(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
2021 年 4 月 22 日 (22.04.2021)

WIPO | PCT

(10) 国际公布号
WO 2021/073263 A1

(51) 国际专利分类号:
G16H 50/30 (2018.01)

(21) 国际申请号: PCT/CN2020/111662

(22) 国际申请日: 2020 年 8 月 27 日 (27.08.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910980973.9 2019年10月15日 (15.10.2019) CN

(71) 申请人: 平安科技（深圳）有限公司 (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN]; 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。

(72) 发明人: 代心灵 (DAI, Xinling); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。

(74) 代理人: 广州三环专利商标代理有限公司 (SCIHEAD IP LAW FIRM); 中国广东省广州市越秀区先烈中路80号汇华商贸大厦1508室, Guangdong 510070 (CN)。

(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, IT, JO, 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,

(54) **Title:** DISEASE SUFFERING RISK PREDICTION METHOD AND DEVICE

(54) **发明名称:** 一种患病风险的预测方法及装置

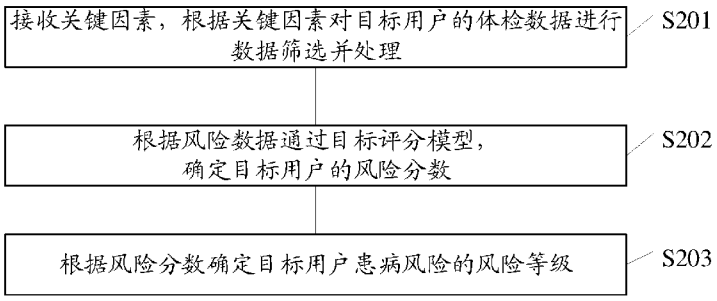


图 2A

S201 RECEIVE A KEY FACTOR AND PERFORM, ACCORDING TO THE KEY FACTOR, DATA SCREENING AND PROCESSING ON PHYSICAL EXAMINATION DATA OF A TARGET USER
S202 DETERMINE A RISK SCORE OF THE TARGET USER BY MEANS OF A TARGET SCORING MODEL ACCORDING TO THE RISK DATA
S203 DETERMINE A RISK LEVEL OF THE DISEASE SUFFERING RISK OF THE TARGET USER ACCORDING TO THE RISK SCORE

(57) **Abstract:** A disease suffering risk prediction method and device. The disease suffering risk prediction method comprises: receiving a key factor and performing, according to the key factor, data screening and processing on physical examination data of a target user to obtain risk data, the risk data comprising one or more of blood pressure data, blood lipid data, and blood sugar level data; determining a risk score of the target user by means of a target scoring model according to the risk data, the target scoring model being a prediction model trained on the basis of an Xgboost algorithm and a K-Folds cross-validation algorithm method; and determining a risk level of the disease suffering risk of the target user according to the risk score. The method can effectively predict the disease suffering risk of the user. Furthermore, it can timely and effectively identify the potential threats of diseases in the body and carry out early screening and intervention of the cause of the diseases as soon as possible so as to delay the manifestation of the diseases and prevent the severity of the chronic disease.

WO 2021/073263 A1

[见续页]

ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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：一种患病风险的预测方法及装置，其中，一种患病风险的预测方法，可包括：接收关键因素，根据所述关键因素对目标用户的体检数据进行数据筛选并处理，获得风险数据，所述风险数据包括血压数据、血脂数据以及血糖数据中的一种或多种；根据所述风险数据通过目标评分模型，确定所述目标用户的风险分数，所述目标评分模型为基于Xgboost算法和K-Folds交叉验证算法方法训练出的预测模型；根据所述风险分数确定所述目标用户患病风险的风险等级。该方法可以有效的预测该用户患病的风险大小，进一步的，可以及时有效的识别身体内造成疾病的潜在威胁，尽早对疾病进行早期筛查和干预，以便延缓疾病显性化和防止慢病重症化。

一种患病风险的预测方法及装置

[0001] 本申请要求于2019年10月15日提交中国专利局、申请号为201910980973.9，发明名称为“一种患病风险的预测方法及装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

[0002] 本申请涉及人工智能技术领域，尤其涉及一种患病风险的预测方法及装置。

背景技术

[0003] 心血管疾病是全球人类健康的首要疾病负担。我国心血管病的发病率和死亡率呈快速上升趋势。据估计高血压、高血脂、高血糖这最常见的三种心血管疾病，也可以说“三高”疾病是其他大多数心脑血管疾病的元凶，因此，“三高”疾病的防治已经成为预防心血管疾病、保障身体健康的重要需求，目前医院的“三高”预防检查是根据传统测量的具体数据和医生的专业知识来获得。但是发明人意识到，对于“三高”疾病患者，仅某项影响因素指标超标时，“三高”疾病的特性有可能未能及时展现出来，导致人们对于患病的风险认知并不清楚，不能及时防治。

[0004] 因此根据传统测量的具体数据和医生的专业知识来诊断防治疾病的方法就显得为时已晚。进一步的，一旦患病后再治疗，对于用户的身体及经济成本都会造成很大的损失。所以，如何及时有效的识别身体内造成疾病的潜在威胁，尽早对疾病产生的原因进行早期筛查和干预，以便延缓疾病显性化和防止慢病重症化。

发明概述

技术问题

[0005] 鉴于上述问题，提出了本申请以便提供一种克服上述问题或者至少部分地解决上述问题的一种患病风险的预测方法及装置。

问题的解决方案

技术解决方案

[0006] 第一方面，本申请实施例提供了一种患病风险的预测方法，可包括：接收关键因素，所述关键因素用于指示目标疾病的疾病类型或指示影响所述目标疾病的影响因子；根据所述关键因素对目标用户的体检数据进行数据筛选并处理，获得风险数据，所述风险数据包括血压数据、血脂数据以及血糖数据中的一种或多种；根据所述风险数据通过目标评分模型，确定所述目标用户的风险分数，所述目标评分模型为基于Xgboost算法和K-Folds交叉验证算法方法训练出的预测模型；根据所述风险分数确定所述目标用户患病风险的风险等级。

[0007] 第二方面，本申请实施例提供了一种患病风险的预测装置，可包括：接收单元，用于接收关键因素，所述关键因素用于指示目标疾病的疾病类型或指示影响所述目标疾病的影响因子；处理单元，用于根据所述关键因素对目标用户的体检数据进行数据筛选并处理，获得风险数据，所述风险数据包括血压数据、血脂数据以及血糖数据中的一种或多种；第一确定单元，用于根据所述风险数据通过目标评分模型，确定所述目标用户的风险分数，所述目标评分模型为基于Xgboost算法和K-Folds交叉验证算法装置训练出的预测模型；第二确定单元，用于根据所述风险分数确定所述目标用户患病风险的风险等级。

[0008] 第三方面，本申请实施例提供了一种患病风险的预测装置，包括存储组件，处理组件和通信组件，存储组件，处理组件和通信组件相互连接，其中，存储组件用于存储计算机程序，通信组件用于与外部设备进行信息交互；处理组件被配置用于调用计算机程序，执行以下步骤：接收关键因素，所述关键因素用于指示目标疾病的疾病类型或指示影响所述目标疾病的影响因子；根据所述关键因素对目标用户的体检数据进行数据筛选并处理，获得风险数据，所述风险数据包括血压数据、血脂数据以及血糖数据中的一种或多种；根据所述风险数据通过目标评分模型，确定所述目标用户的风险分数，所述目标评分模型为基于Xgboost算法和K-Folds交叉验证算法方法训练出的预测模型；根据所述风险分数确定所述目标用户患病风险的风险等级。

