文章编号:1008-1534(2024)02-0099-09

基于机器学习的妊娠期糖尿病智能预测系统设计与实现

马金龙[1],徐立魏[1],杨志芬[2],王胜普[2],郑　芮[2],张永强[1]

(1. 河北科技大学信息科学与工程学院,河北石家庄,050018;2. 河北医科大学第四医院产科,河北石家庄,050035)

摘　要:为帮助有效识别患有妊娠期糖尿病的孕妇,提前干预治疗,降低其潜在风险,设计并开发了妊娠期糖尿病智能预测系统。首先,在对比 10 种机器学习模型的基础上对临床数据集进行数据归一化和特征筛选,减少模型计算复杂度和一些使模型不稳定的特征;其次,使用 Stacking 算法对 10 种不同的机器学习模型进行集成,分别搭建 2 个集成模型 Stacking1 和 Stacking2,比较集成模型的性能;最后,基于 PyQt5 设计妊娠期糖尿病智能预测系统,预测孕妇患妊娠期糖尿病的风险,并给出建议。结果表明:使用 10 个机器学习模型以及 2 个集成模型对 GDM 进行预测,发现 GBDT 的预测结果高于其他机器学习模型,集成模型 Stacking2 将多个异质学习器相结合表现出较高的准确性和可靠性,且评价指标 Accuracy、Precision、Recall、AUC 分别为 0.900 9、0.901 2、0.900 7、0.900 7,均高于同类模型。智能预测系统能有效预测妊娠期糖尿病的风险,能够及早发现易患病人群,并提供妊娠期糖尿病的科普知识,从而加强对易患病人群的健康管理,降低妊娠期糖尿病发生的风险。

关键词:计算机决策支持系统;机器学习;PyQt5;集成学习;妊娠期糖尿病;特征筛选

中图分类号:TP18　　**文献标识码**:A　　**DOI**: 10.7535/hbgykj.2024yx02003

Design and implementation of gestational diabetes mellitus intelligent prediction system based on machine learning

MA Jinlong[1], XU Liwei[1], YANG Zhifen[2], WANG Shengpu[2], ZHENG Rui[2], ZHANG Yongqiang[1]

(1. School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang, Hebei 050018, China; 2. Department of Obstertrics, The Fourth Hospital of Hebei Medical University, Shijiazhuang, Hebei 050035, China)

Abstract: To help to effectively identify pregnant women with gestational diabetes mellitus (GDM), intervene in advance for treatment, and reduce potential risks, a GDM intelligent prediction system was designed and developed. Firstly, based on the comparison of 10 machine learning models, data normalization and feature selection were conducted on the clinical dataset based on these models to reduce model computational complexity and unstable features. Secondly, the Stacking algorithm was used to integrate 10 different machine learning models, and two integrated models, Stacking1 and Stacking2, were built to compare

收稿日期:2023-07-31;修回日期:2024-03-04;责任编辑:王淑霞

基金项目:河北省自然科学基金(H2022206212);河北省医学科学研究课题计划项目(20230775,20230776,20210715)

第一作者简介:马金龙(1981—),男,河北定州人,副教授,博士,主要从事生物信息学方面的研究。

通信作者:杨志芬副主任医师。E-mail: cxsxfy@126.com

马金龙,徐立魏,杨志芬,等.基于机器学习的妊娠期糖尿病智能预测系统设计与实现[J].河北工业科技,2024,41(2):99-107.

MA Jinlong, XU Liwei, YANG Zhifen, et al. Design and implementation of gestational diabetes mellitus intelligent prediction system based on machine learning[J]. Hebei Journal of Industrial Science and Technology, 2024, 41(2): 99-107.

their performance. Finally, based on PyQt5, a GDM intelligent prediction system was designed to predict the risk of GDM in pregnant women and provide suggestions. The results show that using 10 machine learning models and two integrated models to predict GDM, it is found that the prediction results of GBDT are higher than other machine learning models. The integrated model Stacking2, combining multiple heterogeneous learners, exhibits high accuracy and reliability, with evaluation indicators of Accuracy, Precision, Recall, and AUC of 0.900 9, 0.901 2, 0.900 7, and 0.900 7, respectively, all higher than similar models. The intelligent prediction system can effectively predict the risk of GDM, identify susceptible individuals early, and provide popular science knowledge of GDM, thereby strengthen the health management of susceptible individuals and reduce the risk of GDM occurrence.

Keywords: computer decision support system; machine learning; PyQt5; integrated learning; gestational diabetes mellitus; feature selection

随着人们物质生活水平的提高,妊娠期糖尿病(gestational diabetes mellitus, GDM)的患病风险越来越高。GDM是指在妊娠期首次发现或发生的糖代谢异常。目前,研究发现GDM是孕妇异常妊娠的原因之一,异常妊娠包含妊娠高血压、Ⅱ型糖尿[1]等疾病。GDM不仅对孕妇有严重的危害,而且对胎儿也有长期的负面影响,例如婴儿出现肺部发育不良、心血管疾病、尿路感染、巨大儿等问题[2]。甚至有孕妇在患GDM 6年后而引发的Ⅱ型糖尿病其原因仍然未知[3-6]。由此可见,GDM已经成为不容忽视的问题。

