

女性盆底功能障碍性疾病风险预警模型及其构建方法和系统

技术领域

本发明属于医疗预警系统技术领域，具体涉及女性盆底功能障碍性疾病风险预警模型及其构建方法和系统。

背景技术

盆底作为一个结构和功能整体，由骨盆、盆底肌肉、神经、韧带及血管构成，在内分泌、神经和营养代谢系统的调节下，泌尿、生殖、肛肠三大系统协调一致，共同发挥力学支撑功能。盆底功能障碍性疾病（pelvic floor disfunction, PFD）是指各种因素导致盆底支持结构薄弱，造成盆腔脏器位置和功能异常的一组疾病，包括盆腔器官脱垂（pelvic organ prolapse, POP）、压力性尿失禁（stress urinary incontinence, SUI）、性功能障碍（sexual dysfunction, SD）和慢性盆腔疼痛（chronic pelvic pain, CPP）等。其中，盆腔器官脱垂全球发病率高达20%，且呈逐年升高趋势；一项针对发展中国家女性 PFD 发病率的调查显示，盆腔器官脱垂发病率为48.3%。虽然 POP 并不直接威胁生命，但严重影响患者的生活质量和身心健康，对个人、家庭和社会造成严重负担。据预测，2050 年中国 60 岁以上老年人口数量将达到 4.83 亿人，80 岁以上老年人口数量将达到 1.08 亿人，随着三胎政策的放开，PFD 这个随着老龄化而不断发展加重的疾病无疑会更加凸显人民美好生活愿景和医疗服务之间的矛盾。

盆底结构复杂，涉及肌肉、筋膜、韧带、血管和神经等多重组织，构成泌尿、生殖、肛肠等不同器官关联的整体。在直立人体中，盆底承托着整个躯体内脏的压力负荷，并受全身循环、呼吸、营养状况和代谢变化等动力学影响。因此，盆底结构和功能状态是全身状况的局部和动态的体现。妊娠、分娩、肥胖、药物、雌激素缺乏、盆底手术、炎症、创伤、盆腔压力和循环等，通过复杂的相互作用与 PFD 发病关联，不同患者表现为不同的结构和功能状态。同时，PFD 是一个慢性积累的过程，伴随时间推移发生肌肉筋膜结构的生理性退化和相关病理性损伤加重，受各种盆腔内外因素的影响，在初期易被忽略，尤其是盆外因素缺乏关注，导致 PFD 的长程管理不能落实。

目前 PFD 的诊断主要通过临床表型、病史和相关功能性检查，缺乏能反映长期病理过程的功能分子作为早期识别、干预和评价指标；简单的盆底检查，甚至相关手术选择和过程，很大程度上依靠临床医师的经验，而且 PFD 患者具有在不同医疗机构间流动、居家、分散的特点，导致诊疗工作在机构、信息和处置上不连续、同质化差、无法进行有效质控管理。盆底疾病专业性强，专科医师缺乏，同时临床上缺乏简易、精准、统一和多维度的检测和监测

手段，导致无早期预警方法可用、诊断准确性不足、治疗效果不确定等诸多问题，因此，对于 PFD 的早诊早治，迄今没有确切的模式。

在现代医院研究领域，疾病风险预测模型常被用来预测某种疾病未来的发病情况。具体来说就是以疾病的多病因为基础，建立统计模型，用来预测具有某些特征的人群未来某种结局事件发生的概率。近年来，预警模型受到越来越多的关注，并在多个疾病研究领域中得到应用，例如：公开号为 CN111640509A 的发明专利公开的一种宫颈癌术后复发风险预测方法及系统，以及公开号为 CN110200619A 的发明专利公开的一种建立重大心血管疾病风险预警模型方法及装置，等等。目前，仍未见有报道预警模型在 PFD 中的研究和应用。因此，探索及构建 PFD 的早期风险预警模型，已是推动女性盆底功能障碍防治的一项重要措施。

