

## 数据挖掘技术在心血管疾病预后研究中的应用进展

向超益<sup>1,2</sup> 吴亚飞<sup>1,2,3</sup> 方亚<sup>1,2,3</sup>

<sup>1</sup>厦门大学公共卫生学院 361102; <sup>2</sup>卫生技术评估福建省高校重点实验室, 厦门 361102;

<sup>3</sup>厦门大学健康医疗大数据国家研究院 361102

通信作者: 方亚, Email: fangya@xmu.edu.cn

**【摘要】** 数据挖掘技术在心血管疾病预后研究中得到较为广泛的应用。其中, 在脑卒中预后研究中更多涉及干预有效性的预测, 而在其他心血管疾病预后研究中, 主要关注自然预后和干预安全性的预测。相较于传统的统计方法, 机器学习方法尤其是以神经网络为基础的深度学习技术在预测心血管疾病预后方面有更好的性能表现, 值得进一步推广。本文梳理了近些年数据挖掘在心血管疾病预后研究中的应用进展, 并对当前研究不足进行总结、提出展望。

**【关键词】** 数据挖掘; 心血管疾病; 预后; 预测

**基金项目:** 国家自然科学基金(81973144)

### Progress on the application of data mining in the prognosis of cardiovascular disease

Xiang Chaoyi<sup>1,2</sup>, Wu Yafei<sup>1,2,3</sup>, Fang Ya<sup>1,2,3</sup>

<sup>1</sup>School of Public Health, Xiamen University, Xiamen 361102, China; <sup>2</sup>Key Laboratory of Health Technology Assessment of Fujian Province, Xiamen 361102, China; <sup>3</sup>National Institute for Data Science in Health and Medicine, Xiamen University, Xiamen 361102, China

Corresponding author: Fang Ya, Email: fangya@xmu.edu.cn

**【Abstract】** Data mining has been widely used in the study of cardiovascular disease prognosis. For stroke prognosis, the focus was mainly on the prediction of intervention effectiveness. In contrast, the focus was primarily on predicting natural prognostic and intervention safety for other cardiovascular diseases. In addition, compared with traditional statistical methods, machine learning, especially deep learning based on neural networks has much better performance in predicting the prognosis of cardiovascular diseases, which is worthy of further promotion and application. Therefore, this study systematically reviewed the recent application progress of data mining in cardiovascular disease prognosis, summarized the shortcomings of current studies, and put forward future directions.

**【Key words】** Data mining; Cardiovascular disease; Prognosis; Prediction

**Fund program:** National Natural Science Foundation of China (81973144)

心血管疾病是循环系统疾病的统称, 是由环境、生活方式、遗传等因素导致的心脏和血管系统的有关病变, 如脑卒中、高血压、冠心病、心力衰竭等。在全球范围内, 心血管疾病已跃居非传染性疾病死因顺位的首位<sup>[1]</sup>。据估计, 2020年我国心血管疾病的患病人数已达2.9亿, 在过去二十余年里, 其年龄标准化患病率增幅达到14.7%<sup>[2]</sup>。心血管疾病病因复杂、病情迁延, 其预后一般较差, 给患者家庭和社会带来沉重的疾病负担。因此, 科学预测心血管疾病预后

显得极为关键。当前, 数据挖掘技术正逐渐引入医疗卫生领域, 许多研究表明其在心血管疾病预后预测方面具有广阔前景<sup>[3]</sup>。为此, 本文梳理了数据挖掘技术在心血管疾病预后研究中的应用及其局限性, 为未来研究提供参考。

#### 一、心血管疾病预后

心血管疾病预后是指对心血管疾病可能的病程和结局的预测, 根据其在患病后是否接受干预, 分为自然预后和干预预后。在自然预后研究中, 因未施加干预措施, 研究主要

DOI: 10.3760/cma.j.cn112338-20210430-00361

收稿日期 2021-04-30 本文编辑 万玉立

引用本文: 向超益, 吴亚飞, 方亚. 数据挖掘技术在心血管疾病预后研究中的应用进展[J]. 中华流行病学杂志, 2021, 42(12): 2234-2238. DOI: 10.3760/cma.j.cn112338-20210430-00361.



