

DOI:10.3969/j.issn.1005-202X.2022.08.015

医学信号处理与医学仪器

Stacking集成学习算法验证动脉损伤对糖尿病早期检测的意义

张明伟^{1,2}, 张天逸^{1,2}, 钟鸣³, 程云章^{1,2}

1. 上海理工大学健康科学与工程学院, 上海 200093; 2. 上海介入医疗器械工程技术研究中心, 上海 200093; 3. 复旦大学附属中山医院, 上海 200032

【摘要】背景:糖尿病可引起广泛的动脉结构和功能病理变化,导致动脉僵硬增加、顺应性降低和动脉弹性降低。本研究从动脉损伤的角度,实现对尚未出现临床表现但有动脉损伤的糖尿病患者的早期检测。**方法:**动脉损伤会导致血管的力学参数发生变化,而脉搏信号的波形变化与心血管系统的力学参数变化密切相关。通过9级小波对糖尿病患者脉搏信号进行分解,提取cD8、cD7、cD6系数(中高频成分,代表信号细节特征),作为能够反映动脉损伤程度的特征,将特征矩阵输入到10折交叉验证模型的Stacking集成学习模型中,设置第一层的4个基学习器为SVM、Random Forest、XGBoost、Extra Trees,第二层的元学习器是KNN。**结果:**单个机器学习模型可以达到90%以上的准确率。Stacking集成学习算法的准确率比单一机器学习模型高4%~5%,ROC曲线下面积提高1%~6%。**结论:**小波分解得到的脉搏信号cD8、cD7、cD6系数可以有效反映糖尿病引起的动脉损伤程度,因此动脉损伤对糖尿病的早期检测具有一定的指导意义。Stacking集成学习算法将多个模型的优势结合起来生成一个新模型,可以获得比单一模型更好的性能。

【关键词】糖尿病;脉搏信号;小波分解;集成算法;动脉损伤

【中图分类号】R318;R587.1

【文献标志码】A

【文章编号】1005-202X(2022)08-1003-07

Verifying the significance of arterial injury for early detection of diabetes by Stacking ensemble learning algorithm

ZHANG Mingwei^{1,2}, ZHANG Tianyi^{1,2}, ZHONG Ming³, CHENG Yunzhang^{1,2}

1. School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China; 2. Shanghai Interventional Medical Device Engineering Technology Research Center, Shanghai 200093, China; 3. Zhongshan Hospital, Fudan University, Shanghai 200032, China

Abstract: Background Diabetes can cause extensive pathological changes in the structure and function of arteries, leading to increased arterial stiffness, decreased compliance, and decreased arterial elasticity. From the perspective of arterial injury, this study aims to realize the early detection of diabetes in patients who have not yet appeared clinical manifestations of diabetes but have arterial injury. **Methods** Arterial injury leads to mechanical parameters changes in the vascular system. The waveform changes of pulse signals are closely related to mechanical parameters changes in the cardiovascular system. By decomposing the pulse signals of diabetic patients with 9-level wavelet, cD8, cD7 and cD6 coefficients (medium-high frequency components, representing the features in signal details) were extracted as features that reflect the degree of arterial injury. The feature matrix was input into the Stacking ensemble learning algorithm of the 10-fold cross-validation model, with SVM, Random Forest, XGBoost and Extra Trees as the 4 base-learners of the first layer, and KNN as the meta-learner of the second layer. **Results** A single machine learning model could achieve an accuracy higher than 90%. Stacking ensemble learning algorithm was 4%-5% higher than a single machine learning model in accuracy, and 1%-6% higher in area under the ROC curve (AUC). **Conclusion** The cD8, cD7, and cD6 coefficients of pulse signals obtained by wavelet decomposition can effectively reflect the degree of arterial injury caused by diabetes. Therefore, arterial injury has certain guiding significance for the early detection of diabetes. Stacking ensemble learning algorithm that combines the advantages of multiple models to generate a new model can achieve better performance than single models.

Keywords: diabetes; pulse signal; wavelet decomposition; ensemble algorithm; arterial injury

【收稿日期】2022-02-19

【基金项目】上海市公共卫生体系建设3年行动计划(2020-2022年)学科带头人计划项目(GWV-10.2-XD32);上海市“科技创新行动计划”生物医药科技支撑专项(20S31905100);上海工程技术研究中心资助项目(18DZ2250900)

【作者简介】张明伟,硕士研究生,E-mail: 1294851516@qq.com

【通信作者】程云章,教授,博士生导师,E-mail: cyz2008@usst.edu.cn

前言

大量心血管疾病研究数据表明,高血压、高血糖等多种代谢异常是心血管疾病的危险因素^[1-2]。它们通过破坏血管内皮细胞的结构和功能显著加速动脉硬化进程。高血糖可促进晚期糖基化终产物的产生,通过炎症因子破坏血管壁,引起大小动脉结构和功能的广泛病理改变,导致动脉僵硬度增加、顺应性降低、动脉弹性减弱。通过临床测量和数学模型计算发现,随着血管外周阻力、血管壁硬化程度等生理因素的变化,脉搏信号的特征会发生一系列有规律的变化。随着心脏的间歇性收缩和舒张,血压的波动、血流速度和血流,以及血管壁的变形和振动在血管系统中的传播波形,统称为脉搏信号^[3]。脉搏信号的传播特征与心血管系统力学参数的变化密切相关。在外周阻力低、血管壁弹性较好的健康人群中,脉搏信号的升、降支较陡,形成高而尖的主波。由于反射波速度低,冲击主动脉瓣的血液回流波动强度大,重搏波的波峰和波谷明显。随着外周阻力增加,血管壁弹性变差,由于反射波速度逐渐增大,重搏波逐渐变得不明显,重搏波波谷继续上升。图1反映糖尿病患者与正常人脉搏信号波形的差异。

