

• 综述 •

机器学习方法在慢性病研究中的应用进展

吴亚飞^{1,2} 方 亚^{1,2,△}

近年来,我国慢性病发病率呈现逐年上升趋势,由此造成沉重的疾病负担^[1];同时,慢性病并发症所带来的致死和致残现象极为严峻,已成为影响我国居民健康和生存质量的重大威胁^[2]。因而,对慢性病发病风险进行预测,以便及早采取积极的干预措施,对于疾病的防控工作显得尤为重要。

机器学习作为一类经典的预测模型,具有良好的预测性能^[3],在慢性病研究中已经得到了较为广泛的应用^[4-6]。另外,随着大数据的积累以及云计算的普及,基于健康大数据开展深度学习正成为慢性病研究领域的热点,并有望为智能医疗和健康决策提供强有力的支撑。

为此,本文从传统机器学习、集成学习和深度学习三个方面,对近年来机器学习方法在慢性病研究中的应用进行梳理,以期对慢性病的早期筛查和及时干预提供科学借鉴。

传统机器学习

传统机器学习是一类经典的预测模型,在慢性病发病风险识别以及危险因素研究中得到了广泛应用。为此,我们对慢性病研究中常见的几种传统机器学习方法(如回归分析、决策树、支持向量机、人工神经网络、贝叶斯分类、关联规则等)进行梳理。

1. 回归分析

回归分析由于其模型设定简单、结果易于解释等优势,早已成为机器学习方法库中最常用的工具,在慢性病研究中得到较为广泛应用。其中,以线性回归、logistic 回归及正则化回归应用最为广泛。例如,为探究大气污染物、气象因素与哮喘发病的关系, Gabda D 等人利用多重线性回归构建预测模型,结果显示 NO₂、气温和相对湿度是哮喘发病的重要预测因子^[7]。然而,当影响因素与研究结局呈非线性关系时,线性回归预测结果往往存在较大偏差。为此,基于概率思想的非线性 logistic 回归开始得到应用。如郭艳芳等人为探究肥胖对糖尿病风险的影响及其可能存在的性别

差异,采用 logistic 回归对 2015 年深圳市宝安区 3224 名 18~69 岁常住居民开展研究,发现腰围、腰臀比、腰围身高比对糖尿病预测具有重要影响;另外,研究还发现腰臀比对预测男性糖尿病风险相对重要,而腰围身高比则对女性糖尿病风险预测效果更好^[8]。为进一步提高模型的预测准确性,有研究将核函数的思想与回归分析相结合,如 Matsuda K 等人利用核 logistic 回归对慢性丙型肝炎患者继发肝纤维化进行分期,发现该方法预测准确率相较于传统的 logistic 回归有明显提升,同时该法也能对多分类问题进行有效处理^[9]。此外,在实际研究中,慢性病的影响因素往往很多,不可避免地存在多重共线性问题,若缺乏有效的变量筛选技术将会导致预测效能大幅下降,甚至产生错误的结果。为此,研究者开始将正则化技术运用于回归分析中,显著降低了过拟合风险,如 Xin B 等人基于 62 例痴呆、141 例轻度认知功能障碍患者以及 71 例健康者的神经影像学资料,构建 L1 正则化(LASSO) logistic 回归、支持向量机模型,结果显示正则化的 logistic 对于痴呆和轻度认知功能障碍具有更高的预测准确率^[10]。

2. 决策树

决策树最早由 Quinlan 于 1986 年提出,是一种自上而下的树形模型,从根结点出发依次对结点变量进行划分,直至达到叶子结点为止,决策树中的每条路径对应一个分类规则^[11]。由于决策树分类结果直观、易于解释;同时对噪声不太敏感,分类精度较高,在慢性病危险因素及风险预测研究中得到较多应用。

