DOI:10.3969/j.issn.1005-202X.2021.02.023

医学人工智能

基于集成学习的骨质疏松性骨折预测研究

陈婉琦，林勇
上海理工大学医疗器械与食品学院，上海 200093

【摘要】骨质疏松性骨折是老年人发病和死亡的重要原因之一，建立高效的预测模型为老年人尽早提供诊断和治疗建议十分必要。实验利用Stacking构建了一种异构分类器EtDtb-S，将16个相关性较高的特征作为特征向量，选用极端随机树（ET）、基于决策树的装袋集成模型（DTB）作为初级学习器，逻辑回归作为次级学习器进行集成。实验验证将EtDtb-S与单模型、同构分类器进行骨质疏松性骨折预测对比，结果表明异构分类器相对于最优单模型预测精度提高2.8%，相对于最优同构分类器预测精度提高1.5%，具有更高的预测性能。

【关键词】骨质疏松性骨折；机器学习；集成学习；分类预测；十折交叉验证
【中图分类号】R318；R683 【文献标志码】A 【文章编号】1005-202X（2021）02-0254-05

Prediction of osteoporotic fracture based on ensemble learning

CHEN Wanqi, LIN Yong
School of Medical Instrument and Food Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

Abstract: Osteoporotic fracture is one of the important causes of morbidity and death in the elderly. It is necessary to establish an efficient predictive model to provide diagnosis and treatment suggestions for the elderly as soon as possible. In the experiment, Stacking is used to construct a heterogeneous classifier EtDtb-S which uses 16 highly-correlated features as feature vectors, and selects extreme random trees and decision tree-based bagging ensemble models as primary learners, and logistic regression as the secondary learner for ensemble learning. Experimental verification compares EtDtb-S with single model and isomorphic classifiers for osteoporotic fracture prediction. The results show that the prediction accuracy of the heterogeneous classifier is increased by 2.8% and 1.5% as compared with the optimal single model and the optimal isomorphic classifier, respectively. The proposed method has better prediction of osteoporotic fracture.

Keywords: osteoporotic fracture; machine learning; ensemble learning; classification prediction; ten-fold cross validation

前 言

骨质疏松症是骨骼的主要疾病，其特征是骨密度降低和骨组织微结构损坏，进而导致骨折敏感性增加[1]。由骨质疏松症引起的骨折叫骨质疏松性骨折，其给患者带来巨大痛苦，并给社会和医疗系统带来沉重负担[2]。骨质疏松症的发病率在女性中最高，但在接下来的50年中，男性的发病率有可能会翻3倍[3]。因此根据临床变量预测男性骨质疏松性骨折风险对其预防至关重要。

近年来机器学习在医学领域的应用越来越广泛，出现了基于机器学习的骨质疏松性骨折预测研究。章轶立等[4]通过Group Lasso回归算法和Logistic回归模型初步构建骨质疏松性骨折风险评估工具。Villamor等[5]结合临床和生物力学数据通过支持向量机（Support Vector Machine, SVM）对髋部骨折进行有效预测。此类单一模型的预测精度仍有较大提升空间，进而有研究提出采用集成学习方法提高模型预测性能。Kruse等[6]使用逻辑回归、随机森林模型以及Bagging和Boosting集成学习方法预测髋部骨折，研究结果表明集成学习方法预测效果更佳。Kilic等[7]使用Bagging、梯度提升（Gradient Boosting）、随机子空间（Random Subspace）采样等集成学习方法对绝经后妇女进行骨质疏松性骨折预测，结果显示基于随机子空间的随机森林（Random Forest based on Random Subspace, RSM-RF）集成分

【收稿日期】2020-08-26
【基金项目】国家自然科学基金(31301092)
【作者简介】陈婉琦，硕士在读，研究方向：机器学习、生物信息处理，E-mail: chen_wanqi0714@163.com
【通信作者】林勇，博士，副教授，研究方向：机器学习、生物信息处理，E-mail: yong_lynn@163.com

类器模型预测精度最佳。目前使用集成学习模型的研究绝大多数是对相同结构的个体学习器进行集成，使用异构分类器的研究还相对较少。

本研究使用学习法的典型代表Stacking构建异构分类器EtDtb-S，经相关性分析后筛选出16个特征作为特征向量，选用极端随机树、基于决策树的Bagging集成模型（Decision Tree Based on Bagging, DTB）作为初级学习器，逻辑回归作为次级学习器进行集成。实验结果表明集成的异构分类器比同构分类器预测准确性更高。