近年来,利用机器学习预测GDM取得了一定突破。WU等[7]利用机器学习、逻辑回归、k近邻、支持向量机和深度神经网络等方法,对GDM进行了预测。TRAN等[8]采用贝叶斯模型方法建立了预测模型。ZHENG等[9]使用不同的贝叶斯模型方法建立了预测模型。ZHANG等[10]基于特征工程选取了40个重要特征,构建了逻辑回归、Lasso-logistics、GBDT (gradient boosting decision tree)、XGBoost (excreme gradient boosting)、LightGBM (light gradient boosting machine)和CATBoost (categorical boosting)机器学习模型。尽管一些研究者即使对数据进行特征筛选但重要特征还是较多,这会导致模型计算和存储成本增加和过拟合风险的增加。

目前大多数研究使用单一的机器学习模型预测GDM,不仅缺乏泛化能力,还对特定的数据具有较强的依赖性,这使得模型只能处理特定的数据集,对其他新数据集无法较好的适应。鉴于在不同情况下每种机器学习方法各有优缺点,有研究者使用了集成模型,如LU等[11]分别使用LR (logistic regression)、XGBoost、LightGBM、基于LSTM的神经网络模型和集成方法进行了预测。虽然使用了集成模型,但是集成机器学习模型较少,影响了集成模型在数据集上的泛化能力。

本研究首先对数据集进行详尽的处理,随后进行了精确的特征筛选,并搭建了2个集成模型Stacking1、Stacking2。其次,将数据集分别输入到10个单一的机器学习模型和2个集成模型中进行预测,通过比较各模型的性能,确定最优模型。最后,利用这个最优模型,结合PyQt5开发了一个智能预测系统,以期精确预测妊娠期糖尿病的风险,为提前干预治疗提供可能。

1 基于集成学习的GDM数据处理与预测算法

1.1 实验数据描述

数据集来自河北医科大学第四医院共1 002例孕妇的检查记录,是真实的临床数据。GDM患者和非GDM患者各占总样本的50%。表1为GDM患者的相关数据。

表1 GDM患者数据描述

Tab.1 Data description of GDM patients

序号	特征参数	特 征
1	Myear	年龄
2	Height	身高
3	Abnormal	异常妊娠史
4	PCOS	多囊卵巢综合征
5	S1	高血压
6	Gnumber	怀孕次数
7	Pnumber	生产数量
8	Preweight	怀孕前体重
9	Drug	药物治疗
10	Durweight	怀孕期间体重
11	PIH	妊娠期高血压
12	Anemia	妊娠期贫血
13	Hdisease	心脏病
14	HPT1	甲状腺功能亢进
15	HPT2	甲状腺功能减退
16	AF1	羊水过多
17	AF2	羊水过少
18	PP	前置胎盘
19	FGR	FGR
20	DMhistory	糖尿病家族史
21	SN	目标特征(1是患病,2是未患病)

1.2　数据处理

孕前体重和孕期体重差异较大，使得模型的计算量较大，不宜直接输入到模型中。因此，采用数据归一化法将数据映射到[0,1]区间，归一化的数据呈正态分布，提高了机器学习模型的预测效果。通过比较 4 种数据归一化处理方法发现，min-max 归一化法将原始数据等比缩放不仅保留了原始数据之间的相互关系，且有效提高模型的预测精度、缩短模型的计算时间。

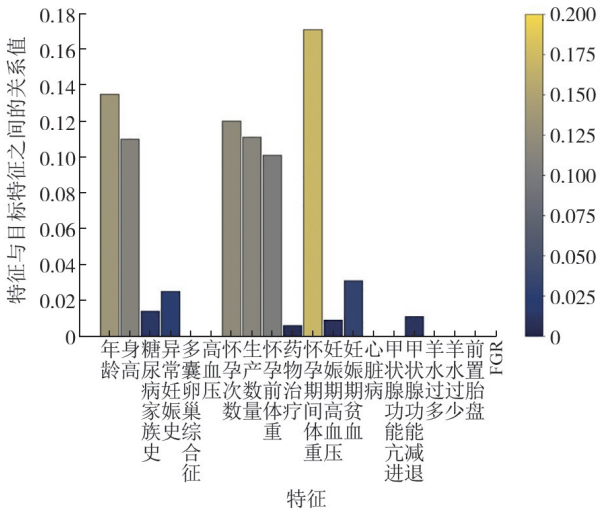
min-max 归一化计算公式为

$$y_i = \frac{x_i - \min\limits_{1 \leqslant j \leqslant n}\{x_j\}}{\max\limits_{1 \leqslant j \leqslant n}\{x_j\} - \min\limits_{1 \leqslant j \leqslant n}\{x_j\}}, \tag{1}$$

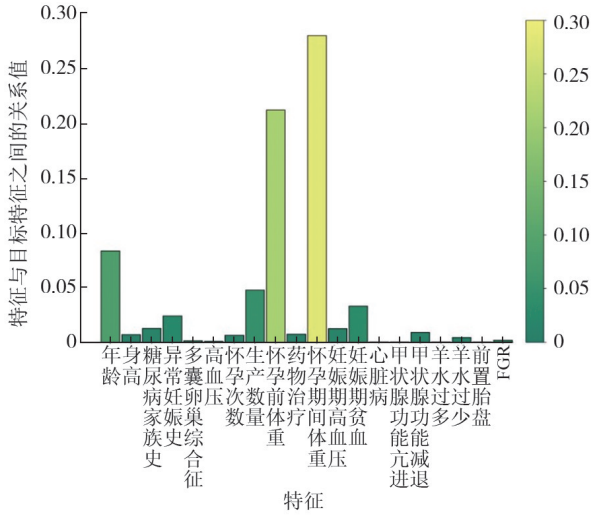
式(1)中，$y_1, y_2, \cdots, y_i \in [0, 1]$ 且无量纲，为归一化后的数据，其中 min 和 max 分别为特征的最小值和最大值。

