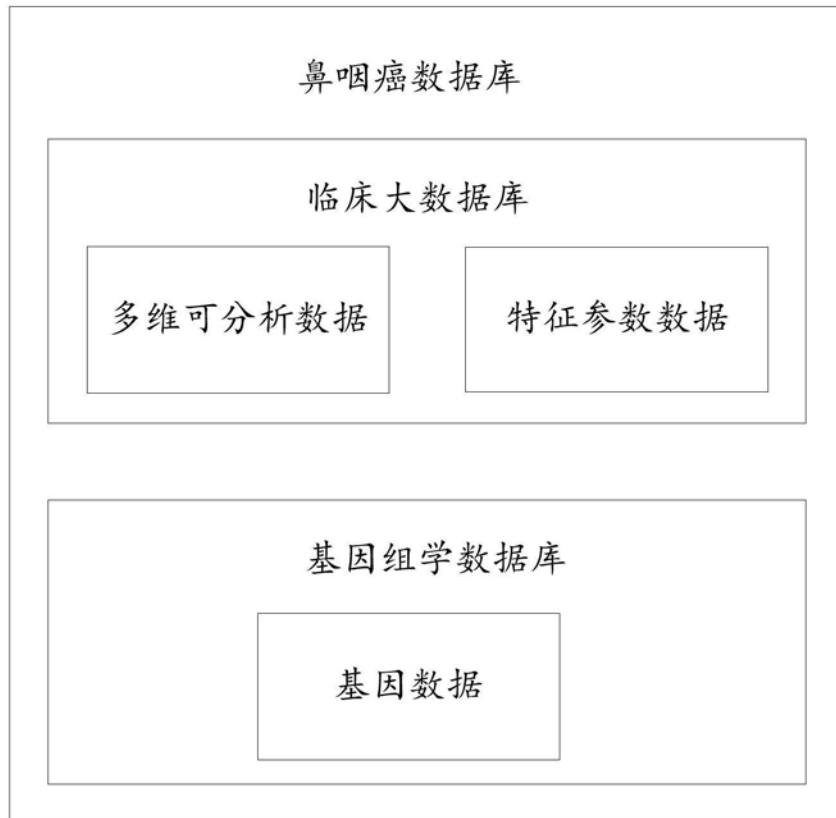


图1

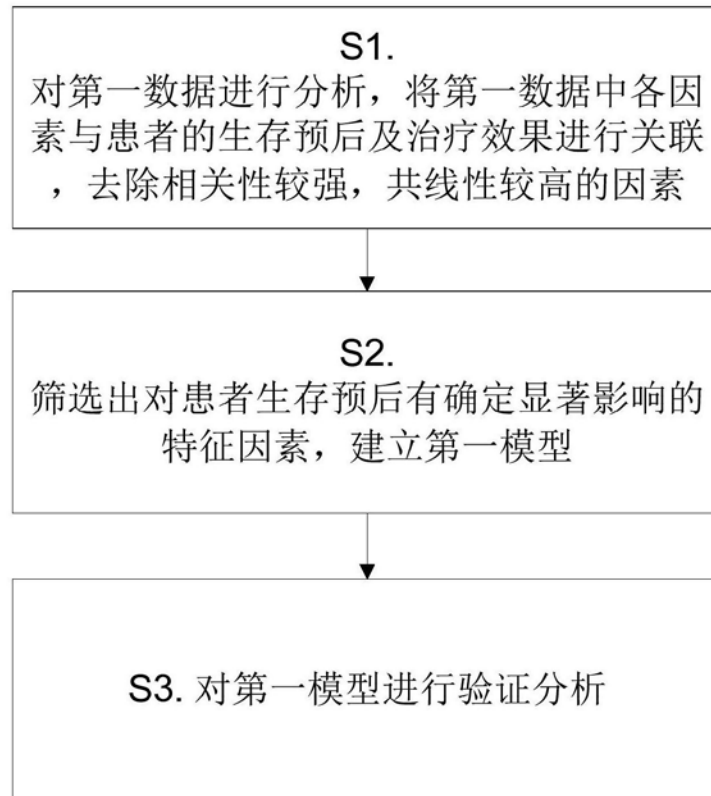


图2