关注患者的自然转归,如生存时间、不良心血管事件(包括非致死性心肌梗死、新发冠心病、脑卒中等心血管并发症和心血管死亡)等。在干预预后研究中,主要关注干预措施的安全性和有效性<sup>[4]</sup>。其中,安全性是对干预措施介入后可能造成的不良风险的评价,主要包括:①全死因死亡;②并发症发生情况:如尿路感染、压疮、脑卒中后抑郁(post-stroke depression, PSD)等。而有效性则是对干预措施可能改善健康状况的评价,主要包括以下几个评估维度:①日常生活活动能力(activities of daily life, ADL),常用 Barthel 指数(Barthel index, BI)和功能独立性评定量表(functional independence measure, FIM)进行测评,前者用于测量患者日常生活活动的功能状态,共计 100 分,后者用来评价日常生活独立能力,包括运动功能得分(FIM motor subscore)和认知功能得分(FIM cognitive subscore)两个子模块,共计 126 分;②认知功能,用于反映学习、记忆、语言、思维、情感等能力,常用简易精神状态检查量表(mini-mental state examination, MMSE)进行测量;③运动功能,常使用 FMA 评测指数(fugl-meyer assessment, FMA)来反映评价对象的整体运动能力;④残疾程度,主要以改良 Rankin 量表(modified Rankin scale, mRS)来测量,由无症状的 0 分到重度残疾的 5 分,共分 6 个等级;⑤吞咽功能,主要通过饮水、唾液吞咽等试验来评估吞咽功能障碍程度;⑥其他,如平均住院日、生活质量等。

## 二、数据挖掘技术及其在心血管疾病预后研究中的应用

信息技术的发展使数据库规模不断扩大,为处理复杂的海量数据,数据挖掘技术应运而生,该技术运用算法发现有价值的信息,以实现知识学习,并对未来实现合理预测<sup>[5]</sup>,主要包括统计方法和机器学习(machine learning),机器学习又可细分为传统机器学习、集成学习和深度学习(deep learning)。下面将对常用数据挖掘方法在心血管疾病预后中的应用进行梳理。

1. 统计方法:在数据分析时使用的回归分析、单变量或多变量分析等技术,可简单快速地处理一些统计问题。其中,逻辑回归(logistic regression, LR)是心血管疾病预后研究中最经典的回归模型,其借助 logit 函数实现对一组自变量与因变量关系的探究,并以概率形式反映风险大小,具有模型设定简单、训练速度快、解释性好的优点。例如,在 ST 段抬高型心肌梗死妇女的住院全死因死亡风险预测研究中, Mansoor 等<sup>[6]</sup>采用后消法纳入 11 个预测变量构建 LR 模型,结果发现 LR 的受试者工作特征曲线下面积(area under the receiver operating characteristic curve, AUROC)为 0.84,模型的区分度较好。由于 LR 要求自变量与因变量之间的关系呈基本的 S 形,因而,当数据结构更复杂或预测变量与结局间呈现非 S 形关系时,LR 的性能则不太理想。如 Jang 等<sup>[7]</sup>对脑卒中后残疾程度开展预测,收集了 6 731 名脑卒中患者的人口学和临床资料,通过构建 LR 和 5 种机器学习模型来预测住院患者 3 个月后的 mRS 情况,结果表明机器学习

模型的区分度和校准度均优于 LR 模型,这提示在面对数量庞大且异质性较强的资料时,机器学习方法的应用将会极大提升预测能力。

2. 传统机器学习方法:是一类经典的预测模型,能够对大规模数据集实现特征提取并自动学习数据内在模式<sup>[8]</sup>,且无需任何假设,更有助于发现高维数据中复杂的非线性关系。在心血管疾病预后研究中常用方法包括 K 近邻(K-nearest neighbor, KNN)、决策树(decision tree, DT)、支持向量机(support vector machine, SVM)、人工神经网络(artificial neural network, ANN)、贝叶斯分类(Bayesian classification)。

