(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
2020 年 9 月 17 日 (17.09.2020)

WIPO | PCT

(10) 国际公布号
WO 2020/181805 A1

(51) 国际专利分类号:
G16H 50/20 (2018.01)

(21) 国际申请号: PCT/CN2019/117217

(22) 国际申请日: 2019 年 11 月 11 日 (11.11.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910185079.2 2019年3月12日 (12.03.2019) 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: 金晓辉 (JIN, Xiaohui); 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)。 阮晓雯 (RUAN, Xiaowen); 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)。 徐亮 (XU, Liang); 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)。 肖京 (XIAO, Jing); 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: 北京中强智尚知识产权代理有限公司 (BEIJING STRONG WISDOM INTELLECTUAL PROPERTY AGENCY LTD.); 中国北京市朝

(54) **Title:** DIABETES PREDICTION METHOD AND APPARATUS, STORAGE MEDIUM, AND COMPUTER DEVICE

(54) **发明名称:** 糖尿病的预测方法及装置、存储介质、计算机设备

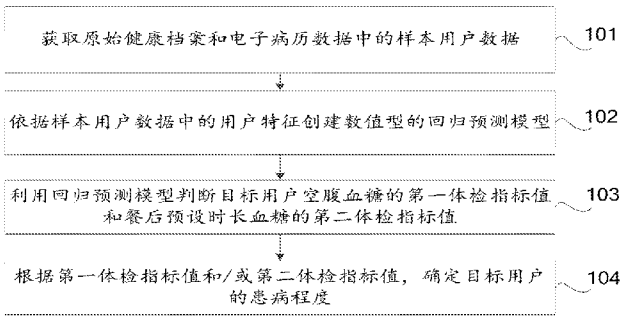


图 1

101 Acquire sample user data from original health profile and electronic medical record data

102 Create, according to user features in the sample user data, a regression prediction model of a numerical type

103 Determine, using the regression prediction model, a first physical examination index value of fasting blood glucose of a target user and a second physical examination index value of blood glucose within a pre-set time period after a meal

104 Determine, according to the first physical examination index value and/or the second physical examination index value, the disease severity of the target user

(57) **Abstract:** The present application discloses a diabetes prediction method and apparatus, a storage medium and a computer device, relating to the field of computer technology, being able to effectively solve the problem in the prior art that it can only be determined whether a user suffers from diabetes, but the severity of the disease thereof cannot be determined. Said method comprises: acquiring sample user data from original health profile and electronic medical record data; creating, according to user features in the sample user data, a regression prediction model of a numerical type; determining, using the regression prediction model, a first physical examina-

WO 2020/181805 A1

[见续页]

阳区北苑东路19号院2号楼2608室, Beijing 100012 (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, 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, 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)。

根据细则4.17的声明：
- 关于申请人有权申请并被授予专利(细则4.17(ii))
- 发明人资格(细则4.17(iv))

本国际公布：
- 包括国际检索报告(条约第21条(3))。

tion index value of fasting blood glucose of a target user and a second physical examination index value of blood glucose within a pre-set time period after a meal; and determining, according to the first physical examination index value and/or the second physical examination index value, the disease severity of the target user. The present application is applicable to the prediction of diabetes, and to the determination of the severity of diabetes.

(57) 摘要：本申请公开了一种糖尿病的预测方法及装置、存储介质、计算机设备，涉及计算机技术领域，能有效解决现有技术中只能判断用户是否患糖尿病，却无法判断其患病的严重程度的问题，其中方法包括：获取原始健康档案和电子病历数据中的样本用户数据；依据所述样本用户数据中的用户特征创建数值型的回归预测模型；利用所述回归预测模型判断目标用户空腹血糖的第一体检指标值和餐后预设时长血糖的第二体检指标值；根据所述第一体检指标值和/或所述第二体检指标值，确定所述目标用户的患病程度。本申请适用于糖尿病的预测，以及对糖尿病患病程度的确定。

糖尿病的预测方法及装置、存储介质、计算机设备

本申请要求与2019年3月12日提交中国专利局、申请号为2019101850792、申请名称为“糖尿病的预测方法及装置、存储介质、计算机设备”的中国专利申请的优先权，其全部内容通过引用结合在申请中。

技术领域

本申请涉及计算机技术领域，尤其涉及到一种糖尿病的预测方法及装置、存储介质、计算机设备。

背景技术

糖尿病是一组以高血糖为特征的代谢性疾病，发病时会导致大血管、微血管受损并危及心、脑、肾、周围神经、眼睛、足等多个部位，还会伴有多种并发症，故加强糖尿病的预测工作是完全必要的。然而随着科技的进步，病种的诊断已经不局限于医生的分析，利用人工智能来预测糖尿病，更能符合如今的发展潮流。

发明人发现目前业内对于糖尿病预测的常见方法是通过收集糖尿病医案，将糖尿病患者数据与健康人群数据进行对比，构建0-1分类模型，通过患者的各类特征维度数据，判断用户是否患糖尿病。然而现有的糖尿病预测的方法只能判断患者是否患糖尿病，却无法判断其患病的严重程度，导致诊断结果不够完善，无法根据患病程度进行配套的控制治疗，进而可能会造成患者病情的加剧恶化。

发明内容

有鉴于此，本申请提供了一种糖尿病的预测方法及装置、存储介质、计算机设备，主要目的在于解决当利用构建的0-1分类模型进行糖尿病的预测时，只能判断用户是否患糖尿病，却无法判断其患病的严重程度，进而导致诊断结果不够完善的问题。

根据本申请的一个方面，提供了一种糖尿病的预测方法，该方法包括：获取原始健康档案和电子病历数据中的样本用户数据；依据所述样本用户数据中的用户特征创建数值型的回归预测模型；利用所述回归预测模型判断目标用户空腹血糖的第一体检指标值和餐后预设时长血糖的第二体检指标值；根据所述第一体检指标值和/或所述第二体检指标值，确定所述目标用户的患病程度。

根据本申请的另一个方面，提供了一种糖尿病的预测装置，该装置包括：

获取单元，用于获取原始健康档案和电子病历数据中的样本用户数据；

创建单元，用于依据所述样本用户数据中的用户特征创建数值型的回归预测模型；

判断单元，用于利用所述回归预测模型判断目标用户空腹血糖的第一体检指标值和餐后预设时长血糖的第二体检指标值；

确定单元，用于根据所述第一体检指标值和/或所述第二体检指标值，确定所述目标用户的患病程度。

根据本申请的又一个方面，提供了一种非易失性可读存储介质，其上存储有计算机可读指令，所述计算机可读指令被处理器执行时实现上述糖尿病的预测方法。根据本申请的再一个方面，提供了一种计算机设备，包括非易失性可读存储介质、处理器及存储在非易失性可读存储介质上并可在处理器上运行的计算机可读指令，所述处理器执行所述计算机可读指令时实现上述糖尿病的预测方法。

借由上述技术方案，本申请提供的一种糖尿病的预测方法及装置、存储介质、计算机设备，与目前利用构建的0-1分类模型预测糖尿病的方法相比，本申请在现有的糖尿病预测模型的基础上，增加了餐后血糖和空腹2h血糖的回归预测模型，可利用回归预测模型判断出目标用户空腹血糖的第一体检指标值和餐后预设时长血糖的第二体检指标值，即可利用体检指标值确定出目标用户是否患糖尿病，并且还能进一步判断出目标用户的患病程度。

上述说明仅是本申请技术方案的概述，为了能够更清楚了解本申请的技术手段，而可依照说明书的内容予以实施，并且为了本申请的上述和其他目的、特征和优点能够更明显易懂，以下特举本申请的具体实施方式。

附图说明

此处所说明的附图用来提供对本申请的进一步理解，构成本申请的一部分，本申请的示意性实施例及其说明用于解释本申请，并不构成对本地申请的不当限定。在附图中：

图1示出了本申请实施例提供的一种糖尿病的预测方法的流程示意图；图2示出了本申请实施例提供的另一种糖尿病的预测方法的流程示意图；图3示出了本申请实施例提供的一种糖尿病的预测装置的结构示意图；图4示出了本申请实施例提供的另一种糖尿病的预测装置的结构示意图。

具体实施方式

下文中将参考实施例并结合附图来详细说明本申请。需要说明的是，在不冲突的情况下，本申请中的实施例及实施例中的特征可以相互结合。针对目前利用构建的0-1分类模型来对糖尿病进行预测时，无法根据用户数据判断出糖尿病患病严重程度的问题，本实施例提供了一种糖尿病的预测方法，如图1所示，该方法包括：

101、获取原始健康档案和电子病历数据中的样本用户数据。其中，样本用户数据可包括患者就诊数据、体检指标数据、用药数据和健康告知数据等。

102、依据样本用户数据中的用户特征创建数值型的回归预测模型。其中，用户特征可包含餐后血糖和空腹2h血糖、血压、皮脂厚度、胰岛素、BMI身体质量指数、糖尿病遗传信息、年龄、诊断结果等多类特征维度数据。在具体的实施方式中，回归预测模型可采用多种基于决策树的不同框架模型进行融合构建，即采用集成学习思想将多个基于决策树的预测模型聚集在一起，以提高预测结果的准确率。决策树是属于机器学习监督学习分类算法中比较简单的一种，决策树是预测模型；它代表的是对象属性与对象值之间的一种映射关系。树中每个节点表示某个对象，而每个分叉路径则代表的某个可能的属性值，而每个叶结点则对应从根节点到该叶节点所经历的路径所表示的对象的值。决策树仅有单一输出，若欲有复数输出，可以建立独立的决策树以处理不同输出。决策树算法有ID3，C4.5，CART算法，共同点为都是贪心算法，区别为度量方式不同，就比如ID3使用了信息获取量作为度量方式，而C4.5使用最大增益率。通过创建得到的回归预测模型可以很好的反应出不同血压、皮脂厚度、胰岛素、BMI身体质量指数、糖尿病遗传信息、年龄、诊断结果等的样本用户分别对应的餐后血糖值和空腹2h血糖值。