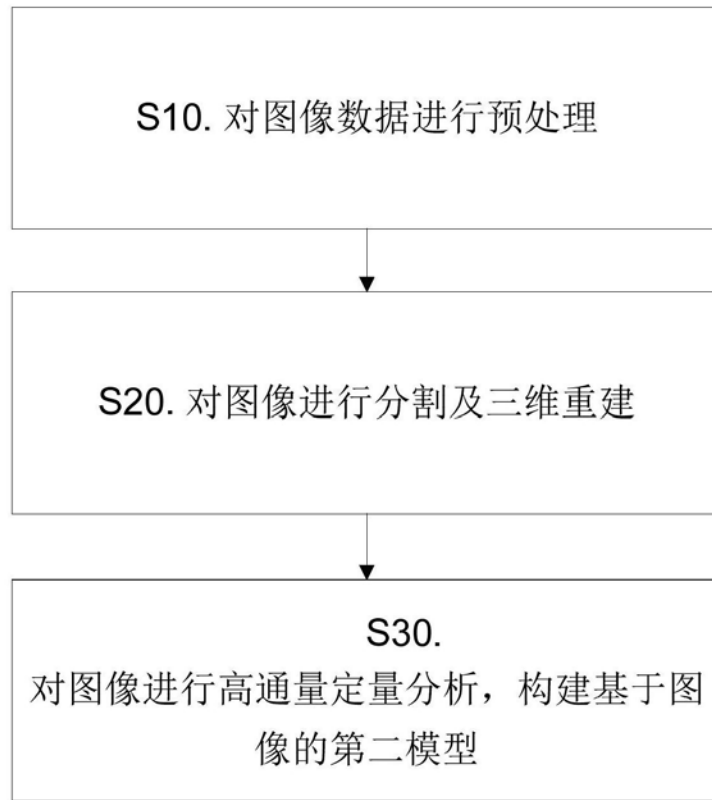


图3

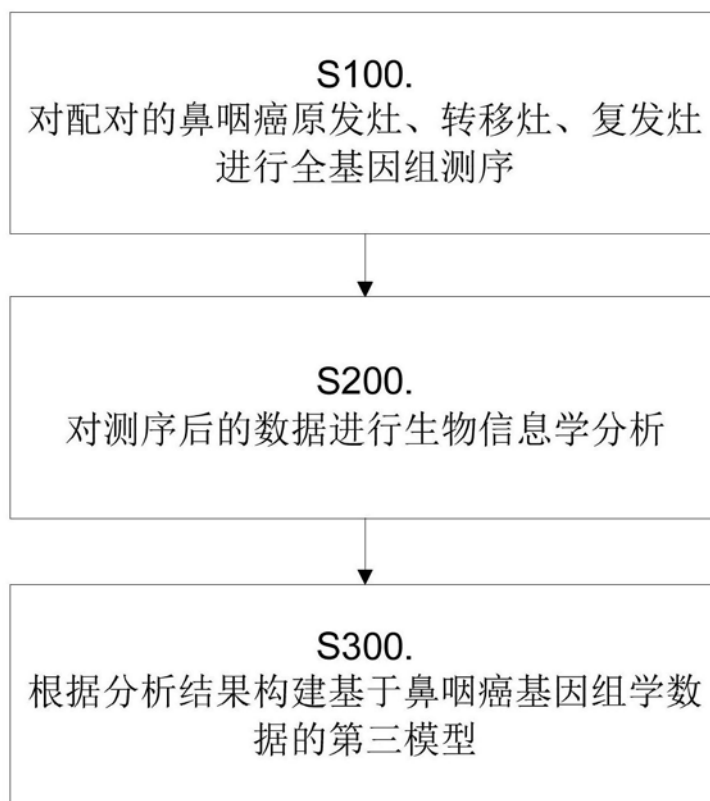


图4