1.3　特征选择

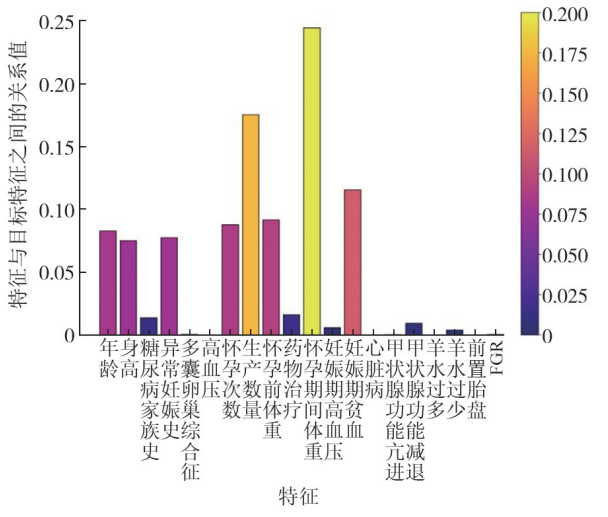
原始数据集包含 20 个自变量特征和 1 个因变量特征。鉴于每个模型的重要特征不同，调用 Python 中 sklearn 库中的 feature_importances_和 conef_函数对 LightGBM、XGBoost、CATBoost、RF (random forest)、LR、GBDT、DT (decition tree)、MLP (multi layer perceptron)、SVM (support vector machine)和 AdaBoost(adaptive boosting)各个模型的重要特征进行选择[13-17, 20-24]。经实验可知，将每个模型的前 9 个重要特征作为输入特征，模型可达到较优的预测效果。10 个模型的特征系数如图 1 所示。