发明内容

本发明的目的在于提供女性盆底功能障碍性疾病风险预警模型及其构建方法和系统，通过对风险因素的相关变量（即高风险因素）进行筛选并构建女性盆底功能障碍性疾病（PFD）的风险预警模型，可以实现 PFD 中阴道前壁膨出、子宫穹隆膨出、阴道后壁膨出、压力性尿失禁、阴道松弛大小、阴道肌力等筛查和预警分析。为此，本发明还提供了实现该 PFD 风险预警模型的计算机系统。

本发明通过下述技术方案实现：女性盆底功能障碍性疾病风险预警模型的构建方法，包括以下步骤：

- S1. 获取区域风险人群的风险因素；
- S2. 筛选风险因素中的相关变量并构建风险预警模型；
- S3. 对风险预警模型进行训练获得模型优化。

所述步骤 S1 中，区域风险人群为选定地区的疾病易感人群。

所述步骤 S1 中，风险因素包括盆外风险因素和盆内风险因素，所述盆外风险因素包括：年龄、BMI、糖尿病、体力劳动、慢性咳嗽、便秘、吸烟；所述盆内风险因素包括孕产次、分娩方式、阴道肌力、肛提肌裂伤。

所述步骤 S2 中，筛选风险因素中的相关变量并构建风险预警模型的步骤包括：

- a. 根据风险因素，通过 Logistic 回归法筛选模型预选变量；
- b. 根据筛选的模型预选变量，通过支持向量机递归特征消除算法筛选相关变量；
- c. 根据筛选的相关变量，通过机器学习方法构建风险预警模型。

所述步骤 a 中，Logistic 回归法的公式如下：

$$\text{logit}(P) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$

其中，P 为阳性结果发生的条件概率， β 为偏回归系数，X 为自变量（暴露因素）， β_0

为常数项，1、2……p 表示偏回归系数和自变量（暴露因素）的编号。

所述步骤 b 中，支持向量机递归特征消除算法的公式如下：

$$w = \sum_k \alpha_k y_k x_k$$

其中，w 为特征权重，α 为特征参数，y 为分类标签，x 为训练样本，k 为编号。

所述步骤 c 中，机器学习方法包括 Logistic 回归法、支持向量机、高斯朴素贝叶斯法、随机森林法、梯度提升树算法、Adaboost 集成学习模型或多层感知机。

所述步骤 c 中，风险预警模型的输出结果包括预测准确度 AUC、召回率和 f1 值。

进一步的，所述步骤 S3 中，通过队列研究，获得不同风险程度的人群，利用人工智能机器学习算法对风险预警模型进行训练，获得模型优化。

一种采用上述方法构建得到的女性盆底功能障碍性疾病风险预警模型。

一种计算机系统，包括储存器、处理器以及存储在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序时实现上述预警模型的构建方法。

术语解释：

预测准确度 AUC 为受试者工作特征曲线下面积（area under the receiver operating characteristic curve），用于评价分类器（预测模型）的预测准确度。

召回率为查全率，即正确预测为正的占全部实际为正的的比例。

f1 值为算数平均数除以几何平均数。

本发明与现有技术相比，具有以下优点及有益效果：

本发明通过构建 PDF 的早期预警体系，对健康人群（非盆底功能障碍性疾病）、风险人群、患者人体画像等开展盆底功能障碍性疾病风险初筛，实现人群的风险度分层，将人群分流为高风险和低风险，通过不同的随访策略，针对重点人群开展辅助诊断、全程管理和分级诊疗。

附图说明

- 图 1 为实施例 2 的技术路线。
- 图 2 为实施例 2 中初步探索数据建模结果图。
- 图 3 为实施例 2 的风险预警模型。
- 图 4 为实施例 3 的分析流程。
- 图 5 为实施例 3 中数据集 1-RFE 变量数据与 auc 值对应曲线。
- 图 6 为实施例 3 中数据集 2-RFE 变量数据与 auc 值对应曲线。
- 图 7 为实施例 3 中数据集 1-ROC 曲线图。