决策树算法众多,其中以 ID3、C4.5、CART 应用最为广泛。如 AL-Dlaeen D 等人基于经典的 ID3 算法构建阿尔茨海默病发生风险的决策树模型,结果发现该预测系统具有较高的准确性,在阿尔茨海默病临床初筛中发挥了较大的助力^[12]。在经典决策树算法基础上,一些新的算法也陆续得到应用,如冯伟等人基于 2010 年中国慢性病与危险因素监测资料,采用卡方自动交互检测(chi-squared automatic interaction detector, CHAID)算法构建决策树,对不同特征人群进行糖尿病风险研究,结果显示腰围增大并伴有糖尿病家族史的人群糖尿病患病率更高,由于该法所需调查的危险因素较少、操作简便,在确定糖尿病优先筛查人群中具

1. 厦门大学公共卫生学院(361102)

2. 厦门大学健康医疗大数据国家研究院

△通信作者:方亚 E-mail:fangya@xmu.edu.cn

有重要的应用价值^[13]。值得注意的是,在应用中,决策树不可避免地存在过拟合的缺陷,为有效降低该缺陷的影响并进而提高预测准确率,对决策树进行剪枝(预剪枝和后剪枝)显得尤为重要。目前,后剪枝方法应用较多,如 Hashem S 等人为实现肝纤维化的早期筛查,基于 39567 名患者的生物标志物等临床资料,构建了肝纤维化预测的决策树模型,结果显示采用误差降低剪枝算法(reduced error pruning, REP)进行后剪枝之后,决策树的预测准确率得到进一步提升^[14]。

3. 支持向量机

1995 年 Vapnik 等人提出一种基于间隔最大化的分类模型:支持向量机(support vector machine, SVM)^[15],由于其分类结果仅依赖于支持向量,因而特别适用于小样本资料研究;同时,在统计学习理论和结构风险最小化准则的保证下,支持向量机能够获得全局最优解,因而对慢性病预测分析具有重要意义。实际上,研究者面临的往往是非线性可分问题,而支持向量机恰能通过核函数将资料映射到高维空间中实现线性划分,具有十分稳定的预测性能,常用的核函数有多项式、sigmoid、高斯核函数等。例如,孙小宇等人在 logistic 回归筛选冠心病危险因素的基础上,结合 24 小时动态心电图资料,利用高斯核函数构建 SVM 对冠心病早期风险进行预测,结果显示该法较单独使用 24 小时动态心电图进行诊断具有更好的分类性能,为冠心病早期筛查提供了参考^[16]。同样地,为探究老年人慢性病用药依从性, Lee SK 等利用高斯核构建 SVM 对 293 例 65 岁及以上老年人进行分析,发现用药依从性与自我效能关系显著,同时,抑郁、健康素养和药物知识对用药依从性亦有影响,为提高老年人用药依从性提供重要参考依据^[17]。进一步地,王勋等人综合考虑环境和遗传因素,同时采用上述三种核函数构建糖尿病发病风险模型,结果显示基于高斯核和 sigmoid 核的 SVM 预测准确率较高(分别为 98.7%、97.3%),而用多项式核构建的 SVM 预测准确率稍低(93.1%),提示 SVM 对糖尿病高危人群筛查及个性化管理意义重大^[18]。

4. 人工神经网络

人工神经网络(artificial neural network, ANN)是一种应用类似大脑神经连接进行资料处理的方法,它的出现始于 1943 年提出的神经元模型,在该模型基础上,通过增加隐层数而得到。人工神经网络一般通过反向传播(back propagation, BP)算法对模型参数进行精细调试,具有很强的自学习能力,泛化性能较好,成为慢性病研究的重要方法之一。例如,为构建一个仅包含少数关键危险因素的高血压预测模型, Wang A 等人借助 logistic 回归进行危险因素初筛,进而采用 BP 算法训练多层前馈神经网络,结果显示神经网络模