(1)KNN:1967 年 Cover 和 Hart<sup>[9]</sup>提出了 KNN 算法,基本思想是通过样本点在特征空间中的 K 个最相似样本的大多数来决定待分类样本的类别,由于该模型思路简单且无输入数据假定,在心血管疾病预后预测方面应用较多,如 Thakkar 等<sup>[10]</sup>使用 KNN 对 239 名接受任务导向性干预(约 30 h)的新发卒中患者的运动功能提升程度进行预测,采用 FMA 评价运动功能,根据干预前后 FMA 得分提升 $\geq 4$ 分(高反应组)和 $< 4$ 分(低反应组)来界定研究结局,结果显示脑卒中发病持续时间、基线独立能力以及基线 FMA 评分对 KNN 预测最为重要,其 AUROC 达到 0.89。

(2)DT:于 1986 年由 Quinlan<sup>[11]</sup>提出,是一种自上而下的树形结构,从根节点开始,经过内部节点的属性划分直至叶子节点,最终输出分类结果,它反映了对象属性和对象值间的映射关系,其本质是基于数据归纳出的分类规则,常用的 DT 算法包括 ID3、C4.5、CART 等。DT 具有抗噪能力强、解释性强的优点,已成为心血管疾病预后研究的重要方法。例如 Iwamoto 等<sup>[12]</sup>以 994 名康复科脑卒中患者为对象,以 mRS 得分衡量残疾程度,采用 DT 构建脑卒中后残疾程度恢复情况预测模型,其 AUROC 达 0.83,同时发现 FIM 转移(床、椅子、轮椅间)、如厕以及洗澡是最重要的 3 个预测变量。此外,DT 也被应用于脑卒中后认知功能预测,如 Marcano-Cedeño 等<sup>[13]</sup>的研究发现 DT 在脑卒中后整体认知功能恢复预测中表现优异,预测准确率达 90.38%。以上研究提示 DT 在心血管疾病预后预测上的性能优势,为临床防治和康复实践提供重要指导。

(3)SVM:最早由 Cortes 和 Vapnik<sup>[14]</sup>在 1995 年提出,是通过最大间隔超平面对数据实现最优划分的监督学习方法,由于 SVM 结果仅取决于支持向量,因而尤其适用于小样本资料的预测。此外,SVM 还可借助核函数将复杂的分类任务映射到高维空间以此实现线性划分,加之该算法考虑了经验风险和结构风险最小化并使用铰链损失函数作为代理损失<sup>[15]</sup>,具有较好的稳定性和稀疏性。SVM 在预测心血管疾病预后方面有着不错的表现,如在脑卒中后 ADL 预测方面,Sale 等<sup>[16]</sup>纳入 55 名康复科的亚急性脑卒中患者,采用人口统计学、临床变量、炎症生物标志物预测其出康复科后的 BI、FIM 总得分、FIM 运动/认知得分,结果显示预测得分与实际得分相关性较高(相关系数:0.75~0.81)。此外,在心机

梗死死亡率研究方面, Wallert 等<sup>[17]</sup>基于 39 个预测因素构建了 SVM 预测心肌梗死患者 2 年死亡风险, 结果显示 SVM 的 AUROC 为 0.84。

(4) ANN: 是 20 世纪 80 年代基于人脑神经网络建立的模型, 通过一定层次排列的神经元和激励函数(activation function), ANN 能对任意非线性问题实现逼近<sup>[18]</sup>。ANN 具有很强的自学习和容错能力, 能高效地寻找复杂非线性问题的优化解, 在心血管疾病预后预测方面有着重要应用价值。如 van OS 等<sup>[19]</sup>通过 1 383 例急性缺血性卒中患者的人口统计学、临床资料和治疗信息构建 ANN 模型, 预测患者经再灌注治疗 3 个月后的 mRS 改善情况, 结果发现 ANN 的 AUROC 达 0.88。类似地, 在 Ramos 等<sup>[20]</sup>的研究中 ANN 对急性缺血性卒中患者再灌注治疗 3 个月后 mRS 的恢复情况预测准确, AUROC 达 0.81。

