结合环论的粒子群优化算法进行冠心病合并慢性心衰预后分析*

张 瑜[1] 田 晶[2] 杨 弘[1] 韩港飞[2] 韩清华[2△] 张岩波[1,3△]

【提 要】 **目的** 采用结合环论的粒子群优化算法(hybridization of ring theory-based evolutionary algorithm and particle swarm optimization,RTPSO)对数据进行均衡化处理,以构建高性能冠心病合并慢性心衰预后模型。**方法** 分别用 SMOTE 算法、RTPSO 算法对数据进行均衡化处理,在均衡化数据集上构建 logistic 回归、随机森林、支持向量机模型。**结果** 本研究共纳入 2229 例冠心病合并慢性心衰患者,依据筛选出的 BMI、射血分数、N 端前脑钠肽等 22 个变量构建模型。用灵敏度、特异度、准确率、F-measure 和 AUC 值评价模型性能,其中 RF、SVM、logistic 回归、RF-RTPSO、SVM-RTPSO、Logistic-RTPSO 灵敏度的中位数分别为 0.0172、0.0773、0.0776、0.7568、0.7640、0.7838;F-measure 的中位数分别为 0.0338、0.1143、0.1283、0.3412、0.3505、0.4545;AUC 的中位数分别为 0.5086、0.5264、0.5313、0.8016、0.7785、0.7985。**结论** RTPSO 算法可以从多数类样本中选择有代表性的少数样本,从而达到数据均衡化,使分类模型具备更高的预测性能,指导临床医生发现高危患者,尽早预防不良事件的发生。

【关键词】 慢性心衰 类不平衡 粒子群优化 随机森林 支持向量机
【中图分类号】 R195.1 【文献标识码】 A **DOI** 10.11783/j.issn.1002-3674.2024.01.011

由于人口增长和老龄化,目前全球心力衰竭患者的总数不断增加[1],影响着近 2% 的成年人口[2]。2017 年调查显示全球有 6430 万人患有心力衰竭[3],其中我国现有心力衰竭患者 890 万[4]。进行心力衰竭风险预测和预后评估,可以有效预防心衰患者不良结局的发生。一般我们收集的临床数据是不平衡的[5],利用此类数据直接建立分类模型的预测结果偏向于多数类,而我们感兴趣的少数类样本往往会被忽略[6],从而降低了预测性能[7]。然而,近年来很多研究提出的方法在从多数类样本中筛选有代表性的少数样本时存在丢失样本重要信息,算法性能依赖于数据结构,使用其他数据表现不佳等问题。因此,本次研究采用灵活、自适应的 RTPSO 算法对心衰数据集进行均衡化处理,在此基础上建立 logistic 回归、随机森林、支持向量机模型,以期建立精准预测模型,准确预测患者结局,降低不良事件发生率。

资料与方法

1.研究资料

(1)研究对象

本次研究选择山西省太原市 2017 年 1 月到 2019 年 8 月期间两所三甲医院诊断为冠心病合并慢性心衰的患者 2229 名,并在出院之后 1 月、3 月、6 月、12 月之后每隔半年随访其生命状态。纳入标准:年龄≥18 岁;有典型的慢性心力衰竭(chronic heart failure, CHF)症状(劳力性或阵发性呼吸困难、乏力、食欲不振)或体征(双下肢水肿、肺部湿啰音、肝颈静脉回流征阳性);纽约心脏病协会(New York Heart Association,NYHA)心功能分级Ⅱ~Ⅳ级;诊断为冠心病的患者。排除标准为:近 2 月有急性心血管事件发生的患者;精神疾病的患者;有其他危及生命的疾病,预期生存时间<1 年的患者;拒绝参加本项目的患者。

(2)资料收集

由课题组成员收集患者一般人口学资料、临床表现、实验室检查、影像学检查、用药情况等患者住院的电子病例信息,追踪随访其结局。使用 EpiData 3.1 软件进行数据双录入并做一致性检验。

(3)数据处理

使用 R(Version 4.1.1)包 ggrandomforests 中的随机森林变量重要性(variable importance,VIMP)和最小深度来筛选变量。其中 VIMP 越大表示该变量越重要,最小深度越小表示变量越重要。在 Python 3.8.0 中使用 SMOTE 算法、RTPSO 算法实现数据均衡化处理。