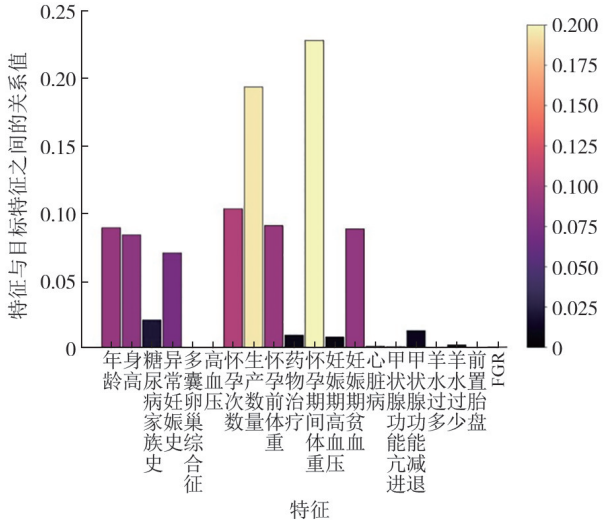
a）LightGBM模型的重要特征



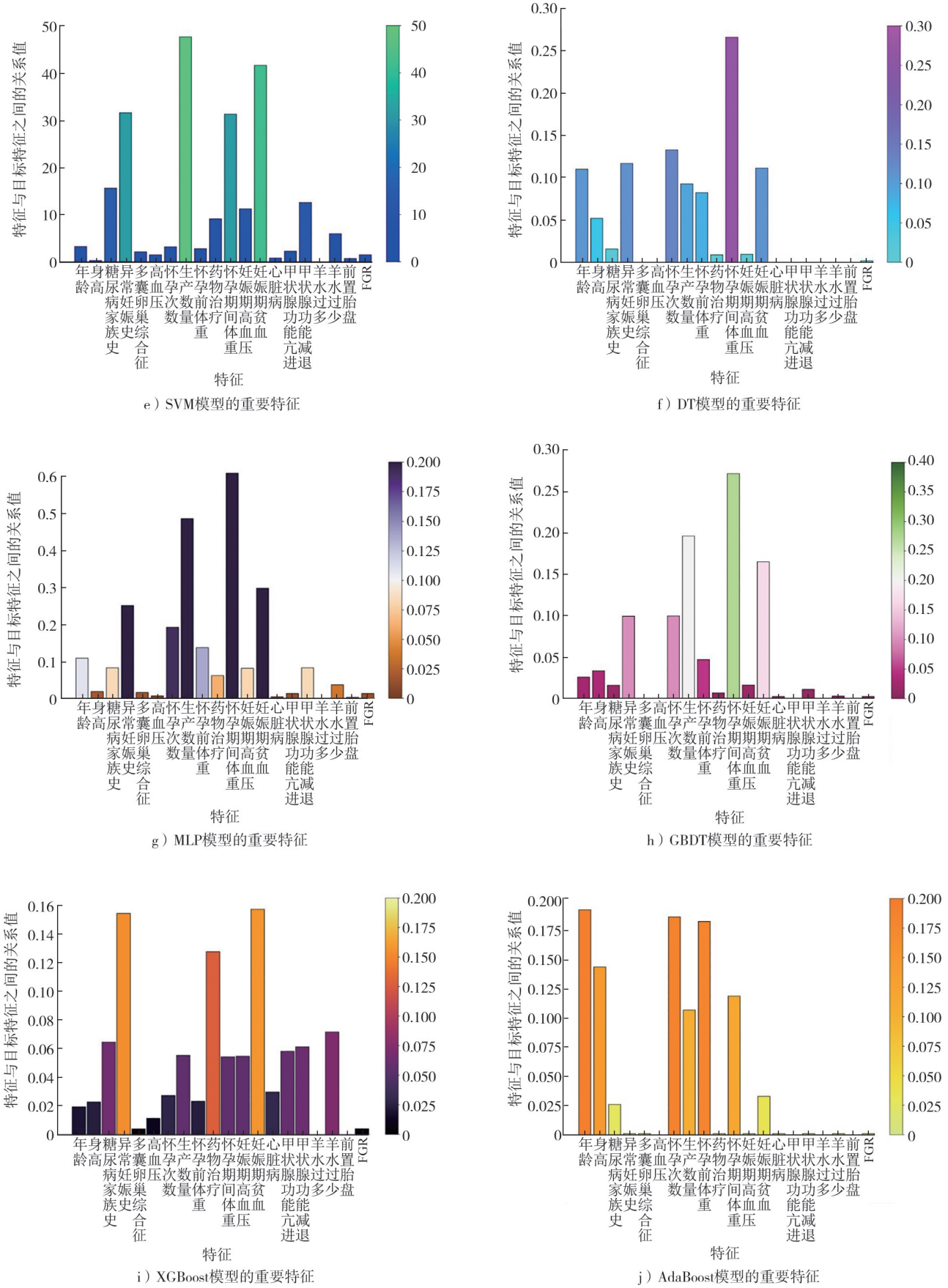
b）LR模型的重要特征



c）RF模型的重要特征



d）CATBoost模型的重要特征



e）SVM模型的重要特征

f）DT模型的重要特征

g）MLP模型的重要特征

h）GBDT模型的重要特征

i）XGBoost模型的重要特征

j）AdaBoost模型的重要特征

图 1 不同模型的特征系数图

Fig. 1 Feature coefficient charts for different models

1.4　机器学习模型的建立

研究中共使用 10 种不同的机器学习模型，机器学习模型的预测流程如图 2 所示。

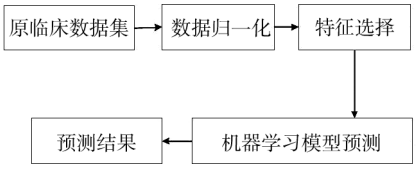


图 2　机器学习预测流程图

Fig. 2　Machine learning prediction flowchart

首先，对原始临床数据进行 min-max 归一化处理。其次，进行特征选择，将 10 个模型的不同重要特征分为 10 个不同的数据子集。最后，将数据集按照 7∶3 的比例分为训练集和测试集，将训练集输入到机器学习模型中对其进行训练，然后将测试集分别输入到训练好的机器学习模型中进行 GDM 的预测。

1.5　集成模型的建立

Stacking 集成方法是将一系列模型（基模型）的输出结果作为新特征输入到其他模型（元模型），实现了模型的层叠[19]。Stacking 分为 2 层，分别称为基模型和元模型。使用 Stacking 集成方法搭建 2 个集成模型，其中 Stacking1 选取 4 个预测效果较好的模型作为基模型，LR 作为元模型。Stacking2 则将 9 个模型作为基模型，LR 作为元模型。

将数据集以 7∶3 的比例分为训练集、测试集与验证集，在训练模型时使用五折交叉验证训练，如图 3 所示。交叉验证步骤如下。

1）以基模型中第 1 个模型（模型 1）为例，需要将训练集分为 5 份，模型 1 进行 5 次训练，每次在训练集中挑选 1 份作为验证集，其他 4 份作为训练集，经过一次训练后，将在验证集上的输出记作 $a1$，在测试集上的输出记作 $b1$。经过第 2 次训练后，将在验证集上的输出记作 $a2$，在测试集上的输出记作 $b2$。以此类推，模型 1 经过五折交叉验证后得到 $a1$、$a2$、$a3$、$a4$、$a5$ 和 $b1$、$b2$、$b3$、$b4$、$b5$，即 model1 训练后在验证集上的输出结果，将它们拼接在一起得到 $A1(C1)$，为 model1 训练后在数据集上的预测结果。$b1$、$b2$、$b3$、$b4$、$b5$ 为 model1 在测试集上的输出结果，将其相加后求平均值得到 $B1(D1)$，即 model1 完成五折交叉验证。