[0009] 第四方面，本申请实施例提供了一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序被处理器执行以实现以下步骤：接收关键因素，所述关键因素用于指示目标疾病的疾病类型或指示影响所述目标

疾病的影响因子；根据所述关键因素对目标用户的体检数据进行数据筛选并处理，获得风险数据，所述风险数据包括血压数据、血脂数据以及血糖数据中的一种或多种；根据所述风险数据通过目标评分模型，确定所述目标用户的风险分数，所述目标评分模型为基于Xgboost算法和K-Folds交叉验证算法方法训练出的预测模型；根据所述风险分数确定所述目标用户患病风险的风险等级。

发明的有益效果

有益效果

[0010] 采用本申请实施例可以有效的预测用户患病的风险大小，进一步的，可以及时有效的识别身体内因“三高”因素等造成慢性疾病的潜在威胁。

对附图的简要说明

附图说明

[0011] 为了更清楚地说明本申请实施例或背景技术中的技术方案，下面将对本申请实施例或背景技术中所需要使用的附图进行说明。

[0012] 图1是本申请实施例提供的一种患病风险的预测系统架构的示意图。

[0013] 图2A是本申请实施例提供的一种患病风险的预测方法流程的示意图。

[0014] 图2B是本申请实施例提供的一种终端设备的应用界面示意图。

[0015] 图3是本申请实施例提供的一种患病风险的预测方法流程的示意图。

[0016] 图4是本申请实施例提供的一种患病风险的预测装置的结构示意图。

[0017] 图5是本申请实施例提供的另一种患病风险的预测装置的结构示意图。

发明实施例

本发明的实施方式

[0018] 下面将结合本申请实施例中的附图，对本申请实施例进行描述。

[0019] 本申请的技术方案可应用于人工智能、智慧城市和/或数字医疗技术领域，可预测用户患病的风险，以进行健康管理，实现智慧医疗。

[0020] 本申请的说明书和权利要求书及所述附图中的术语“第一”、“第二”和“第三”等是用于区别不同对象，而不是用于描述特定顺序。此外，“包括”和“具有”以及它们任何变形，意图在于覆盖不排他的包含。例如包含了一系列步

骤或单元的过程、方法、系统、产品或设备没有限定于已列出的步骤或单元，而是可选地还包括没有列出的步骤或单元，或可选地还包括对于这些过程、方法、产品或设备固有的其它步骤或单元。

[0021] 在本文中提及“实施例”意味着，结合实施例描述的特定特征、结构或特性可以包含在本申请的至少一个实施例中。在说明书中的各个位置出现该短语并不一定均是指相同的实施例，也不是与其它实施例互斥的独立的或备选的实施例。本领域技术人员显式地和隐式地理解的是，本文所描述的实施例可以与其它实施例相结合。

[0022] 在本申请中使用的术语“服务器”、“单元”、“系统”等用于表示计算机相关的实体、硬件、固件、硬件和软件的组合、软件、或执行中的软件。例如，服务器可以是但不限于，处理器，数据处理平台，计算设备，计算机，两个或更多个计算机等。

[0023] 首先，对本申请中的部分用语进行解释说明，以便于本领域技术人员理解。

[0024] （1）XGBoost算法，全称是XtremeGradient Boosting，其中回归树(regressiontree，CART)是XGBoost最基本的组成部分，回归树是指其根据训练特征及训练数据构建分类树，判定每条数据的预测结果。其中，构建树需要使用吉尼系数（Gini index）计算增益，即进行构建树的特征选取。

[0025] （2）K-Folds交叉验证算法，又称循环验证。它将原始数据分成K组(K-Fold)，将每个子集数据分别做一次验证集，其余的K-1组子集数据作为训练集，这样会得到K个模型。这K个模型分别在验证集中评估结果，最后的误差MSE(Mean Squared Error)加和平均就得到交叉验证误差。交叉验证有效利用了有限的数据，并且评估结果能够尽可能接近模型在测试集上的表现，可以做为模型优化的指标使用。

[0026] （3）“三高”疾病，是指高血压、高血糖（糖尿病）和高脂血症。它们是现代社会所派生出来的“富贵病”，可能单独存在，也可能相互关联。如：糖尿病人很容易同时患上高血压或高血脂症，而高血脂又是动脉硬化形成和发展的主要因素，动脉硬化患者血管弹性差加剧血压升高。所以，出现这三种疾患中

的任何一种，后期都易形成了“三高”疾病。

[0027] 其次，对本申请实施例所基于的其中一种患病风险的预测系统架构进行描述。请参考附图1，图1是本申请实施例提供的一种患病风险的预测系统架构的示意图，包括：患病风险的预测装置101和终端装置102。

[0028] 患病风险的预测装置101可以包括但不限于后台服务器、组件服务器、数据处理服务器等，当上述患病风险的预测装置101为服务器时，所述服务器可以通过互联网与多个终端进行通信，接收来自终端设备的体检数据，为终端的用户提供患病风险预测的服务。例如，服务器可以接收关键因素，根据关键因素对目标用户的体检数据进行数据筛选并处理，获得风险数据，所述风险数据包括血压数据、血脂数据以及血糖数据中的一种或多种；根据所述风险数据通过目标评分模型，确定所述目标用户的风险分数，所述目标评分模型为基于Xgboost算法和K-Folds交叉验证算法方法训练出的预测模型；根据所述风险分数确定所述目标用户患病风险的风险等级。

[0029] 终端设备102可以安装并运行相关的应用。应用是指与服务器相对应，为客户提供本地服务的程序。这里，该本地服务可包括但不限于：向服务器发送体检数据和接收服务器发送的预测报告（例如：预测报告中可以包含用户的风险分数和风险等级）以及其他共享信息等等。本方案实施例中的终端可以包括但不限于任何一种基于智能操作系统的电子产品，其可与用户通过键盘、虚拟键盘、触摸板、触摸屏以及声控设备等输入设备来进行人机交互，诸如智能手机、平板电脑、个人电脑等。其中，智能操作系统包括但不限于任何通过向移动设备提供各种移动应用来丰富设备功能的操作系统，诸如：安卓(Android™)、iOS™、Windows Phone™等。

[0030] 还可以理解的是，图1的患病风险的预测系统架构只是本申请实施例中的部分示例性的实施方式，本申请实施例中的患病风险的预测系统架构包括但不仅限于以上患病风险的预测系统架构。

[0031] 参考附图2A，图2A是本申请实施例提供的一种患病风险的预测方法流程的示意图。可应用于上述图1中的系统，下面将结合图2A从患病风险的预测装置101的单侧进行描述。该方法可以包括以下步骤S201-步骤S203。

[0032] 步骤S201：接收关键因素，根据关键因素对目标用户的体检数据进行数据筛选并处理，获得风险数据。

[0033] 具体地，患病风险的预测装置101可以接收关键因素，根据所述关键因素对目标用户的体检数据进行数据筛选并处理，获得风险数据，其中，所述关键因素用于指示目标疾病的疾病类型或指示影响所述目标疾病的影响因子，例如：所述关键因素可以包括关键字或者关键因子，所述风险数据包括血压数据、血脂数据以及血糖数据中的一种或多种。可以理解的是，所述关键因素包括的影响所述目标疾病的影响因子是关键因子是可以是在体检数据中确定某些影响目标疾病的风险数据的影响因子还可以是在体检数据中影响所述目标疾病的影响因子，因此关键因素主要是指示能够对目标疾病造成影响的因素，或者是指示目标疾病的疾病类型，进一步的，可以通过该疾病类型可以筛选出目标疾病造成影响的因素。还可以理解的是，因为体检数据多种多样（例如：文字形式、图片形式、曲线形式等），所以不能够直接输入目标评分模型中；又因为体检数据往往是全面的身体检查数据，如果需要预估某一类型的疾病需要把与该疾病类型相关的体检数据筛选出来后，在针对筛选出的体检数据做患病风险预测，由此得到的风险分数才会更准确。因此，本申请实施需要对所述体检数据进行数据处理，再将处理后的数据输入目标评分模型中。