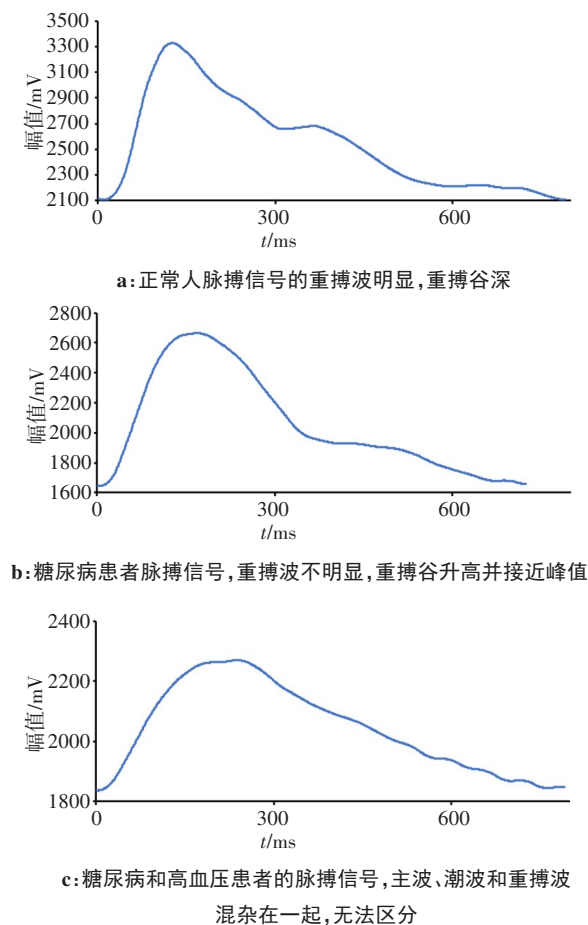


图1 正常人与糖尿病患者脉搏信号波形特征比较
Figure 1 Comparison of waveform features of the pulse signals of normal control and diabetic patients

随着我国2型糖尿病发病率逐年升高^[4],预计其导致的血管并发症也将随之增多^[5]。多项研究发现血糖调节障碍患者的脉搏波形会产生变化,且随着糖尿病病程的增加,脉搏波形特征的变化会更加明显^[6-7]。多项研究表明脉搏波速度(Pulse Wave Velocity, PWV)对2型糖尿病血管并发症具有较好的预测价值^[8-10]。王维敏等^[9]发现2型糖尿病患者的踝臂脉搏波速度(BaPWV)与糖尿病病程呈正相关($P<0.05$),发现病程超过10年的患者BaPWV明显高于首发患者、病程 <5 年患者、5~10岁患者($P<0.05$)。有研究表明2型糖尿病患者的手臂和脚踝的PWV明显高于正常人^[10],2型糖尿病患者BaPWV与糖尿病微血管病变可用于评估糖尿病并发症的危险因素^[11]。Yokoyama等^[12]发现早期糖尿病肾病患者手臂和脚踝脉搏波速度增加,提示动脉硬化的发病机制可能与糖尿病肾病有关。同时相关研究显示2型糖尿病患者合并微量白蛋白尿者和大量白蛋白尿者BaPWV显著升高($P<0.01$),其中大量白蛋白尿者BaPWV又显著高于微量白蛋白尿者($P<0.01$)^[13]。有研究发现一些糖尿病肾病患者的蛋白尿正常,但肾小球滤过率下降^[14],同时研究发现BaPWV为肾小球滤过率值快速下降的独立危险因素;基线BaPWV值越高,肾小球滤过率下降越快^[15]。董晓云等^[16]发现2型糖尿病患者组BaPWV及趾肱指数(TBI)异常,说明患者存在不同程度的动脉弹性下降及末梢动脉血液循环障碍,因此动脉弹性检测可作为早期预测糖尿病血管并发症的方法,利于早期干预治疗。然而,单个脉搏信号特征有时不能有效反映受检者的生理状态,如脉搏信号分析中经常使用的K值。在一些研究中发现,不同的脉搏信号具有相同的K值,而重搏波谷点和重搏波点等特征不同意味着受试者的生理状态是不一样的^[17]。近年来,研究人员利用信号处理技术提取功率谱峰值^[18]、脉搏波传导时间^[19]、近似熵等作为脉搏信号的特征,进行了血管健康状态的识别,取得了良好的识别效果^[20]。同时增加信号的特征信息,也会成倍增加分析的难度和复杂度,如主成分分析法利用降维思想将较多的评价指标转化为较少的综合指标,在解决许多实际问题中取得了较好的效果。传统的信号分析在提取特征的过程中过于主观,无法提取所有有效特征,而且这个过程不能体现智能医学检测的优势,自动化程度低。因此,本文对单周期脉搏波信号进行9层的小波分解,将小波分解后的cD8、cD7和cD6系数直接作为反映糖尿病患者动脉损伤的特征,将特征矩阵输入到10折交叉验证的Stacking集成学习模型中,从而实现对没有临床表现但有动脉损伤的早期糖尿病患者的检测。