1 材料与方法

1.1 实验材料

本研究采用MrOs Online（https://mrosdata.sfcc-cpmc.net/）上的美国男性骨质疏松性骨折研究数据，数据包含5 994例男性病例样本，病例均为年龄在65岁以上的非卧床男子，其中有12.13%（727名）的患者主要部位（髋部、颈椎、腰椎、胸椎、腕部、肩部）发生过骨折。

选取MrOs数据集中的骨相关数据作为基线数据，包括临床数据、骨密度数据、骨小梁评分数据、腹主动脉钙化数据以及病例骨折情况记录数据。每项基线数据均包含若干特征，如骨密度数据中包含髋部骨密度、股骨骨密度、腰椎骨密度等特征。对这些数据进行特征相关性分析，提取与骨折相关性较高的特征。部分基线数据描述如表1所示。

表1 MrOs数据集简要描述

Tab.1 Brief description of MrOs dataset

基线数据	数据描述
AR1FEB3	腹主动脉钙化数据
B1AUG16	骨密度数据
BS1AUG17	骨小梁评分数据
FAAUG19	骨折情况记录数据
V1FEB14	临床测量结果数据

1.2 特征选择

特征选择是选择相关特征子集以用于模型构建的过程。本研究选用的相关数据文件中均包含众多特征，其中有许多冗余或不相关特征，它们会使得机器学习算法的训练速度降低，增加模型的复杂性，产生模型过拟合现象并会影响预测模型的准确性。因此对数据进行特征选择，考虑到所用学习算法较多，且对模型进行了集成学习，采用过滤式特征选择方法：通过数据的内在属性来估计特征的差异性，根据特征的差异性评分进行排序，并选取评分较高的一部分特征作为特征子集输入到分类算法上。过滤式方法计算简单快速，独立于分类算法，适用于不同的分类算法[8]。笔者选用过滤式中基于皮尔逊（Pearson）相关系数的算法。

Pearson相关系数是衡量向量相似度的一种方法。输出范围为-1~+1，0代表无相关性，负值为负相关，正值为正相关。其公式为：

$$\rho(X,Y)=\frac{\sum_1^n(X_i-\mu_X)(Y_i-\mu_Y)}{\sqrt{\sum_{i=1}^n(X_i-\mu_X)^2}\sqrt{\sum_{i=1}^n(Y_i-\mu_Y)^2}} \quad (1)$$

其中，n为样本个数，X_i为选取的特征数据集，Y_i为标签数据集，μ_X表示随机变量X的均值，μ_Y表示随机变量Y的均值。从临床数据中提取身高、体质量、体重指数（BMI）等数据；骨密度数据中提取髋部骨密度、股骨骨密度、腰椎骨密度等数据；腹主动脉钙化数据中提取腹主动脉钙化评分数据；骨小梁评分数据中提取L_1~L_4腰椎段的骨小梁评分数据；并从骨折情况数据中提取病例主要部位骨折数据标签。经过相关性分析后，剔除与骨折数据标签Pearson相关性低于0.6的特征，对于存在高度相关的特征组（本研究取Pearson相关性高于0.9）每组仅保留一个特征。最终筛选出骨小梁评分、腹主动脉钙化评分、身体质量指数、髋部骨密度、股骨骨密度、颈部骨密度T评分等共16个相关性较高的特征纳入模型中。

1.3 数据类别不平衡校正

本研究数据中只有12.13%的患者主要部位骨折，数据类别失衡较严重，若直接使用不平衡的数据进行实验，则多数类与少数类之间的不平衡将导致机器学习产生偏差，影响模型的性能。目前重采样技术是处理类不平衡问题的常用方法，例如过采样，欠采样和综合采样。其中过采样少数类虽可以平衡本文数据的类分布但无法解决数据集中存在的类重叠问题，并在使用分类器后易产生过拟合现象。本研究将过采样方法Smote与数据清除方法Tomek links相结合可以解决上述问题[9]。Smote+Tomek方法不仅可以平衡数据，还能消除决策边界错误一侧的嘈杂示例，最适用于本研究这种具有少量正样本的数据集。