型对高血压预测准确率超过 72%,提示将 logistic 回归和 ANN 进行有效结合,对于高血压快速初筛具有重要参考意义^[19]。值得注意的是,隐层及隐层结点数对神经网络预测性能具有重要影响,目前关于这些参数的设定,研究结论尚不一致,基本上是在众多组合中选择最优隐层和结点数。如为探讨 logistic 回归、决策树、ANN 在 2 型糖尿病发病风险预测性能上的差异,宋健等人利用人群基本资料和实验室指标构建模型,结果显示含有一个隐层的 ANN 预测效能最好^[20]。同样地,为实现结核病的早诊早治, Er O 等人基于土耳其某医院 50 名结核病患者以及 100 名健康个体的 38 个实验室指标构建 ANN 模型,发现当隐层数为 2 时,模型预测效果最好(准确率 $\geq 92.3\%$)^[21]。此外,神经网络学习算法众多,不同的学习算法对模型结果也会产生重要影响。如曹文君等人利用单隐层 BP 网络、径向基函数网络和概率神经网络对北京市 6938 名 45 岁及以上人群多个慢性病的危险因素进行探究,采用逐步回归结合遗传算法进行变量筛选,进而利用上述三种算法构建高血压、糖尿病、冠心病预测模型,结果显示:年龄、体质指数(body mass index, BMI)、总胆固醇、家族史等是慢性病的主要危险因素,其中年龄影响最大^[22]。该研究结论提示我们在慢性病防治中尤其要关注老年患者。

5. 贝叶斯分类

贝叶斯分类也是一种基于概率的学习方法。在小样本资料中分类性能出色,当数据量很大时,能借助增量方式进行分批学习;同时,也能够处理多分类任务,对缺失数据不敏感,性能稳定,因而在慢性病研究中也得到较多应用。目前,在慢性病研究中常见的贝叶斯分类模型有两种:朴素贝叶斯和贝叶斯网络。例如, Wei W 等人从基因水平对阿尔茨海默病发生风险进行预测,利用 1411 例老年人队列研究的全基因组数据构建基于平均思想的朴素贝叶斯模型(model-averaged naive Bayes, MANB),同时与朴素贝叶斯(naive Bayes, NB)和经特征筛选的朴素贝叶斯(naive Bayes with feature selection, FSNB)进行比较,结果显示 MANB 通过对高维基因数据建模结果平均后,不仅节省了运算时间,也提升了预测准确率,为基因层面阿尔茨海默病的筛查提供了指导^[23]。进一步地,在慢性病研究中为充分展现各影响因素间的依赖关系,以概率图为基础的贝叶斯网络开始得到应用。如为从中医角度探究糖尿病并发症患者的血糖指标与中医证候间的关系,龚燕冰等人采用贝叶斯网络对糖尿病合并高血压、脑病、肾病者的空腹血糖、餐后两小时血糖以及糖化血红蛋白与舌、脉等中医证候进行网络关系分析,结果发现不同并发症患者的中医证候存在较大差异,为糖尿病并发症的中医诊断提供了新的思路^[24]。类似

地 Sambo F 等人基于 7807 名欧洲糖尿病人群队列研究资料构建贝叶斯网络,发现代谢综合征、餐后两小时血糖和教育程度是预测人群糖尿病发病风险的重要因素^[25]。此外,贝叶斯网络也应用于儿童哮喘研究中,如 Demissie K 等人采用该法对蒙特利尔 952 名 6~12 岁儿童进行横断面分析,结果发现相对于运动测试,采用问卷调查预测的准确率更高,为儿童哮喘筛查提供了更快捷的方法^[26]。