图 8 为实施例 3 中数据集 1-变量相关性 SHAP 图-RF 模型。

图 9 为实施例 3 中数据集 2-ROC 曲线图。

图 10 为实施例 3 中数据集 2-变量相关性 SHAP 图-RF 模型。

具体实施方式

下面将本发明的发明目的、技术方案和有益效果作进一步的说明。

应该指出，以下详细说明都是示例性的，旨在对所要求的本发明提供进一步的说明，除非另有说明，本文使用的所有技术和科学术语具有与本发明所属技术领域的普通技术人员通常理解的含义。

实施例 1：风险预警模型构建

本实施例所述风险预警模型主要用于女性盆底功能障碍性疾病的筛查和预警分析，主要包括：阴道前壁膨出、子宫穹隆膨出、阴道后壁膨出、压力性尿失禁、阴道松弛大小、阴道肌力，等等。

首先，获取区域风险人群（即选定地区的疾病易感人群）的风险因素，如盆外风险因素和盆内风险因素，所述盆外风险因素包括：年龄、BMI、糖尿病、体力劳动、慢性咳嗽、便秘、吸烟；所述盆内风险因素包括孕产次、分娩方式、阴道肌力、肛提肌裂伤，等等。

然后，根据风险因素，通过 Logistic 回归法筛选模型预选变量，如下式所示：

$$\text{logit}(P) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

其中， P 为阳性结果发生的条件概率， β 为偏回归系数， X 为自变量（暴露因素）， β_0 为常数项， $1, 2, \dots, p$ 表示偏回归系数和自变量（暴露因素）的编号。

再根据筛选的模型预选变量，通过支持向量机递归特征消除算法筛选相关变量，如下式所示：

$$w = \sum_k \alpha_k y_k x_k$$

其中， w 为特征权重， α 为特征参数， y 为分类标签， x 为训练样本， k 为编号。

然后再根据筛选的相关变量，通过机器学习方法，如 Logistic 回归法、支持向量机、高斯朴素贝叶斯法、随机森林法、梯度提升树算法、Adaboost 集成学习模型或多层感知机等，构建风险预警模型，并获得预测准确度 AUC、召回率和 f1 值。

在一个具体的实施例中，根据筛选的相关变量，通过随机森林法获得高风险预警模型，如下式所示：

$$VIM_j = \frac{VIM_j}{\sum_{i=1}^n VIM_i}$$

其中，VIM 为变量重要性评分（variable importance measures）， j 为特征编号， i 为随机

森林树编号。

最后，为增加模型的预测准确度，通过队列研究，获得不同风险程度的人群，利用人工智能机器学习算法对风险预警模型进行训练，获得模型优化。

在一个具体的实施例中，可以是一种包括上述构建方法的计算机系统，具体包括储存器、处理器以及存储在所述处理器上运行的计算机程序，处理器执行该计算机程序时实现上述预警模型的构建方法，使得本实施例所述女性盆底功能障碍性疾病的筛查和预警分析工作均可通过计算机系统即可实现。

下面以两个典型实施例来进一步列举说明本发明的具体实施方式，当然，本发明的保护范围并不局限于以下的具体实施方式。

实施例 2：构建阴道松弛分析的风险预警模型

在本实施例中，针对区域风险人群，采集盆外风险因素（包括年龄、BMI、糖尿病、体力劳动、慢性咳嗽、便秘、吸烟等因素）和盆内风险因素（包括孕产次、分娩方式、阴道肌力/肌电、肛提肌裂伤等因素），通过 Logistic 和随机森林法，获得风险因素排序，构建风险预警模型，再通过人工智能机器学习、队列验证获得模型优化，具体的技术路线如图 1 所示。

本实施例对本地区就诊女性盆腔器官脱垂的流行病学调查显示:POP 是最为常见的 PFD, 年龄≥55 岁和自然分娩史，是导致盆腔器官脱垂加重的两个独立危险因素，如下表 1 所示。