1.4 基于集成学习的骨质疏松性骨折预测模型

集成学习是一种使用多个基础学习器来提高预测准确性的机器学习技术。对分类器进行集成的思想是将一组分类器使用选定的结合策略通过多种方法（例如投票和平均）对新样本进行分类[10]。目前行

之有效的集成技术是Bagging，Boosting，Stacking和随机子空间（Random subspace）方法[11]。本文选用极端随机树（Extremely Randomized Trees, ET）、DTB作为初级学习器，逻辑回归作为次级学习器，使用Stacking算法对上述不同个体学习器进行集成，构建异构分类器以进一步提高模型预测精度。集成时初级学习器ET、DTB的个数均取1，且ET、DTB中决策树的集成度均为40。

Stacking先从初始数据集训练出几个不同的初级学习器，并通过训练一个次级学习器来结合这些初级学习器[12]。用于训练次级学习器的数据集是一个新数据集。在这个新数据集中，初级学习器的输出被当作样例输入特征，而初始样本的标记仍被当作样例标记[13]。我们需要定义初级学习器以及次级学习器来构建Stacking。本文Stacking框架如下所述。

设基学习算法为L_k，L_k分别为ET、DTB。设基学习器为C_k：

$$C_k = L_k(S),\ k \in [1, 2] \tag{2}$$

其中，S表示本文骨质疏松患者训练数据集，且S中的样本为S_i：

$$S_i = (\boldsymbol{X}_i, y_i) \tag{3}$$

其中，$\boldsymbol{X}_i$为筛选出的16维特征向量，y_i为主要部位是否发生过骨折。

使用交叉验证方式，用训练初级学习器未使用的样本来产生次级学习器的训练样本。本文将数据集S分割为20份。此时设C_k^j为数据集S中去除第j份数据子集后使用第k个基学习算法训练出的基学习器，其表示为：

$$C_k^j = L_k(S - S_j),\ j \in [1, 20] \tag{4}$$

其中，S_j为第j份数据子集。将第j份数据子集中的特征向量作为基学习器的测试集来预测是否会发生骨折，预测结果表示为：

$$C_k^j(X_j) = (\hat{z}_j^k, \hat{o}_j^k) \tag{5}$$

其中，$\hat{z}_j^k$表示第j份患者数据集特征向量预测后不发生骨折的概率，$\hat{o}_j^k$则表示发生骨折的概率，令$\hat{y}_j^k = (\hat{z}_j^k, \hat{o}_j^k)$。

为每一份数据子集预测出患者发生骨折和不发生骨折的概率，得出基学习器的预测结果集为$(\hat{y}_j^1, \hat{y}_j^2, \hat{y}_j^3)$，将$((\hat{y}_j^1, \hat{y}_j^2, \hat{y}_j^3), y_j)$作为次级学习器的数据集，其中$y_j$为该患者是否骨折的初始样本标记。

设$(\hat{y}_j^1, \hat{y}_j^2, \hat{y}_j^3)$为$x_l$，$y_i$为$y_l$，则次级学习器的数据集表示为$S_l = (x_l, y_l)$。

本文采用逻辑回归作为Stacking的次级学习器算法，其模型可以表示为：

$$\hat{y}_i = \left\lfloor h_\theta(x_i) + \frac{1}{2} \right\rfloor \tag{6}$$

其中，$x_i \in x_l$，θ为需学习的参数，$h_\theta(x)$为逻辑回归的假设函数，其公式为：

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}} \tag{7}$$

将式（7）代入式（6）可得：

$$\hat{y}_i = \left\lfloor \frac{1}{1 + e^{-\theta^T x_i}} + \frac{1}{2} \right\rfloor \tag{8}$$

该模型的目标函数可以定义为：

$$\mathrm{obj}(\theta) = \min(J(\theta)) \tag{9}$$

其中，$J(\theta)$为逻辑回归模型的代价函数，本文使用交叉熵作为代价函数，其公式为：

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} \left(y^{(i)} \log\left(h_\theta(x^{(i)})\right) + (1 - y^{(i)}) \log\left(1 - h_\theta(x^{(i)})\right) \right) \tag{10}$$

其中，m为训练样本的个数，y为样本的标签值。将基学习器学习所得的是否发生骨折的结果集x_l作为逻辑回归模型训练样本的特征数据，将初始样本中病例是否骨折的标记y_l作为逻辑回归模型的标签，最终训练得到本文基于Stacking的异构分类器EtDtb-S。