2.研究方法

(1)随机森林

随机森林是 Breiman 和 Adele Cutler 在 2001 年提出的一种 Bagging 集成算法[8]。本次研究使用 Python 3.8.0 中 sklearn 库的“*RandomForestClassifier*”函数建立随机森林模型,参数设置为默认值。

(2)支持向量机

支持向量机是由 Vapnik 等人于 1995 年提出的以结构风险最小化为理论基础的一种机器学习方法[9]。本次研究使用 Python 3.8.0 中 sklearn 库的“*SVC*”函数建立支持向量机模型,以线性核为核函数。

* 基金项目:国家自然科学基金(81872714;82173631)
1. 山西医科大学公共卫生学院流行病与卫生统计学教研室(030001)
2. 山西医科大学附属第一医院心内科
3. 重大疾病风险评估山西省重点实验室
△通信作者:张岩波,E-mail:sxmuzyb@126.com;韩清华,E-mail:syhqh@sohu.com

(3)结合环理论的粒子群优化算法(hybridization of ring theory-based evolutionary algorithm and particle swarm optimization, RTPSO)

粒子群优化算法(particle swarm optimization, PSO)是由 Kennedy 和 Eberhart 提出的一种群体智能优化算法[10]。该算法通过初始化一群随机粒子,迭代找到最优解,在每次迭代中更新个体极值 Pbest 和群体极值 Gbest,同时每个粒子根据公式(1)和(2)更新其速度和位置,以向全局最优位置移动。

$$v_i^{k+1}=wv_i^k+c_1r_1(P_i^k-X_i^k)+c_2r_2(P_g^k-X_i^k) \quad (1)$$

$$X_i^{k+1}=X_i^k+v_i^{k+1} \quad (2)$$

其中 c_1、c_2 为加速度因子,r_1、r_2 是 0~1 之间的随机数。

基于环论的进化算法(ring theory(RT)-based evolutionary algorithm, RTEA)[11]是最近提出的利用代数理论解决组合优化问题的方法,该算法使用全局探索算子(R-GEO)和局部开发算子(R-LDO)生成新的个体,按照贪婪策略选择个体。

粒子群优化算法有广泛的局部开发能力,但是缺乏适当的全局搜索能力,而 RTEA 具有很强的全局搜索能力。两者结合可以很好的解决各自算法的缺点,达到很好的运算结果。

其中,为了评估该算法在每次迭代中的性能,本算法选用 ROC-AUC 作为适应度函数,用于计算 RTPSO 的适应度值。

(4)模型构建

本研究按患者结局是否死亡不设种子数随机抽取训练集∶验证集∶测试集=6∶2∶2 的数据集,重复 50 次。其中训练集用于训练分类器以进行模型拟合;验证集用于评估所选样本的代表性,即计算 RTPSO 的适应度值来评估其性能;测试集用于评估分类模型性能,测试模型的泛化能力。将筛选出的变量作为输入变量,是否死亡作为结局变量,采用均衡后的数据建立 logistic 回归、随机森林、支持向量机模型,并与未采用 RTPSO 算法处理的数据建立的模型进行比较。

结　果

1.一般情况

2229 例冠心病合并心衰患者中,死亡患者为 233(10.45%)例,存活患者为 1996(89.55%)例,两者比例接近 1∶9,是一个不平衡数据集。男性有 1456(65.32%)例,女性有 773(34.68%)例。平均年龄为(69.42±11.04)岁。研究对象的一般情况见表 1。