6. 关联规则

关联规则 (association rule, AR) 最早由 Rakesh Agrawal 等人于 1993 年提出,是一种用于挖掘数据资料背后有意义的关联方法,由于其对数据分布不做任何假设,应用更为灵活,在医疗诊断尤其是慢性病研究中受到关注。其中,Apriori 算法因其优异的性能成为应用最为普遍的关联规则方法。例如,为提高阿尔茨海默病的预测效果,Chaves R 等人应用 Apriori 算法从脑图像中提取与阿尔茨海默病高度相关的区域作为研究对象,进而采用 SVM 进行分类,结果发现经 AR 筛选后的模型预测准确率明显提升^[27]。类似地,郭玉婷等将 AR 用于慢性阻塞性肺疾病 (COPD) 患者超时住院费用的影响因素探究,发现与超时住院费用具有较大关联的因素有住院天数、医院级别和年龄。这对控制住院费用、减轻 COPD 患者医疗负担具有重要的借鉴意义^[28]。进一步,为探究多种慢性病间的关联性及关联强度,任仙龙等基于 4585 名调查者的健康资料对常见慢性病进行探究,结果发现高血压、高血脂和糖尿病之间存在较强的关联性,同时这些疾病与其他疾病之间也存在广泛关联,提示在慢性病的诊治中,应提高对疾病间相关性的关注^[29]。

除上述常用方法外,聚类分析、遗传算法、主成分分析等在慢性病危险因素研究中也有一些应用。如为探究不同区域间高血压患者的血压控制情况和影响因素分布差异,孙丽等人对我国 132 家医院的 4110 例高血压患者健康资料进行聚类分析,结果发现高血压患者血压控制情况、影响因素存在较大的地区差异,提示在高血压防治中,应当因地制宜采取防控措施^[30]。同样地,Shahbakhti M 等人为实现帕金森病的早期筛查,利用主成分分析对影响因素进行降维,得到若干个综合变量,在此基础上,采用 SVM 构建预测模型,对帕金森病实行早期筛查^[31]。

集成学习

集成学习是一种将多个性能较弱的学习模型进行组合优化的方法,集成思想最早由 Dasarathy 和 Sheela 于 1979 年提出^[32]。1990 年,有研究者提出基于神经网络的集成学习方法,集成方法开始凸显其优越的预测性能。随后,大量集成学习方法开始出现并得到应用。

由于集成学习能够针对实际问题设计合适的学习方案,从而得到性能更优的解决策略^[33],因而近年来在慢性病研究中得到越来越多应用。目前常用的集成方法有 Boosting、Bagging、随机森林等。

Boosting 是一种串联式的学习方法,根据上一次预测结果对训练样本进行权重调整,错误分类的样本赋予更大的权重,通过多轮迭代后对结果进行集成,其中应用最为广泛的是自适应增强算法 (AdaBoost)。例如, Savio A 等人为实现阿尔茨海默病的早诊早治,利用患者脑磁共振图像构建 Adaboost 预测模型,发现该法在阿尔茨海默病早期诊断方面具有较高的准确率 (85%)^[34]。此外,在 AdaBoost 基础上衍生出多种性能优异的增强算法,如 Chen P 等采用 Adaboost M1 和 LogitBoost 对 35669 名糖尿病患者资料构建模型,通过交叉验证发现后者的预测准确率稍高 (92.60% vs 93.93%)^[35],说明 Boosting 在临床糖尿病早期筛查中具有重要意义。

有别于 Boosting, Bagging 算法是一种并行式的学习方法,它能同时生成多个基分类器,最后采用投票方式进行集成。如为早期识别糖尿病视网膜病变,进而采取干预措施预防失明, Syna Sreng 等从 1200 张视网膜图像中提取 208 个特征,采用混合模拟退火选取最优的影响因素集合,接着运用 Bagging 算法构建糖尿病视网膜病变预测模型,结果显示该系统对视网膜病变预测准确率高达 97.08%,为糖尿病眼部并发症的早期筛查提供了有效方法^[36]。进一步地,在 Bagging 基础上,研究者将特征随机选择技术引入模型中得到随机森林,有利于构造更多样化的基分类器,从而进一步提高集成模型的预测性能。如于大海等为比较随机森林与传统机器学习在预测性能上的差异,采用随机森林、logistic 回归、决策树对 942 例肝硬化患者并发上消化道出血风险进行研究,通过比较发现随机森林预测性能最佳,对肝硬化患者并发消化道出血的预后评估具有重要的应用价值^[37]。