2 实验验证与结果分析

2.1 验证方法

为验证本研究的有效性，本文采用准确率（Accuracy）、精确率（Precision）、召回率（Recall）和F_1值（F_1）对各分类模型进行评估。

$$\mathrm{Accuracy} = \frac{(\mathrm{TP} + \mathrm{TN})}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{TN}} \tag{11}$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{12}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{13}$$

$$F_1 = \frac{2\mathrm{Recall}\ \mathrm{Precision}}{\mathrm{Recall} + \mathrm{Precision}} \tag{14}$$

其中，TP为真阳性，表示实际骨折患者预测结果也为骨折；TN为真阴性，表示实际健康男性预测结果也为健康；FP为假阳性，表示实际健康男性预测结果为骨折患者；FN为假阴性，表示实际骨折患者预测结果为健康。

本文将构建的EtDtb-S模型与单模型以及同构分类器共8种模型进行对比，设标签为骨折的样本为正类样本，不骨折的样本为负类样本。用于实验对比的模型分为2类：（1）单独使用ET模型和DTB模

型；(2)使用不同集成学习方法(Bagging、Boosting、Stacking)分别对ET模型、DTB模型进行同构集成。

对比实验采用十折交叉验证，验证过程中每一折内类别标签比例随机，为减少因样本划分不同而引入的差别，本文重复进行10次十折交叉验证再取均值，得出以上共9种模型的准确率(Accuracy)、精确率(Precision)、召回率(Recall)、F_1值(F_1)以及相应标准差。

2.2 实验结果与分析讨论

本文数据集在不同模型下的分类预测结果及标准差如表2所示。

表2 9种模型预测结果及标准差比较($\bar{x} \pm s$)

Tab.2 Comparison of prediction results obtained by 9 models (*Mean±SD*)

模型	准确率	精确率	召回率	F_1值
ET	0.904±0.012	0.890±0.019	0.927±0.013	0.904±0.012
DTB	0.903±0.088	0.918±0.020	0.901±0.167	0.895±0.090
ET-Bagging	0.880±0.009	0.857±0.026	0.918±0.019	0.883±0.012
DTB-Bagging	0.889±0.056	0.888±0.027	0.885±0.167	0.873±0.109
ET-Boosting	0.902±0.012	0.894±0.017	0.923±0.011	0.905±0.013
DTB-Boosting	0.917±0.060	0.920±0.027	0.924±0.123	0.907±0.090
ET-Stacking	0.908±0.013	0.885±0.022	0.935±0.010	0.911±0.010
DTB-Stacking	0.905±0.063	0.910±0.022	0.906±0.165	0.886±0.119
EtDtb-S	0.932±0.021	0.918±0.024	0.957±0.053	0.929±0.022