[0034] 可选的，体检数据可以包括外科基本检查数据、内科基本检查数据、口腔检查数据、超声数据、心电图数据、眼科检查数据、妇科检查数据、生化检查数据、血压测量结果、血脂测量结果以及血糖测量结果中的一种或多种。

[0035] 步骤S202：根据风险数据通过目标评分模型，确定目标用户的风险分数。

[0036] 具体地，患病风险的预测装置101可以根据所述患病风险数据通过目标评分模型，确定所述目标用户对应的风险分数，其中，所述目标评分模型为基于Xgboost算法和K-Folds交叉验证算法方法训练出的预测模型。

[0037] 可选的，所述关键因素包括M个关键字，所述M个关键字用于指示所述目标疾病的疾病类型，其中，M为大于或等于1的正整数；因此接收M个关键字，根据所述M个关键字对所述体检数据进行数据筛选，确定筛选出的与所述M个关键字对应的体检数据为所述风险数据。例如：关键字是血压、血糖或血脂等，则体检数

据中如眼科、口腔等相关特征数据可以删除，而关于血压测量结果、血脂测量结果以及血糖测量结果的体检数据可以保留。

[0038] 可选的，所述关键因素包括M个关键字，所述M个关键字用于指示所述目标疾病的疾病类型，其中，M为大于或等于1的正整数；因此接收M个关键字，将所述M个关键字对应的所述体检数据按照正则表达方式处理为0-1特征数据，获得所述风险数据。例如：关键字是血压、血糖或血脂等，则体检数据中，一些像＂脂肪肝＂，＂高血糖＂这样的文字写正则构成0-1特征，即高血糖认为1，其中血糖测量的数值型特征可以保持不变；又或者，体检数据中“血压测量结果正常”的文字也可以写正则构成0-1特征，即血压测量结果正常为0。

[0039] 步骤S203：根据风险分数确定目标用户患病风险的风险等级。

[0040] 具体地，患病风险的预测装置101可以根据所述风险分数确定所述目标用户患病风险的风险等级。

[0041] 可选的，患病风险的预测装置101根据所述风险分数确定患病风险预测报告，所述患病风险预测报告包括所述风险分数对应的所述目标用户患病风险的风险等级。请参考附图2B，图2B是本申请实施例提供的一种终端设备的应用界面示意图。目标用户可以通过终端设备102将获取的体检报告进行患病风险预测，以获取该用户的患病风险预测报告，及时了解自身的可能会患的疾病的等级，以便提早就医预防。其中，所述患病风险预测报告还可以包含所述风险等级对应的用户在该阶段需要的预防措施或治疗手段。例如：可以根据用户体检数据基于目标评分模型确定风险分数后，再根据风险分数确定患病风险等级，最后生成患病风险预测报告，根据该报告得到每个用户所处阶段需要的预防措施或治疗手段，尤其对于一些潜在患者，根据以上算法找到的一些潜在威胁因子有针对性地进行预防及干预。

[0042] 实施本申请实施例，可以接收关键因素，根据所述关键因素对目标用户的体检数据进行数据筛选并处理，获得血压数据、血脂数据以及血糖数据中的一种或多种；根据处理好的数据，通过基于Xgboost算法和K-Folds交叉验证算法方法训练出的目标评分模型，确定所述目标用户对应的风险分数；根据所述风险分数确定所述目标用户患病风险的风险等级。将处理好的血压数据、血脂数据以

及血糖数据中的一种或多种输入到风险预测模型中后，风险预测模型可以有效的预测该用户患病的风险大小，进一步的，可以及时有效的识别身体内因“三高”因素等造成慢性疾病（例如：高血压、高血糖、高血脂、心脑血管疾病等）的潜在威胁，尽早对造成该类型疾病的因素进行早期筛查和干预，以便延缓疾病显性化和防止慢病重症化。

[0043] 参考附图3，图3是本申请实施例提供的一种患病风险的预测方法流程的示意图。可应用于上述图1中的系统，下面将结合图3从患病风险的预测装置101的单侧进行描述。该方法可以包括以下步骤S301-步骤S308。

[0044] 步骤S301：获取N个用户的体检数据为样本体检数据。

[0045] 具体地，患病风险的预测装置101可以从体检中心或者其他数据平台获取N个用户的体检数据为样本体检数据，因为N个用户的体检数据是作为训练目标评分模型的样本数据，所以所述N个用户的体检数据不包括目标用户的体检数据，N为大于1的正整数。

[0046] 步骤S302：根据K-Folds交叉验证算法将获取的N个样本体检数据分成K-1个训练集及1个测试集。

[0047] 具体地，患病风险的预测装置101根据所述K-Folds交叉验证算法将获取的N个样本体检数据分成K-1个训练集及1个测试集，其中，K为大于1的正整数。通过K-Folds交叉验证算法在N个样本体检数据中得到K个训练集及测试集：将数据集分成K份，以满足每次用其中K-1份训练模型，剩余1份数据验证模型效果。

[0048] 步骤S303：基于Xgboost算法，在所述K-1个训练集上建立i个回归树训练预测模型。

[0049] 具体地，患病风险的预测装置101基于Xgboost算法，在所述K-1个训练集上建立i个回归树训练预测模型，所述预测模型用于对患病的风险大小进行评分预测，其中，i为大于1的正整数。即，在训练集上建立m个回归树进行预测：每一个回归树都是一个弱学习器，通过一步步加入新的弱学习器，来努力纠正前面所有弱学习器的残差，因此，最终这m个学习器相加在一起可以用来进行最终患病风险的预测。

[0050] 可选的，在患病风险的预测装置101基于Xgboost算法训练模型的过程中，可以

优化该训练过程中的目标函数：使得树群的预测值尽量接近真实值（即：准确率）而且有尽量大的泛化能力。需要说明的是，要满足目标函数中预测误差尽量小，即树的叶子节点尽量少，节点数值尽量不极端分布。Xgboost算法的优化过程中，找到使其损失函数最小的特征值一步步进行分裂，每个特征值的分裂节点由使得特征值的均方根误差最小的节点确定。

[0051] 步骤S304：根据1个测试集测试，通过训练后的预测模型对患病的风险大小进行评分预测。

[0052] 具体地，患病风险的预测装置101根据1个测试集测试，通过训练后的预测模型对患病的风险大小进行评分预测。即K-1个训练集训练出预测模型后需要将剩余的1个测试集数据代入训练好的预测模型中进行患病的风险大小进行评分预测，若评分预测的结果不在预设误差范围内，则所述预测模型不能够准确预测用户的患病风险，则该预测模型还需要进一步的训练。

[0053] 步骤S305：若评分预测的结果在预设误差范围内，则确定预测模型为目标评分模型。

[0054] 具体地，患病风险的预测装置101将剩余的1个测试集数据代入训练好的预测模型中进行患病的风险大小进行评分预测时，若评分预测的结果在预设误差范围内，则确定所述预测模型为所述目标评分模型，证明所述预测模型可以准确预测用户的患病风险。