(5) 贝叶斯分类: 是一种非规则的概率分类方法<sup>[21]</sup>, 具有最大概率的类别为最终类别, 主要分为朴素贝叶斯和贝叶斯网络, 适用于小样本资料预测。同时, 该算法也能以增量方式对大样本数据分批学习, 对于多分类问题也同样很有效, 因此在心血管疾病预后预测方面受到较多关注。如在卒中后不良事件预测方面, Easton 等<sup>[22]</sup>采用朴素贝叶斯预测卒中后不同时段(1~7 d 和 8~93 d)的死亡风险, 发现两个不同时段模型的灵敏度和特异度分别为 92.6% 和 69.4%, 83.3% 和 70.8%。此外, Lee 等<sup>[23]</sup>通过性别、临床吞咽困难分级、管饲状态、基底神经节或放射冠或内囊双侧病变等 10 个预测因素, 构建增强型朴素贝叶斯网络(tree-augmented naive bayes, TAN)、半朴素贝叶斯网络(semi-naive network model)和 SVM 对 137 例伴有吞咽功能障碍的卒中患者在 6 个月后的吞咽情况进行预测, 其结果表明 TAN 模型的预测性能最佳(AUROC: 0.80)。以上研究说明贝叶斯方法在心血管疾病预后风险识别方面具有一定优势。

3. 集成学习: 传统预测方法由于仅使用单个弱分类器进行分析, 在解决复杂问题时性能有限。1979 年 Dasarathy 和 Sheela<sup>[24]</sup>基于集成思想提出集成学习方法, 其基本思想是通过组合多个同质或异质的弱分类器以实现综合预测, 有助于提升模型性能。其中, 同质集成是对多个相同的弱分类器进行集成预测, 如 Boosting 和 Bagging 常用于心血管疾病的预测研究。异质集成则是通过多个不同的弱分类器来开展集成预测, Stacking 是其典型代表, 但目前该方法在心血管疾病预后中应用很少。

(1) Boosting: 通过串行方式不断叠加相同的弱学习器, 并根据学习器表现调整样本分布, 最终通过投票或加权平均等形式产生一个强学习器。Boosting 在心血管疾病预后研究中具有广阔前景, 如 Motwani 等<sup>[25]</sup>利用 logitBoost 对疑似冠心病患者 5 年内的全死因死亡情况开展预测, 结果显示相较于传统的节段狭窄评分(SSS)、节段受累评分(SIS)和改良 Duke 指数(DI)等工具, logitBoost 预测效果更佳(AUROC 分别为 logitBoost: 0.79; SSS: 0.64; SIS: 0.64; DI: 0.62)。Xie 等<sup>[26]</sup>利用 513 名脑卒中患者的入院及 24 h 的

物标志物、脑卒中评分等资料, 构建 XGB(extreme gradient boosting)和 GBM(gradient boosting machine)模型来预测患者 90 d 后残疾程度的恢复情况, 结果发现 XGB 和 GBM 的 AUROC 分别为 0.88 和 0.87, 预测效果良好。

(2) Bagging: 与 Boosting 相反, Bagging 是通过并行方式综合多个同质弱学习器结果实现预测。在心血管疾病预后中, 以 DT 为弱分类器的随机森林(random forest, RF)是最为常用的 Bagging 集成方法, RF 具有抗过拟合能力强, 自动完成特征选择的优势<sup>[27]</sup>, 对预防心血管疾病不良预后和探索危险因素有着重要意义。如罗晓舟等<sup>[28]</sup>的研究关注了 PSD, 通过收集 688 例卒中患者的病历资料, 在变量选择基础上构建 RF 模型预测 PSD 发生风险, 预测准确率达 99.47%, 研究同时发现抑郁家族史、病灶位置以及中医药干预是预测 PSD 发生最重要的 3 个因素。另外, Lacson 等<sup>[29]</sup>借助人口学和临床等信息构建 RF 模型, 预测高血压患者[SBP $\geq$ 130 mmHg(1 mmHg=0.133 kPa)]37 个月内发生其他心血管事件的风险, 结果表明 RF 的 AUROC 为 0.71, 尿蛋白/肌酐比率、估算肾小球滤过率、年龄是最重要的 3 个预测变量。