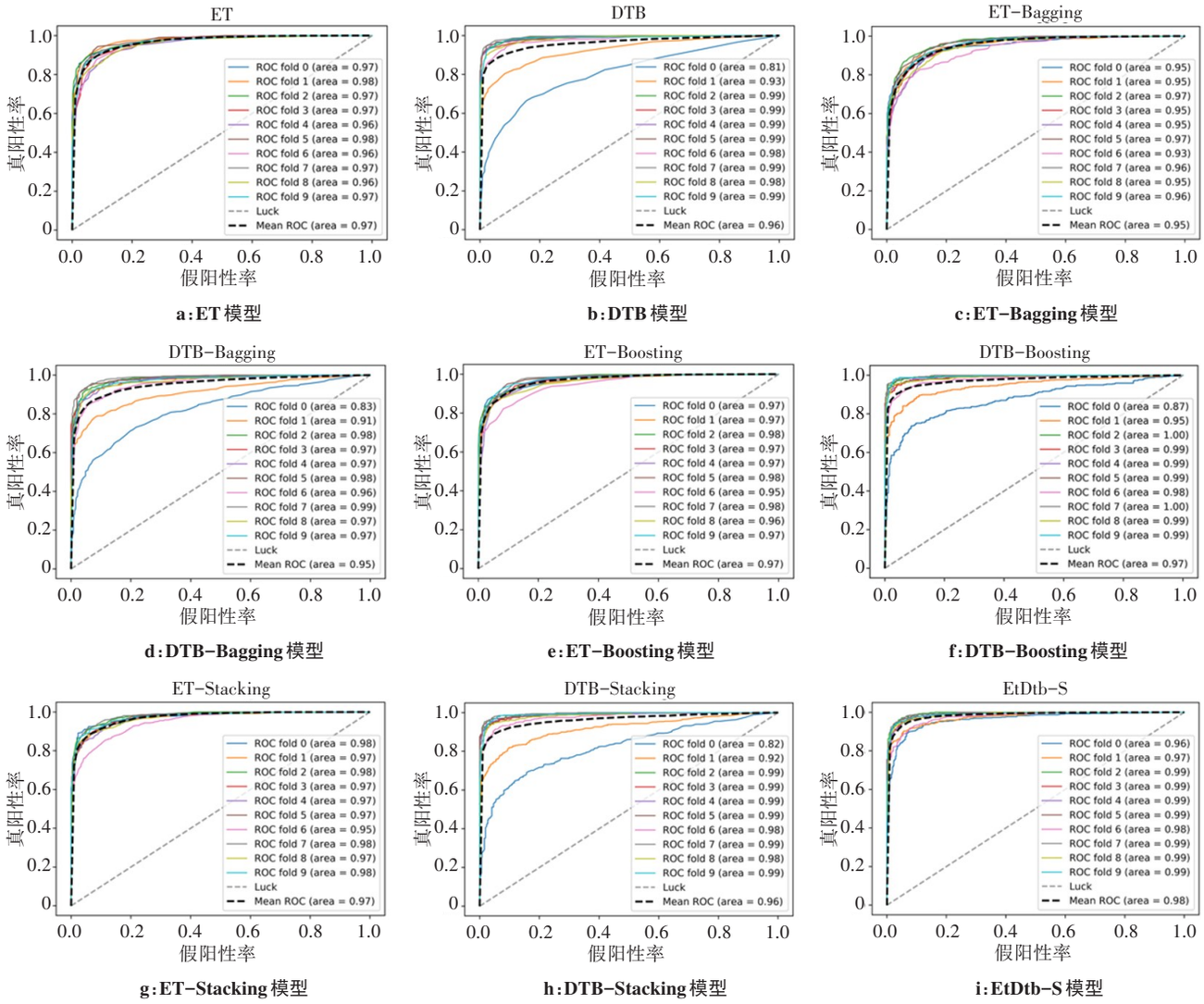
由表2可以发现，本文异构分类器EtDtb-S的分类精度为0.932，相较单独使用ET的分类精度0.904和单独使用DTB的分类精度0.903，分别提高2.8%和2.9%。基于Bagging对ET、DTB分别进行集成的同构分类器ET-Bagging、DTB-Bagging分类精度分别为0.880、0.889；基于Boosting对ET、DTB分别进行集成的同构分类器ET-Boosting、DTB-Boosting分类精度分别为0.902、0.917；基于Stacking对ET、DTB分别进行集成的同构分类器ET-Stacking、DTB-Stacking分类精度分别为0.908、0.905。本文的异构分类器相较上述同构分类器的分类精度提高1.5%~5.2%。由此可得出，本文异构分类器的分类精度优于单模型和同构分类器，分类效果最佳。

为比较上述分类器的性能，绘制出各分类器基于十折交叉验证的ROC曲线，在得出每一折交叉验证的ROC曲线下面积(AUC)值后求出AUC的均值，结果如图1所示。ROC曲线越靠近左上角边界，即AUC越大，表示分类器性能越好。由图1可以看出，在ROC曲线中，ET-Bagging和DTB-Bagging的AUC均值为0.95，DTB和DTB-Stacking的AUC均值为0.96，ET、ET-Boosting、DTB-Boosting和ET-Stacking的AUC均值为0.97，本文异构分类器EtDtb-S的AUC均值为0.98。以上数据说明，本文提出的基于Stacking的异构分类器EtDtb-S相较于单模型和同构分类器分类性能最好。