表 1　患者基本情况

基本情况	死亡	存活	t/H	P
血红蛋白(g/L)	127.87±21.21	135.81±19.40	5.855	<0.001
红细胞宽度(%)	14.50(13.89,15.50)	13.86(13.30,14.70)	−7.841	<0.001
血小板(g/L)	177.00(143.50,216.50)	182.00(149.00,223.00)	−1.393	0.164
白蛋白(g/L)	40.78±5.00	43.97±5.78	8.071	<0.001
谷丙转氨酶(U/L)	21.08±15.33	24.53±16.55	3.029	0.002
γ-谷氨酰转移酶(%)	29.00(19.00,53.00)	27.00(19.00,44.39)	−1.255	0.210
血清总胆红素(μmol/L)	14.70(10.50,19.90)	14.50(11.03,20.03)	−0.126	0.900
血清直接胆红素(μmol/L)	5.00(3.10,6.70)	3.40(2.30,5.11)	−7.656	<0.001
血清间接胆红素(μmol/L)	10.50(7.10,14.55)	11.30(8.30,15.10)	−2.444	0.015
碱性磷酸酶(U/L)	77.00(59.00,97.50)	77.00(64.00,94.00)	−0.441	0.659
尿素氮(mmol/L)	7.20(5.50,9.85)	6.10(4.97,7.80)	−5.839	<0.001
肌酐(mmol/L)	91.00(72.85,115.40)	78.80(66.00,93.78)	−7.078	<0.001
胱抑素 C(mg/L)	1.38±0.55	1.23±0.48	−4.621	<0.001
N 端前脑钠肽(mg/L)	3279.00(1659.18,6504.22)	1279.50(657.51,3179.25)	−10.110	<0.001
血糖(mmol/L)	6.13±2.79	5.76±2.13	−2.386	0.017
射血分数(%)	44.40±13.41	50.51±13.52	6.539	<0.001
收缩压(mmHg)	126.00(110.00,142.00)	130.00(118.00,140.00)	−1.737	0.082
BMI(kg/m^2)	23.42(20.85,25.45)	24.77(22.54,27.06)	−5.863	<0.001
血浆总胆固醇	3.85±1.15	4.11±1.16	3.204	0.001
低密度脂蛋白胆固醇(μmol/L)	2.16(1.69,2.80)	2.36(1.91,2.96)	−3.893	<0.001
高密度脂蛋白胆固醇(μmol/L)	0.96(0.82,1.12)	0.96(0.82,1.14)	−0.603	0.547
血清总胆汁酸(μmol/L)	4.00(2.00,7.00)	4.00(2.00,7.00)	−0.301	0.763

*:正态分布数据采用均数±标准差描述,非正态分布数据采用中位数和四分位数进行描述

2.变量筛选

在 RStudio 中使用 ggRandomForests 包筛选自变量,根据变量重要性(VIMP)对变量进行排序。其中,剔除掉 VIMP 为零和负值的变量,只选择 VIMP 为正值的变量;使用最小深度分布的平均值作为阈值,将最小深度低于阈值的变量定义为对结果有重要影响的变量。通过取排名最高的最小深度变量(低于选择阈值),并匹配 VIMP 排名来综合考虑两种方法筛选出的变量。最终筛选出 22 个变量纳入模型(如图 1 所示),变量赋值如表 2 所示。