4. 深度学习: 由 Hinton 和 Salakhutdinov<sup>[30]</sup>提出, 是当下的热点领域, 它借助复杂的网络结构实现对任意非线性问题的逼近, 相较于传统机器学习方法和集成学习更具自我驱动性, 但中间过程不透明, 训练成本较大, 实现也相对复杂。随着近些年深度学习算法的发展和计算机算力的提升, 深度学习也逐步在心血管疾病预后研究中得到应用, 以卷积神经网络(convolutional neural networks, CNN)和长短期记忆网络(long short-term memory, LSTM)应用最为广泛。

(1) CNN: 是一种包括卷积、池化等计算在内的前馈神经网络, 卷积层通过局部连接和权重共享实现局部特征的提取, 池化层则能有效降低输入维度, CNN 具备表征学习能力, 在图像、音频等信息处理方面有强大优势, 在心血管疾病预后研究中也有越来越多的应用。如 Liu 等<sup>[31]</sup>使用心力衰竭患者的非结构化临床记录训练 CNN 模型, 该研究将结局分为一般再入院和 30 d 再入院, 其中只要发生再入院即为一般再入院, 而首次入院和再入院的时间间隔 $\leq$ 30 d 即为 30 d 再入院, 采用  $F_1$  得分综合评价模型性能, 结果表明 CNN 在预测一般再入院和 30 d 再入院时的  $F_1$  值分别为 0.75 和 0.73, 同时研究也发现与服用药物和手术有关的变量在再入院预测中尤为重要。Diller 等<sup>[32]</sup>的研究则关注了先天性心脏畸形疾病—法洛四联症的预后, 利用 372 名法洛四联症患者心脏影像学资料来训练 5 种基于 CNN 的深度学习模型, 对患者不良结局(死亡、心脏停搏和记录有 $\geq$ 3 次的室性心动过速)的发生情况进行预测, 发现 5 种 CNN 模型对识别不良结局均有一定参考价值(AUROC $>$ 0.6)。CNN 由于能够自动获得成像参数, 极大简化影像分析过程, 并能达到较好的预测性能, 未来在心血管疾病预后预测中将会发挥更大的应用价值。

(2) LSTM: 现实研究中, 以纵向形式收集的时序性资料更为常见, 这类数据的主要特征是具有很强的次序关系, 而

CNN 等模型由于无法充分利用时序信息,在解决该类问题时具有一定局限性。循环神经网络(recurrent neural network, RNN)能够对时序性资料进行有效处理,尤其是 LSTM 在心血管疾病预后中得到一定应用。LSTM 由 Hochreiter 和 Schmidhuber<sup>[33]</sup>于 1997 年引入,其实质仍是 RNN 模型,通过输入门、遗忘门和输出门接受当前时刻的输入和上个时刻神经元的输出,并能通过控制这 3 个门来实现神经元记忆状态的修改和输出,有效解决了长期依赖问题,适用于处理时间间隔和延迟很长的事件。例如 Park 等<sup>[34]</sup>纳入 74 535 例原发性高血压患者,收集患者人口学和医疗就诊信息,建立 LR、SVM、DT、RF、多层感知机和 LSTM 预测其 5 年内发生其他心血管疾病的风险,结果表明与其他模型相比,LSTM 对于纵向时序性资料具有显著的预测优势, $F_1$ 得分达 0.77。