103、利用回归预测模型判断目标用户空腹血糖的第一体检指标值和餐后预设时长血糖的第二体检指标值。其中，目标用户为需要进行糖尿病病情预测的用户；第一体检指标值对应目标用户空腹血糖的数据检验结果；第二体检指标值对应目标用户餐后预设时长血糖的数据检验结果；预设时长可根据实际需求确定。对于本实施例，基于不同特征的样本用户反应出的餐后血糖值和空腹2h血糖值，将目标用户的特征与样本用户的特征进行匹配，找到匹配样本用户特征对应的餐后血糖值和空腹2h血糖值。

104、根据第一体检指标值和/或第二体检指标值，确定目标用户的患病程度。

在具体的应用场景中，可根据第一体检指标值判断目标用户空腹血糖是否正常，根据第二体检指标值判断目标用户餐后预设时长血糖是否正常，当第一体检指标值和/或第二体检指标值显示异常时，即可判断出用户患有糖尿病，并且还可以通过与临界值进行比较，进一步判断患者的患病程度。通过本实施例中糖尿病的预测方法，可以根据样本用户数据中的用户特征创建数值型的回归预测模型，利用回归预测模型判断目标用户空腹血糖的第一体检指标值和餐后预设时长血糖的第二体检指标值，并根据第一体检指标值和/或第二体检指标值，确定出目标用户是否患病以及患病的严重程度，使病情诊断结果更为精准，诊断内容更加完善，便于根据糖尿病的不同发展程度进行及时有效的配套治疗，进而遏制病情发展。

进一步的，作为上述实施例具体实施方式的细化和扩展，为了完整说明本申请实施例中的具体实施过程，提供了另一种糖尿病的预测方法，如图2所示，该方法包括：

201、获取原始健康档案和电子病历数据中的样本用户数据。例如，在原始健康档案和电子病历数据中共获取100项左右用户特征完整的样本用户数据，然后对样本用户数据做进一步分析处理。

下面分两种预测方式进行说明，一种是利用空腹血糖值这一体检指标值进行预测（即步骤202a至205a所示过程），另一种是利用餐后两小时血糖值这一体检指标值进行预测（即步骤202b至205b所示过程）。

202a、将样本用户的用户特征中空腹血糖值作为标签信息Y1，并将样本用户除空腹血糖值和餐后两小时血糖值以外的目标特征数据作为特征信息X1，创建第一模型训练集。其中，用户特征是利用正则表达式从样本用户数据中提取的，目标特征数据至少包括样本用户的患病史数据、住院数据、就诊用药数据、体检数据、健康告知数据中的一项或多项，例如可包括用户的患病史记录、住院记录、用药情况、体检情况、健康告知等相关信息。创建得到的第一模型训练集中包含各个特征信息X1，以及各自对应的标签信息Y1。即不同患病史记录、住院记录、用药情况、体检情况、健康告知等的样本用户分别对应的空腹血糖值。

203a、通过第一模型训练集基于预设回归预测算法训练用于判断第一体检指标值的第一识别模型。

其中，预设回归预测算法由随机森林（Random Forest）、梯度提升决策树（Gradient Boosting Decision Tree，GBDT）、Xgboost、LightGBM四种算法融合得到，第一识别模型的评估采用平均绝对百分比误差（MAPE）指标，当第一识别模型对应的MAPE指标值小于预置标准比较阈值时，确定第一识别模型符合评估标准。MAPE指标用于评估模型预测值和真实值之间的误差。常见的回归模型评估指标有MAP、

MSE、RMSE和MAPE，但MAP、MSE和RMSE只考虑误差的值，MAPE还考虑了误差与真实值之间的比例，其计算公式为：

$$MAPE = \sum_{t=1}^{N}\left|\frac{X-Y}{X}\right| * \frac{100}{N}$$

，

在上面公式中，N为样本总数，X为实测值，Y为模拟值。MAPE值越小，说明模型预测值和真实值之间的误差越小，在具体实施方式中，可根据实际情况设定标准比较阈值，当MAPE小于标准比较阈值时，说明第一识别模型符合评估标准。通过符合评估标准的识别模型进行预测，可保证预测结果的准确性。通过符合评估标准的第一识别模型可确定特征信息X1和标签信息Y1之间的第一映射关系。为了说明利用上述四种算法融合得到的预设回归预测算法训练得到第一识别模型的过程，作为一种可选方式，该过程具体可包括：

（1）采用随机采样方式从第一模型训练集中分别获取第一训练样本集、第二训练样本集、第三训练样本集、第四训练样本集，例如从第一模型训练集中随机抽取n个训练样本，共进行四轮抽取，得到四个训练集。（四个训练集之间相互独立，元素可以有重复）；

（2）基于第一训练样本集利用随机森林算法，训练得到第一分类器；基于第二训练样本集利用GBDT算法，训练得到第二分类器；基于第三训练样本集利用Xgboost算法，训练得到第三分类器；基于第四训练样本集利用LightGBM算法，训练得到第四分类器；其中，每个训练样本集中都包含有不同的特征信息X1，以及各自对应的标签信息Y1，这四种分类器的训练过程可基于各自对应的模型训练算法训练得到，且得到的这四种分类器都可单独进行用户糖尿病的预测，即输入待测用户的特征数据（具体内容对应特征信息X1），通过分类器找到相对应的标签信息Y1。

对于第一分类器的具体训练过程：①从第一训练样本集中使用Bootstraping方法随机有放回采样选出m个样本，共进行n次采样，生成n个训练集；②对于n个训练集，分别训练n个决策树模型（可通过ID3算法、C4.5算法、CART算法等现有算法构建）；③对于单个决策树模型，假设训练样本特征的个数为n，那么每次分裂时根据信息增益/信息增益比/基尼指数选择最好的特征进行分裂；④每棵树都一直这样分裂下去，直到该节点的所有训练样例都属于同一类，在决策树的分裂过程中不需要剪枝；⑤将生成的多棵决策树组成随机森林。对于回归问题，由多棵树预测值的均值决定最终预测结果，即作为第一分类器的预测结果。

对于第二分类器的具体训练过程：输入第二训练样本T={(x1,y1),(x2,y2),...(xm,ym)}T={(x1,y1),(x2,y2),...(xm,ym)}，最大迭代次数T，损失函数L。输出是强学习器f(x)：

A) 初始化弱学习器

$$f_0(x) = \underbrace{\arg\min}_{c}\sum_{i=1}^{m} L(y_i, c)$$

，其中，c为设定常数。

B) 对迭代轮数t=1,2,...T有：

a)对样本i=1,2，...m，计算负梯度r_{ti}

$$r_{ti} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f(x)=f_{t-1}(x)}$$

b)利用(xi,rti)(i=1,2,..m)，拟合一颗CART回归树，得到第t颗回归树，其对应的叶子节点区域为Rtj,j=1,2,...,J。其中J为回归树t的叶子节点的个数。

c) 对叶子区域j =1,2,..J,计算最佳拟合值c_{tj}

$$c_{tj} = \underbrace{\arg\min}_{c}\sum_{x_i \in R_{tj}} L(y_i, f_{t-1}(x_i) + c)$$

d) 更新强学习器

$$f_t(x) = f_{t-1}(x) + \sum_{j=1}^{J} c_{tj} I(x \in R_{tj})$$

，其中，I为所有叶子节点区域Rtj的训练样本的集合。

C) 得到强学习器f(x)的表达式

$$f(x) = f_T(x) = f_0(x) + \sum_{t=1}^{T}\sum_{j=1}^{J} c_{tj} I(x \in R_{tj})$$

基于上述强学习器f(x)，训练得到第二分类器。

对于第三分类器的具体训练过程：

A)建立初始模型，具体如下式：

$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in F$，其中，k表示树的个数，F表示构建的每个树结构，xi表示第i个样本，xi在每个树上的得分值的和就是xi的预测值，$\hat{y}_i$为预测值。

该初始模型的目标函数为

$Obj(\Theta)^t = \sum_{i}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$，yi为xi对应的样本实际值。

B）随着树的增长，通过t轮的公式递推，得到最终目标函数为

$Obj^t \approx \sum_{j=1}^{T}[G_j w_j + \frac{1}{2}(H_j + \lambda) w_j^2] + \gamma T$，其中，$G_j = \sum_{j=1}^{T} g_i$，$H_j = \sum_{i \in I_j} h_i$，$g_i = \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)})$，$h_i = \partial^2_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)})$，$I_j$ 表示：第 j 个叶子中包括的所有样本，wj表示第 j 个叶子的权重，γT 对应叶子的个数。

C）利用上述初始模型代入第三训练样本集数据进行拟合训练，并利用上述最终目标函数衡量模型拟合训练数据的好坏程度（即利用目标函数计算损失函数，损失函数越小说明模型能够较好的拟合训练数据），使得模型的偏差和方差得到标准要求，即最终训练得到第三分类器。