[0055] 可选的，患病风险的预测装置101还可以根据所述N个样本体检数据和所述目标评分模型，确定所述N个样本体检数据对应的风险分数，根据确定所述N个样本体检数据对应的风险分数，划分风险等级并确定所述风险分数与所述风险等级的映射关系。可以理解的是，风险分数越高，器对应的风险等级越高，用户可能会患得疾病的概率也越大。因此，根据该映射关系，可以确定目标用户的风险分数对应的风险等级，进而可以判断出目标用户患病的概率。

[0056] 步骤S306：根据关键因素对目标用户的体检数据进行数据筛选并处理，获得风险数据。

[0057] 具体地，患病风险的预测装置101基于Xgboost算法，在所述K个训练集上建立i个回归树，确定所述体检数据中影响所述风险数据的s个不同的关键因子，其中

，s为大于1的正整数。所述关键因素包括所述s个不同的关键因子；所述根据所述关键因素对目标用户的体检数据进行数据筛选并处理，获得风险数据，包括：确定在所述s个关键因子的影响力值中所述影响力值最大的前j个关键因子对应的所述目标用户的体检数据，其中，j为大于1且小于s的正整数；对所述前j个关键因子对应的目标用户的体检数据进行数据筛选，获得所述风险数据。可以理解的是，可以在目标用户的体检数据中筛选出前j个关键因子对应的所述目标用户的体检数据，然后在进行数据处理后获得最终的风险数据，该风险数据可以是按照正则表达方式处理为0-1特征数据或者将文字数据转化为的数值数据。

[0058] 步骤S307：根据风险数据通过目标评分模型，确定目标用户的风险分数。

[0059] 步骤S308：根据风险分数确定目标用户患病风险的风险等级。

[0060] 具体地，步骤S307-步骤S308的实现还可以对应参照上述图2A中步骤S202-步骤S203的相关描述，此处不再赘述。

[0061] 实施本申请实施例，可以首先对用户的体检数据进行数据处理，使得体检数据为数值形式，以便计算，然后将处理好的用户数据输入到风险预测模型中，风险预测模型预测该用户患病的风险大小。其中，该风险预测模型是基于海量的样本数据利用单变量分析相关性或决策树确定样本数据中影响Y特征数据的关键因子，再将其目标关键因子以及根据K-Folds交叉验证算法以及N个样本体检数据，得到的K个训练集及测试集输入xgboost算法建立的“三高”疾病风险预测模型中训练得到的预测模型，其次可以训练好的风险评分模型确定目标用户患病的风险大小，并且可以发现一些“三高”疾病的内在规律，控制疾病影响因素，预防疾病形成，或者抑制病情发展，为解决国家医保控费难题，提升医疗卫生服务水平提供AI支撑。

[0062] 上述详细阐述了本申请实施例的方法，下面提供了与本申请实施例的相关患病风险的预测装置，患病风险的预测装置10可以是一种通过获取、处理、分析和提取有价值的数据，为第三方使用带来各种便利的服务设备。请参考附图4，图4是本申请实施例提供的一种患病风险的预测装置的结构示意图。可以包括接收单元401，处理单元402，第一确定单元403，第二确定单元404，还可以包括：

模型训练单元405，第三确定单元406，第四确定单元407。

[0063] 接收单元401，用于接收关键因素，所述关键因素用于指示目标疾病的疾病类型或指示影响所述目标疾病的影响因子。

[0064] 处理单元402，用于根据所述关键因素对目标用户的体检数据进行数据筛选并处理，获得风险数据，所述风险数据包括血压数据、血脂数据以及血糖数据中的一种或多种。

[0065] 第一确定单元403，用于根据所述风险数据通过目标评分模型，确定所述目标用户的风险分数，所述目标评分模型为基于Xgboost算法和K-Folds交叉验证算法装置训练出的预测模型。

[0066] 第二确定单元404，用于根据所述风险分数确定所述目标用户患病风险的风险等级。

[0067] 在一种可能实现的方式中，所述体检数据包括外科基本检查数据、内科基本检查数据、口腔检查数据、超声数据、心电图数据、眼科检查数据、妇科检查数据、生化检查数据、血压测量结果、血脂测量结果以及血糖测量结果中的一种或多种。

[0068] 在一种可能实现的方式中，所述关键因素包括M个关键字，所述M个关键字用于指示所述目标疾病的疾病类型，其中，M为大于或等于1的正整数；所述处理单元402具体用于：根据所述M个关键字对所述体检数据进行数据筛选，确定筛选出的与所述M个关键字对应的体检数据为所述风险数据。

[0069] 在一种可能实现的方式中，所述关键因素包括M个关键字，所述M个关键字用于指示所述目标疾病的疾病类型，其中，M为大于或等于1的正整数；所述处理单元402具体用于：将所述M个关键字对应的所述体检数据按照正则表达方式处理为0-1特征数据，获得所述风险数据。

[0070] 在一种可能实现的方式中，所述装置还包括：模型训练单元405，用于根据所述风险数据通过目标评分模型，确定所述目标用户的风险分数之前，获取N个用户的体检数据为样本体检数据，其中，N个用户不包括所述目标用户，N为大于1的正整数；根据所述K-Folds交叉验证算法将获取的N个样本体检数据分成K-1个训练集及1个测试集，其中，K为大于1的正整数；基于Xgboost算法，在所述K-1

个训练集上建立i个回归树训练预测模型，所述预测模型用于对患病的风险大小进行评分预测，其中，i为大于1的正整数；根据所述1个测试集测试，通过训练后的所述预测模型对患病的风险大小进行评分预测；若评分预测的结果在预设误差范围内，则确定所述预测模型为所述目标评分模型。

[0071] 在一种可能实现的方式中，所述装置还包括：第三确定单元406，用于基于Xgboost算法，在所述K个训练集上建立i个回归树，确定所述体检数据中影响所述风险数据的s个不同的关键因子，其中，s为大于1的正整数；所述关键因素包括所述s个不同的关键因子；所述处理单元402具体用于：确定在所述s个关键因子的影响力值中所述影响力值最大的前j个关键因子对应的所述目标用户的体检数据，其中，j为大于1且小于s的正整数；对所述前j个关键因子对应的目标用户的体检数据进行数据筛选，获得所述风险数据。

[0072] 在一种可能实现的方式中，所述装置还包括：第四确定单元407，用于根据所述N个样本体检数据和所述目标评分模型，确定所述N个样本体检数据对应的风险分数；根据确定所述N个样本体检数据对应的风险分数，划分风险等级并确定所述风险分数与所述风险等级的映射关系。

[0073] 需要说明的是，各个操作的实现还可以对应参照图2A-图3所示的方法实施例的相应描述，此处不再赘述。

[0074] 如图5所示，图5是本申请实施例提供的另一种患病风险的预测装置的结构示意图，该装置20包括至少一个处理器501，至少一个存储器502、至少一个通信接口503。此外，该设备还可以包括天线等通用部件，在此不再详述。

[0075] 处理器501可以是通用中央处理器（CPU），微处理器，特定应用集成电路（application-specific integrated circuit，ASIC)，或一个或多个用于控制以上方案程序执行的集成电路。

[0076] 通信接口503，用于与其他设备或通信网络通信，如以太网，无线接入网（RAN），核心网，无线局域网(Wireless Local Area Networks，WLAN)等。

[0077] 存储器502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。存储器可以是独立存在，通过总线与处理器相连接。存储器也可以和处理器集成在一起。