2）$A1$ 作为元模型的训练集，$B1$ 作为元模型的测试集，对最终结果进行预测。

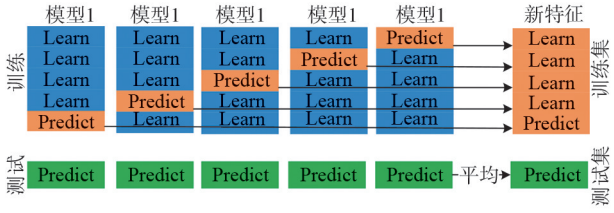


图 3　五折交叉验证流程图

Fig. 3　Five-fold cross-validation flowchart

Stacking1 重复上述过程可得到 $C1$、$C2$、$C3$、$C4$ 将其合并作为元模型的训练集，$D1$、$D2$、$D3$、$D4$ 作为测试集。Stacking2 重复上述过程可得到 $A1$、$A2$、$A3$、$A4$、$A5$、$A6$、$A7$、$A8$、$A9$ 将其合并作为元模型的训练集，$B1$、$B2$、$B3$、$B4$、$B5$、$B6$、$B7$、$B8$、$B9$ 作为测试集。这里只是将训练好的模型进行封装，所以测试集暂时不输入模型测试。模型流程图如图 4 和图 5 所示。