表1 盆腔器官脱垂的独立危险因素分析

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-8.0398	197.8	0.0017	0.9676
年龄	≥55	1	0.8969	0.2111	18.0453	<.0001
BMI	≥25	1	0.1603	0.1874	0.7312	0.3925
职业	脑力劳动者	1	-0.1622	0.0916	3.1381	0.0765
分娩次数	2	1	0.0356	0.0683	0.2721	0.602
分娩方式	顺产	1	0.4318	0.0671	41.457	<.0001
糖尿病	有	1	-0.0357	0.1369	0.0679	0.7945
高血压	有	1	0.1988	0.226	0.7738	0.379
遗传	有	1	-0.1648	0.1455	1.2828	0.2574
吸烟	有	1	-6.539	197.8	0.0011	0.9736
慢性咳嗽	有	1	0.271	0.2212	1.5008	0.2206
便秘	有	1	-0.0673	0.0761	0.7818	0.3766

备注：Estimate 为最大似然估计，是由似然函数得到，用于反映灵敏度和特异度的复合指标；Standard Error 为标准误，通过标准差除以样本大小 N 的平方根，用来衡量抽样误差；Wald Chi-Square 为卡方统计量，为实测频率与预期频率的比值，表明检验数据的相关性；Pr > ChiSq 为统计学显著性，基于卡方统计量确定，反映两个变量之间是否存在关联。

本实施例基于收录的 2019 年 1 月至 2020 年 8 月门诊病例数据记录，获取身高、体重、职业类型、职业、文化程度、怀孕次数、分娩次数、胎数、疾病史和家族疾病史信息等基本信息，以及包括前壁 3cm 点 Aa、前壁最低点 Ba、前穹隆或阴道残端 C、生殖道裂隙 gh、前静息阶段-测试值、手测 I、A3 反馈、尿容量、最大尿流量、盆底肌静态张力、盆底肌动态张力、阴道前壁膨出、子宫/穹隆膨出、阴道后壁膨出等信息在内的多次检查结果在内的检查信息。依次利用数据挖掘技术进行数据预处理，开展变量描述性统计，变量重要性排序、个体化的变量贡献度统计，通过大量病人数据样本的训练对模型进行充分的参数调整，建立重度盆底功能障碍人工智能预测模型。

例如：利用数据挖掘技术进行数据预处理时，可以对数据进行去重清洗，删除无效变量，删除时间变量，构造目标变量，并为后续分析整理自变量和目标变量，使一个目标变量对应多个自变量；开展变量描述性统计时，运用统计检验方法分析自变量与目标变量是否有显著相关性，实现自变量与目标变量的相关性分析，获得检验统计量和统计学显著性；在变量重

要性排序时，可通过随机森林法，获得变量重要性排序，明确自变量与目标变量的相关性大小，以获得自变量的重要度；在个体化的变量贡献度统计时，可通过随机森林法，获得变量贡献度，明确自变量与目标变量的相关性方向和大小，从而获得自变量的贡献度。

原始数据中共选出 64 个变量作为模型预选变量（表 2），包括 44 个分类变量和 20 个连续变量。

表 2 64 个模型预选变量

身高	浅层手术 I
尿路感染	浅层手术 II
便秘	阴道松弛
尿失禁	急性尿失禁
痔疮	混合性尿失禁
心脏起搏器	外阴白斑
心脏植入金属	压力性尿失禁
是否吸烟	阴道后壁膨出
高血压	子宫/穹窿膨出
糖尿病	腹直肌是否分离
慢性咳嗽	盆腔总阶段-得分
腰椎间盘突出	快组 (2 类纤维) 阶段-得分
康复治疗情况	慢组 (1 类纤维) 阶段-得分不能为空
文化程度	盆腔总阶段-得分
母亲或姐妹有尿失禁的情况	总得分
母亲或姐妹有盆腔器官脱落	阴道肌力
现在体重	阴道松弛大小
就诊部门	腹直肌分离宽度
引产次数	腹肌参与度
孕前尿失禁	腹直肌分离
孕期尿失禁	BMI 体质指数 (kg/m ²)
产后尿失禁	盆腔联合分离
妊娠糖尿病	盆腔部疼痛
怀孕总次数	盆腔联合分离疼痛
分娩次数	扁平足
阴道总长 TVL	体态
临产次数	压力性尿失禁程度
是否顺产	是否做尿禁
阴道前壁膨出	尿禁类型
孕期母亲体重增加 kg	尿量
深层手术 I	FBF
深层手术 II	基础代谢