对于第四分类器的具体训练过程： A)使用现有的LightGBM算法拟合第四样本训练集中的数据，并对每次拟合后得到模型利用从第四样本训练集中选出的测试集进行测试，得到相应的决定系数以及均方误差值； B）在该决定系数大于一定阈值和均分误差值小于一定阈值时，确定拟合后的模型符合标准，并将符合标准的模型确定为第四分类器。

（3）最后将第一分类器、第二分类器、第三分类器、第四分类器利用套袋法（bagging）进行融合处理，得到第一识别模型。具体融合处理方式是通过投票表决的过程，即采用大多数原则，少数服从多数。例如，对于这四种分类器，在输入待测用户的患病史数据、住院数据、就诊用药数据、体检数据、健康告知数据后，如果四个分类器中有三个分类器得到的预测结果对应的空腹血糖值符合患有糖尿病的标准，那么可确定待测用户患有糖尿病；如果只有一个分类器得到的预测结果对应的空腹血糖值符合患有糖尿病的标准，另外三个分类器对应的空腹血糖值不符合该标准，那么可确定待测用户没有患糖尿病。

需要说明的是，如果第一识别模型的MAPE指标值大于预置标准比较阈值，即训练得到的第一识别模型不符合评估标准，那么可重新划分第一模型训练集，得到新的第一训练样本集、第二训练样本集、第三训练样本集、第四训练样本集，然后利用新的第一训练样本集继续训练第一分类器，且利用新的第二训练样本集继续训练第二分类器，且利用新的第三训练样本集继续训练第三分类器，且利用新的第四训练样本集继续训练第四分类器，然后再通过新训练得到的这四个分类器融合处理，判定新的第一识别模型的MAPE指标值是否小于预置标准比较阈值，如果仍大于预置标准比较阈值，则重复执行上述重复划分模型训练集、以及更新训练分类器的过程，直至最新得到的第一识别模型的MAPE指标值大于预置标准比较阈值，即符合评估标准。

204a、将目标用户的特征信息输入到第一识别模型中与特征信息X1进行相似度匹配。其中，目标用户的特征信息对应目标用户除空腹血糖值和餐后两小时血糖值以外的目标特征数据。作为一种可选方式，步骤204a具体可包括：将目标用户的特征信息经过数据清洗、特征提取、缺失值填充、异常值处理，得到结构化数据的特征信息；将结构化数据的特征信息与特征信息X1进行相似度匹配。

由于目标用户的特征信息有时是包含无用数据、和/或存在缺失值、和/或存在异常值的，即不适合利用第一识别模型直接进行预测的非结构化数据。因此，首先可对目标用户的特征信息进行数据清洗，清除无用数据（如去除用户现住处所在地、户口所在地等数据，只保留患病史数据、住院数据、就诊用药数据、体检数据、健康告知数据等）；再对保留的数据进行特征提取（如提取患病史数据、住院数据、

就诊用药数据、体检数据、健康告知数据等）；如果提取的特征数据中存在缺失值时可利用0值进行填充（如用户体检数据中身高和体重一项空缺，可利用0值填充，这样后续与模型中特征信息X1匹配时保证具有可比性，避免特征匹配时产生无法匹配的错误）；如果提取的特征数据中存在异常值可参考实际情况进行修正处理（如住院时长一项为99999天，明显存在异常，可进一步通过住院开始时间和结束时间计算正确的住院时长，然后进行修改处理）。通过本可选方式中的数据清洗、特征提取、缺失值填充、异常值处理等一系列处理，可保证得到与第一识别模型中特征信息X1匹配时具有可比性的结构化数据，避免特征匹配时产生无法匹配的错误，去除异常值，提高特征匹配的精确度。

205a、利用相似度大于预设阈值、且相似度最高的特征信息X1和第一映射关系，确定目标用户对应的第一体检指标值。其中，预设阈值可根据实际需求预先设置。例如，预设阈值设置得越大，相应的特征匹配精度越高，如果相似度为100%，则说明特征完全匹配。例如，在第一识别模型输入目标用户的患病史数据、住院数据、就诊用药数据、体检数据、健康告知数据后，相当于将这些数据分别输入到步骤203a的四个分类器中并与分类器各自对应的特征信息进行相似度匹配，分别找到最相似且大于一定阈值的特征信息，进而通过这四个分类器分别求出各自对应的体检指标值，即目标用户的空腹血糖值，如果这四个空腹血糖值中有三个空腹血糖值符合患有糖尿病的标准，那么可确定目标用户患有糖尿病，并计算这三个空腹血糖值的平均值作为第一识别模型计算得到的第一体检指标值；如果这四个空腹血糖值中有两个空腹血糖值不符合患有糖尿病的标准，另外两个空腹血糖值符合患有糖尿病的标准，那么计算这四个空腹血糖值的平均值作为第一识别模型计算得到的第一体检指标值，并依据这个平均值判定目标用户是否患有糖尿病。

与步骤202a并列的步骤202b、将样本用户的用户特征中餐后两小时血糖值作为标签信息Y2，并将样本用户除空腹血糖值和餐后两小时血糖值以外的目标特征数据作为特征信息X2，创建第二模型训练集。需要说明的是，步骤202b与步骤202a类似，样本用户的目标特征数据至少包括样本用户的患病史数据、住院数据、就诊用药数据、体检数据、健康告知数据中的一项或多项。并且创建得到的第二模型训练集中包含各个特征信息X2，以及各自对应的标签信息Y2。即不同患病史记录、住院记录、用药情况、体检情况、健康告知等的样本用户分别对应的餐后两小时血糖值。

203b、通过第二模型训练集基于预设回归预测算法训练用于判断第二体检指标值的第二识别模型。其中，第二识别模型的评估同样采用MAPE指标，当第二识别模型对应的MAPE指标值小于预定标准比较阈值时，确定第二识别模型符合评估标准，通过符合评估标准的第二识别模型可确定特征信息X2和标签信息Y2之间的第二映射关系。作为一种可选方式，该203b步骤的具体过程可包括：（1）采用随机采样方式从第二模型训练集中分别获取第五训练样本集、第六训练样本集、第七训练样本集、第八训练样本集；（2）基于第五训练样本集利用随机森林算法，训练得到第五分类器；基于第六训练样本集利用GBDT算法，训练得到第六分类器；基于第七训练样本集利用Xgboost算法，训练得到第七分类器；基于第八训练样本集利用LightGBM算法，训练得到第八分类器；（3）将第五分类器、第六分类器、第七分类器、第八分类器利用套袋法进行融合处理，得到第二识别模型。

与步骤203a中的可选方式类似，具体融合处理方式也是通过投票表决的过程，即采用大多数原则，少数服从多数。例如，对于这四种分类器，在输入待测用户的患病史数据、住院数据、就诊用药数据、体检数据、健康告知数据后，如果四个分类器中有三个分类器得到的预测结果对应的餐后两小时血糖值符合患有糖尿病的标准，那么可确定待测用户患有糖尿病；如果只有一个分类器得到的预测结果对应的餐后两小时血糖值符合患有糖尿病的标准，另外三个分类器对应的餐后两小时血糖值不符合该标准，那么可确定待测用户没有患糖尿病。需要说明的是，如果第二识别模型的MAPE指标值大于预置标准比较阈值，即训练得到的第二识别模型不符合评估标准，那么可重新划分第一模型训练集，得到新的第五训练样本集、第六训练样本集、第七训练样本集、第八训练样本集，然后利用新的第五训练样本集继续训练第五分类器，且利用新的第六训练样本集继续训练第六分类器，且利用新的第七训练样本集继续训练第七分类器，且利用新的第八训练样本集继续训练第八分类器，然后再通过新训练得到的这四个分类器融合处理，判定新的第二识别模型的MAPE指标值是否小于预置标准比较阈值，如果仍大于预置标准比较阈值，则重复执行上述重复划分模型训练集、以及更新训练分类器的过程，直至最新得到的第二识别模型的MAPE指标值大于预置标准比较阈值，即符合评估标准。

204b、将目标用户的特征信息输入到第二识别模型中与特征信息X2进行相似度匹配。在本步骤中，目标用户的特征信息对应目标用户除空腹血糖值和餐后两小时血糖值以外的目标特征数据。作为一种可选方式，步骤204b具体可包括：将目标用户的特征信息经过数据清洗、特征提取、缺失值填充、异常值处理，得到结构化数据的特征信息；将结构化数据的特征信息与特征信息X2进行相似度匹配。

与步骤204a中的可选方式类似，通过本可选方式中的数据清洗、特征提取、缺失值填充、异常值处理等一系列处理，可保证得到与第二识别模型中特征信息X2匹配时具有可比性的结构化数据，避免特

征匹配时产生无法匹配的错误，去除异常值，提高特征匹配的精确度。

205b、利用相似度大于预定阈值、且相似度最高的特征信息X2和第二映射关系，确定目标用户对应的第二体检指标值。其中，预定阈值可根据实际需求预先设置。例如，预定阈值设置得越大，相应的特征匹配精度越高，如果相似度为100%，则说明特征完全匹配。例如，在第二识别模型输入目标用户的患病史数据、住院数据、就诊用药数据、体检数据、健康告知数据后，相当于将这些数据分别输入到步骤203b的四个分类器中并与分类器各自对应的特征信息进行相似度匹配，分别找到最相似且大于一定阈值的特征信息，进而通过这四个分类器分别求出各自对应的体检指标值，即目标用户的餐后两小时血糖值，如果这四个餐后两小时血糖值中有三个餐后两小时血糖值符合患有糖尿病的标准，那么可确定目标用户患有糖尿病，并计算这三个餐后两小时血糖值的平均值作为第二识别模型计算得到的第二体检指标值；如果这四个餐后两小时血糖值中有两个餐后两小时血糖值不符合患有糖尿病的标准，另外两个餐后两小时血糖值符合患有糖尿病的标准，那么计算这四个餐后两小时血糖值的平均值作为第二识别模型计算得到的第二体检指标值，并依据这个平均值判定目标用户是否患有糖尿病。