[0078] 其中，所述存储器502用于存储执行以上方案的应用程序代码，并由处理器501来控制执行。所述处理器501用于执行所述存储器502中存储的应用程序代码。

[0079] 存储器502存储的代码可执行以上图2A-图3提供的患病风险的预测方法，比如，当装置20为患病风险的预测101时，可以接收关键因素，根据所述关键因素对目标用户的体检数据进行数据筛选并处理，获得风险数据，所述风险数据包括血压数据、血脂数据以及血糖数据中的一种或多种；根据所述风险数据通过目标评分模型，确定所述目标用户的风险分数，所述目标评分模型为基于Xgboost算法和K-Folds交叉验证算法方法训练出的预测模型；根据所述风险分数确定所述目标用户患病风险的风险等级。

[0080] 需要说明的是，本申请实施例中所描述的患病风险的预测装置20中各功能单元的功能可参照图2A-图3所示的方法实施例的相应描述，此处不再赘述。

[0081] 本申请实施例还提供了一种计算机可读存储介质，其上存储有计算机程序，计算机程序被处理器执行时实现上述实施例中方法的步骤，或者，计算机程序被处理器执行时实现上述实施例中装置的各模块/单元的功能，这里不再赘述。可选的，本申请涉及的存储介质如计算机可读存储介质可以是非易失性的，也可以是易失性的。

[0082] 在本申请中，所述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申请实施例方案的目的。

[0083] 另外，在本申请各个实施例中的各功能组件可以集成在一个组件也可以是各个

组件单独物理存在，也可以是两个或两个以上组件集成在一个组件中。上述集成的组件既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。

[0084] 所述集成的组件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，本申请的技术方案本质上或者说对现有技术做出贡献的部分，或者该技术方案的全部或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）执行各个本申请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）、磁碟或者光盘等各种可以存储程序代码的介质。

[0085] 以上所述，仅为本申请的具体实施方式，但本申请的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本申请揭露的技术范围内，可轻易想到各种等效的修改或替换，这些修改或替换都应涵盖在本申请的保护范围之内。因此，本申请的保护范围应以权利要求的保护范围为准。

[0086] 应理解，在本申请的各种实施例中，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，而不应对本申请实施例的实施过程构成任何限定。尽管在此结合各实施例对本申请进行了描述，然而，在实施例所要求保护的本申请过程中，本领域技术人员可理解并实现公开实施例的其他变化。

权利要求书

[权利要求 1] 一种患病风险的预测方法，其中，包括：

接收关键因素，所述关键因素用于指示目标疾病的疾病类型或指示影响所述目标疾病的影响因子；

根据所述关键因素对目标用户的体检数据进行数据筛选并处理，获得风险数据，所述风险数据包括血压数据、血脂数据以及血糖数据中的一种或多种；

根据所述风险数据通过目标评分模型，确定所述目标用户的风险分数，所述目标评分模型为基于Xgboost算法和K-Folds交叉验证算法方法训练出的预测模型；

根据所述风险分数确定所述目标用户患病风险的风险等级。

[权利要求 2] 根据权利要求1所述方法，其中，所述体检数据包括外科基本检查数据、内科基本检查数据、口腔检查数据、超声数据、心电图数据、眼科检查数据、妇科检查数据、生化检查数据、血压测量结果、血脂测量结果以及血糖测量结果中的一种或多种。

[权利要求 3] 根据权利要求1所述方法，其中，所述关键因素包括M个关键字，所述M个关键字用于指示所述目标疾病的疾病类型，其中，M为大于或等于1的正整数；

所述根据所述关键因素对目标用户的体检数据进行数据筛选并处理，获得风险数据，包括：

根据所述M个关键字对所述体检数据进行数据筛选，确定筛选出的与所述M个关键字对应的体检数据为所述风险数据。

[权利要求 4] 根据权利要求1所述方法，其中，所述关键因素包括M个关键字，所述M个关键字用于指示所述目标疾病的疾病类型，其中，M为大于或等于1的正整数；

所述根据所述关键因素对目标用户的体检数据进行数据筛选并处理，获得风险数据，包括：

将所述M个关键字对应的所述体检数据按照正则表达方式处理为0-1特

征数据，获得所述风险数据。

[权利要求 5] 根据权利要求1所述方法，其中，所述根据所述风险数据通过目标评分模型，确定所述目标用户的风险分数之前，还包括：

获取N个用户的体检数据为样本体检数据，其中，N个用户不包括所述目标用户，N为大于1的正整数；

根据所述K-Folds交叉验证算法将获取的N个样本体检数据分成K-1个训练集及1个测试集，其中，K为大于1的正整数；

基于Xgboost算法，在所述K-1个训练集上建立i个回归树训练预测模型，所述预测模型用于对患病的风险大小进行评分预测，其中，i为大于1的正整数；

根据所述1个测试集测试，通过训练后的所述预测模型对患病的风险大小进行评分预测；

若评分预测的结果在预设误差范围内，则确定所述预测模型为所述目标评分模型。

[权利要求 6] 根据权利要求5所述方法，其中，所述方法还包括：

基于Xgboost算法，在所述K个训练集上建立i个回归树，确定所述体检数据中影响所述风险数据的s个不同的关键因子，其中，s为大于1的正整数；

所述关键因素包括所述s个不同的关键因子；所述根据所述关键因素对目标用户的体检数据进行数据筛选并处理，获得风险数据，包括：

确定在所述s个关键因子的影响力值中所述影响力值最大的前j个关键因子对应的所述目标用户的体检数据，其中，j为大于1且小于s的正整数；

对所述前j个关键因子对应的目标用户的体检数据进行数据筛选，获得所述风险数据。

[权利要求 7] 根据权利要求5所述方法，其中，所述方法还包括：

根据所述N个样本体检数据和所述目标评分模型，确定所述N个样本体检数据对应的风险分数；

根据确定所述N个样本体检数据对应的风险分数，划分风险等级并确定所述风险分数与所述风险等级的映射关系。

[权利要求 8] 一种患病风险的预测装置，其中，包括：

接收单元，用于接收关键因素，所述关键因素用于指示目标疾病的疾病类型或指示影响所述目标疾病的影响因子；

处理单元，用于根据所述关键因素对目标用户的体检数据进行数据筛选并处理，获得风险数据，所述风险数据包括血压数据、血脂数据以及血糖数据中的一种或多种；

第一确定单元，用于根据所述风险数据通过目标评分模型，确定所述目标用户的风险分数，所述目标评分模型为基于Xgboost算法和K-Folds交叉验证算法方法训练出的预测模型；

第二确定单元，用于根据所述风险分数确定所述目标用户患病风险的风险等级。

[权利要求 9] 一种患病风险的预测装置，其中，包括处理组件、存储组件和通信模组件，处理组件、存储组件和通信组件相互连接，其中，存储组件用于存储计算机程序，通信组件用于与外部设备进行信息交互；处理组件被配置用于调用计算机程序，执行以下步骤：