1 数据来源

本研究使用来自桂林市人民医院的PPG信号数据集^[21],PPG信号是通过LED光源和探测器测量人体血管和组织反射和吸收的衰减光,反映血管的搏动状态,最终得到脉搏信号。该PPG信号数据库包括高血压、糖尿病、脑血管病患者,年龄21~86岁,中位年龄58岁,男性占48%。采集过程中使用型号为SEP9AF-2(韩国SMPLUS)的PPG传感器,双LED[波长为660 nm(红光)和905 nm(红外光)],采样率为1 kHz,12位ADC,采集信号通过0.5~12 Hz带通滤波器。使用数据集中包含37名正常人和29名2型糖尿病患者的PPG信号,每个人包含3段处于稳定状态下的2.1 s的PPG信号。

2 实验方法

2.1 有效信号提取步骤

(1)确定PPG段内所有的主波的有效波峰,采用双阈值法,高阈值为信号段中的最大值,低阈值为最大值*0.8。(2)检测信号段内所有的峰值点,形成集合S1。为了消除重搏波、潮波和运动噪声引起的峰值点,只提取集合S1中落在高低阈值区间内的点,形成集合S2。(3)因为2.1 s的PPG信号主波的正确峰值数最多为3个,所有处理后的信号段最多只有3个完整的PPG信号周期。如果集合S2中的点数超过3个,则依次比较两个峰值之间的时间间隔,如果T<0.5 s(所有信号的心跳周期在0.6~1.0 s),则取较小的峰值。(4)选取PPG信号的主波峰值点作为基点,利用滑动窗口分别向后采集150个点、向前采集460个点,取610个点(约0.61 s)作为信号样本。该样本可以很好地描述主波、潮波和重搏波的特征。为增加样本数量,同时以主波峰值点左右两侧间隔5、10、15、20 ms的点作为基点,然后以相同的窗口值收集信号样本。最后平均每个个体采集27个样本,然后对信号进行z-score归一化处理,处理后的信号均值为0,标准差为1,归一化公式如式(1)所示:

$$x^* = \frac{x - \bar{x}}{\sigma}$$

(1)

其中, \bar{x} 是原始数据的平均值, σ 是原始数据的标准差,提取的单周期有效信号如图2所示。

2.2 提取小波系数

小波变换是时间频率的局部变换,信号的多尺度细节分析可以通过缩放和平移实现^[22-23]。任何函数都可以在任何尺度上分解为细节部分和近似部分,小波分解相当于信号通过双通道滤波器,滤波器分别具有高通和低通特性,滤波后的输出对应离

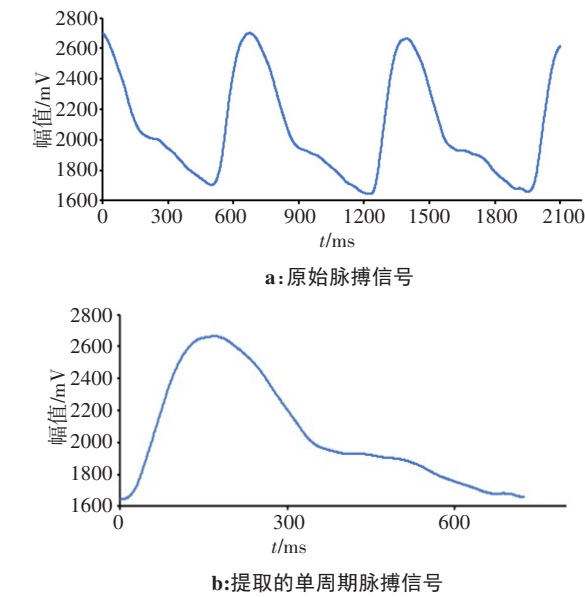


图2 提取的有效信号

Figure 2 Extracted effective signals

散信号的低频和高频细节。对不同尺度进行双通道滤波,就可以实现原始离散信号的多分辨率分解。Daubechies小波是常用的小波函数之一,db后面的数字代表消失矩,一般来说消失矩数越大,小波越平滑,频域的定位能力越强,分频效果越好,但也会削弱时域的紧凑性,大大增加计算时间。使用具有仅支持正交性、双正交性和绝对规律性的db6小波对脉搏波信号进行9层小波分解,得到cA9、cD9、cD8、cD7、cD6、cD5、cD4、cD3、cD2、cD1共10个小波系数,分解过程如图3所示,通过小波分解后的系数对脉搏波信号进行重构,发现cD5、cD4、cD3、cD2、cD1代表脉搏波信号中的高频噪声,而cA9和cD9为低频信号,代表脉搏信号的一般特征,因此使用cD8、cD7、cD6中高频信号系数作为反映糖尿病患者动脉损伤的特征。重建的信号如图4所示。