206、根据第一体检指标值和/或第二体检指标值，确定目标用户的患病程度。

作为一种可选方式，步骤206具体可包括：若目标用户对应的第一体检指标值大于等于第一预设阈值，和/或第二体检指标值大于等于第二预设阈值，则确定目标用户患有糖尿病；然后通过第一体检指标值所处的第一数值区间、和/或第二体检指标值所处的第二数值区间，判断目标用户的患病程度。其中，第一预设阈值为根据空腹血糖判断糖尿病的设定标准来确定的，如7.0mmol/L；第二预设阈值为根据餐后两小时血糖判断糖尿病的设定标准来确定的，如11.1mmol/L。例如，若确定目标用户第一预设阈值为8.0mmol/L，第二预设阈值为7.6mmol/L，因第一体检指标值大于第一预设阈值，故可确定目标用户患有糖尿病；若确定目标用户第一预设阈值为5.7mmol/L，第二预设阈值为11.9mmol/L，因第二体检指标值大于第二预设阈值，故可确定目标用户患有糖尿病；若确定目标用户第一预设阈值为8.3mmol/L，第二预设阈值为11.7mmol/L，因第一体检指标值大于第一预设阈值，第二体检指标值大于第二预设阈值，故可确定目标用户患有糖尿病。

而对于糖尿病的患病程度，下面分三种情况进行讨论：（1）只用第一体检指标值判断，即通过第一体检指标值所处的第一数值区间，判断所述目标用户的患病程度，具体可包括：划分大于第一预设阈值，且按照预定数值规律递增的多个数值区间；创建多个数值区间与糖尿病患病程度之间的第三映射关系；确定第一体检指标值对应处于多个数值区间中的第一数值区间；根据第三映射关系以及第一数值区间，判断目标用户的第一糖尿病患病程度。例如，设定大于第一预设阈值7.0mmol/L的多个数值区间以及第三映射关系中对应的糖尿病患病程度分别为：轻度糖尿病：7.0~8.4mmol/L，中度糖尿病：8.4~10.1mmol/L，重度糖尿病：大于10.11mmol/L。若确定第一体检指标值为9.6mmol/L，则可判断出第一体检指标值处于的第一数值区间为：8.4~11.1mmol/L，则根据第三映射关系以及第一数值区间，可判断出目标用户的糖尿病的患病程度为中度糖尿病。（2）只用第二体检指标值判断，即通过第二体检指标值所处的第二数值区间，判断目标用户的患病程度，具体包括：划分大于第二预设阈值，且按照预定数值规律递增的多个数值区间；创建多个数值区间与糖尿病患病程度之间的第四映射关系；确定第二体检指标值对应处于多个数值区间中的第二数值区间；根据第四映射关系以及第二数值区间，判断目标用户的第二糖尿病患病程度。例如，设定大于第二预设阈值11.1mmol/L的多个数值区间以及第四映射关系中对应的糖尿病患病程度分别为：中度糖尿病：11.1~16.7mmol/L，重度糖尿病：大于16.7mmol/L（当大于16.7mmol/L时容易出现酮症酸中毒的现象）。若确定第二体检指标值为12.6mmol/L，则可判断出第二体检指标值处于的第二数值区间为：11.1~16.7mmol/L，则根据第四映射关系以及第二数值区间，可判断出目标用户的糖尿病的患病程度已经为中度糖尿病。（3）结合第一体检指标值和第二体检指标值进行综合判定（这种判定方式由于考虑多方面因素，因此预测精度相对较高），即通过第一体检指标值所处的第一数值区间、和第二体检指标值所处的第二数值区间，判断目标用户的患病程度，具体包括：若第一糖尿病患病程度和第二糖尿病患病程度相同，则按照二者相同的糖尿病患病程度确定最终患病程度。若第一糖尿病患病程度和第二糖尿病患病程度不同，则依据用户对通过第一识别模型和第二识别模型这两种预测方式反馈的准确率或采纳率，分别获取第一识别模型对应的第一权重和第二识别模型对应的第二权重；在第一权重大于第二权重时，将第一糖尿病患病程度确定为目标用户的患病程度；在第二权重大于第一权重时，将第二糖尿病患病程度确定为目标用户的患病程度。

在本实施例中，两种预测方式各自对应的权重可依据用户反馈的准确率或采纳率进行设定。具体可统计不同的准确率或采纳率分别对应的权重值，然后通过统计得到的映射关系，找到预测方式相应的权重。对于本实施例，依据用户反馈的准确率或采纳率，可准确反映出哪种预测方式预测精度更高，进而选用预测精度更高的预测方式得到的预测结果作为最终判定结果，更加准确。除此之外，还可人为根据实际情况预先设定两种预测方式各自对应的权重。例如，根据用户反馈的结果，发现利用第一体检指标

值预测糖尿病患病程度的准确率较高，则可为第一体检指标值预测方式配置的权重为70%，为第二体检指标值预测方式配置的权重为30%，当两种预测产生的结果不同时可将第一体检指标值预测的结果反馈给目标用户，作为最终诊断结果。假设第一体检指标值预测的为中度糖尿病，第二体检指标值预测的为重度糖尿病，则根据配置的权重高低，最终确定目标用户的糖尿病患病程度为中度糖尿病。后续在获得目标用户的实际空腹血糖值和餐后两小时血糖值后，还可作为新的样本训练集对本实施例中的两个识别模型继续进行训练，以达到预测精度更高的效果。通过上述糖尿病的预测方法，可以通过对模型训练集进行训练，确定特征信息和标签信息之间的映射关系，将目标用户的结构化数据与回归预测模型进行匹配，进而通过映射关系确定空腹血糖的第一体检指标值和餐后两小时的第二体检指标值，通过与第一预设阈值和第二预设阈值的数值比较，即可判断出用户是否患有糖尿病，从糖尿病确诊指标出发，不但能预测用户是否患病，还能通过第一体检指标值所处的第一数值区间、和/或第二体检指标值所处的第二数值区间，判断目标用户的患病程度，使诊断结果更加完善。

进一步的，作为图1和图2所示方法的具体体现，本申请实施例提供了一种糖尿病的预测装置，如图3所示，该装置包括：获取单元31、创建单元32、判断单元33、确定单元34。获取单元31，可用于获取原始健康档案和电子病历数据中的样本用户数据；创建单元32，可用于依据样本用户数据中的用户特征创建数值型的回归预测模型；判断单元33，可用于利用回归预测模型判断目标用户空腹血糖的第一体检指标值和餐后预设时长血糖的第二体检指标值；确定单元34，可用于根据第一体检指标值和/或第二体检指标值，确定目标用户的患病程度。

在具体实施应用场景中，为了依据样本用户数据中的用户特征创建数值型的回归预测模型，如图4所示，创建单元32，具体可包括：创建模块321、训练模块322、确定模块323。创建模块321，具体可用于将所述用户特征中空腹血糖值作为标签信息Y1，并将样本用户除所述空腹血糖值和所述餐后两小时血糖值以外的目标特征数据作为特征信息X1，创建第一模型训练集，所述目标特征数据至少包括所述样本用户的患病史数据、住院数据、就诊用药数据、体检数据、健康告知数据中的一项或多项；训练模块322，具体可用于通过所述第一模型训练集基于预设回归预测算法训练用于判断所述第一体检指标值的第一识别模型，其中，所述预设回归预测算法由随机森林、梯度提升决策树GBDT、Xgboost、LightGBM四种算法融合得到，所述第一识别模型的评估采用平均绝对百分比误差MAPE指标，当所述第一识别模型对应的MAPE指标值小于预置标准比较阈值时，确定所述第一识别模型符合评估标准；确定模块323，具体可用于通过符合评估标准的所述第一识别模型可确定所述特征信息X1和所述标签信息Y1之间的第一映射关系；创建模块321，具体还可用于将所述用户特征中餐后两小时血糖值作为标签信息Y2，并将所述样本用户的所述目标特征数据作为特征信息X2，创建第二模型训练集；训练模块322，具体还可用于通过所述第二模型训练集基于所述预设回归预测算法训练用于判断所述第二体检指标值的第二识别模型，其中，所述第二识别模型的评估采用MAPE指标，当所述第二识别模型对应的MAPE指标值小于预定标准比较阈值时，确定所述第二识别模型符合评估标准；确定模块323，具体还可用于通过符合评估标准的所述第二识别模型可确定所述特征信息X2和所述标签信息Y2之间的第二映射关系。