值得注意的是,上述集成学习方法均是基于同一种算法进行集成,性能提高有限^[38]。为进一步提高集成模型的预测效能,同时对多种不同算法进行集成的 stacking 方法开始得到应用,但目前尚未见该方法在慢性病研究中的应用,提示在未来研究中可以将 stacking 应用在慢性病领域。另外,在 Boosting 基础上提出的梯度提升树 (gradient boosting decision tree, GBDT) 算法,由于其能够灵活处理连续和离散资料且具有较高的预测性能,在多个领域已得到应用,在慢性病研究中,仅吴健等人基于医疗体检数据,采用 GBDT 结合 logistic 回归构建糖尿病患病风险模型,为临床诊断提供有效辅助,也便于病人了解自己的患病风险^[39]。

深度学习

随着大数据的发展以及云计算的普及,近年来深度学习得到越来越多关注。上述传统机器学习方法和集成学习方法大多为浅层学习,虽然在疾病预测和影响因素研究中发挥了一定作用,但当面对海量的数据资料时,上述方法往往无法有效完成高维数据、庞大运算等任务。为此,2006年,Hinton等人提出了深度学习^[40](deep learning, DL)这一概念,随后DL经历了快速的发展。深度学习作为一类深度学习模型,实质上是神经网络的发展,通过多个非线性模块组合构成复杂的深度网络结构^[41],对多水平特征进行表示学习。传统的神经网络模型由于运算能力受限,网络的层数和神经元个数往往较少,网络复杂度也较为有限,这也在很大程度上限制了模型性能的提升。然而,随着算力的提升,DL中的网络层数和神经元个数可以根据实际需求灵活设定,同时,借助各种学习算法,深度学习能够实现众多参数的快速更新。因而,深度学习预测性能得到大幅提升,正成为慢性病预测研究中的热点。

深度学习算法众多,在慢性病研究领域,目前主要集中在深度神经网络(deep neural network, DNN)、卷积神经网络(convolutional neural network, CNN)方面^[42]。DNN相比于传统的ANN,层数显著增多,网络复杂度增加。同时,DNN采用ReLU等作为激活函数,有效克服了梯度消失问题,使得DNN在疾病预测中得到广泛应用。如为探究吸烟对机体老化的影响,Mamoshina P等人基于149000名研究对象的健康资料进行DNN分析,结果发现血生化指标和细胞计数对吸烟状态的预测性能较高,同时定量分析显示:与非吸烟者相比,女性吸烟者的生物年龄是实际年龄的两倍,而男性仅为1.5倍,提示吸烟对女性影响更大^[43]。近年来,基于医学影像资料进行疾病诊断发展迅速,其中在慢性病研究中应用最普遍的方法是CNN。CNN通过卷积、池化等操作实现样本资料的降采样,同时结合权重共享等手段有效增强了对图像局部特征的利用,对于减少网络参数、优化网络结构具有重要意义,已成为目前图像识别领域最高效的深度学习模型^[44]。如李琼等人基于患者的视网膜图像资料,采用CNN对视网膜病变进行分类预测,结果显示CNN在视网膜病变预测中具有较高的准确性(93.0%),既避免了人工提取特征和图像分类的困扰,也为早期筛查、预防失明起到重要帮助^[45]。同样地,Ting DSW等人基于多种族人群的视网膜资料构建CNN模型,不仅实现了糖尿病视网膜病变的早期筛查,发现该法对其他眼部疾病也具有很高的预测性能。^[46]另外,也有研究者基于CNN对心血管疾病风

险进行预测,通过对284335名患者的视网膜眼底图像资料进行分析,发现年龄、性别、吸烟、糖化血红蛋白、收缩压等是心血管事件高风险的重要影响因素^[47]。

除此之外,循环神经网络(recurrent neural network, RNN)在医疗健康中也开始引起关注。医疗数据往往是纵向或呈现时间序列性质,DNN和CNN等深度学习模型对此效果一般不好,而RNN对于具有依赖性的时序资料,恰能通过其网络内部的定向循环结构对资料形成一定记忆,可以有效解决输入资料间存在相互联系的问题,目前在电子医疗记录(electronic health record, EHR)数据中得到一些应用,为海量医疗文本挖掘提供了有效的技术支撑^[42, 48],但目前尚缺少RNN在慢性病中的应用研究。