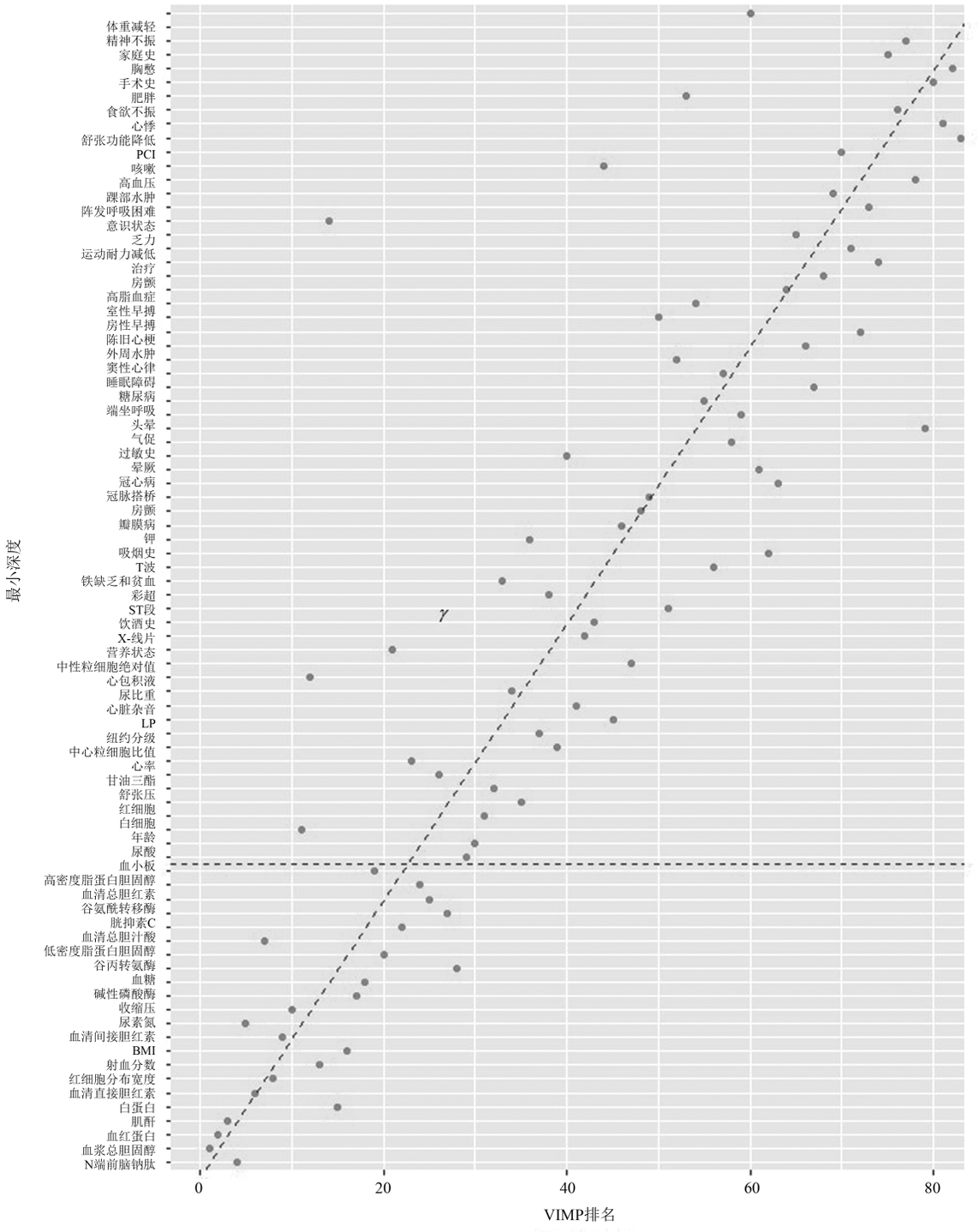

图 1 随机森林最终变量筛选图

表 2 变量赋值表

变量	变量类型	赋值/单位	变量	变量类型	赋值/单位
死亡	分类	0=无 1=有	肌酐	数值	mmol/L
血红蛋白	数值	g/L	胱抑素 C	数值	mg/L
红细胞宽度	数值	%	N 端前脑钠肽	数值	mg/L
血小板	数值	10^9/L	血糖	数值	mmol/L
白蛋白	数值	g/L	射血分数	数值	%
谷丙转氨酶	数值	U/L	收缩压	数值	mmHg
γ-谷氨酰转移酶	数值	%	BMI	数值	kg/m^2
血清总胆红素	数值	mol/L	血浆总胆固醇	数值	mmol/L
血清直接胆红素	数值	mol/L	低密度脂蛋白胆固醇	数值	mol/L
血清间接胆红素	数值	mol/L	高密度脂蛋白胆固醇	数值	mol/L
碱性磷酸酶	数值	U/L	血清总胆汁酸	数值	mol/L
尿素氮	数值	mmol/L			

3.SMOTE 算法处理结果

在 Python 3.8.0 中对训练数据集进行正负样本均衡化处理，将生成样本与初始少数类样本相组合并进行模型训练，在测试集上对模型性能进行测试，各模型均重复 50 次。结果如表 3 所示。

表 3　SMOTE 预处理下不同分类模型评价指标比较

模型	灵敏度	特异度	准确率	F-measure	AUC
RF	0.0172(0,0.0345)	1(0.9990,1)	0.8978(0.8951,0.8987)	0.0338(0,0.0653)	0.5086(0.4995,0.5167)
SVM	0.0773(0.0430,0.0862)	0.9845(0.9820,0.9965)	0.8879(0.8839,0.8919)	0.1143(0.0533,0.1399)	0.5264(0.5103,0.5346)
Logistic	0.0776(0.0689,0.1163)	0.9880(0.9820,0.9887)	0.8906(0.8875,0.8965)	0.1283(0.1134,0.1889)	0.5313(0.5259,0.5514)
RF-SMOTE	0.3620(0.2586,0.5258)	0.8760(0.8720,0.8790)	0.8172(0.8118,0.8414)	0.2916(0.2222,0.4081)	0.6160(0.5672,0.7019)
SVM-SMOTE	0.7499(0.7197,0.7887)	0.6325(0.6255,0.6425)	0.6505(0.6379,0.6631)	0.3082(0.2999,0.3199)	0.6940(0.6824,0.7111)
Logistic-SMOTE	0.6379(0.5819,0.7241)	0.7430(0.7360,0.7473)	0.7339(0.7284,0.7419)	0.3341(0.3071,0.3636)	0.6925(0.6614,0.7310)