相应的，为了判断出目标用户空腹血糖的第一体检指标值和餐后预设时长血糖的第二体检指标值，如图4所示，判断单元33，具体可包括：匹配模块331、确定模块332。匹配模块331，具体可用于将所述目标用户的特征信息输入到所述第一识别模型中与所述特征信息X1进行相似度匹配，所述目标用户的特征信息对应所述目标用户除所述空腹血糖值和所述餐后两小时血糖值以外的所述目标特征数据；确定模块332，具体可用于利用相似度大于预设阈值、且相似度最高的所述特征信息X1和所述第一映射关系，确定所述目标用户对应的第一体检指标值；匹配模块331，具体还可用于将所述目标用户的特征信息输入到所述第二识别模型中与所述特征信息X2进行相似度匹配；确定模块332，具体还可用于利用相似度大于预定阈值、且相似度最高的特征信息X2和所述第二映射关系，确定所述目标用户对应的第二体检指标值。在具体的应用场景中，为了根据第一体检指标值和/或第二体检指标值，确定出目标用户的患病程度，如图4所示，确定单元34，具体可包括：确定模块341、判断模块342。

确定模块341，可用于若目标用户对应的第一体检指标值大于等于第一预设阈值，和/或第二体检指标值大于等于第二预设阈值，则确定目标用户患有糖尿病；判断模块342，可用于通过第一体检指标值所处的第一数值区间、和/或第二体检指标值所处的第二数值区间，判断目标用户的患病程度。在具体的应用场景中，为了准确的判断出目标用户的患病程度，判断模块342，具体还用于划分大于第一预设阈值，且按照预定数值规律递增的多个数值区间；创建多个数值区间与糖尿病患病程度之间的第三映射关系；确定第一体检指标值对应处于多个数值区间中的第一数值区间；根据第三映射关系以及第一数值区间，判断目标用户的糖尿病患病程度。划分大于第二预设阈值，且按照预定数值规律递增的多个数值区间；创建多个数值区间与糖尿病患病程度之间的第四映射关系；确定第二体检指标值对应处于多个数值区间中的第二数值区间；根据第四映射关系以及第二数值区间，判断目标用户的糖尿病患病程度；

判断模块342，具体还用于若所述第一糖尿病患病程度和所述第二糖尿病患病程度不同，则依据用户对通过所述第一识别模型和所述第二识别模型这两种预测方式反馈的准确率或采纳率，分别获取所述第一识别模型对应的第一权重和所述第二识别模型对应的第二权重；在所述第一权重大于所述第二权重时，将所述第一糖尿病患病程度确定为所述目标用户的患病程度；在所述第二权重大于所述第一权重时，将所述第二糖尿病患病程度确定为所述目标用户的患病程度。在具体的应用场景中，匹配模块331，具体可用于将所述目标用户的特征信息经过数据清洗、特征提取、缺失值填充、异常值处理，得到结构化数据的特征信息；将结构化数据的特征信息与所述特征信息X1进行相似度匹配；匹配模块331，具体可用于将所述目标用户的特征信息经过数据清洗、特征提取、缺失值填充、异常值处理，得到结构化数据的特征信息；将结构化数据的特征信息与所述特征信息X2进行相似度匹配。

在具体的应用场景中，训练模块322，具体可用于采用随机采样方式从所述第一模型训练集中分别获取第一训练样本集、第二训练样本集、第三训练样本集、第四训练样本集；基于所述第一训练样本集利用随机森林算法，训练得到第一分类器；基于所述第二训练样本集利用GBDT算法，训练得到第二分类器；基于所述第三训练样本集利用Xgboost算法，训练得到第三分类器；基于所述第四训练样本集利用LightGBM算法，训练得到第四分类器；将所述第一分类器、所述第二分类器、所述第三分类器、所述第四分类器利用套袋法进行融合处理，得到所述第一识别模型；训练模块322，具体还可用于采用随机采样方式从所述第二模型训练集中分别获取第五训练样本集、第六训练样本集、第七训练样本集、第八训练样本集；基于所述第五训练样本集利用随机森林算法，训练得到第五分类器；基于所述第六训练样本集利用GBDT算法，训练得到第六分类器；基于所述第七训练样本集利用Xgboost算法，训练得到第七分类器；基于所述第八训练样本集利用LightGBM算法，训练得到第八分类器；将所述第五分类器、所述第六分类器、所述第七分类器、所述第八分类器利用套袋法进行融合处理，得到所述第二识别模型。

需要说明的是，本实施例提供的一种糖尿病的预测装置所涉及各功能单元的其它相应描述，可以参考图1至图2中的对应描述，在此不再赘述。基于上述如图1和图2所示方法，相应的，本申请实施例还提供了一种存储介质，其上存储有计算机可读指令，该计算机可读指令被处理器执行时实现上述如图1和图2所示的糖尿病的预测方法。基于这样的理解，本申请的技术方案可以以软件产品的形式体现出来，该软件产品可以存储在一个非易失性存储介质（可以是CD-ROM，U盘，移动硬盘等）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）执行本申请各个实施场景的方法。

基于上述如图1、图2所示的方法，以及图3、图4所示的虚拟装置实施例，为了实现上述目的，本申请实施例还提供了一种计算机设备，具体可以为个人计算机、服务器、网络设备等，该实体设备包括存储介质和处理器；存储介质，用于存储计算机可读指令；处理器，用于执行计算机可读指令以实现上述如图1和图2所示的糖尿病的预测方法。可选地，该计算机设备还可以包括用户接口、网络接口、摄像头、射频（Radio Frequency，RF）电路，传感器、音频电路、WI-FI模块等等。用户接口可以包括显示屏（Display）、输入单元比如键盘（Keyboard）等，可选用户接口还可以包括USB接口、读卡器接口等。网络接口可选的可以包括标准的有线接口、无线接口（如蓝牙接口、WI-FI接口）等。

本领域技术人员可以理解，本实施例提供的计算机设备结构并不构成对该实体设备的限定，可以包括更多或更少的部件，或者组合某些部件，或者不同的部件布置。非易失性可读存储介质中还可以包括操作系统、网络通信模块。操作系统是管理糖尿病的预测的实体设备硬件和软件资源的程序，支持信息处理程序以及其它软件和/或程序的运行。网络通信模块用于实现非易失性可读存储介质内部各组件之间的通信，以及与该实体设备中其它硬件和软件之间通信。通过以上的实施方式的描述，本领域的技术人员可以清楚地了解到本申请可以借助软件加必要的通用硬件平台的方式来实现，也可以通过硬件实现。通过应用本申请的技术方案，与目前现有技术相比，本申请可以在检测出目标用户患有糖尿病的基础上，进一步判断患病的严重程度，可以使诊断结果更加完善，进而可及时跟踪了解目标用户的病情发展情况，并进行相应的配套治疗。

本领域技术人员可以理解附图只是一个优选实施场景的示意图，附图中的模块或流程并不一定是实施本申请所必须的。本领域技术人员可以理解实施场景中的装置中的模块可以按照实施场景描述进行分布于实施场景的装置中，也可以进行相应变化位于不同于本实施场景的一个或多个装置中。上述实施场景的模块可以合并为一个模块，也可以进一步拆分成多个子模块。上述本申请序号仅仅为了描述，不代表实施场景的优劣。以上公开的仅为本申请的几个具体实施场景，但是，本申请并非局限于此，任何本领域的技术人员能思之的变化都应落入本申请的保护范围。

权 利要求书

1、一种糖尿病的预测方法，包括：

获取原始健康档案和电子病历数据中的样本用户数据；

依据所述样本用户数据中的用户特征创建数值型的回归预测模型；

利用所述回归预测模型判断目标用户空腹血糖的第一体检指标值和餐后预设时长血糖的第二体检指标值；

根据所述第一体检指标值和/或所述第二体检指标值，确定所述目标用户的患病程度。

2、根据权利要求 1 所述的方法，所述用户特征是利用正则表达式从所述样本用户数据中提取的，所述预设时长为两小时；所述依据所述样本用户数据中的用户特征创建数值型的回归预测模型，具体包括：将所述用户特征中空腹血糖值作为标签信息 Y1，并将样本用户除所述空腹血糖值和所述餐后两小时血糖值以外的目标特征数据作为特征信息 X1，创建第一模型训练集，所述目标特征数据至少包括所述样本用户的患病史数据、住院数据、就诊用药数据、体检数据、健康告知数据中的一项或多项；通过所述第一模型训练集基于预设回归预测算法训练用于判断所述第一体检指标值的第一识别模型，其中，所述预设回归预测算法由随机森林、梯度提升决策树 GBDT、Xgboost、LightGBM 四种算法融合得到，所述第一识别模型的评估采用平均绝对百分比误差 MAPE 指标，当所述第一识别模型对应的 MAPE 指标值小于预置标准比较阈值时，确定所述第一识别模型符合评估标准，通过符合评估标准的所述第一识别模型可确定所述特征信息 X1 和所述标签信息 Y1 之间的第一映射关系；将所述用户特征中餐后两小时血糖值作为标签信息 Y2，并将所述样本用户的所述目标特征数据作为特征信息 X2，创建第二模型训练集；通过所述第二模型训练集基于所述预设回归预测算法训练用于判断所述第二体检指标值的第二识别模型，其中，所述第二识别模型的评估采用 MAPE 指标，当所述第二识别模型对应的 MAPE 指标值小于预定标准比较阈值时，确定所述第二识别模型符合评估标准，通过符合评估标准的所述第二识别模型可确定所述特征信息 X2 和所述标签信息 Y2 之间的第二映射关系。