接收关键因素，所述关键因素用于指示目标疾病的疾病类型或指示影响所述目标疾病的影响因子；

根据所述关键因素对目标用户的体检数据进行数据筛选并处理，获得风险数据，所述风险数据包括血压数据、血脂数据以及血糖数据中的一种或多种；

根据所述风险数据通过目标评分模型，确定所述目标用户的风险分数，所述目标评分模型为基于Xgboost算法和K-Folds交叉验证算法方法训练出的预测模型；

根据所述风险分数确定所述目标用户患病风险的风险等级。

[权利要求 10] 根据权利要求9所述装置，其中，所述关键因素包括M个关键字，所述M个关键字用于指示所述目标疾病的疾病类型，其中，M为大于或等于

1的正整数；

所述处理组件在根据所述关键因素对目标用户的体检数据进行数据筛选并处理，获得风险数据时，具体用于执行以下步骤：

根据所述M个关键字对所述体检数据进行数据筛选，确定筛选出的与所述M个关键字对应的体检数据为所述风险数据。

[权利要求 11] 根据权利要求9所述装置，其中，所述关键因素包括M个关键字，所述M个关键字用于指示所述目标疾病的疾病类型，其中，M为大于或等于1的正整数；

所述处理组件在根据所述关键因素对目标用户的体检数据进行数据筛选并处理，获得风险数据时，具体用于执行以下步骤：

将所述M个关键字对应的所述体检数据按照正则表达方式处理为0-1特征数据，获得所述风险数据。

[权利要求 12] 根据权利要求9所述装置，其中，所述根据所述风险数据通过目标评分模型，确定所述目标用户的风险分数之前，所述处理组件还用于执行以下步骤：

获取N个用户的体检数据为样本体检数据，其中，N个用户不包括所述目标用户，N为大于1的正整数；

根据所述K-Folds交叉验证算法将获取的N个样本体检数据分成K-1个训练集及1个测试集，其中，K为大于1的正整数；

基于Xgboost算法，在所述K-1个训练集上建立i个回归树训练预测模型，所述预测模型用于对患病的风险大小进行评分预测，其中，i为大于1的正整数；

根据所述1个测试集测试，通过训练后的所述预测模型对患病的风险大小进行评分预测；

若评分预测的结果在预设误差范围内，则确定所述预测模型为所述目标评分模型。

[权利要求 13] 根据权利要求12所述装置，其中，所述处理组件还用于执行以下步骤：

基于Xgboost算法，在所述K个训练集上建立i个回归树，确定所述体检数据中影响所述风险数据的s个不同的关键因子，其中，s为大于1的正整数；

所述关键因素包括所述s个不同的关键因子；所述处理组件在根据所述关键因素对目标用户的体检数据进行数据筛选并处理，获得风险数据时，具体用于执行以下步骤：

确定在所述s个关键因子的影响力值中所述影响力值最大的前j个关键因子对应的所述目标用户的体检数据，其中，j为大于1且小于s的正整数；

对所述前j个关键因子对应的目标用户的体检数据进行数据筛选，获得所述风险数据。

[权利要求 14] 根据权利要求12所述装置，其中，所述处理组件还用于执行以下步骤：

根据所述N个样本体检数据和所述目标评分模型，确定所述N个样本体检数据对应的风险分数；

根据确定所述N个样本体检数据对应的风险分数，划分风险等级并确定所述风险分数与所述风险等级的映射关系。

[权利要求 15] 一种计算机可读存储介质，其中，所述计算机可读存储介质存储有计算机程序，所述计算机程序被处理器执行以实现以下步骤：

接收关键因素，所述关键因素用于指示目标疾病的疾病类型或指示影响所述目标疾病的影响因子；

根据所述关键因素对目标用户的体检数据进行数据筛选并处理，获得风险数据，所述风险数据包括血压数据、血脂数据以及血糖数据中的一种或多种；

根据所述风险数据通过目标评分模型，确定所述目标用户的风险分数，所述目标评分模型为基于Xgboost算法和K-Folds交叉验证算法计算机可读存储介质训练出的预测模型；

根据所述风险分数确定所述目标用户患病风险的风险等级。

[权利要求 16] 根据权利要求15所述计算机可读存储介质，其中，所述关键因素包括M个关键字，所述M个关键字用于指示所述目标疾病的疾病类型，其中，M为大于或等于1的正整数；

所述根据所述关键因素对目标用户的体检数据进行数据筛选并处理，获得风险数据时，具体实现以下步骤：

根据所述M个关键字对所述体检数据进行数据筛选，确定筛选出的与所述M个关键字对应的体检数据为所述风险数据。

[权利要求 17] 根据权利要求15所述计算机可读存储介质，其中，所述关键因素包括M个关键字，所述M个关键字用于指示所述目标疾病的疾病类型，其中，M为大于或等于1的正整数；

所述根据所述关键因素对目标用户的体检数据进行数据筛选并处理，获得风险数据时，具体实现以下步骤：

将所述M个关键字对应的所述体检数据按照正则表达方式处理为0-1特征数据，获得所述风险数据。

[权利要求 18] 根据权利要求15所述计算机可读存储介质，其中，所述根据所述风险数据通过目标评分模型，确定所述目标用户的风险分数之前，所述计算机程序被处理器执行还用于实现以下步骤：