重度盆底功能障碍的数据建模结果如图 2 所示，64 个预选变量中筛选出 23 个相关变量，相关程度排序依次为：文化程度、年龄、现在体重、阴道总长 TVL、身高、怀孕总次数、痔疮、母亲或姐妹有尿失禁的情况、便秘、流产次数、分娩次数、引产次数、胎数量、是否有手术史、妊娠糖尿病、康复治疗情况、腰椎疾病、尿路感染、母亲或姐妹有盆腔器官脱落、慢性咳嗽、高血压、糖尿病、是否吸烟；其中，文化程度、阴道总长 TVL、身高呈负相关，其余变量呈正相关。模型的预测准确度 AUC 为 0.634，表明筛选变量可以解释 63.4%的重度盆底功能障碍变异；模型在测试集中得到的 f1 值为 0.674，召回率为 0.613。

进一步的，本实施例完成了成都地区 5 个社区 3 万例女性的盆底疾病高风险因素调查，利用 Logistic 和随机森林方法，获得风险因素排名，初步构建阴道前壁膨出、子宫穹隆膨出、阴道后壁膨出、压力性尿失禁、阴道松弛大小、阴道肌力的风险预警模型（图 3）；其中，压力性尿失禁的随机森林模型准确度最低，为 0.70，阴道松弛大小的经典决策树模型准确度

最高，为 0.97。当然，

由此可以证明，不同 PFD 的风险预警模型的准确度大小不一，但阴道松弛大小可以较准确的预测，进一步的可以证明，本实施例所述风险预警模型的输出结果可以判断女性 PFD 的高风险/非高风险。

实施例 3：构建阴道前壁脱垂分析的风险预警模型

按图 4 所示流程进行分析和构建风险预警模型。

（一）数据处理：

主要对盆底 app 中自 2014 年 1 月 1 日至今、生胎数量大于 0 的数据进行处理，主要分为三步：

第一步是对“胎数”进行去重清洗，最后得到最多的胎数是 5 胎，但拥有 4 胎和 5 胎的患者数量极少，本次分析主要最多选择了 3 胎数据。

第二步是对就诊日期、末次月经时间等时间变量进行删除，并对多分类变量进行伪编码处理。

第三步删除记录数量不足 3000 的变量和无效变量(无效变量指的是只有一种类型记录的变量，无信息差异)，具体数据缺失情况和变量类型见附表 1-变量类型及缺失情况(建模表)。

最终得到有效变量 151 个，12456 个有效样本。

（二）统计检验与一元 logistic 回归

运用统计检验方法分析自变量与“阴道前壁是否脱垂”(目标变量)是否有显著相关性，结果详见附表 2-数据统计检验表，另外用一元 logistic 回归分析自变量与“阴道前壁是否脱垂”(目标变量)是否有显著相关性，结果详见附表 3-一元 logistic 回归结果表。