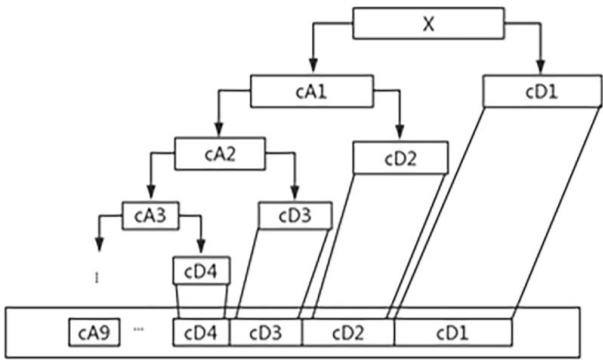


图3 9层小波的分解过程

Figure 3 Process of 9-level wavelet decomposition

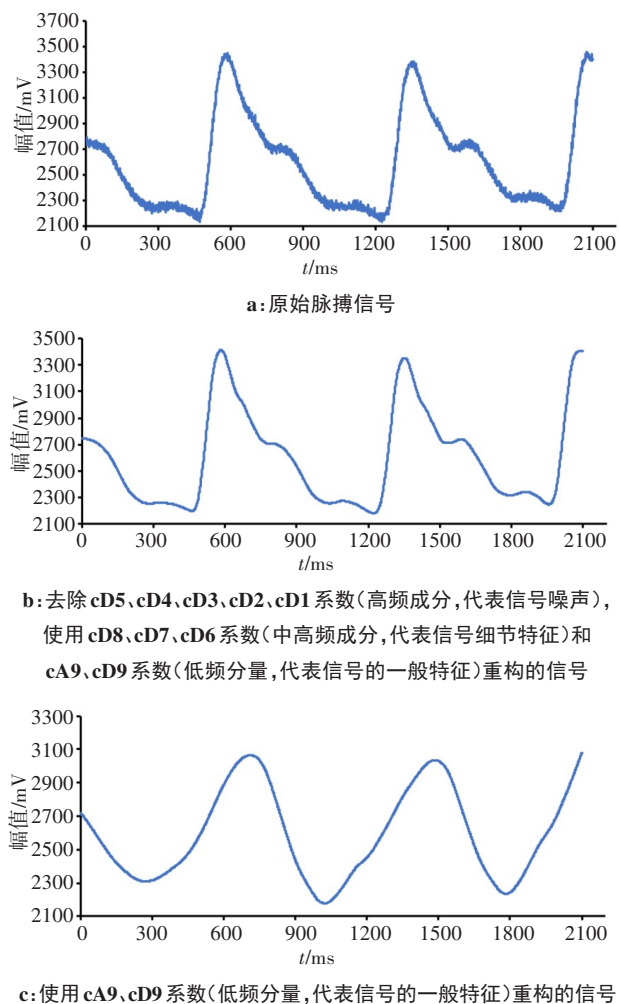


图 4 不同小波系数重构信号

Figure 4 Reconstructed signals with different wavelet coefficients

2.3 SMOTE-Tomek 过采样算法

对于正负样本量的不平衡分布,采用 SMOTE-Tomek 过采样算法,利用 K-最近邻算法原理。具体来说,对少数样本使用 K-最近邻法(预先确定 K 的值)寻找距离最近的 K 个样本,其中距离定义为样本间 N 维特征空间的欧氏距离。然后随机选择 K 个最近邻中的一个,使用公式(2)生成一个新的样本:

$$X_{new} = X_i + (\hat{X} - X_i) \times \delta$$
 (2)

其中, \hat{X} 是选择的 K 个最近邻, $\delta \in [0, 1]$ 是一个随机数。然而 SMOTE 算法在没有考虑其相邻样本的分布特征

的情况下,为每个原始少数样本生成相同数量的合成样本,增加了类别之间重叠的可能性。Tomek Link 表示不同类别之间最接近的一对样本,即两个样本是彼此最近的邻居,但属于不同类别。如果两个样本形成一个 Tomek Link,其中一个是噪声,或者两个样本都在边界附近,这样,通过去除 Tomek Link,可以清理类别之间重叠的样本,使彼此最近邻的样本属于同一类别,从而可以增加类别区分度。