获取N个用户的体检数据为样本体检数据，其中，N个用户不包括所述目标用户，N为大于1的正整数；

根据所述K-Folds交叉验证算法将获取的N个样本体检数据分成K-1个训练集及1个测试集，其中，K为大于1的正整数；

基于Xgboost算法，在所述K-1个训练集上建立i个回归树训练预测模型，所述预测模型用于对患病的风险大小进行评分预测，其中，i为大于1的正整数；

根据所述1个测试集测试，通过训练后的所述预测模型对患病的风险大小进行评分预测；

若评分预测的结果在预设误差范围内，则确定所述预测模型为所述目标评分模型。

[权利要求 19] 根据权利要求18所述计算机可读存储介质，其中，所述计算机程序被处理器执行还用于实现以下步骤：

基于Xgboost算法，在所述K个训练集上建立i个回归树，确定所述体检数据中影响所述风险数据的s个不同的关键因子，其中，s为大于1的正整数；

所述关键因素包括所述s个不同的关键因子；所述根据所述关键因素对目标用户的体检数据进行数据筛选并处理，获得风险数据时，具体实现以下步骤：

确定在所述s个关键因子的影响力值中所述影响力值最大的前j个关键因子对应的所述目标用户的体检数据，其中，j为大于1且小于s的正整数；

对所述前j个关键因子对应的目标用户的体检数据进行数据筛选，获得所述风险数据。

[权利要求 20] 根据权利要求18所述计算机可读存储介质，其中，所述计算机程序被处理器执行还用于实现以下步骤：

根据所述N个样本体检数据和所述目标评分模型，确定所述N个样本体检数据对应的风险分数；

根据确定所述N个样本体检数据对应的风险分数，划分风险等级并确定所述风险分数与所述风险等级的映射关系。

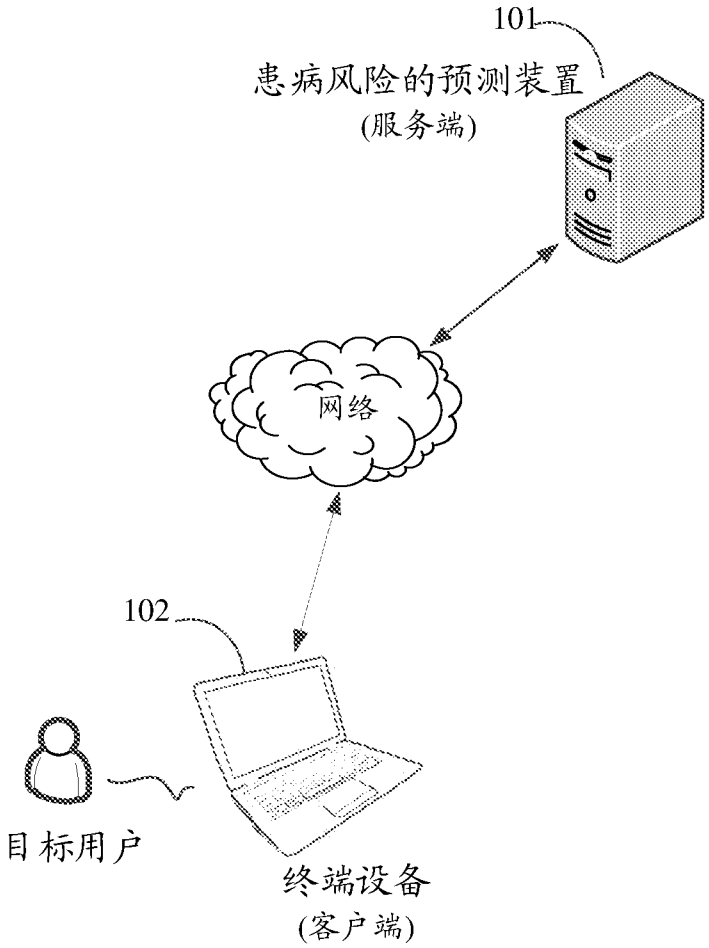

101
患病风险的预测装置
(服务端)
网络
102
目标用户
终端设备
(客户端)

图 1

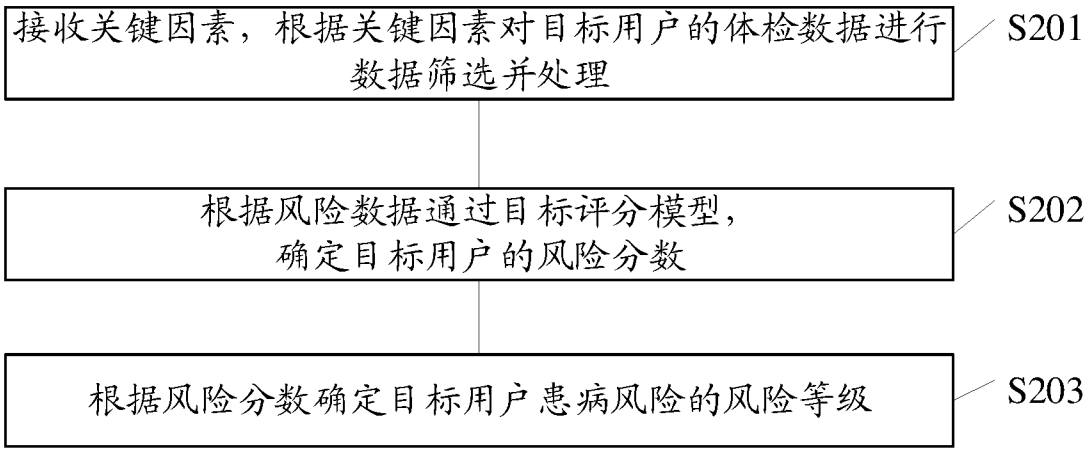

接收关键因素，根据关键因素对目标用户的体检数据进行数据筛选并处理
S201
根据风险数据通过目标评分模型，确定目标用户的风险分数
S202
根据风险分数确定目标用户患病风险的风险等级
S203

图 2A

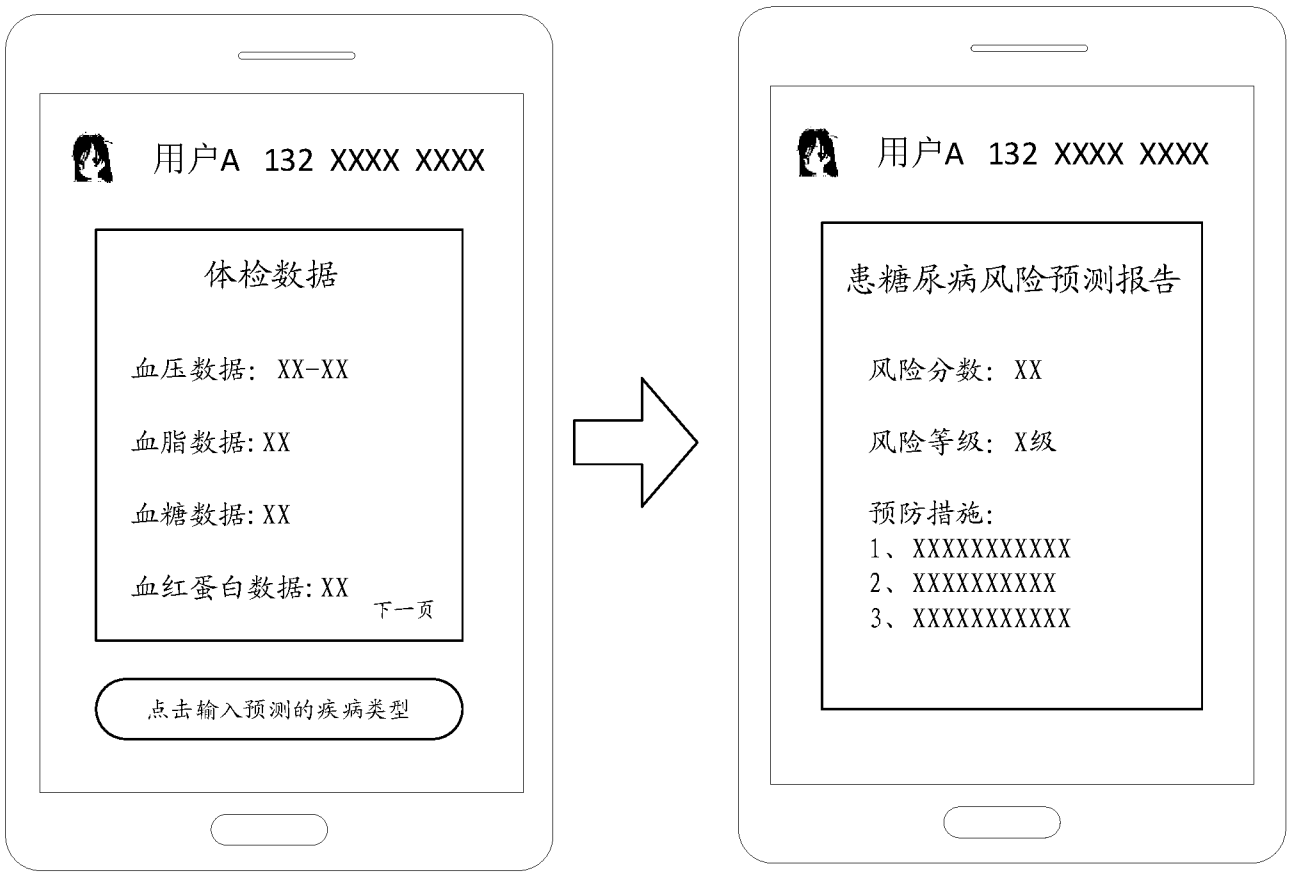

用户A 132 XXXX XXXX
体检数据
血压数据: XX-XX
血脂数据: XX
血糖数据: XX
血红蛋白数据: XX
下一页
点击输入预测的疾病类型
用户A 132 XXXX XXXX
患糖尿病风险预测报告
风险分数: XX
风险等级: X级
预防措施:
1、XXXXXXXXXX
2、XXXXXXXXX
3、XXXXXXXXXX