表 3 统计检验信息表样例

变量名	ALL	Group 0	Group 1	Statistics	P_value	sample_n
尿路感染	677 (5.435%)	172 (4.36%)	505 (5.933%)	12.681	<0.001	12456
便秘	2704 (21.708%)	856 (21.698%)	1848 (21.713%)	0	0.996	12456
粪失禁	62 (0.498%)	9 (0.228%)	53 (0.623%)	7.696	0.006	12456
慢性咳嗽	228 (1.83%)	64 (1.622%)	164 (1.927%)	1.228	0.268	12456
前静息阶段_腹肌_均值	0.89 (0.64)	0.89 (0.63)	0.9 (0.65)	0.692	0.405	8219
慢肌阶段_腹肌_均值	0.89 (0.64)	0.89 (0.63)	0.9 (0.65)	0.692	0.405	8219
后静息阶段_腹肌_均值	0.89 (0.64)	0.89 (0.63)	0.9 (0.65)	0.692	0.405	8219
II 类肌纤维疲劳度	-0.33 (0.64)	-0.34 (0.66)	-0.33 (0.63)	0.038	0.845	3452
分娩时是否使用胎头吸引器_2	82.0 (2.468%)	18.0 (2.095%)	64.0 (2.598%)	0.477	0.49	3322
分娩时是否使用产钳_2	87.0 (2.619%)	19.0 (2.212%)	68.0 (2.761%)	0.553	0.457	3322

	)	)	)			
分娩时是否有撕裂伤_2	883.0(26.58%)	117.0(13.62%)	766.0(31.1%)	98.821	<0.001	3322
BMI 体质指数(kg/m2)	22.35(2.49)	22.3(2.49)	22.36(2.49)	0.274	0.601	3143
婴儿出生体重_2	3.29(0.43)	3.29(0.41)	3.3(0.44)	0.057	0.811	3030

备注：sample_n 表示对应变量不缺失的记录样本量(下同)，应用该样本量分析对应变量与“前壁是否脱垂”的相关性，P_value 小于 0.05 表示有显著相关性，否则无显著相关性。

表 4 逻辑回归结果信息表

变量名	Uni OR	Lower Bound	Upper Bound	Uni P Value	sample_n
尿路感染	1.384	1.159	1.652	0	12456
便秘	1.001	0.913	1.097	0.985	12456
浅层手测 II	0.961	0.907	1.017	0.168	12228
深层手测 II	0.957	0.904	1.013	0.129	12227
阴道总长 TVL	1.042	0.96	1.13	0.327	12149
后壁最低点 Bp	5.509	4.884	6.213	0	12145
身高	1.008	1	1.016	0.062	12068
耻骨联合分离	432177149	0	inf	0.999	3974
扁平足	0.992	0.13	7.577	0.994	3837
分娩时是否有撕裂伤_2	2.863	2.314	3.542	0	3322
BMI 体质指数(kg/m2)	1.011	0.978	1.045	0.532	3143
婴儿出生体重_2	1.03	0.853	1.243	0.76	3030

分别应用各自变量与“前壁是否脱垂”建立一元 logistic 回归，Uni P Value 小于 0.05 说明显著相关，Uni OR 大于 1 表示该变量与“前壁是否脱垂”成正相关关系，否则成负相关关系。

统计检验筛选得到 84 个与“前壁是否脱垂”显著相关的变量，logistic 回归筛选得到 78 个“前壁是否脱垂”显著相关的变量(全部包含在统计检验筛选的变量中)，因为统计量的计算方法不一致，结果会有些差异，为了不减少信息，选择 84 个变量作为后续分析的基础变量。

对得到的 84 个变量查看变量完整度信息，详细完整度信息表见附表 4-数据表完整度信息。

表 5 显著变量缺失信息表

变量名	Missing_Count	Missing_Percentage
尿路感染	0	0
粪失禁	0	0
痔疮	0	0
总得分	3788	0.304110469
快肌阶段_盆底_最大值	3891	0.312379576
前静息阶段_盆底_均值	3921	0.314788054
慢肌阶段_盆底_均值	3921	0.314788054
后静息阶段_盆底_均值	3921	0.314788054

快肌(2 类纤维)阶段-得分	3983	0. 319765575
后静息阶段-测试值(盆底/腹肌) (变异性)	4723	0. 379174695
腹直肌分离宽度	6552	0. 526011561
是否频发	8070	0. 647880539
腹直肌分离	8216	0. 659601798
I 类肌纤维肌力	8996	0. 722222222
分娩时是否有撕裂伤_2	9134	0. 73330122
孕期母亲体重增加_2	9461	0. 759553629