3、根据权利要求 2 所述的方法，所述利用所述回归预测模型判断目标用户空腹血糖的第一体检指标值和餐后预设时长血糖的第二体检指标值，具体包括：将所述目标用户的特征信息输入到所述第一识别模型中与所述特征信息 X1 进行相似度匹配，所述目标用户的特征信息对应除所述空腹血糖值和所述餐后两小时血糖值以外的所述目标特征数据；利用相似度大于预设阈值、且相似度最高的所述特征信息 X1 和所述第一映射关系，确定所述目标用户对应的第一体检指标值；将所述目标用户的特征信息输入到所述第二识别模型中与所述特征信息 X2 进行相似度匹配；利用相似度大于预定阈值、且相似度最高的特征信息 X2 和所述第二映射关系，确定所述目标用户对应的第二体检指标值。

4、根据权利要求 3 所述的方法，所述根据所述第一体检指标值和/或所述第二体检指标值，确定所述目标用户的患病程度，具体包括：若所述目标用户对应的所述第一体检指标值大于等于第一预设阈值，和/或所述第二体检指标值大于等于第二预设阈值，则确定所述目标用户患有糖尿病；通过所述第一体检指标值所处的第一数值区间、和/或所述第二体检指标值所处的第二数值区间，判断所述目标用户的患病程度。

5、根据权利要求 4 所述的方法，通过所述第一体检指标值所处的第一数值区间，判断所述目标用户的患病程度，具体包括：划分大于所述第一预设阈值，且按照预定数值规律递增的多个数值区间；创建所述多个数值区间与糖尿病患病程度之间的第三映射关系；确定所述第一体检指标值对应处于所述多个数值区间中的所述第一数值区间；根据所述第三映射关系以及所述第一数值区间，判断所述目标用户的第一糖尿病患病程度；

通过所述第二体检指标值所处的第二数值区间，判断所述目标用户的患病程度，具体包括：划分大于所述第二预设阈值，且按照预定数值规律递增的多个数值区间；创建所述多个数值区间与糖尿病患病程度之间的第四映射关系；确定所述第二体检指标值对应处于所述多个数值区间中的所述第二数值区间；根据所述第四映射关系以及所述第二数值区间，判断所述目标用户的第二糖尿病患病程度；

通过所述第一体检指标值所处的第一数值区间、和所述第二体检指标值所处的第二数值区间，判断所述目标用户的患病程度，具体包括：若所述第一糖尿病患病程度和所述第二糖尿病患病程度不同，则依据用户对通过所述第一识别模型和所述第二识别模型这两种预测方式反馈的准确率或采纳率，分别获取所述第一识别模型对应的第一权重和所述第二识别模型对应的第二权重；在所述第一权重大于所述第二权重时，将所述第一糖尿病患病程度确定为所述目标用户的患病程度；在所述第二权重大于所述第一权重时，将所述第二糖尿病患病程度确定为所述目标用户的患病程度。

6、根据权利要求 3 所述的方法，所述将所述目标用户的特征信息输入到所述第一识别模型中与所述特征信息 X1 进行相似度匹配，具体包括：将所述目标用户的特征信息经过数据清洗、特征提取、缺失值填充、异常值处理，得到结构化数据的特征信息；将结构化数据的特征信息与所述特征信息 X1 进

行相似度匹配；

所述将所述目标用户的特征信息输入到所述第二识别模型中与所述特征信息X2进行相似度匹配，具体包括：将所述目标用户的特征信息经过数据清洗、特征提取、缺失值填充、异常值处理，得到结构化数据的特征信息；将结构化数据的特征信息与所述特征信息X2进行相似度匹配。

7、根据权利要求2所述的方法，所述通过所述第一模型训练集基于预设回归预测算法训练用于判断所述第一体检指标值的第一识别模型，具体包括：采用随机采样方式从所述第一模型训练集中分别获取第一训练样本集、第二训练样本集、第三训练样本集、第四训练样本集；基于所述第一训练样本集利用随机森林算法，训练得到第一分类器；基于所述第二训练样本集利用GBDT算法，训练得到第二分类器；基于所述第三训练样本集利用Xgboost算法，训练得到第三分类器；基于所述第四训练样本集利用LightGBM算法，训练得到第四分类器；将所述第一分类器、所述第二分类器、所述第三分类器、所述第四分类器利用套袋法进行融合处理，得到所述第一识别模型；

所述通过所述第二模型训练集基于所述预设回归预测算法训练用于判断所述第二体检指标值的第二识别模型，具体包括：采用随机采样方式从所述第二模型训练集中分别获取第五训练样本集、第六训练样本集、第七训练样本集、第八训练样本集；基于所述第五训练样本集利用随机森林算法，训练得到第五分类器；基于所述第六训练样本集利用GBDT算法，训练得到第六分类器；基于所述第七训练样本集利用Xgboost算法，训练得到第七分类器；基于所述第八训练样本集利用LightGBM算法，训练得到第八分类器；将所述第五分类器、所述第六分类器、所述第七分类器、所述第八分类器利用套袋法进行融合处理，得到所述第二识别模型。

8、一种糖尿病的预测装置，包括：

获取单元，用于获取原始健康档案和电子病历数据中的样本用户数据；

创建单元，用于依据所述样本用户数据中的用户特征创建数值型的回归预测模型；

判断单元，用于利用所述回归预测模型判断目标用户空腹血糖的第一体检指标值和餐后预设时长血糖的第二体检指标值；

确定单元，用于根据所述第一体检指标值和/或所述第二体检指标值，确定所述目标用户的患病程度。

9、根据权利要求8所述的装置，所述创建单元，具体包括：创建模块、训练模块、确定模块；

所述创建模块，具体用于将所述用户特征中空腹血糖值作为标签信息Y1，并将样本用户除所述空腹血糖值和所述餐后两小时血糖值以外的目标特征数据作为特征信息X1，创建第一模型训练集，所述目标特征数据至少包括所述样本用户的患病史数据、住院数据、就诊用药数据、体检数据、健康告知数据中的一项或多项；

所述训练模块，具体用于通过所述第一模型训练集基于预设回归预测算法训练用于判断所述第一体检指标值的第一识别模型，其中，所述预设回归预测算法由随机森林、梯度提升决策树GBDT、Xgboost、LightGBM四种算法融合得到，所述第一识别模型的评估采用平均绝对百分比误差MAPE指标，当所述第一识别模型对应的MAPE指标值小于预置标准比较阈值时，确定所述第一识别模型符合评估标准；

所述确定模块，具体用于通过符合评估标准的所述第一识别模型可确定所述特征信息X1和所述标签信息Y1之间的第一映射关系；

所述创建模块，具体还用于将所述用户特征中餐后两小时血糖值作为标签信息Y2，并将所述样本用户的所述目标特征数据作为特征信息X2，创建第二模型训练集；

所述训练模块，具体还用于通过所述第二模型训练集基于所述预设回归预测算法训练用于判断所述第二体检指标值的第二识别模型，其中，所述第二识别模型的评估采用MAPE指标，当所述第二识别模型对应的MAPE指标值小于预定标准比较阈值时，确定所述第二识别模型符合评估标准；

所述确定模块，具体还用于通过符合评估标准的所述第二识别模型可确定所述特征信息X2和所述标签信息Y2之间的第二映射关系。

10、根据权利要求9所述的装置，所述判断单元，具体包括：匹配模块、确定模块；

所述匹配模块，具体用于将所述目标用户的特征信息输入到所述第一识别模型中与所述特征信息X1进行相似度匹配，所述目标用户的特征信息对应所述目标用户除所述空腹血糖值和所述餐后两小时血糖值以外的所述目标特征数据；

所述确定模块，具体用于利用相似度大于预设阈值、且相似度最高的所述特征信息X1和所述第一映射关系，确定所述目标用户对应的第一体检指标值；

所述匹配模块，具体还用于将所述目标用户的特征信息输入到所述第二识别模型中与所述特征信息X2进行相似度匹配；

所述确定模块，具体还用于利用相似度大于预定阈值、且相似度最高的特征信息X2和所述第二映

射关系，确定所述目标用户对应的第二体检指标值。

11、根据权利要求 10 所述的装置，所述确定单元，具体包括：确定模块、判断模块；

所述确定模块，用于若目标用户对应的第一体检指标值大于等于第一预设阈值，和/或第二体检指标值大于等于第二预设阈值，则确定目标用户患有糖尿病；

所述判断模块，用于通过第一体检指标值所处的第一数值区间、和/或第二体检指标值所处的第二数值区间，判断目标用户的患病程度。

12、根据权利要求 11 所述的装置，所述判断模块，具体还用于划分大于第一预设阈值，且按照预定数值规律递增的多个数值区间；创建多个数值区间与糖尿病患病程度之间的第三映射关系；确定第一体检指标值对应处于多个数值区间中的第一数值区间；根据第三映射关系以及第一数值区间，判断目标用户的糖尿病患病程度。划分大于第二预设阈值，且按照预定数值规律递增的多个数值区间；创建多个数值区间与糖尿病患病程度之间的第四映射关系；确定第二体检指标值对应处于多个数值区间中的第二数值区间；根据第四映射关系以及第二数值区间，判断目标用户的糖尿病患病程度；

所述判断模块，具体还用于若所述第一糖尿病患病程度和所述第二糖尿病患病程度不同，则依据用户对通过所述第一识别模型和所述第二识别模型这两种预测方式反馈的准确率或采纳率，分别获取所述第一识别模型对应的第一权重和所述第二识别模型对应的第二权重；在所述第一权重大于所述第二权重时，将所述第一糖尿病患病程度确定为所述目标用户的患病程度；在所述第二权重大于所述第一权重时，将所述第二糖尿病患病程度确定为所述目标用户的患病程度。