2.4 Stacking 集成学习算法

Stacking 集成学习是一种基于统计学习理论的多模型集成机器学习算法。通常情况下,对于单一的预测模型,其预测精度呈边际效用递减趋势。Stacking 集成学习算法是一种模型集成技术,将来自多个预测模型的信息结合起来生成新模型。不同的机器学习算法以级联的方式组合,以获得比单一算法更好的性能。本研究使用 4 种机器学习算法: Random Forest (RF)、XGBoost、SVM 和 Extra Tree (ET)作为 Stacking 集成学习的第一层基学习器。其中,SVM 在更高维空间中构造一组超平面,创建类与类之间数据点的最大距离,实现的边缘距离越大,泛化误差越低^[24]。RF 通过在训练时构建多个决策树并根据单个树的平均预测输出分类进行操作^[25]。XGBoost 是一种可扩展的端到端的树提升系统,是为了使用较少的资源解决现实较大规模的问题^[26]。与 RF 相比,ET 进一步增加了随机性,不仅特征是从特征子集中随机选择的,阈值也是随机选择的,这通常允许更大程度地减少模型的方差。RF 和 XGBoost 分别使用 Bagging 和 Boosting 的集成学习方法,都具有出色的学习能力。同时,RF 具有可视化内部树结构的优势,使得机器学习在临床使用中具有可解释性机制。本研究中使用 RF 训练数据并获得了可视化的内部树结构,具体结构如图 5 所示。使用鲁棒性和泛化能力强的 KNN 作为第二层元学习器,构建了一个使用 Stacking 集成学习算法的糖尿病检测模型。Stacking 集成学习算法如图 6 所示,在 Stacking 集成学习算法下,整个模型的构建分为两个阶段:第一层基学习器的预测结果通过级联的方式传递到第一层

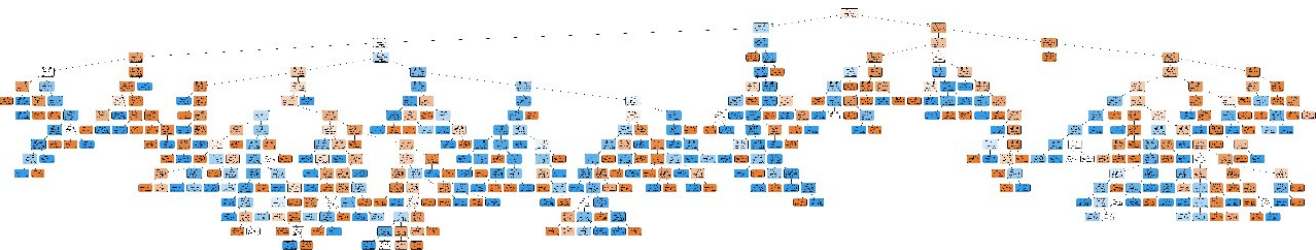


图 5 由本次训练数据集得到的 RF 的可视化内部树结构

Figure 5 Visualized internal tree structure of the Random Forest obtained from this data training

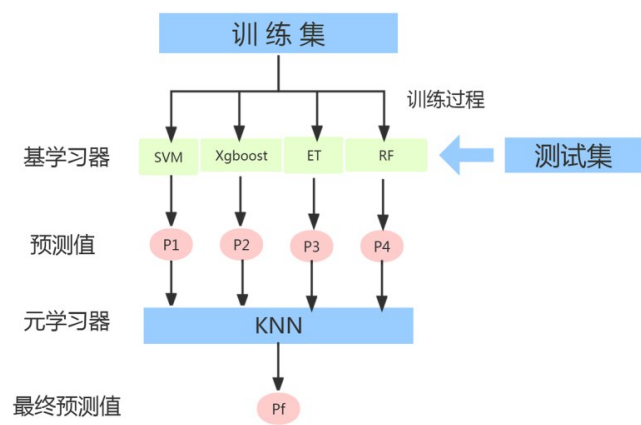


图6 Stacking集成学习算法的具体架构

Figure 6 Specific architecture of Stacking ensemble learning algorithm

元学习器中,再经过元学习器的训练,再次提高检测精度^[27]。

使用的Stacking集成学习算法流程图如图7所示,具体步骤如下:(1)将原始数据集按一定比例划分为训练集和测试集。(2)将训练集分成M份,对每个基本学习器进行M折交叉验证,每次交叉验证时用M-1份作为模型的训练集,剩下的作为验证集。每次交叉验证完成后,使用训练好的基学习器模型对验证集和测试集进行预测。(3)单个基本学习器完成M次交叉验证后,得到每个验证集的预测值,同时得到对测试集的M组预测值,然后将每个验证集的预测值整合为一个矩阵,记为A1,将测试集的M组预测值按行平均得到B1。(4)第一层基本学习器完成训练后,将得到训练集输出特征矩阵(A1,A2,⋯,An)作为第二层元学习器的训练集。(5)第二层元学习器训练完成后,将第一层得到的测试集输出特征矩阵(B1,B2,⋯,Bn)作为测试集,输出最终的预测结果。