### 三、小结与展望

数据挖掘方法在心血管疾病预后研究中前景广阔,相较于统计方法,机器学习具有更大的优势。当然,数据挖掘技术在发展中也面临一些问题:①样本量问题:数据挖掘方法往往更适用于大型数据集,而在心血管疾病预后研究中,由于伦理、安全性等原因,收集到的样本量通常较小;②数据质量问题:临床数据的收集会受医疗器械、检测人员水平的限制,同时对医学数据进行标注时常存在错误标注或重复标注等问题,加之医学数据的非平衡性问题,都制约着机器学习的效率;③缺乏外部验证:当前大量研究仅开展了内部验证,缺乏有效的外部验证来评估模型泛化能力;④模型的可解释性:绝大多数机器学习模型属于黑盒模型,在中间过程和结果解释方面存在缺陷,如今兴起的可解释性机器学习将是未来一大发展方向;⑤临床应用问题:绝大多数预测模型仍停留在理论研究,尚未转化为临床应用。

从研究内容来看,脑卒中预后是心血管疾病预后研究的核心,且多集中于干预后的转归,不仅关注了全死因死亡、并发症等安全性评价,还重点对运动能力、认知能力等干预有效性情况开展预测,而对于其他心血管疾病,主要关注自然预后和干预安全性,如心脏衰竭患者的死亡及再入院预测等。未来研究可从以下方面做更多深入探索:①安全性结合有效性:同时考虑干预的安全性和有效性问题,对预后情况做更全面的预测;②预测人群细分:由于个体发病后差异较大,因而对人群分层开展风险预测,更有助于精准预防;③预测结局细分:细化量表/临床诊断等,开展更精细维度预测,如将 FIM 细化为运动功能和认知功能,将 FMA 量表细化为上肢和下肢功能等;④利用康复干预的初、中、末期评价或临床干预的多期随访结果,对患者的动态预后开展预测。

利益冲突 所有作者均声明不存在利益冲突

### 参 考 文 献

[1] Balakumar P, Maung-UK, Jagadeesh G. Prevalence and

prevention of cardiovascular disease and diabetes mellitus[J]. *Pharmacol Res*, 2016, 113: 600-609. DOI: 10.1016/j.phrs.2016.09.040.