13、根据权利要求 10 所述的装置，所述匹配模块，具体用于将所述目标用户的特征信息经过数据清洗、特征提取、缺失值填充、异常值处理，得到结构化数据的特征信息；将结构化数据的特征信息与所述特征信息 X1 进行相似度匹配；

所述匹配模块，具体用于将所述目标用户的特征信息经过数据清洗、特征提取、缺失值填充、异常值处理，得到结构化数据的特征信息；将结构化数据的特征信息与所述特征信息 X2 进行相似度匹配。

14、根据权利要求 9 所述的装置，所述训练模块，具体用于采用随机采样方式从所述第一模型训练集中分别获取第一训练样本集、第二训练样本集、第三训练样本集、第四训练样本集；基于所述第一训练样本集利用随机森林算法，训练得到第一分类器；基于所述第二训练样本集利用 GBDT 算法，训练得到第二分类器；基于所述第三训练样本集利用 Xgboost 算法，训练得到第三分类器；基于所述第四训练样本集利用 LightGBM 算法，训练得到第四分类器；将所述第一分类器、所述第二分类器、所述第三分类器、所述第四分类器利用套袋法进行融合处理，得到所述第一识别模型；

所述训练模块，具体还用于采用随机采样方式从所述第二模型训练集中分别获取第五训练样本集、第六训练样本集、第七训练样本集、第八训练样本集；基于所述第五训练样本集利用随机森林算法，训练得到第五分类器；基于所述第六训练样本集利用 GBDT 算法，训练得到第六分类器；基于所述第七训练样本集利用 Xgboost 算法，训练得到第七分类器；基于所述第八训练样本集利用 LightGBM 算法，训练得到第八分类器；将所述第五分类器、所述第六分类器、所述第七分类器、所述第八分类器利用套袋法进行融合处理，得到所述第二识别模型。

15、一种非易失性可读存储介质，其上存储有计算机可读指令，所述计算机可读指令被处理器执行时实现糖尿病的预测方法，包括：获取原始健康档案和电子病历数据中的样本用户数据；依据所述样本用户数据中的用户特征创建数值型的回归预测模型；利用所述回归预测模型判断目标用户空腹血糖的第一体检指标值和餐后预设时长血糖的第二体检指标值；根据所述第一体检指标值和/或所述第二体检指标值，确定所述目标用户的患病程度。

16、根据权利要求 15 所述的非易失性可读存储介质，所述用户特征是利用正则表达式从所述样本用户数据中提取的，所述预设时长为两小时；所述计算机可读指令被处理器执行时实现所述依据所述样本用户数据中的用户特征创建数值型的回归预测模型，具体包括：将所述用户特征中空腹血糖值作为标签信息 Y1，并将样本用户除所述空腹血糖值和所述餐后两小时血糖值以外的目标特征数据作为特征信息 X1，创建第一模型训练集，所述目标特征数据至少包括所述样本用户的患病史数据、住院数据、就诊用药数据、体检数据、健康告知数据中的一项或多项；通过所述第一模型训练集基于预设回归预测算法训练用于判断所述第一体检指标值的第一识别模型，其中，所述预设回归预测算法由随机森林、梯度提升决策树 GBDT、Xgboost、LightGBM 四种算法融合得到，所述第一识别模型的评估采用平均绝对百分比误差 MAPE 指标，当所述第一识别模型对应的 MAPE 指标值小于预置标准比较阈值时，确定所述第一识别模型符合评估标准，通过符合评估标准的所述第一识别模型可确定所述特征信息 X1 和所述标签信息 Y1 之间的第一映射关系；将所述用户特征中餐后两小时血糖值作为标签信息 Y2，并将所述样本用户的所述目标特征数据作为特征信息 X2，创建第二模型训练集；通过所述第二模型训练集基于所述预设回归预测算法训练用于判断所述第二体检指标值的第二识别模型，其中，所述第二识别模型的评估采用 MAPE 指标，当所述第二识别模型对应的 MAPE 指标值小于预定标准比较阈值时，确定所述

第二识别模型符合评估标准，通过符合评估标准的所述第二识别模型可确定所述特征信息 X2 和所述标签信息 Y2 之间的第二映射关系。

17、根据权利要求 16 所述的非易失性可读存储介质，所述计算机可读指令被处理器执行时实现所述利用所述回归预测模型判断目标用户空腹血糖的第一体检指标值和餐后预设时长血糖的第二体检指标值，具体包括：将所述目标用户的特征信息输入到所述第一识别模型中与所述特征信息 X1 进行相似度匹配，所述目标用户的特征信息对应除所述空腹血糖值和所述餐后两小时血糖值以外的所述目标特征数据；利用相似度大于预设阈值、且相似度最高的所述特征信息 X1 和所述第一映射关系，确定所述目标用户对应的第一体检指标值；将所述目标用户的特征信息输入到所述第二识别模型中与所述特征信息 X2 进行相似度匹配；利用相似度大于预定阈值、且相似度最高的特征信息 X2 和所述第二映射关系，确定所述目标用户对应的第二体检指标值。

18、一种计算机设备，包括非易失性可读存储介质、处理器及存储在非易失性可读存储介质上并可在处理器上运行的计算机可读指令，所述处理器执行所述计算机可读指令时实现糖尿病的预测方法，包括：获取原始健康档案和电子病历数据中的样本用户数据；依据所述样本用户数据中的用户特征创建数值型的回归预测模型；利用所述回归预测模型判断目标用户空腹血糖的第一体检指标值和餐后预设时长血糖的第二体检指标值；根据所述第一体检指标值和/或所述第二体检指标值，确定所述目标用户的患病程度。

19、根据权利要求 18 所述的计算机设备，所述用户特征是利用正则表达式从所述样本用户数据中提取的，所述预设时长为两小时；所述计算机可读指令被处理器执行时实现所述依据所述样本用户数据中的用户特征创建数值型的回归预测模型，具体包括：将所述用户特征中空腹血糖值作为标签信息 Y1，并将样本用户除所述空腹血糖值和所述餐后两小时血糖值以外的目标特征数据作为特征信息 X1，创建第一模型训练集，所述目标特征数据至少包括所述样本用户的患病史数据、住院数据、就诊用药数据、体检数据、健康告知数据中的一项或多项；通过所述第一模型训练集基于预设回归预测算法训练用于判断所述第一体检指标值的第一识别模型，其中，所述预设回归预测算法由随机森林、梯度提升决策树 GBDT、Xgboost、LightGBM 四种算法融合得到，所述第一识别模型的评估采用平均绝对百分比误差 MAPE 指标，当所述第一识别模型对应的 MAPE 指标值小于预置标准比较阈值时，确定所述第一识别模型符合评估标准，通过符合评估标准的所述第一识别模型可确定所述特征信息 X1 和所述标签信息 Y1 之间的第一映射关系；将所述用户特征中餐后两小时血糖值作为标签信息 Y2，并将所述样本用户的所述目标特征数据作为特征信息 X2，创建第二模型训练集；通过所述第二模型训练集基于所述预设回归预测算法训练用于判断所述第二体检指标值的第二识别模型，其中，所述第二识别模型的评估采用 MAPE 指标，当所述第二识别模型对应的 MAPE 指标值小于预定标准比较阈值时，确定所述第二识别模型符合评估标准，通过符合评估标准的所述第二识别模型可确定所述特征信息 X2 和所述标签信息 Y2 之间的第二映射关系。

20、根据权利要求 19 所述的计算机设备，所述计算机可读指令被处理器执行时实现所述利用所述回归预测模型判断目标用户空腹血糖的第一体检指标值和餐后预设时长血糖的第二体检指标值，具体包括：将所述目标用户的特征信息输入到所述第一识别模型中与所述特征信息 X1 进行相似度匹配，所述目标用户的特征信息对应除所述空腹血糖值和所述餐后两小时血糖值以外的所述目标特征数据；利用相似度大于预设阈值、且相似度最高的所述特征信息 X1 和所述第一映射关系，确定所述目标用户对应的第一体检指标值；将所述目标用户的特征信息输入到所述第二识别模型中与所述特征信息 X2 进行相似度匹配；利用相似度大于预定阈值、且相似度最高的特征信息 X2 和所述第二映射关系，确定所述目标用户对应的第二体检指标值。

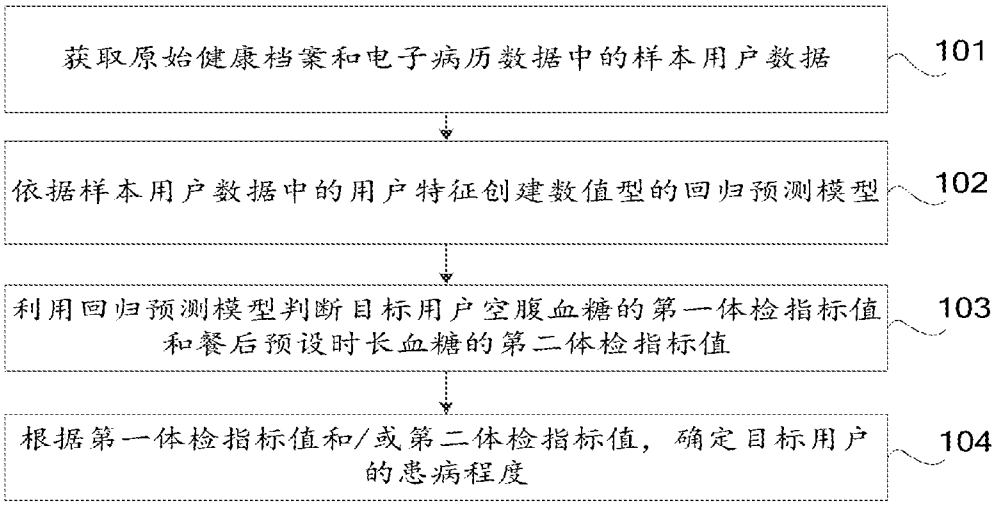

获取原始健康档案和电子病历数据中的样本用户数据
101
依据样本用户数据中的用户特征创建数值型的回归预测模型
102
利用回归预测模型判断目标用户空腹血糖的第一体检指标值和餐后预设时长血糖的第二体检指标值
103
根据第一体检指标值和/或第二体检指标值，确定目标用户的患病程度
104