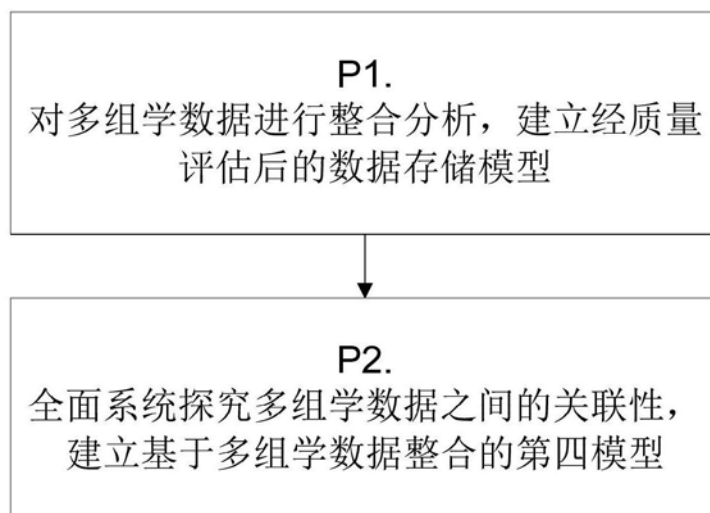


图5

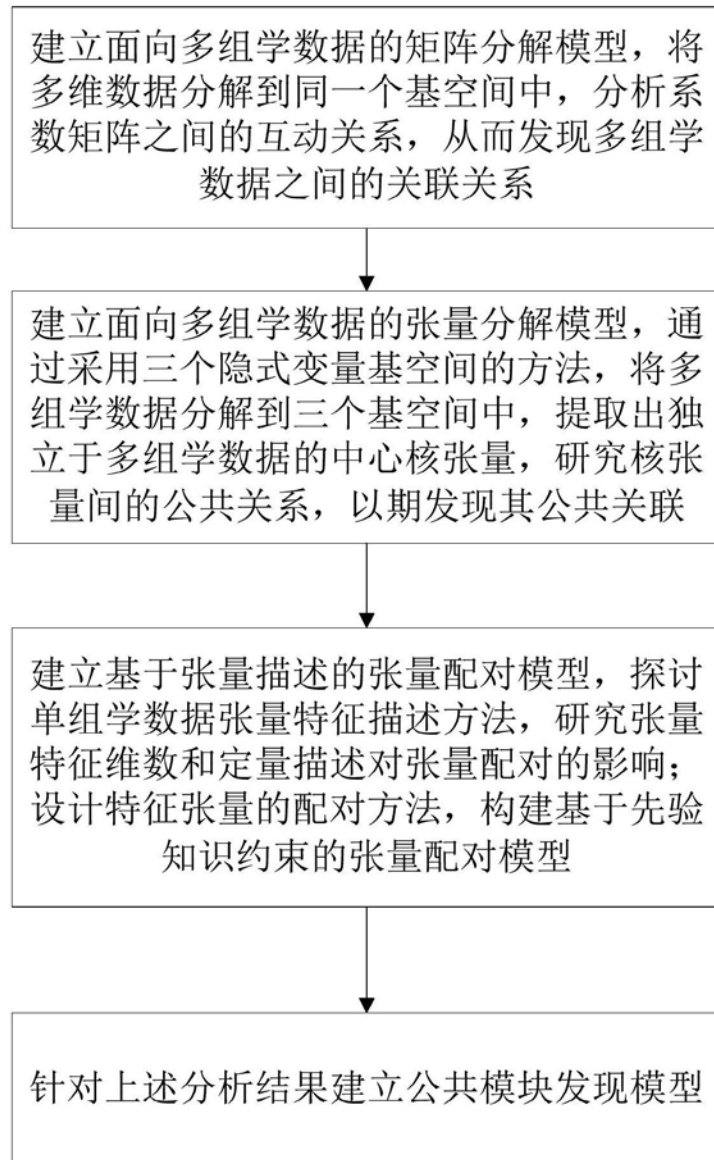


图6