表中加粗变量的缺失较多，若以这些变量不缺失为标准筛选小样本数据集，会损失其他变量的很多信息，为此，后续分析将数据拆成两个数据集进行分析，缺失率小于 40%的变量与目标变量组成数据集 1，大于 40%的变量与目标变量组成数据集 2，分别对数据集 1 和数据集 2 进行分析研究各变量与“前壁是否脱垂”的相关性。

（三）RFE 进行模型变量筛选

数据集 1(含有 12456 样本)

选择 SVM 作为基础模型，以 AUC 指标作为评价指标应用 RFE 方法结合五折交叉验证进行变量筛选，最后选择了 20 个变量，获取 RFE 变量数据与 auc 值的对应曲线，如图 5 所示。当变量为 20 个时，模型 auc 值达到最高，此 20 个变量与目标变量相关性最密切，可运用这 20 个变量建模预测目标变量。

数据集 2(将缺失记录超过 2/3 的样本剔除，剩余 4447 样本)

选择 SVM 作为基础模型，以 AUC 指标作为评价指标应用 RFE 方法结合五折交叉验证进行变量筛选，最后选择了 5 个变量，获取 RFE 变量数据与 auc 值的对应曲线，如图 6 所示。

（四）建模分析

应用 1000 次 bootstrap，对数据进行抽样划分训练集和测试集，并得到变量相关性的 SHAP 图，参见下表 6、表 7 以及图 7 至图 10。

数据集 1 结果：

表 6 测试集结果

mode 1	AUC	Accuracy	Sensitivity	Specificity	F1
LR	0. 91 (0. 902, 0. 919)	0. 805 (0. 794, 0. 819)	0. 758 (0. 742, 0. 775)	0. 909 (0. 892, 0. 925)	0. 842 (0. 832, 0. 854)
SVC	0. 913 (0. 903, 0. 921)	0. 804 (0. 791, 0. 816)	0. 74 (0. 722, 0. 757)	0. 945 (0. 932, 0. 957)	0. 839 (0. 826, 0. 85)
GNB	0. 887 (0. 877, 0. 898)	0. 734 (0. 721, 0. 748)	0. 64 (0. 623, 0. 659)	0. 939 (0. 926, 0. 952)	0. 767 (0. 754, 0. 781)
RF	0. 914 (0. 904, 0. 922)	0. 835 (0. 823, 0. 846)	0. 899 (0. 887, 0. 91)	0. 694 (0. 669, 0. 721)	0. 882 (0. 873, 0. 89)
GBM	0. 922 (0. 914, 0. 93)	0. 849 (0. 837, 0. 861)	0. 872 (0. 858, 0. 886)	0. 798 (0. 775, 0. 821)	0. 888 (0. 879, 0. 897)

	0.93)	0.86)	0.885)	0.821)	0.897)
ADA	0.92(0.91, 0.927)	0.842(0.829, 0.853)	0.871(0.858, 0.884)	0.777(0.754, 0.8)	0.883(0.874, 0.892)
MLP	0.915(0.907, 0.923)	0.833(0.821, 0.844)	0.88(0.868, 0.891)	0.731(0.705, 0.755)	0.879(0.869, 0.888)

数据集 2 结果:

表 7 测试集结果

models	AUC	Accuracy	Sensitivity	Specificity	F1
LR	0.669(0.632, 0.702)	0.494(0.468, 0.522)	0.425(0.398, 0.454)	0.85(0.8, 0.894)	0.585(0.558, 0.614)
SVC	0.726(0.695, 0.758)	0.475(0.448, 0.503)	0.383(0.353, 0.41)	0.955(0.924, 0.981)	0.55(0.519, 0.578)
GNB	0.715(0.685, 0.747)	0.788(0.765, 0.81)	0.901(0.884, 0.918)	0.196(0.143, 0.251)	0.877(0.862, 0.891)
RF	0.731(0.701, 0.76)	0.839(0.819, 0.858)	1.0(1.0, 1.0)	0.0(0.0, 0.0)	0.912(0.9, 0.924)
GBM	0.72(0.687, 0.754)	0.836(0.817, 0.855)	0.996(0.993, 1.0)	0.009(0.0, 0.023)	0.911(0.899, 0.922)
ADA	0.73(0.697, 0.761)	0.839(0.819, 0.857)	1.0(1.0, 1.0)	0.0(0.0, 0.0)	0.912(0.9, 0.923)
MLP	0.717(0.685, 0.748)	0.839(0.819, 0.857)	1.0(1.0, 1.0)	0.0(0.0, 0.0)	0.912(0.9, 0.923)

综合上表 6，GBM 的效果最佳，综合上表 7，SVC 的效果最佳。

由图 5 可知，I 级阴道后壁膨出与目标变量相关性最强，且为正相关，即出现阴道后壁膨出，越可能出现阴道前壁脱垂，其次是后壁最低点 Bp，成正相关关系，即最低点 Bp 越大，越容易脱垂。

以上所述，仅是本发明的较佳实施例，并非对本发明做任何形式上的限制，凡是依据本发明的技术实质对以上实施例所作的任何简单修改、等同变化，均落入本发明的保护范围之内。

权利要求书

1、女性盆底功能障碍性疾病风险预警模型的构建方法，其特征在于：包括以下步骤：

S1.获取区域风险人群的风险因素；

S2.筛选风险因素中的相关变量并构建风险预警模型；

S3.对风险预警模型进行训练获得模型优化。

2、根据权利要求1所述的构建方法，其特征在于：所述步骤S1中，区域风险人群为选定地区的疾病易感人群。

3、根据权利要求1所述的构建方法，其特征在于：所述步骤S1中，风险因素包括盆外风险因素和盆内风险因素，所述盆外风险因素包括：年龄、BMI、糖尿病、体力劳动、慢性咳嗽、便秘、吸烟；所述盆内风险因素包括孕产次、分娩方式、阴道肌力、肛提肌裂伤。

4、根据权利要求1所述的构建方法，其特征在于：所述步骤S2中，筛选风险因素中的相关变量并构建风险预警模型的步骤包括：

a. 根据风险因素，通过Logistic回归法筛选模型预选变量；

b. 根据筛选的模型预选变量，通过支持向量机递归特征消除算法筛选相关变量；

c. 根据筛选的相关变量，通过机器学习方法构建风险预警模型。

5、根据权利要求4所述的构建方法，其特征在于：所述步骤a中，Logistic回归法的公式如下：

$$\text{logit}(P) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$

其中，P为阳性结果发生的条件概率， β 为偏回归系数，X为自变量， β_0 为常数项，1、2……p表示偏回归系数和自变量的编号。

6、根据权利要求4所述的构建方法，其特征在于：所述步骤b中，支持向量机递归特征消除算法的公式如下：

$$w = \sum_k \alpha_k y_k x_k$$

其中，w为特征权重， α 为特征参数，y为分类标签，x为训练样本，k为编号。

7、根据权利要求4所述的构建方法，其特征在于：所述步骤c中，机器学习方法包括Logistic回归法、支持向量机、高斯朴素贝叶斯法、随机森林法、梯度提升树算法、Adaboost集成学习模型或多层感知机。

8、根据权利要求4所述的构建方法，其特征在于：所述步骤c中，风险预警模型的输出结果包括预测准确度AUC、召回率和f1值。

9、一种采用权利要求1~8任一项所述方法构建得到的女性盆底功能障碍性疾病风险预警模型。

10、一种计算机系统，其特征在于：包括储存器、处理器以及存储在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序时实现如权利要求 1~8 任一项所述预警模型的构建方法。