总结与展望

机器学习作为一类性能优异的预测方法,对慢性病的早期筛查具有重要意义。在小样本健康资料研究中,传统的机器学习方法是一种重要的探索技巧;进一步,集成学习通过对多个弱学习方法进行集成,预测性能有了大幅提升,为慢性病筛查和健康管理提供了重要帮助。近年来,基于大数据平台和云计算技术,越来越多的深度学习开始应用于慢性病预测中,为智能医疗以及临床决策提供了强大支撑。随着健康中国战略的深入发展以及健康医疗大数据系统的日趋成熟,机器学习尤其是深度学习在疾病预测方面将具有越来越重要的作用。同时,随着研究的不断深入,未来将呈现以下发展趋势:首先,在国家健康大数据战略的积极推动下,势必促成医疗健康数据的大融合和共享;其次,当前的健康医疗大数据质量不容乐观,对模型预测准确性和精度具有重要影响,为此,有必要规范数据采集标准;再次,需要进一步探索机器学习与可视化的深度融合,使研究结果得到更直观的呈现;此外,随着机器学习方法的应用和推广,未来将不仅局限于慢性病预测研究,而是着眼于人群的大健康研究;最后,在大数据和深度学习的推动下,医疗诊断和决策将变得更智能、更精准,医疗资源也将得到更合理的分配。

参考文献

- [1] 王丽敏, 陈志华, 张梅, 等. 中国老年人群慢性病患者状况和疾病负担研究. 中华流行病学杂志. 2019, 40(3): 277-283.
- [2] 喻思思, 童琪, 王丹, 等. 老年综合评估在老年慢性病中的研究进展. 中国老年学杂志. 2019, 39(1): 245-248.
- [3] Horvitz E, Mulligan D. Machine learning: Trends, perspectives, and prospects. Science. 2015, 349(6245): 253-255.
- [4] Waljee AK, Wallace BI, Cohen-Mekelburg S, et al. Development and Validation of Machine Learning Models in Prediction of Remission in Patients With Moderate to Severe Crohn Disease. JAMA Network Open. 2019, 2(5): e193721.
- [5] Jamshidi A, Pelletier J, Martel-Pelletier J. Machine-learning-based patient-specific prediction models for knee osteoarthritis. Nature Re-