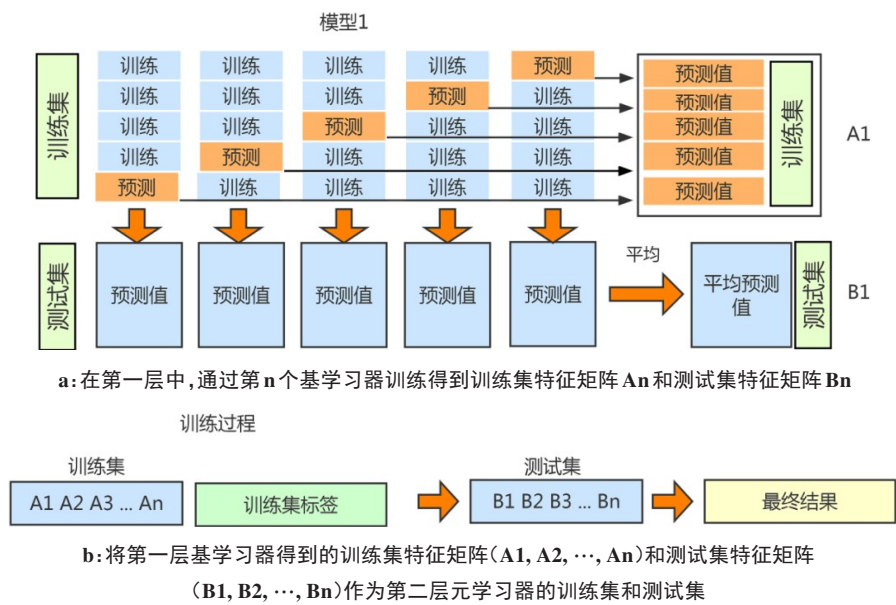


图7 Stacking集成学习算法

Figure 7 Stacking ensemble learning algorithm

3 结果

通过有效信号提取算法,提取了3259个有效信号(定长脉搏信号段),其中正常脉搏信号1767个,糖尿病患者脉搏信号1492个。但是正负样本数量不平衡,使用SMOTE-Tomek过采样算法,使总样本数达到3334个,其中正常脉搏信号样本数1667个,糖尿病患者脉搏信号样本数1667个,经过9级小波分解在脉搏信号中提取cD8、cD7、cD6系数(中高频成分,代表信号细节特征)作为反映动脉损伤程度的特征,系数数量总和为48个。因此,得到总信号特征矩阵为[3334,48],分为训练集特征矩阵[2660,48]和测试集特征矩阵[674,48]。将所有的特征矩阵输入到基于10折交叉验证模型的

Stacking集成学习算法中。通过第一层4个基学习器(SVM、RF、XGBoost、ET)的并行训练,得到特征矩阵[2660,4]作为第二层元学习器的训练集,特征矩阵[674,4]作为第二层元学习器的测试集。元学习器在测试集上的预测结果如表1和图8所示。这里使用的评价指标:准确率(Accuracy)、精确率(Precision)、召回率(Recall)、F1分数(F1-Score)、ROC曲线下面积(Area Under the Curve, AUC)。

从表1可以看出,单一的机器学习算法可以达到90%以上的检测准确率,因此可以证明9级小波变换(中高频)得到的cD8、cD7、cD6系数成分(代表信号细节的特征)可以有效反映糖尿病对动脉的损伤程度。在检测精度方面,Stacking集成学习算法比单一

表1 单一模型与集成模型对糖尿病诊断性能比较
Table 1 Comparison of diagnostic performance of single models and ensemble model for diabetes

方法	指标	分类		准确率/%
		正常	糖尿病	
RF	精确率	0.90	0.92	91.10
	召回率	0.92	0.90	
	F1分数	0.91	0.91	
XGBoost	精确率	0.92	0.91	91.25
	召回率	0.91	0.92	
	F1分数	0.91	0.91	
SVM	精确率	0.90	0.91	90.80
	召回率	0.91	0.90	
	F1分数	0.91	0.91	
ET	精确率	0.91	0.92	91.40
	召回率	0.92	0.91	
	F1分数	0.91	0.91	
Stacking	精确率	0.96	0.95	95.10
	召回率	0.95	0.96	
	F1分数	0.95	0.95	

机器学习算法提高4%~5%,AUC提高1%~6%,证明通过集成多种算法优势,生成的新模型可以获得比单一模型更优越的性能。

4 讨论

本研究中提出了一种检测糖尿病的新方法。脉搏波信号经9级小波分解,小波分解后的系数cD8、cD7、cD6作为反映糖尿病患者动脉损伤的特征。Stacking集成学习算法从大量数据集中学习2型糖尿病患者与正常人在动脉特征上的差异,从而实现对没有临床表现但动脉弹性异常的早期糖尿病患者的检测。本研究结果的高检测精度证明该方法用于糖尿病患者早期检测的可行性。希望脉搏信号特征能够成为临床早期发现糖尿病的便捷、无创的检查方法,从而给予患者早期干预。本次采用的Stacking集成学习算法,有效结合不同学习器的学习能力,获得最佳检测效果。Stacking集成学习算法可以提高建模的准确性,但由于集成模型具有集成多种算法进行建模的特点,整体建模效果会受到单个不良算法的影响,同时,如果第一层有多个相似的算法会导致