- [2] 基层心血管病综合管理实践指南 2020[J]. *中国医学前沿杂志:电子版*, 2020, 12(8):1-73. DOI:10.12037/YXQY.2020.08-01.  
Beijing Hypertension Association, Beijing Diabetes Prevention and Treatment Association, Beijing Research Association for Chronic Diseases Control and Health Education. Practice Guide for comprehensive management of basic cardiovascular disease 2020[J]. *Chin J Front Med Sci:Electron Vers*, 2020, 12(8):1-73. DOI:10.12037/YXQY.2020.08-01.
- [3] Sevakula RK, Au-Yeung WTM, Singh JP, et al. State-of-the-art machine learning techniques aiming to improve patient outcomes pertaining to the cardiovascular system [J]. *J Am Heart Assoc*, 2020, 9(4):e013924. DOI:10.1161/JAHA.119.013924.
- [4] 王清, 史慧玲, 薛俐俐, 等. 早期活动对急性脑卒中患者预后影响的 Meta 分析 [J]. *中华护理杂志*, 2016, 51(12): 1443-1450. DOI:10.3761/j.issn.0254-1769.2016.12.007.  
Wang Q, Shi HL, Xue LL, et al. Effects of early mobilization on prognosis for acute stroke patients:a Meta-analysis[J]. *Chin J Nurs*, 2016, 51(12): 1443-1450. DOI: 10.3761/j.issn.0254-1769.2016.12.007.
- [5] Chen LYA, Fawcett TN. Using data mining strategies in clinical decision making: a literature review[J]. *Comput Inform Nurs*, 2016, 34(10): 448-454. DOI: 10.1097/CIN.0000000000000282.
- [6] Mansoor H, Elgendy IY, Segal R, et al. Risk prediction model for in-hospital mortality in women with ST-elevation myocardial infarction: a machine learning approach[J]. *Heart Lung*, 2017, 46(6): 405-411. DOI: 10.1016/j.hrtlng.2017.09.003.
- [7] Jang SK, Chang JY, Lee JS, et al. Reliability and clinical utility of machine learning to predict stroke prognosis: comparison with logistic regression[J]. *J Stroke*, 2020, 22(3):403-406. DOI:10.5853/jos.2020.02537.
- [8] Sengupta PP, Shrestha S. Machine learning for data-driven discovery: the rise and relevance[J]. *JACC Cardiovasc Imaging*, 2019, 12(4):690-692. DOI:10.1016/j.jcmg.2018.06.030.
- [9] Cover T, Hart P. Nearest neighbor pattern classification[J]. *IEEE Trans Inform Theory*, 1967, 13(1): 21-27. DOI: 10.1109/tit.1967.1053964.
- [10] Thakkar HK, Liao WW, Wu CY, et al. Predicting clinically significant motor function improvement after contemporary task-oriented interventions using machine learning approaches[J]. *J Neuroeng Rehabil*, 2020, 17(1): 131. DOI:10.1186/s12984-020-00758-3.
- [11] Quinlan JR. Induction of decision trees[J]. *Mach Learn*, 1986, 1(1):81-106. DOI:10.1007/BF00116251.
- [12] Iwamoto Y, Imura T, Tanaka R, et al. Development and validation of machine learning-based prediction for dependence in the activities of daily living after stroke inpatient rehabilitation: a decision-tree analysis[J]. *J Stroke Cerebrovasc Dis*, 2020, 29(12): 105332. DOI: 10.1016/j.jstrokecerebrovasdis.2020.105332.
- [13] Marcano-Cedeño A, Chausa P, García A, et al. Data mining applied to the cognitive rehabilitation of patients with acquired brain injury[J]. *Expert Syst Appl*, 2013, 40(4):