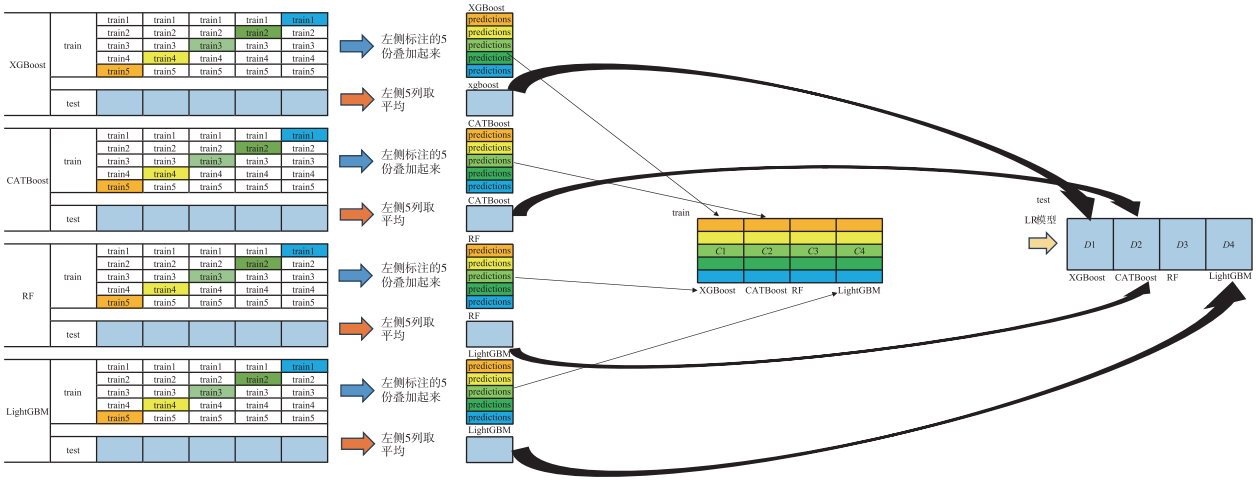


图 4　Stacking1 集成模型流程图

Fig. 4　Stacking1 integration model flowchart

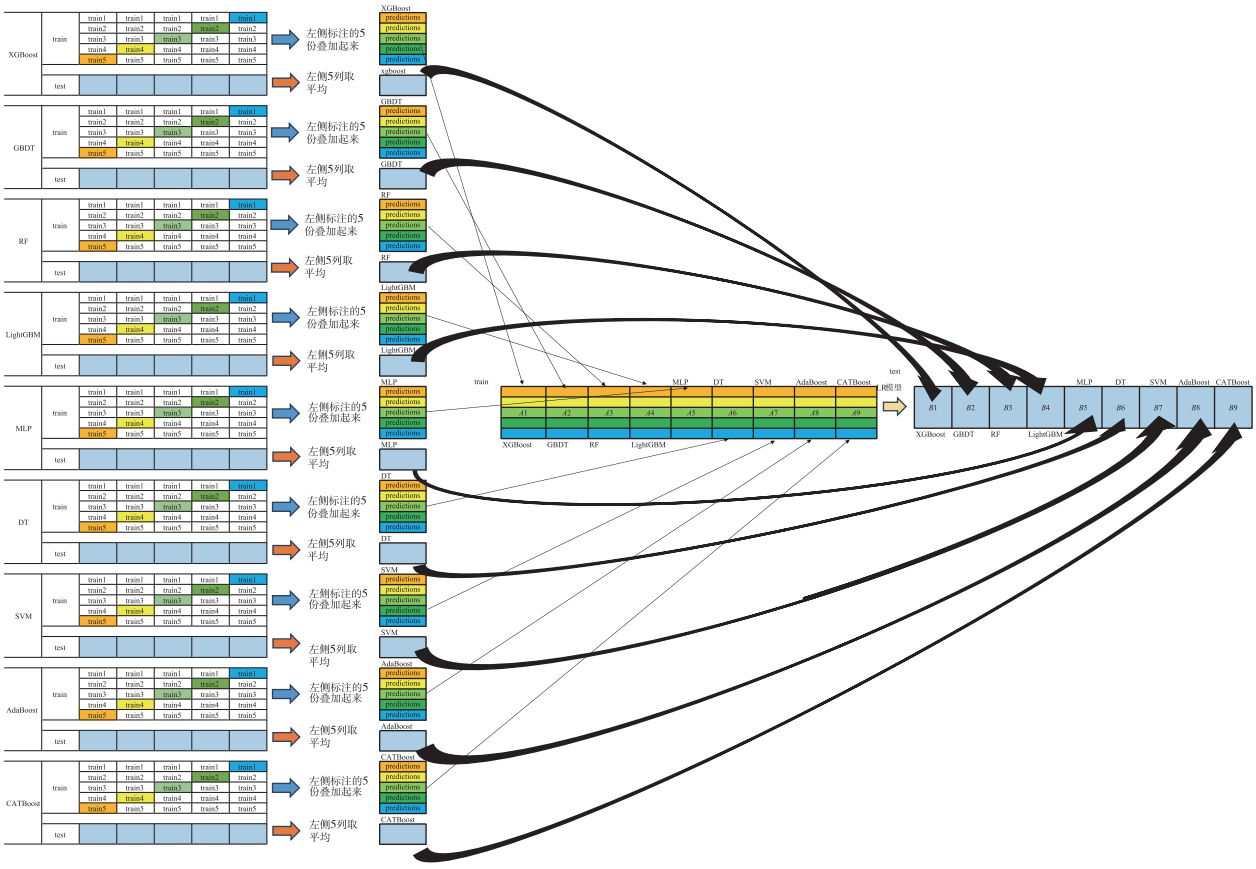


图 5 Stacking2 集成模型流程图

Fig. 5 Stacking2 integration model flowchart

2 评估方法与结果分析

2.1 评估方法

机器学习方法的有效性和准确性可以使用性能指标进行评估。当数据被归类为 GDM 时，会出现积极的分类，当数据没有被归类为 GDM 时，有一个负面的分类。模型评估的 4 种评价指标，如式(2)—式(5)：

$$\text{Accuracy}=\frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}}, \tag{2}$$

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}}, \tag{3}$$

$$\text{Recall}=\text{True positive rate}=\frac{\text{TP}}{\text{TP}+\text{FN}}, \tag{4}$$

$$\text{False positive rate}=\frac{\text{FP}}{\text{TN}+\text{FP}}, \tag{5}$$

式中：Accuracy，为准确率，表示在整体样本中被正确分类的样本所占比例；TP 表示真阳性，即模型正确识别为患 GDM；TN 表示真阴性，即模型正确识别为未患 GDM；FP 表示假阳性，即模型识别非 GDM 患者为 GDM 患者；FN 表示假阴性，即模型识别 GDM 患者为非 GDM 患者；Precision，即精准率，表示在模型预测为正样本的集合中，实际为正样本的比例；Recall 为召回率；True positive rate 为真阳性率，表示在整体实际为正样本中，被预测为正样本的比例，召回率与真阳性率相同；False positiver ate，为假阳性率，表示负样本被错误分类为正样本的比例。

AUC(area under the curve)为 ROC(Receiver operating characteristic)曲线下的面积，ROC 曲线以 False positive rate 为横坐标，以 Ture positive rate 为纵坐标，绘制曲线以重新评估模型在不同分类阈值下的性能。AUC 值越大表示模型性能越好。

2.2 结果分析

表 2 为 10 种单一机器学习模型，在 4 种评价指标下对 GDM 的预测结果。可以看出，GBDT 的 AUC 值最高，达到了 0.868 4。与其他单一模型相比，GBDT 具有较高的预测性能。

尽管 GBDT 模型的 AUC 值已经达到 0.868 4，为了进一步提高预测准确性，采用集成学习模型 Stacking1、Stacking2 融合了各基模型的优点，提升了模型的性能，且通过交叉验证能够有效避免模型出现过拟合现象。单一机器学习 GBDT 与 2 个集成学习模型的预测对比结果如表 3 所示。Stacking1 集成了 5 个模型，Stacking2 集成了 10 个模型。从表 3 中可以发现，Stacking2 的 AUC 值最高，达到了 0.900 7，相比最优的单一机器学习模型

GBDT 提高了 0.032 3 个百分点，相比于集成较少机器学习模型的 Stacking1 提高了 0.024 8 个百分点。因此，Stacking2 结果得到了有效提高。通过采用集成多个模型的方法，实验成功提高了预测准确性。

表 2　10 种模型预测结果

Tab. 2　Prediction results of 10 models

模　型	Accuracy	Precision	Recall	AUC
GBDT	0.867 9	0.870 9	0.868 4	0.868 4
RF	0.858 0	0.856 0	0.858 2	0.858 2
XGBoost	0.825 0	0.824 8	0.824 9	0.824 9
CATBoost	0.831 6	0.832 5	0.832 2	0.832 2
LightGBM	0.851 4	0.852 5	0.851 7	0.851 7
MLP	0.778 8	0.779 4	0.779 0	0.779 0
DT	0.752 4	0.752 7	0.752 6	0.752 6
SVM	0.739 2	0.745 9	0.740 0	0.740 0
AdaBoost	0.798 6	0.802 3	0.799 2	0.799 2
LR	0.620 4	0.621 6	0.621 6	0.621 6

表 3　GBDT 与集成学习模型的预测结果对比

Tab. 3　Comparison of prediction results between GBDT and ensemble learning models

模　型	Accuracy	Precision	Recall	AUC
GBDT	0.867 9	0.870 9	0.868 4	0.868 4
Stacking1	0.879 1	0.878 2	0.875 9	0.875 9
Stacking2	0.900 9	0.901 2	0.900 7	0.900 7