由表 3 可知，采用原始数据构建模型特异度很高，而其他评价指标较低，表示预测结果偏向于多数类，不能准确的预测少数类。经过 SMOTE 均衡化处理后模型的特异度降低，但是灵敏度、F-measure、AUC 均有显著提高，说明 SMOTE 均衡化处理数据后模型预测结果不再偏向于少数类，两类预测结果趋于平衡。

4.RTPSO 算法结果

采用 RTPSO 算法从多数类样本中选出最具代表性的样本，并与原始少数样本合并来训练模型，将其结果与 SMOTE 算法比较。各结果均重复 50 次，模型在测试集上结果如表 4 所示。

表 4　最终模型性能比较结果

模型	灵敏度	特异度	准确率	F-measure	AUC
RF	0.0172(0,0.0345)	1(0.9990,1)	0.8978(0.8951,0.8987)	0.0338(0,0.0653)	0.5086(0.4995,0.5167)
SVM	0.0773(0.0430,0.0862)	0.9845(0.9820,0.9965)	0.8879(0.8839,0.8919)	0.1143(0.0533,0.1399)	0.5264(0.5103,0.5346)
Logistic	0.0776(0.0689,0.1163)	0.9880(0.9820,0.9887)	0.8906(0.8875,0.8965)	0.1283(0.1134,0.1889)	0.5313(0.5259,0.5514)
RF-SMOTE	0.3620(0.2586,0.5258)	0.8760(0.8720,0.8790)	0.8172(0.8118,0.8414)	0.2916(0.2222,0.4081)	0.6160(0.5672,0.7019)
SVM-SMOTE	0.7499(0.7197,0.7887)	0.6325(0.6255,0.6425)	0.6505(0.6379,0.6631)	0.3082(0.2999,0.3199)	0.6940(0.6824,0.7111)
Logistic-SMOTE	0.6379(0.5819,0.7241)	0.7430(0.7360,0.7473)	0.7339(0.7284,0.7419)	0.3341(0.3071,0.3636)	0.6925(0.6614,0.7310)
RF-RTPSO	0.7568(0.7297,0.7568)	0.7400(0.7300,0.7500)	0.8010(0.7832,0.8065)	0.3624(0.3571,0.3714)	0.8016(0.7852,0.8098)
SVM-RTPSO	0.7640(0.7526,0.7787)	0.6900(0.6725,0.7025)	0.7809(0.7712,0.8040)	0.3505(0.3484,0.3541)	0.7785(0.7612,0.8089)
Logistic-RTPSO	0.7838(0.7297,0.7838)	0.7250(0.7000,0.7525)	0.7947(0.7908,0.8095)	0.4545(0.4454,0.4706)	0.7985(0.7876,0.8084)

从表 4 可以看出，传统模型特异度和准确率均达到 85% 以上，但是灵敏度均在 10% 以下。经过 SMOTE 预处理后各模型灵敏度均有所提高，SVM 灵敏度由原来的 7.73%提高到 74.99%，灵敏度和特异度测量值之间的差距缩小，证明均衡化数据集后模型预测心衰患者死亡的能力提升显著，但是各模型的 AUC 值仍然在 0.7 以下，效果不是十分理想。但是经过 RTPSO 处理后，各模型 AUC 均有很大程度提高，RF 达到 0.8016，SVM 达到 0.7785，logistic 达到 0.7985，logistic 模型的 F-measure 甚至达到 0.4545。综上所述，RTPSO 算法比 SMOTE 算法在处理类不平衡问题上有更好的性能，其中经 RTPSO 预处理的 logistic 模型性能最优。

讨　论

心力衰竭是当今社会一项沉重的医疗负担，是 65 岁以上患者住院的主要原因，每年护理支出超过 310 亿美元[12]。研究显示心衰患者发展到晚期，5 年绝对死亡率约为 50%[13]。因此，早期诊断和预后识别在临床领域越来越受到重视。