- 1054-1060. DOI:10.1016/j.eswa.2012.08.034.
- [14] Cortes C, Vapnik V. Support-vector networks[J]. *Mach Learn*, 1995, 20(3):273-297. DOI:10.1007/BF00994018.
- [15] Chauhan VK, Dahiya K, Sharma A. Problem formulations and solvers in linear SVM: a review[J]. *Artif Intell Rev*, 2019, 52(2):803-855. DOI:10.1007/s10462-018-9614-6.
- [16] Sale P, Ferriero G, Ciabattini L, et al. Predicting motor and cognitive improvement through machine learning algorithm in human subject that underwent a rehabilitation treatment in the early stage of stroke[J]. *J Stroke Cerebrovasc Dis*, 2018, 27(11): 2962-2972. DOI: 10.1016/j.jstrokecerebrovasdis.2018.06.021.
- [17] Wallert J, Tomasoni M, Madison G, et al. Predicting two-year survival versus non-survival after first myocardial infarction using machine learning and Swedish national register data[J]. *BMC Med Inform Decis Mak*, 2017, 17(1):99. DOI:10.1186/s12911-017-0500-y.
- [18] Yamamura S. Clinical application of artificial neural network (ANN) modeling to predict pharmacokinetic parameters of severely ill patients[J]. *Adv Drug Deliv Rev*, 2003, 55(9): 1233-1251. DOI: 10.1016/s0169-409x(03)00121-2.
- [19] van Os HJA, Ramos LA, Hilbert A, et al. Predicting outcome of endovascular treatment for acute ischemic stroke: potential value of machine learning algorithms[J]. *Front Neurol*, 2018, 9:784. DOI:10.3389/fneur.2018.00784.
- [20] Ramos LA, Kappelhof M, van Os HJA, et al. Predicting poor outcome before endovascular treatment in patients with acute ischemic stroke[J]. *Front Neurol*, 2020, 11:580957. DOI:10.3389/fneur.2020.580957.
- [21] Zellner A. Bayesian and non-Bayesian approaches to statistical inference and decision-making[J]. *J Comput Appl Math*, 1995, 64(1/2):3-10. DOI:10.1016/0377-0427(95)00002-X.
- [22] Easton JF, Stephens CR, Angelova M. Risk factors and prediction of very short term versus short/intermediate term post-stroke mortality: a data mining approach[J]. *Comput Biol Med*, 2014, 54: 199-210. DOI: 10.1016/j.compbimed.2014.09.003.
- [23] Lee WH, Lim MH, Seo HG, et al. Development of a novel prognostic model to predict 6-month swallowing recovery after ischemic stroke[J]. *Stroke*, 2020, 51(2): 440-448. DOI:10.1161/Strokeaha.119.027439.
- [24] Dasarathy BV, Sheela BV. A composite classifier system design: Concepts and methodology[J]. *Proc IEEE*, 1979, 67(5):708-713. DOI:10.1109/PROC.1979.11321.
- [25] Motwani M, Dey D, Berman DS, et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis[J]. *Eur Heart J*, 2017, 38(7): 500-507. DOI:10.1093/eurheartj/ehw188.
- [26] Xie YA, Jiang B, Gong EH, et al. Use of gradient boosting machine learning to predict patient outcome in acute ischemic stroke on the basis of imaging, demographic, and clinical information[J]. *Am J Roentgenol*, 2019, 212(1):44-51. DOI:10.2214/Ajr.18.20260.
- [27] Scornet E. Random forests and kernel methods[J]. *IEEE Trans Inform Theory*, 2016, 62(3): 1485-1500. DOI: 10.1109/Tit.2016.2514489.
- [28] 罗晓舟, 温小鹏, 何家扬, 等. 基于机器学习的卒中后抑郁影响因素分析[J]. *中医杂志*, 2017, 58(17): 1478-1481. DOI:10.13288/j.11-2166/r.2017.17.011.
- [28] Luo XZ, Wen XP, He JY, et al. Analysis of the influencing factors of post-stroke depression: based on machine learning[J]. *J Tradit Chin Med*, 2017, 58(17): 1478-1481. DOI:10.13288/j.11-2166/r.2017.17.011.
- [29] Lacson RC, Baker B, Suresh H, et al. Use of machine-learning algorithms to determine features of systolic blood pressure variability that predict poor outcomes in hypertensive patients[J]. *Clin Kidney J*, 2019, 12(2):206-212. DOI:10.1093/ckj/sfy049.
- [30] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks[J]. *Science*, 2006, 313(5786):504-507. DOI:10.1126/science.1127647.
- [31] Liu X, Chen Y, Bae J, et al. Predicting heart failure readmission from clinical notes using deep learning[C]// 2019 IEEE international conference on bioinformatics and biomedicine (BIBM). San Diego: IEEE, 2019: 2642-2648. DOI:10.1109/BIBM47256.2019.8983095.
- [32] Diller GP, Orwat S, Vahle J, et al. Prediction of prognosis in patients with tetralogy of Fallot based on deep learning imaging analysis[J]. *Heart*, 2020, 106(13): 1007-1014. DOI:10.1136/heartjnl-2019-315962.
- [33] Hochreiter S, Schmidhuber J. Long short-term memory[J]. *Neural Comput*, 1997, 9(8): 1735-1780. DOI: 10.1162/neco.1997.9.8.1735.
- [34] Park J, Kim JW, Ryu B, et al. Patient-level prediction of cardio-cerebrovascular events in hypertension using nationwide claims data[J]. *J Med Internet Res*, 2019, 21(2):e11757. DOI:10.2196/11757.

## 本刊 2021 年第 1 期文章基金编号的更正

本刊 2021 年第 1 期发表的文章《广西壮族自治区多地区长寿人群分布特征及其相关因素研究》(第 99 ~ 105 页)和《广西壮族自治区南宁市上林地区长寿人群分布特征和影响因素研究》(第 106 ~ 112 页), 国家自然科学基金编号 3176029 改为 31760299。特此更正。