- views Rheumatology 2019 ,15: 49-60.
- [6] Zhan A ,Mohan S ,Tarolli C. Using Smartphones and Machine Learning to Quantify Parkinson Disease Severity: The Mobile Parkinson Disease Score. JAMA Neurology 2018 ,75(7) : 876-880.
 - [7] Gabda D ,Jubok ZH ,Budín K ,et al. Multiple Linear Regression in Forecasting the Number of Asthmatics. WSEAS Transactions on Information Science and Applications 2008 ,5(6) : 972-977.
 - [8] 郭艳芳 ,唐瑛 ,刘峥 ,等. 不同肥胖指标对糖尿病风险预测效果比较的性别差异. 中国热带医学 2015 ,15(9) : 1062-1065.
 - [9] Matsuda K ,Ohsaki M ,Katagiri S ,et al. Application of Kernel Logistic Regression to the Prediction of Liver Fibrosis Stages in Chronic Hepatitis C. Joint International Conference on Soft Computing & Intelligent Systems 2012.
 - [10] Xin B ,Yoshinobu ,Kawaharay ,et al. Efficient Generalized Fused Lasso and Its Application to the Diagnosis of Alzheimer's Disease. Proceeding AAAI'14 Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence 2014.
 - [11] Quinlan JR. Induction of Decision Trees. Machine Learning ,1986 ,1(1) : 81-106.
 - [12] AL-Dlaen D ,Alashqur A. Using decision tree classification to assist in the prediction of Alzheimer's disease. 2014 6th International Conference on Computer Science and Information Technology (CSIT) , 2014.
 - [13] 冯伟 ,董红军 ,俞敏 ,等. 社区无症状 II 型糖尿病优先筛查人群研究. 国际流行病学传染病学杂志 2015 ,42(1) : 37-40.
 - [14] Hashem S ,Esmat G ,Elakel W ,et al. Accurate Prediction of Advanced Liver Fibrosis Using the Decision Tree Learning Algorithm in Chronic Hepatitis C Egyptian Patients. Gastroenterology Research and Practice 2016 ,2016: 1-7.
 - [15] Cortes C ,Vapnik V. Support-Vector Networks. Machine Learning , 1995 ,20(3) : 273-297.
 - [16] 孙小宇 ,姚晨 ,康晓平. 支持向量机在建立冠心病早期诊断模型中的应用. 中国卫生统计 2011 ,28(2) : 122-125.
 - [17] Lee S K ,Kang B ,Kim H ,et al. Predictors of Medication Adherence in Elderly Patients with Chronic Diseases Using Support Vector Machine Models. Healthcare Informatics Research 2013 ,19(1) : 33.
 - [18] 王勋 ,陈大方. 支持向量机在建立 2 型糖尿病预测模型中的应用. 中国慢性病预防与控制 2010 ,18(6) : 560-562.
 - [19] Wang A ,An N ,Xia Y ,et al. A Logistic Regression and Artificial Neural Network-Based Approach for Chronic Disease Prediction: A Case Study of Hypertension. 2014 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber ,Physical and Social Computing (CPSCom) 2014.
 - [20] 宋健 ,吴学森 ,张杰 ,等. 三种统计学模型在糖尿病个体患病风险预测中的应用. 中国卫生统计 2017 ,34(2) : 312-314.
 - [21] Er O ,Temurtas F ,Tanrikulu AC. Tuberculosis Disease Diagnosis Using Artificial Neural Networks. Journal of Medical Systems 2010 ,34(3) : 299-302.
 - [22] 曹文君 ,徐勇勇 ,谭志军 ,等. 基于人工神经网络模型的多个慢性病主要危险因素筛查研究. 中国全科医学 2015 ,18(25) : 3050-3053.
 - [23] Wei W ,Visweswaran S ,Cooper GF. The application of naive Bayes model averaging to predict Alzheimer's disease from genome-wide data. Journal of the American Medical Informatics Association 2011 ,18(4) : 370-375.
 - [24] 龚燕冰 ,倪青 ,高思华 ,等. 2 型糖尿病不同并发症中医证候与血糖相关性的贝叶斯网络分析. 北京中医药大学学报 2009 ,32(12) : 815-818.
 - [25] Sambo F ,Di Camillo B ,Franzin A ,et al. A Bayesian Network analysis of the probabilistic relations between risk factors in the predisposition to type 2 diabetes ,United States 2015.