3　GDM 预测系统实现

GDM 预测系统主要包含前端界面交互系统和后端模型预测算法，如图 6 所示。用户交互系统包含辅助功能模块和显示模块，当使用此系统进行特征输入时如果输入有误，可以点击清除数据重新输入；如果输入正确可以点击开始预测，预测完成后将由显示模块显示结果和温馨提示。完成后可以点击返回上一页也可退出系统。

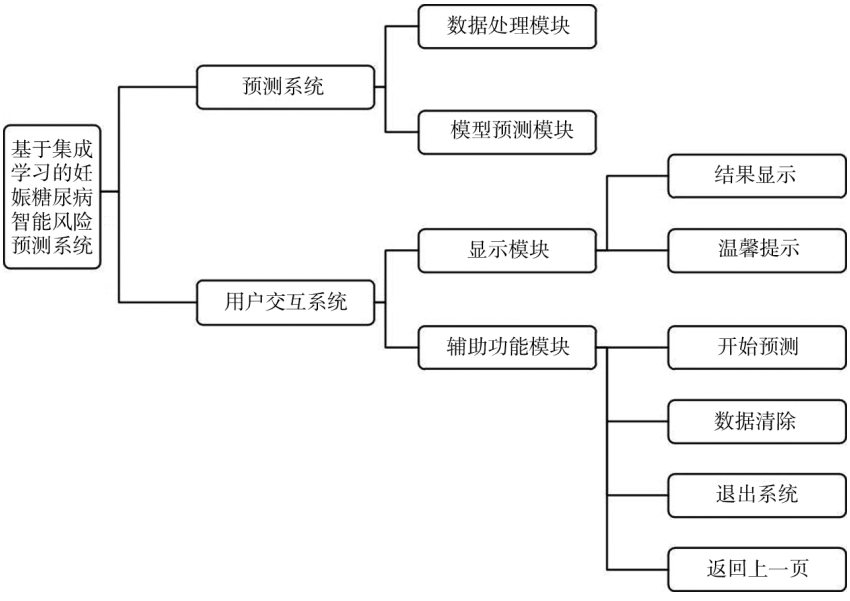


图 6　GDM 预测系统整体构架图

Fig. 6　Overall architecture diagram of GDM prediction system

3.1　信息采集界面

将每个模型的重要特征进行整合放在信息采集界面中，使用 PyQt5 中的 LineEdit 作为文本输入框。设置数据采集页面的确定按钮交互用户输入数据后，点击确定按钮将调用 MySQL 将数据存入建立好的表中。信息采集界面如图 7 所示。

设置清空按钮交互，其作用是如果多个特征输入错误可以同时删除，方便输入与删除。

3.2　模型预测界面

模型预测界面由显示模块和辅助功能模块组成，

图 7　信息采集界面

Fig. 7　Information acquisition interface

如图 8 所示。其中,显示模块主要采用 PyQt5 平台中的 QgraphicsView 作为显示控件,其显示内容包括预测结果以及系统根据预测结果所提出的建议。PyQt5 为开发者提供了 QVBoxLayout、QHBoxLayout、QGridLayout 等基本布局管理类,这些基本布局管理类可以通过相互嵌套实现自适应窗口大小的功能,本系统采用了 QVBoxLayout 和 QHBoxLayout 的嵌套。该系统中包含预测、清除数据、退出程序和返回上一页等智能化功能。将训练好的模型与数据归一化程序进行封装,然后利用 PyQt5 提供的信号和槽对其进行调用。

设置清除数据按钮,为保护用户的隐私,点击清除数据按钮后将删除 MySQL 中存储的关于用户的所有信息。设置返回上一页按钮交互,其功能是错误输入后点击了确定,不能保证预测准确性,需要返回上一页。同时返回上一页按钮中带有删除数据的隐藏功能。如果返回上一页,第 1 次输入的数据将在存入的表中被删除,需要用户重新输入特征保证输入数据的准确性。

安装 Pyinstaller 将 python 程序打包为. exe 脚本,以提高系统的可移植性。

图 8 预测界面

Fig. 8 Prediction interface

4 结 语

为提高 GDM 的诊断准确性,辅助医生制定个性化治疗方案,从而保障母婴健康,降低并发症风险,本文使用不同的多个机器学习和集成学习对 GDM 进行预测。

1)根据 10 个模型的不同重要特征,建立了 10 个数据子集,其中每个数据子集中只包含 9 个特征,减少了模型的计算和存储成本。使用 10 个机器学习模型对 GDM 进行预测,发现 GBDT 的预测结果高于其他机器学习模型。

2)利用 10 个机器模型搭建了 2 个集成模型分别为 Stacking1、Stacking2。其中 Stacking1 因集成的模型数量较少,未能达到最好的泛化能力,Stacking2 集成了更多个模型,其最优 AUC 为 0.900 7 均高于其他模型,有更好的泛化能力和更准确的预测结果。

3)开发的系统调用训练好的 Stacking2 模型进行 GDM 的预测,用户可根据系统按钮进行数据采集和最终的预测,预测结果中包含了对用户的温馨提示,为患者提供更优质的服务,为风险孕妇提前干预治疗作准备。

通过实验发现,集成较多数量的机器学习模型能够有效提高集成模型的预测性能。而集成数据数量较少,训练后的模型达不到最优的泛化能力。因此,后期应采集较多的实验数据对模型进行训练,以达到模型的最优泛化能力,进一步提高模型的预测性能。

参考文献/References:

[1] LEEUWEN M V,OPEMEER B C,ZWEERS E,el at. Estimating the risk of gestational diabetes mellitus:A clinical prediction model based on patient characteristics and medical history [J]. BJOG,2010,117(1):69-75.