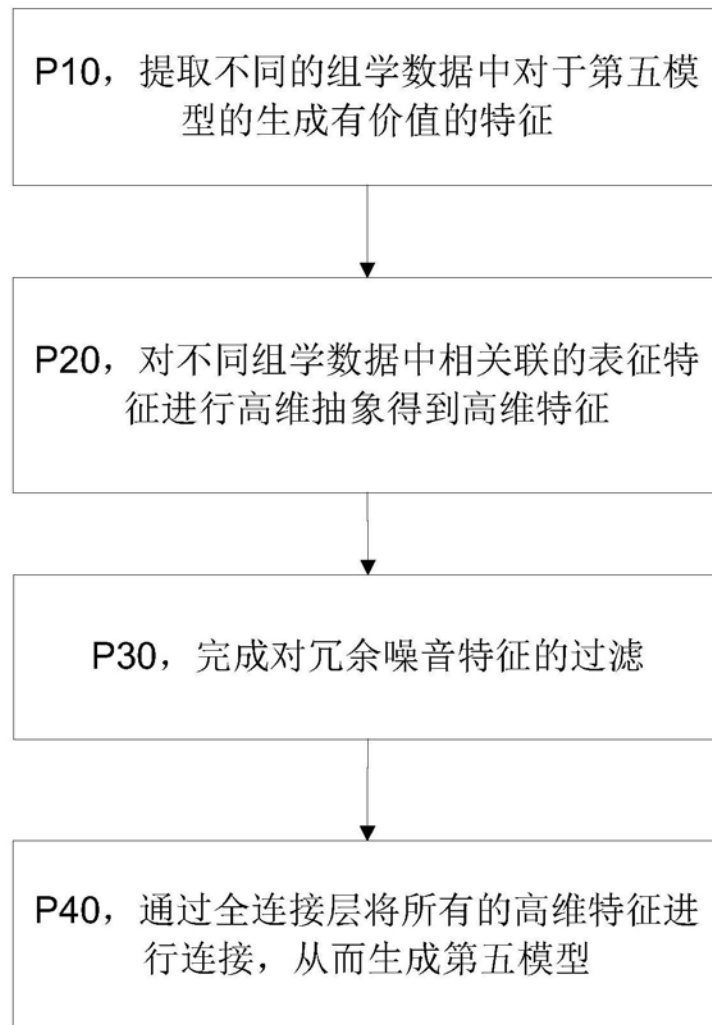


图7

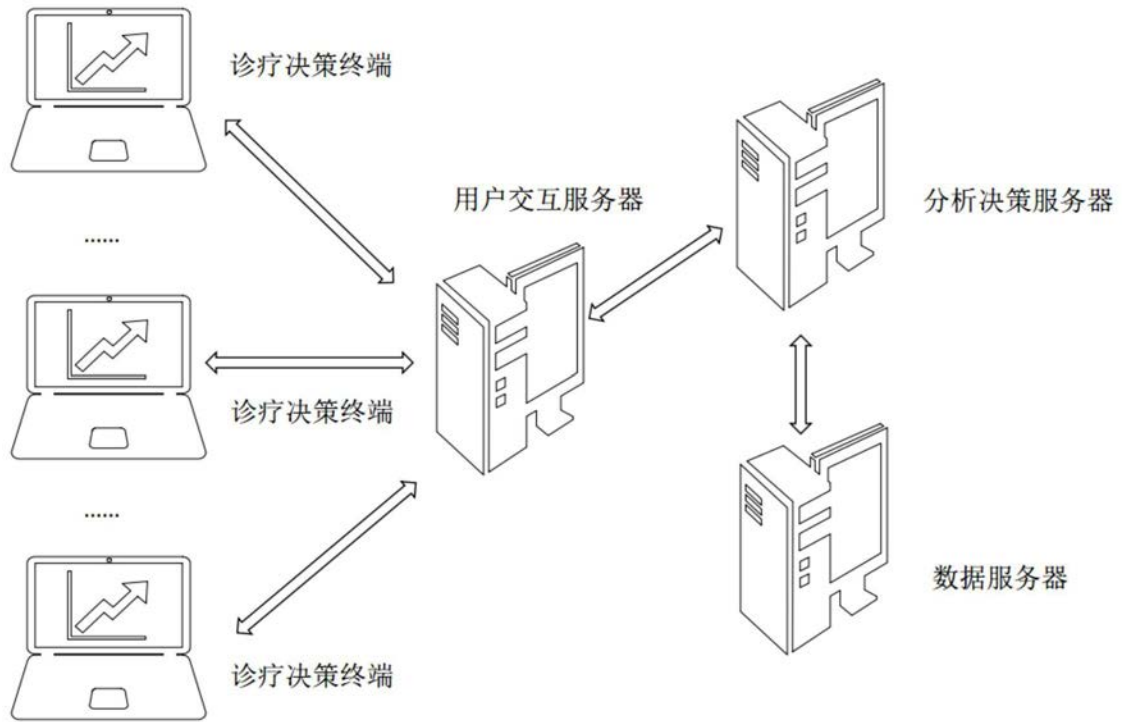


图8



图9