 - [26] Demissie K ,White N ,Joseph L ,et al. Bayesian Estimation of Asthma Prevalence and Comparison of Exercise and Questionnaire Diagnostics in the Absence of a Gold Standard. Annals of Epidemiology , 1998 ,8(3) : 201-208.
 - [27] Chaves R ,Ramírez J ,Górriz JM ,et al. FDG and PIB Biomarker PET Analysis for the Alzheimer's Disease Detection Using Association Rules: 2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC) 2012.
 - [28] 郭玉婷 ,刘豫 ,张丽丽 ,等. 基于 Apriori 算法的慢性阻塞性肺疾病超限住院费用关联规则数据挖掘. 中国慢性病预防与控制 2017 ,25(4) : 245-248.
 - [29] 任仙龙 ,胡冬梅 ,王文娟 ,等. 关联规则在社区居民慢性病患者率分析中的应用. 中国卫生统计 2013 ,30(6) : 818-820.
 - [30] 孙丽 ,王继伟 ,余金明. 门诊高血压患者血压控制及其影响因素的聚类分析. 中华心血管病杂志 2015 ,43(1) : 62-67.
 - [31] Shabbakhti M ,Taherifar D ,Zareei Z. Combination Of PCA And SVM For Diagnosis Of Parkinson's Disease. 2013 2nd International Conference on Advances in Biomedical Engineering 2013.
 - [32] Dasarathy BV ,Sheela BV. A composite classifier system design: Concepts and methodology. Proceedings of the IEEE ,1979 ,67(5) : 708-713.
 - [33] 徐继伟 ,杨云. 集成学习方法: 研究综述. 云南大学学报(自然科学版) 2018 ,40(6) : 1082-1092.
 - [34] Savio A ,García-Sebastián M ,Graña M ,et al. Results of an Adaboost approach on Alzheimer's Disease detection on MRI. Proceeding IWI-NAC '09 Proceedings of the 3rd International Work-Conference on The Interplay Between Natural and Artificial Computation: Part II. Bioinspired Applications in Artificial and Natural Computation , 2009 ,5602: 114-123.
 - [35] Chen P ,Pan C. Diabetes classification model based on boosting algorithms. BMC Bioinformatics 2018 ,19(1) : 109-117.
 - [36] Sreng S ,Maneerat N ,Hamamoto K ,et al. Automated Diabetic Retinopathy Screening System Using Hybrid Simulated Annealing and Ensemble Bagging Classifier. Applied Sciences 2018 ,8(7) : 1198.
 - [37] 于大海 ,李金 ,罗艳虹 ,等. 随机森林模型和决策树模型在肝硬化上消化道出血预后中的应用. 中国卫生统计 2019 ,36(2) : 162-166.
 - [38] 鲁莹 ,郝少智. Stacking 学习与一般集成方法的比较研究. 中国科技论文在线精品论文 2018 ,11(4) : 372-379.
 - [39] 吴健 ,周立水 ,顾盼 ,等. 一种基于医疗体检数据的糖尿病预测方法. 中国 ,CN106682412A. 2017-5-17.
 - [40] Hinton GE ,Salakhutdinov RR. Reducing the Dimensionality of Data with Neural Networks. Science 2006 ,313(5786) : 504-507.
 - [41] LeCun Y ,Bengio Y ,Hinton G. Deep learning. Nature ,2015 ,521(7553) : 436-444.
 - [42] Esteva A ,Robicquet A ,Ramsundar B ,et al. A guide to deep learning in healthcare. Nature Medicine 2019 ,25(1) : 24-29.
 - [43] Mamoshina P ,Kochetov K ,Cortese F ,et al. Blood Biochemistry Analysis to Detect Smoking Status and Quantify Accelerated Aging in Smokers. Scientific Reports 2019 ,9(1) : 142-151.
 - [44] LeCun Y ,Kavukcuoglu K ,Farabet C. Convolutional Networks and Applications in Vision. Proceedings of 2010 IEEE International Symposium on Circuits and Systems 2010.
 - [45] 李琼 ,柏正尧 ,刘莹芳. 糖尿病性视网膜膜图像的深度学习分类方法. 中国图象图形学报 2018 ,23(10) : 1594-1603.
 - [46] Ting DSW ,Cheung CY ,Lim G ,et al. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. JAMA 2017 ,318(22) : 2211-2223.
 - [47] Poplin R ,Varadarajan AV ,Blumer K ,et al. Predicting Cardiovascular Risk Factors from Retinal Fundus Photographs using Deep Learning. Nature Biomedical Engineering 2018(2) : 158-164.
 - [48] Shickel B ,Tighe PJ ,Bihorac A ,et al. Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. IEEE Journal of Biomedical and Health Informatics 2018 ,22(5) : 1589-1604.

(责任编辑: 刘 壮)