[2] PATIL B M,JOSHI R C,TOSHNIWAL D. Association rule for classification of type-2 diabetic patients[C]//2010 Second International Conference on Machine Learning and Computing. Bangalore, India:IEEE,2010:330-334.

[3] MADHAVAN A,KUMARI B,SANAL M G. A pilot study on the usefulness of body mass index and waist hip ratio as a predictive tool for gestational diabetes in Asian Indians[J]. Gynecological Endocrinology,2008, 24(12):701-707.

[4] SESHIAH V,BALAJIV,BALAJI M S,et al. Glycemic level at the first visit and prediction of GDM[J]. Journal of the Association of Physicians of India,2007,55:630-632.

[5] CAMPOS M A A,REICHELT A A J,FAÇANHA C,et al. Evaluation of a 1-h 75-g oral glucose tolerance test in the diagnosis of gestational diabetes[J]. Brazilian Journal of Medical and Biological Research,2008, 41(8):684-688.

[6] AKINCI B,CELTIK A,YENER S,et al. Is fasting glucose level during oral glucose tolerance test an indicator of the insulin need in gestational diabetes? [J]. Diabetes Research and Clinical Practice,2008,82(2):219-225.

[7] WU Yanting,ZHANG Chenjie,MOL B W,et al. Early prediction of gestational diabetes mellitus in the Chinese population via advanced machine learning[J]. The Journal of Clinical Endocrinology & Metabolism,2021,106(3):e1191-e1205.

[8] TRAN T S,HIRST J E,DO M A T,et al. Early prediction of gestational diabetes mellitus in Vietnam:Clinical impact of currently recommended diagnostic criteria[J]. Diabetes Care, 2013,36(3):618-624.

[9] ZHENG Tao, YE Weiping, WANG Xipeng, et al. A simple model to predict risk of gestational diabetes mellitus from 8 to 20 weeks of gestation in Chinese women[J]. BMC Pregnancy and Childbirth,2019, 19(1):252.

[10] ZHANG Jie，WANG Fang. Prediction of gestational diabetes mellitus under cascade and ensemble learning algorithm[J]. Computational Intelligence and Neuroscience，2022，2022：3212738.

[11] LU Xinxi，WANG Jikai，CAI Junxia，et al. Prediction of gestational diabetes and hypertension based on pregnancy examination data[J]. Journal of Mechanics in Medicine and Biology，2022，22(3)：2240001.

[12] 冯鑫磊，俞凯，袁贞明. 改进 Stacking 算法在妊娠期糖尿病预测中的应用[J]. 杭州师范大学学报(自然科学版)，2023，22(2)：126-134.
FENG Xinlei，YU Kai，YUAN Zhenming. Application of optimized stacking algorithm in prediction of gestational diabetes mellitus[J]. Journal of Hangzhou Normal University(Natural Sciences Edition)，2023，22(2)：126-134.

[13] KE Guolin，MENG Qi，FINLEY T，et al. LightGBM：A highly efficient gradient boosting decision tree[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach：NIPS，2017：3149-3157.

[14] CHEN Tianqi，GUESTRIN C. XGBoost：A scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco：Association for Computing Machinery，2016：785-794.

[15] DOROGUSH A V，ERSHOV V，GULIN A. CatBoost：Gradient boosting with categorical features support[EB/OL]. (2018-10-24)[2023-08-29]. https://arxiv. org/abs/1810.11363.

[16] BREIMAN L. Random forests[J]. Machine Learning，2001，45(1)：5-32.

[17] ADENOMON M O，JOHN DO. Modelling hypertension and risk factors among adults using ordinal logistics regression model[EB/OL]. (2020-01-24)[2023-08-29]. https://europepmc. org/article/ PPR/ PPR110756.

[18] MARK S. Qt 高级编程[M]. 白建平. 译. 北京：电子工业出版社，2011.

[19] 王鹏，曹丽惠，阮冬茹. 基于 Stacking 模型融合的店铺销量预测[J]. 河北工业科技，2022，39(3)：204-209.
WANG Peng，CAO Lihui，RUAN Dongru. Store sales forecast based on Stacking model fusion[J]. Hebei Journal of Industrial Science and Technology，2022，39(3)：204-209.

[20] FRIEDMAN J H. Greedy function approximation：A gradient boosting machine[J]. Annals of Statistics，2001，29(5)：1189-1232.

[21] QUINLAN J R. Induction of decision trees[J]. Machine Learning，1986，1(1)：81-106.

[22] LIPPMANN R. An introduction to computing with neural nets[J]. IEEE ASSP Magazine，1987，4(2)：4-22.

[23] BURGES C J C. A tutorial on support vector machines for patternrecognition[J]. Data Mining and Knowledge Discovery，1998，2(2)：121-167.

[24] FREUND Y，SCHAPIRE R E. A decision-theoretic generalization of on-line learning and an application to boosting[J]. Journal of Computer and System Sciences，1997，55(1)：119-139.