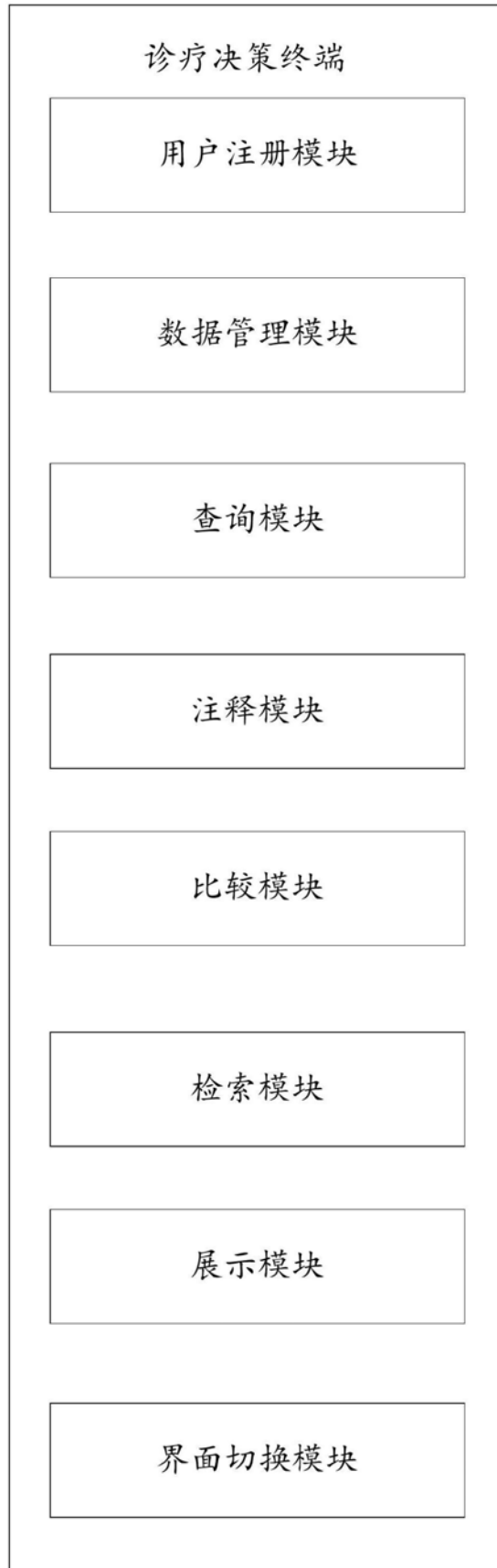


图10

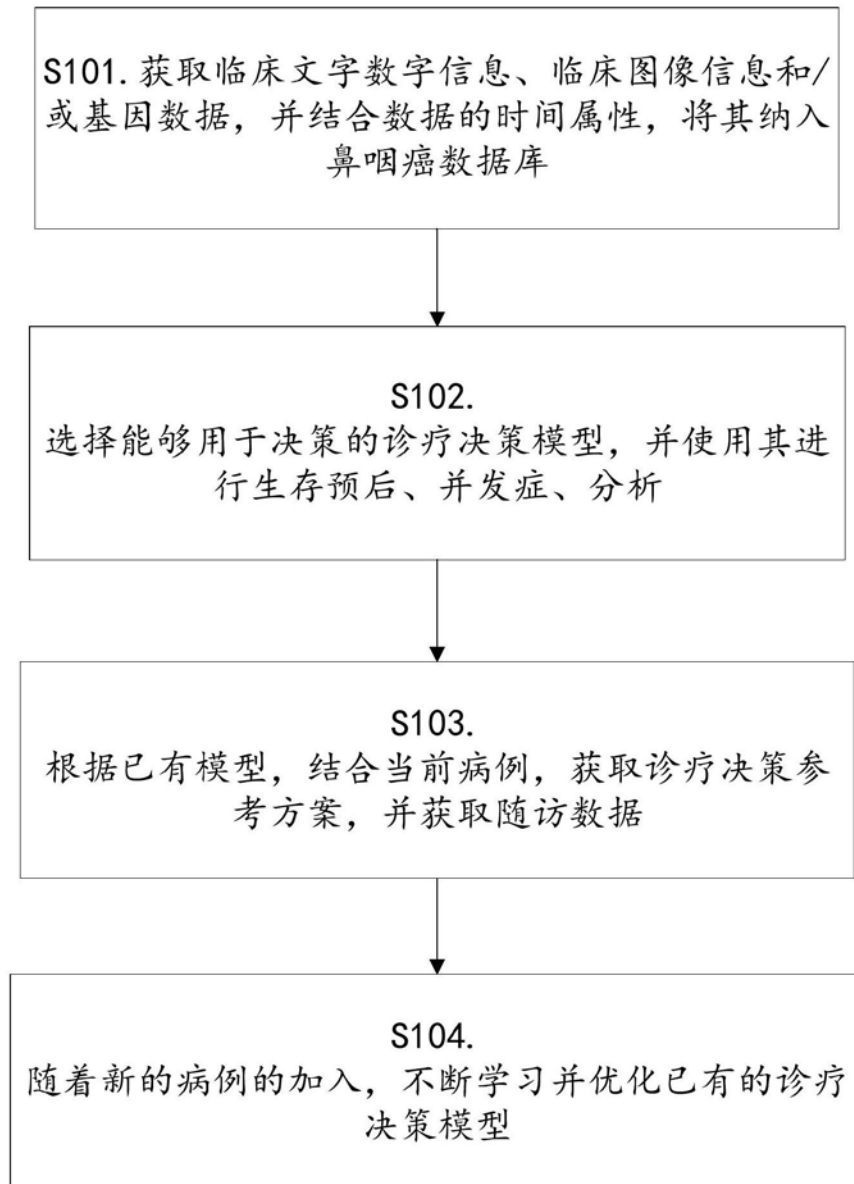


图11