虽然已有很多模型对心衰患者死亡进行了预测，但是这些算法都是在假定数据平衡的情况下进行的，当数据不平衡时，这些算法的预测精度可能会受到不同程度的影响[7]。近年来，许多研究者开发出了不同的方法来解决以上问题，随机欠采样（random under sampler，RUS）通过从多数类样本中随机的移除一定数量的少数样本以达到数据的平衡[14]，但是该方法可能会丢失样本重要信息；Gong 等人[15]提出了一种混合采样方法 RHSBoost 来解决随机欠采样丢失重要信息问题；Cai 等人[16]提出的 Bernoulli-based RUS 方法通过考虑每个样本可能包含的信息，为每个样本分配不同的剔除概率。但是上述算法对样本的分布具有一定的选择性，因此具有一定自适应性、灵活的优化算法被提出，如粒子群优化算法[17-19]，蚁群优化算法[20]等。

本研究使用的 RTPSO 算法结合了 PSO 和 RTEA 算法的优势，可以自适应的从多数类样本中选择出有代表性的样本，而不损失重要信息。因此本研究通过回顾性的收集冠心病合并心衰患者的电子病历信息，考虑数据不平衡，采用 RTPSO 算法从多数类样本中筛选出有代表性的少数样本，再进行分类模型的构建，使分类结果不再倾向于多数类样本，而是使心衰死亡患者预测与存活患者达到平衡，提高了对死亡患者的识别率，从而保证了分类器能得到稳定的分类性能，能够指导临床医生早期发现高危患者，尽早预防不良事件的发生。

Wang Ke[21]等人的研究也显示,经 SMOTE 预处理之后 SVM、RF 模型性能较未进行平衡处理的模型性能均有所提高,这在此次研究中也得到了明确体现。这也侧面反映了数据平衡会对各分类器的预测精度产生重要影响。因此,对于不平衡数据集的分类问题,在建模前对数据进行预处理是相当必要的。SMOTE 是一种经典的处理类不平衡问题的算法,但是其会引入不必要的噪声而影响分类准确性[22]。而此研究使用的 RTPSO 算法对不同数据均有自适应能力,从本研究结果也可看出,经 RTPSO 算法预处理后的数据在灵敏度、AUC 等评价指标上均优于 SMOTE 算法,对不平衡问题更加有效。

本文发现尿素氮、肌酐、N 端前脑钠肽、射血分数、收缩压、BMI、血红蛋白、白蛋白、血糖是患者死亡的重要影响因素,这与其他研究结果一致[23-25]。血小板也是心衰死亡的一个重要影响因素,有研究[25]表明这可能是因为该指标与肾功能损害或心脏损伤有关,当与表示肾功能的关键指标(肌酐等)相结合时,发挥重要的作用。国外一项研究也发现低水平的低密度脂蛋白胆固醇、总胆固醇是死亡的危险因素[27],但这一观点现在仍存在争议。本研究还发现其他研究未发现的危险因素,如血清直接胆红素、血清间接胆红素等表示肝脏功能的指标,这可能与心衰常累及肝脏功能有关,在今后的研究中需引起重点关注。

另外,本研究还存在一些不足之处,①样本量只来源于山西省内三甲医院,比较单一,存在一定的选择偏倚,之后可以进一步扩大数据收集范围,并可以采用其他数据评估改算法的普适性;②只采用简单的分类器对结局进行预测,之后可以选择更加高级、先进的分类器与此算法结合,以获得更高的预测精度。

综上所述,RTPSO 算法可以有效的解决类不平衡问题,并可与各分类模型相结合,提高对死亡患者的预测,指导临床工作者尽早采取干预措施,延缓疾病进展。

参 考 文 献

[1] Groenewegen A, Rutten FH, Mosterd A, et al. Epidemiology of heart failure. Eur J Heart Fail,2020,22(8):1342-1356.

[2] Smeets M, Vaes B, Mamouris P, et al. Burden of heart failure in Flemish general practices: a registry-based study in the Intego database.BMJ Open,2019,9:e022972.

[3] Ouyang G, Pan G, Liu Q, et al. The global, regional, and national burden of pancreatitis in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. BMC Med,2020,18(1):388.

[4]《中国心血管健康与疾病报告 2020》概述.中国心血管病研究,2021,19(7):582-590.