图 2B


获取N个用户的体检数据为样本体检数据
S301
根据K-Folds交叉验证算法
将获取的N个样本体检数据分成K-1个训练集及1个测试集
S302
基于Xgboost算法，在所述
K-1个训练集上建立i个回归树训练预测模型
S303
根据1个测试集测试，通过训练后的预测模型对患病的风险大小进行评分预测
S304
若评分预测的结果在预设误差范围内，则确定预测模型为目标评分模型
S305
根据关键因素对目标用户的体检数据
进行数据筛选并处理，获得风险数据
S306
根据风险数据通过目标评分模型，
确定目标用户的风险分数
S307
根据风险分数确定目标用户患病风险的风险等级
S308

图 3

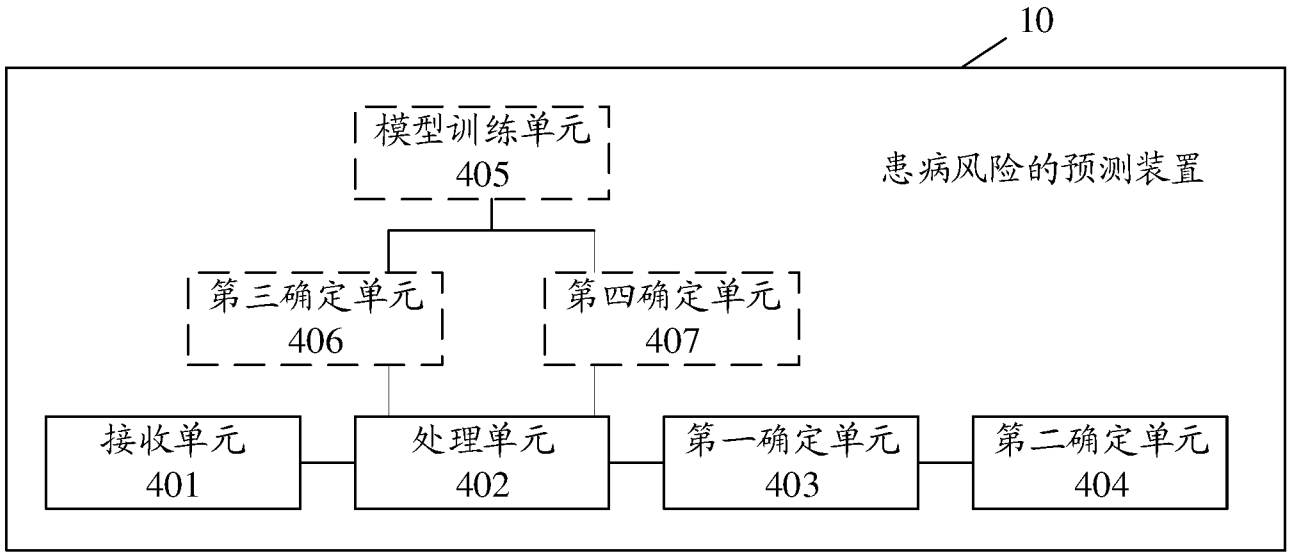

10
患病风险的预测装置
模型训练单元
405
第三确定单元
406
第四确定单元
407
接收单元
401
处理单元
402
第一确定单元
403
第二确定单元
404

图 4

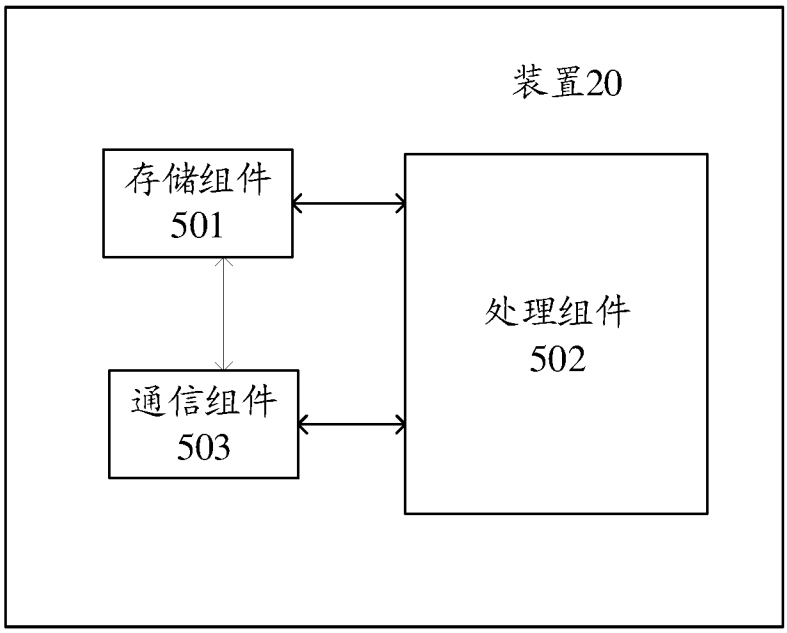

装置20
存储组件
501
处理组件
502
通信组件
503

图 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2020/111662

A. CLASSIFICATION OF SUBJECT MATTER

G16H 50/30(2018.01)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)

G16H

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)

CNKI, CNPAT, WPI, EPODOC, GOOGLE, IEEE: 患病, 疾病, 发病, 风险, 预测, 评分, 等级, 模型, 算法, 筛选, risk, disease, health, model, xgboost, sample, train

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110838366 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 25 February 2020 (2020-02-25) claims 1-10	1-20
X	CN 108257675 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 06 July 2018 (2018-07-06) description, paragraphs [0034]-[0096], and figures 1-5	1-20
A	CN 109785976 A (QINGDAO ZHONGKE HUIKANG TECHNOLOGY CO., LTD. et al.) 21 May 2019 (2019-05-21) entire document	1-20
A	US 2019192010 A1 (MANE, Viraj et al.) 27 June 2019 (2019-06-27) entire document	1-20

☐ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"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)
"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
"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
"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
"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
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 October 2020	**26 November 2020**
Name and mailing address of the ISA/CN **China National Intellectual Property Administration (ISA/CN)** **No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088** **China** Facsimile No. **(86-10)62019451**	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2020/111662

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 110838366 A	25 February 2020	None	
CN 108257675 A	06 July 2018	WO 2019153595 A1	15 August 2019
CN 109785976 A	21 May 2019	None	
US 2019192010 A1	27 June 2019	None	

国际检索报告

国际申请号
PCT/CN2020/111662

A. 主题的分类

G16H 50/30(2018.01)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)

G16H

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））

CNKI, CNPAT, WPI, EPODOC, GOOGLE, IEEE: 患病，疾病，发病，风险，预测，评分，等级，模型，算法，筛选，risk, disease, health, model, xgboost, sample, train

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
PX	CN 110838366 A (平安科技深圳有限公司) 2020年 2月 25日 (2020-02-25) 权利要求1-10	1-20
X	CN 108257675 A (平安科技深圳有限公司) 2018年 7月 6日 (2018-07-06) 说明书[0034]-[0096]段，附图1-5	1-20
A	CN 109785976 A (青岛中科慧康科技有限公司 等) 2019年 5月 21日 (2019-05-21) 全文	1-20
A	US 2019192010 A1 (MANE, Viraj 等) 2019年 6月 27日 (2019-06-27) 全文	1-20

☐ 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	国际检索报告邮寄日期
2020年 10月 30日	2020年 11月 26日
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451	受权官员 吴瑶 电话号码 86-10-53961339

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2020/111662

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 110838366 A	2020年 2月 25日	无	
CN 108257675 A	2018年 7月 6日	WO 2019153595 A1	2019年 8月 15日
CN 109785976 A	2019年 5月 21日	无	
US 2019192010 A1	2019年 6月 27日	无	