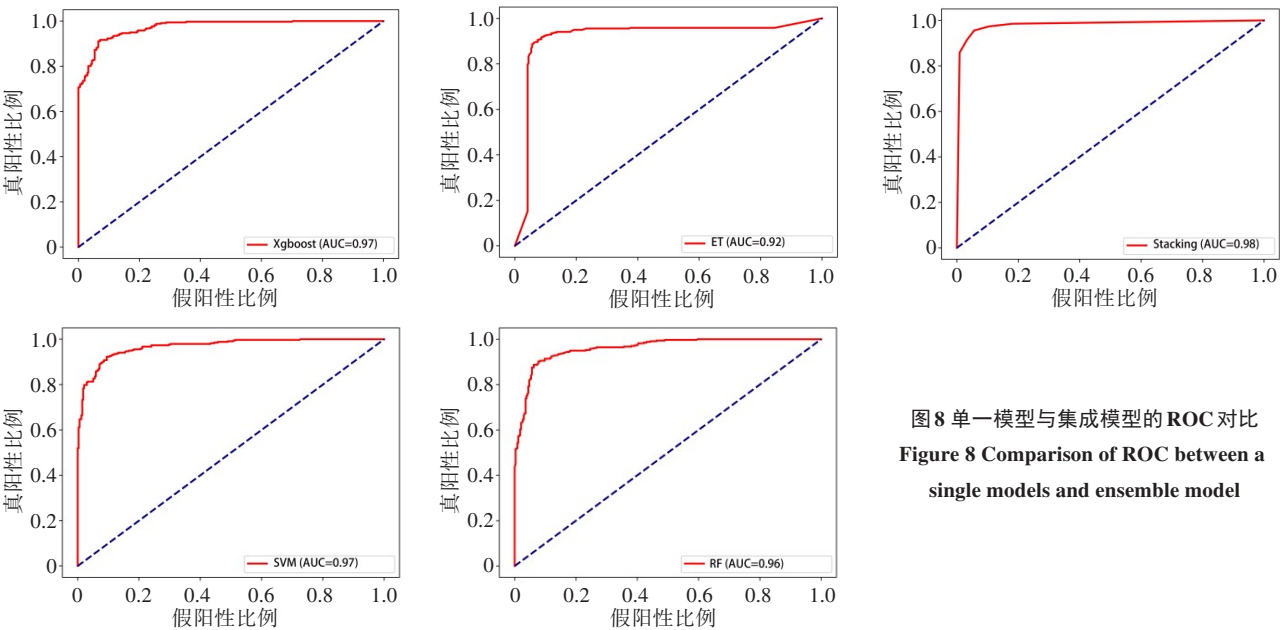


图8 单一模型与集成模型的ROC对比
Figure 8 Comparison of ROC between a single models and ensemble model

过拟合。所以分别使用基于Bagging、Boosting的RF和XGBoost算法,使用鲁棒性和泛化能力强的KNN作为第二层元学习器,最终构建了级联多种机器学习算法的糖尿病检测模型。
这项研究有几个局限性。由于本工作是一项探索性研究,其结果还需要大量临床数据进一步证实。因此,在未来的研究中,需要增加受试者的数量,以获得更多的实验数据,并充分对脉搏信号进行更多

的数据挖掘,提取出更能反映糖尿病患者动脉状态的脉搏波信号特征,进一步建立糖尿病患者动脉状态评价理论。

【参考文献】

[1] Ho CT, Lin CC, Hsu HS, et al. Arterial stiffness is strongly associated with insulin resistance in Chinese: A population-based study (Taichung Community Health Study, TCHS)[J]. J Atheroscler Thromb, 2011, 18 (2): 122-130.