[5] Hassan AKI, Abraham A. Modeling insurance fraud detection using imbalanced data classification. Advances in Nature and Biologically Inspired Computing, 2016,419:117-127.

[6] Galar M, Fernandez A, Barrenechea E, et al. A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches. IEEE Trans Syst Man Cybern Part C (Applications and Reviews),2012, 42(4):463-484.

[7] Chawla NV. Data Mining for Imbalanced Datasets: An Overview. Data Mining and Knowledge Discovery Handbook,2009:875-886.

[8] Breiman L.Random Forests. Machine Learning, 2001, 45(1): 5-32.

[9] Cortes C, Vapnik V. Support-Vector Networks. Machine Learning, 1995,20(3):273-297.

[10] Kennedy J, Eberhart R. Particle swarm optimization. Proceedings of ICNN'95-International Conference on Neural Networks, 1995, 4(8):1942-1948.

[11] He Y, Wang X, Gao S.Ring theory-based evolutionary algorithm and its application to d|0-1| KP. Applied Soft Computing Journal, 2019,77:714-722.

[12] Virani SS, Alonso A, Benjamin EJ, et al. Heart disease and stroke statistics-2020update: a report from the American Heart Association. Circulation,2020,141(9):e139-e596.

[13] Yancy CW, Jessup M, Bozkurt B, et al. 2013 ACCF/AHA guideline for the management of heart failure: A report of the american college of cardiology foundation/american heart association task force on practice guidelines. Circulation, 2013, 128(16): e31829-e8776.

[14] Batista GE, Prati RC, Monard MC. A study of the behavior of several methods for balancing machine learning training data. Acm Sigkdd Explorations Newsletter,2004, 6(1):20-29.

[15] Gong J, Kim H.RHSBoost: Improving classifification performance in imbalance data. Computational Statistics and Data Analysis, 2017, 111:1-13.

[16] Cai R, Zhao Q, She D, et al. Bernoulli-based random undersampling schemes for 2D seismic data regularization. Applied Geophysics, 2014,11(3):321-330.

[17] Gao M, Hong X, Chen S, et al. A combined SMOTE and PSO based RBF classifier for two-class imbalanced problems. Neurocomputing, 2011,74(17):3456-3466.

[18] Samma H, Lim C, Ngah UK. A hybrid PSO-FSVM model and its application to imbalanced classification of mammograms. Lecture Notes in Computer Science, 2013,7802(1):275-284.

[19] Aydogan EK, Ozmen M, Delice Y. CBR-PSO: cost-based rough particle swarm optimiza-tion approach for high-dimensional imbalanced problems. Neural Comput Appl,2018,31(10):6345-6363.

[20] Li M, Xiong A, Wang L, et al. ACO Resampling: Enhancing the performance of oversampling methods for class imbalance classification. Knowledge-BasedSystems,2020,196:105818.

[21] Wang K, Tian J, Zheng C, et al. Improving Risk Identification of Adverse Outcomes in Chronic Heart Failure Using SMOTE+ENN and Machine Learning. Risk Manag Healthc Policy,2021,14:2453-2463.

[22] Mi Y. Imbalanced classification based on active learning SMOTE. Res j Applied Sci, Engineering Technol,2013,5(3):944-949.

[23] Angraal S, Mortazavi BJ, Gupta A, et al. Machine Learning Prediction of Mortality andHospitalization in Heart Failure With Preserved Ejection Fraction. JACC Heart Fail,2020,8(1):12-21.

[24] Voors AA, Ouwerkerk W, Zannad F, et al. Development and validation of multivariablemodels to predict mortality and hospitalization in patients with heart failure.Eur J HeartFail,2017,19(5):627-634.

[25] Kim W, Park JJ, Lee HY, et al. Predicting survival in heart failure: a risk score basedon machine-learning and change point algorithm. Clin Res Cardiol,2021,110(8):1321-1333.

[26] Asif N, Nadim A, Farhan S H. Survival prediction of heart failure patients using machine learning techniques. Informatics in Medicine Unlocked,2021,26:100772.

[27] Johannesen CDL, Langsted A, Mortensen MB, et al. Association between low density lipoprotein and all cause and cause specific mortality in Denmark: prospective cohort study.BMJ,2020,8(371): m4266.

(责任编辑:邓 妍)