3 总结与展望

本文介绍了一种用于骨质疏松性骨折预测的新的集成方法，使用Stacking对ET、DTB模型进一步集成构建出异构分类器EtDtb-S。首先，提出了基于机器学习理论的Stacking集成方法的模型构建过程；其次，使用不同集成学习方法对本文集成方法中所采用的初级学习器分别进行集成，将单独使用初级学习器的模型、集成后的同构分类器与本研究的异构分类器分别对选取的特征变量进行训练；最后，通过十折交叉验证得出的准确率、精确率、召回率、ROC曲线比较各模型在测试集上预测的性能，验证本文提出方法的有效性。用Stacking集成时初级学习器ET和DTB在模型结构和分类偏差上的差异性改善了集成后异构分类器的预测精度。实验结果表明，本文基于Stacking的异构分类器能够正确预测骨质疏松性骨折的大部分病例，并且比单模型和集成的同构分类器预测准确性更高，具有最好的分类性能。

本文在运用Stacking进行模型融合的过程中将数据集分割成20份，叶子结点最少样本数为1，内部结点再划分所需最小样本数为2，决策树集成度为40，后续还将调整这些参数，以进一步提高模型性能。本研究还尝试过加入其它分类模型例如神经网络作为集成模型的基学习器，但最终预测准确率并不理想，后续将基于本文现有个体学习器的特征，尝试加入不同神经网络作为个体学习器，进一步提高模型的准确性和通用性。



a：ET模型 **b：DTB模型** **c：ET−Bagging模型**

d：DTB−Bagging模型 **e：ET−Boosting模型** **f：DTB−Boosting模型**

g：ET−Stacking模型 **h：DTB−Stacking模型** **i：EtDtb−S模型**

图1 9种模型ROC曲线对比图

Fig.1 Comparison of receiver operating characteristic curves of 9 models

【参考文献】

[1] FAULKNER K G, CUMMINGS S R, BLACK D, et al. Simple measurement of femoral geometry predicts hip fracture: the study of osteoporotic fractures[J]. J Bone Miner Res, 1993, 8(10): 1211-1217.

[2] JOHNELL O, KANIS J. Epidemiology of osteoporotic fractures[J]. Osteoporos Int, 2005, 16(Suppl 2): S3-S7.

[3] GULLBERG B, JOHNELL O, KANIS J A. World-wide projections for hip fracture[J]. Osteoporos Int, 1997, 7(5): 407-413.

[4] 章轶立, 魏戍, 聂佩芸, 等. 基于Group Lasso的Logistic回归模型构建绝经后骨质疏松性骨折初发风险评估工具[J]. 中国骨质疏松杂志, 2018, 24(8): 994-999.
ZHANG Y L, WEI X, NIE P Y, et al. Establishment of risk assessment tool for postmenopausal osteoporotic fractures based on Group Lasso's logistic regression model[J]. Chinese Journal of Osteoporosis, 2018, 24(8): 994-999.

[5] VILLAMOR E, MONSERRAT C, DEL RIO L, et al. Prediction of osteoporotic hip fracture in postmenopausal women through patient-specific FE analyses and machine learning[J]. Comput Methods Programs Biomed, 2020, 193: 105484.

[6] KRUSE C, EIKEN P, VESTERGAARD P. Machine learning principles can improve hip fracture prediction[J]. Calcif Tissue Int, 2017, 100(4): 348-360.

[7] KILIC N, HOSGORMEZ E. Automatic estimation of osteoporotic fracture cases by using ensemble learning approaches[J]. J Med Syst, 2016, 40(3): 61.

[8] 杜冲, 周长银. 基因表达数据特征子集的冗余研究[J]. 统计与信息论坛, 2019, 34(5): 10-15.
DU C, ZHOU C Y. Redundant study on feature subset of gene expression data[J]. Statistics & Information Forum, 2019, 34(5): 10-15.

[9] BATISTA G E, PRATI R C, MONARD M C. A study of the behavior of several methods for balancing machine learning training data[J]. ACM SIGKDD Explorations Newsletter, 2004, 6(1): 20-29.

[10] DIETTERICH T G. Ensemble methods in machine learning[C]. International Workshop on Multiple Classifier Systems. Springer, 2000: 1-15.

[11] MERT A, KILIÇ N, AKAN A. Evaluation of bagging ensemble method with time-domain feature extraction for diagnosing of arrhythmia beats[J]. Neural Comput Appl, 2012, 24(2): 317-326.

[12] WOLPERT D H. Stacked generalization[J]. Neural Networks, 1992, 5(2): 241-259.

[13] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016: 183-184.
ZHOU Z H. Machine learning[M]. Beijing: Tsinghua University Press, 2016: 183-184.

（编辑：薛泽玲）