- [2] Levin G, Kestenbaum B, Chen YI, et al. Glucose, insulin, and incident hypertension in the multi-ethnic study of atherosclerosis[J]. *Am J Epidemiol*, 2010, 172(10): 1144-1154.
- [3] Jagannath M. *Biofluid Mechanics*[M]. World Scientific, 2016.
- [4] 中华医学会糖尿病学分会. 中国2型糖尿病防治指南(2017年版)[J]. 中国实用内科杂志, 2018, 38(4): 292-344. Chinese Diabetes Society. Guidelines for the prevention and control of type 2 diabetes in China (2017 Edition)[J]. *Chinese Journal of Practical Internal Medicine*, 2018, 38(4): 292-344.
- [5] Saeedi P, Petersohn I, Salpea P, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition[J]. *Diabetes Res Clin Pract*, 2019, 157: 107843.
- [6] Chen Y, Huang XM, Zhang W, et al. Association of arterial stiffness with HbA1c in 1,000 type 2 diabetic patients with or without hypertension[J]. *Endocrine-Basingstoke*, 2009, 36(2): 262-267.
- [7] Lukich E, Matas Z, Boaz M, et al. Increasing derangement of glucose homeostasis is associated with increased arterial stiffness in patients with diabetes, impaired fasting glucose and normal controls[J]. *Diabetes Metab Res Rev*, 2010, 26(5): 365-370.
- [8] Bouchi R, Babazono T, Mugishima M, et al. Arterial stiffness is associated with incident albuminuria and decreased glomerular filtration rate in type 2 diabetic patients[J]. *Diabetes Care*, 2011, 34(12): 2570-2575.
- [9] 王维敏, 胡云, 李莉蓉, 等. 2型糖尿病患者脉搏波传导速度及其相关危险因素分析[J]. 临床内科杂志, 2011, 28(10): 672-674. Wang WM, Hu Y, Li LR, et al. Analysis of pulse wave velocity and related risk factors in patients with type 2 diabetes[J]. *Journal of Clinical Internal Medicine*, 2011, 28(10): 672-674.
- [10] 韦秀莫, 韦汐, 郑杨杨. 臂踝脉搏波传导速度对早期糖尿病肾病的预测价值[J]. 广西医科大学学报, 2012, 29(1): 91-93. Wei XY, Wei X, Zheng YY. Predictive value of brachial-ankle pulse wave velocity in early diabetic nephropathy[J]. *Journal of Guangxi Medical University*, 2012, 29(1): 91-93.
- [11] Aso K, Miyata M, Kubo T, et al. Brachial-ankle pulse wave velocity is useful for evaluation of complications in type 2 diabetic patients[J]. *Hypertens Res*, 2003, 26(10): 807-813.
- [12] Yokoyama H, Hirasawa K, Aoki T, et al. Brachial-ankle pulse wave velocity measured automatically by oscillometric method is elevated in diabetic patients with incipient nephropathy[J]. *Diabet Med*, 2003, 20(11): 942-945.
- [13] 龚伟, 俞一飞, 何敏, 等. 2型糖尿病合并慢性肾脏病且尿白蛋白正常者脉搏波传导速度变化及其危险因素分析[J]. 中华糖尿病杂志, 2020, 12(10): 6. Gong W, Yu YF, He M, et al. Analysis of pulse wave velocity changes and risk factors in patients with type 2 diabetes mellitus complicated with chronic kidney disease and normal urinary albumin[J]. *Chinese Journal of Diabetes*, 2020, 12(10): 6.
- [14] Chen C, Wang C, Hu C, et al. Normoalbuminuric diabetic kidney disease[J]. *Front Med*, 2017, 11(3): 310-318.
- [15] Sheen YJ, Lin JL, Li TC, et al. Peripheral arterial stiffness is independently associated with a rapid decline in estimated glomerular filtration rate in patients with type 2 diabetes[J]. *Biomed Res Int*, 2013, 2013: 309294.
- [16] 董晓云, 张磊, 李娟娟, 等. 2型糖尿病患者血清踝肱指数、趾肱指数和踝踝脉搏波速度改变及其影响因素的观察[J]. 中国糖尿病杂志, 2013, 21(7): 597-598. Dong XY, Zhang L, Li JJ, et al. The changes of serum ankle-brachial index, toe-brachial index and brachial-ankle pulse wave velocity in patients with type 2 diabetes and their influencing factors[J]. *Chinese Journal of Diabetes*, 2013, 21(7): 597-598.
- [17] 杨琳, 张松, 杨益民, 等. 基于重搏波谷点的脉搏波波形特征量分析[J]. 北京生物医学工程, 2008, 27(3): 229-233. Yang L, Zhang S, Yang YM, et al. Analysis of pulse wave waveform characteristics based on dicrotic trough point[J]. *Beijing Biomedical Engineering*, 2008, 27(3): 229-233.
- [18] 杨凤霞. 基于脉搏信号的亚健康状态识别方法的研究[D]. 兰州: 兰州理工大学, 2006. Yang FX. Research on sub-health state recognition method based on pulse signal[D]. Lanzhou: Lanzhou University of Technology, 2006.
- [19] 王琦, 张爱华. 心电脉搏信息融合在亚健康状态识别中的应用[J]. 生物医学工程学杂志, 2009, 26(6): 1181-1185. Wang Q, Zhang AH. The application of electrocardiographic pulse information fusion in sub-health state recognition[J]. *Journal of Biomedical Engineering*, 2009, 26(6): 1181-1185.
- [20] 任亚莉, 张爱华, 孔令杰. 脉搏信号和主成分分析在亚健康状态识别中的应用[J]. 计算机应用与软件, 2013, 30(3): 200-202. Ren YL, Zhang AH, Kong LJ. Application of pulse signal and principal component analysis in sub-health state recognition[J]. *Computer Applications and Software*, 2013, 30(3): 200-202.
- [21] Liang YB, Chen ZC, Liu GY, et al. A new short-recorded photoplethysmogram dataset for blood pressure monitoring in China[J]. *Sci Data*, 2018, 5: 180020.
- [22] Milchevski A, Gusev M. Improved pipelined wavelet implementation for filtering ECG signals[J]. *Pattern Recognit Lett*, 2017, 95: 85-90.
- [23] Sahoo S, Kanungo B, Behera S, et al. Multiresolution wavelet transform based feature extraction and ECG classification to detect cardiac abnormalities[J]. *Measurement*, 2017, 108: 55-66.
- [24] Bastanlar Y, Ozuysal M. Introduction to machine learning[J]. *Methods Mol Biol*, 2014, 1107: 105-128.
- [25] Ho TK. The random subspace method for constructing decision forests[J]. *IEEE Trans Pattern Anal Mach Intell*, 1998, 20: 832-844.
- [26] Chen T, Guestrin C. XGBoost: A scalable tree boosting system[J]. *arXiv*, 2016, arXiv: 1603. 02754.
- [27] 周星, 丁立新, 万润泽, 等. 分类器集成算法研究[J]. 武汉大学学报(理学版), 2015, 61(6): 503-508. Zhou X, Ding LX, Wan RZ, et al. Research on classifier ensemble algorithm[J]. *Journal of Wuhan University (Science Edition)*, 2015, 61(6): 503-508.

(编辑:黄开颜)