图 1


获取原始健康档案和电子病历数据中的样本用户数据
201
将样本用户的用户特征中空腹血糖值作为标签信息Y1，并将样本用户除空腹血糖值和餐后两小时血糖值以外的目标特征数据作为特征信息X1，创建第一模型训练集
202a
将样本用户的用户特征中餐后两小时血糖值作为标签信息Y2，并将样本用户除空腹血糖值和餐后两小时血糖值以外的目标特征数据作为特征信息X2，创建第二模型训练集
202b
通过第一模型训练集基于预设回归预测算法训练用于判断第一体检指标值的第一识别模型
203a
通过第二模型训练集基于预设回归预测算法训练用于判断第二体检指标值的第二识别模型
203b
将目标用户的特征信息输入到第一识别模型中与特征信息X1进行相似度匹配
204a
将目标用户的特征信息输入到第二识别模型中与特征信息X2进行相似度匹配
204b
利用相似度大于预设阈值、且相似度最高的特征信息X1和第一映射关系，确定目标用户对应的第一体检指标值
205a
利用相似度大于预定阈值、且相似度最高的特征信息X2和第二映射关系，确定目标用户对应的第二体检指标值
205b
根据第一体检指标值和/或第二体检指标值，确定目标用户的患病程度
206

图 2

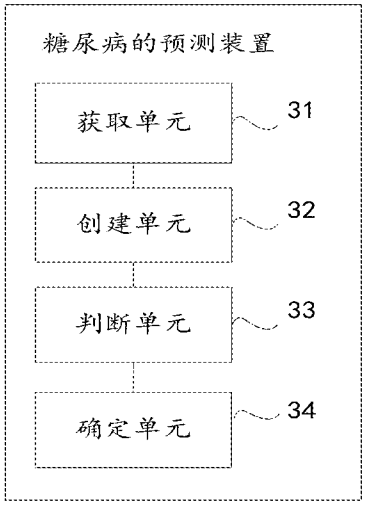

糖尿病的预测装置
获取单元
31
创建单元
32
判断单元
33
确定单元
34

图 3

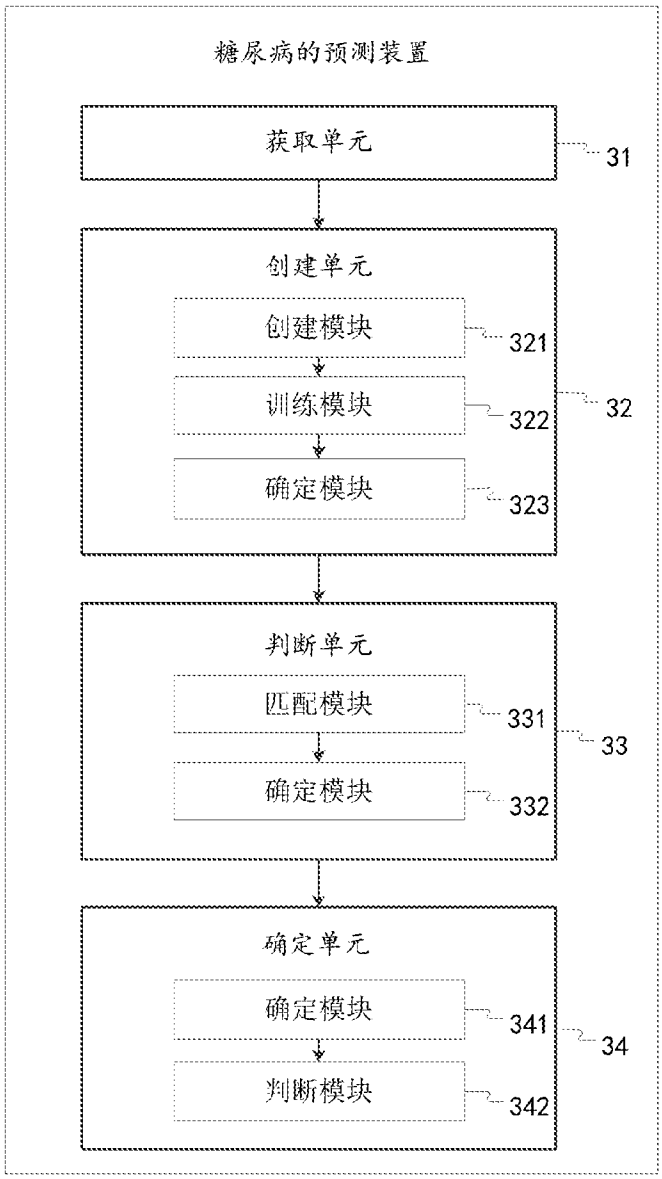

糖尿病的预测装置
获取单元
31
创建单元
创建模块
321
训练模块
322
确定模块
323
32
判断单元
匹配模块
331
确定模块
332
33
确定单元
确定模块
341
判断模块
342
34

图 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2019/117217

A. CLASSIFICATION OF SUBJECT MATTER

G16H 50/20(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,G06K,G06F

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)

CNPAT, CNKI, WPI, EPODOC: 糖尿病, 血糖, 体检, 数据, 回归, 预测, 指标, 空腹, 餐后, diabetes, blood 1w glucose, physical 1w examination, data, regression, prediction, limosis, postprandial

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110197720 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 03 September 2019 (2019-09-03) claims 1-10	1-20
Y	CN 106682412 A (ZHEJIANG UNIVERSITY) 17 May 2017 (2017-05-17) description, paragraphs [0005]-[0009], and figure 2	1, 8, 15, 18
Y	CN 109378072 A (SUN YAT-SEN UNIVERSITY) 22 February 2019 (2019-02-22) description, paragraphs [0025]-[0030] and [0056]-[0064]	1, 8, 15, 18
A	CN 109308545 A (PING AN LIFE INSURANCE COMPANY OF CHINA) 05 February 2019 (2019-02-05) entire document	1-20
A	US 2018158552 A1 (UNIVERSITY OF SOUTHERN CALIFORNIA) 07 June 2018 (2018-06-07) 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
08 January 2020	**06 February 2020**

Name and mailing address of the ISA/CN	Authorized officer
China National Intellectual Property Administration **No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088** **China**	
Facsimile No. **(86-10)62019451**	Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2019/117217

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	白江梁等 (BAI, Hongliang et al.). "某医院体检人群糖尿病预测模型研究 (Diabetes Prediction Models Based on Health Check-Up Population in a Hospital)." *实用预防医学 (Practical Preventive Medicine).*, Vol. 25, No. (1), 31 January 2018 (2018-01-31), pp. 116-118	1-20

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2019/117217

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 110197720 A	03 September 2019	None	
CN 106682412 A	17 May 2017	None	
CN 109378072 A	22 February 2019	None	
CN 109308545 A	05 February 2019	None	
US 2018158552 A1	07 June 2018	None	

国际检索报告

国际申请号
PCT/CN2019/117217

A. 主题的分类

G16H 50/20(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,G06K,G06F

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））

CNPAT, CNKI, WPI,EPODOC:糖尿病，血糖，体检，数据，回归，预测，指标，空腹，餐后， diabetes, blood 1w glucose, physical 1w examination, data, regression, prediction, limosis, postprandial

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 110197720 A (平安科技深圳有限公司) 2019年 9月 3日 (2019-09-03) 权利要求1-10	1-20
Y	CN 106682412 A (浙江大学) 2017年 5月 17日 (2017-05-17) 说明书第[0005]-[0009]段以及图2	1、8、15、18
Y	CN 109378072 A (中山大学) 2019年 2月 22日 (2019-02-22) 说明书第[0025]-[0030]、[0056]-[0064]段	1、8、15、18
A	CN 109308545 A (中国平安人寿保险股份有限公司) 2019年 2月 5日 (2019-02-05) 全文	1-20
A	US 2018158552 A1 (UNIVERSITY OF SOUTHERN CALIFORNIA) 2018年 6月 7日 (2018-06-07) 全文	1-20
A	白江梁等. "某医院体检人群糖尿病预测模型研究." 实用预防医学., 第25卷，第1期，2018年 1月 31日 (2018-01-31)， 第116-118页	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 1月 8日	2020年 2月 6日

ISA/CN的名称和邮寄地址	受权官员
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	韩雪莲
传真号 (86-10)62019451	电话号码 86-(10)-53961800

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2019/117217

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 110197720 A	2019年 9月 3日	无	
CN 106682412 A	2017年 5月 17日	无	
CN 109378072 A	2019年 2月 22日	无	
CN 109308545 A	2019年 2月 5日	无	
US 2018158552 A1	2018年 6月 7日	无	