

· 综述 ·

机器学习在脑卒中风险预测中的应用进展

万红燕¹ 郝舒欣² 刘婕² 刘悦²¹东南大学附属中大医院江北院区介入与血管外科, 南京 210048; ²中国疾病预防控制中心环境与健康相关产品安全所, 北京 100021

通信作者: 刘悦, Email: liuyue@nieh.chinacdc.cn

【摘要】 脑卒中临床表现多样, 病因复杂, 具有高发病率、高致残率、高复发率、高病死率和高经济负担五大特征。目前传统临床诊疗方法由于人力、时间等限制, 存在患病和预后预测困难、诊断精确度低、治疗缓慢等问题。随着人工智能领域的深入研究和在医疗领域的应用, 利用机器学习模型不仅能够较为准确地进行脑卒中预测和诊断, 还可以识别危险因素, 确定高危人群。本研究综述了机器学习算法的研究现状、脑卒中的危险因素识别和脑卒中预测常见的机器学习算法及在脑卒中风险预测中的研究现状和脑卒中风险预测的效果, 为早期识别高危人群、采取有效的预防措施以及制定精确的治疗方案提供科学依据。

【关键词】 卒中; 机器学习; 预测; 危险因素; 决策树; 支持向量机; 神经网络; 诊断; 综述

基金项目: 东南大学附属中大医院护理科研基金项目 (KJZC-HL-202201)

Advances in the application of machine learning for stroke risk prediction

Wan Hongyan¹, Hao Shuxin², Liu Jie², Liu Yue²¹Department of Interventional and Vascular Surgery, Jiangbei Branch, Zhongda Hospital, Southeast University, Nanjing 210048, Jiangsu Province, China; ²National Institute of Environmental Health, Chinese Center for Disease Control and Prevention, Beijing 100021, China

Corresponding author: Liu Yue, Email: liuyue@nieh.chinacdc.cn

【Abstract】 Stroke has diverse clinical manifestations and complex causes, characterized by high incidence, high disability rate, high recurrence rate, high mortality rate, and high economic burden. Currently, conventional clinical diagnostic and treatment methods face challenges such as difficulty in predicting disease and prognosis, low diagnostic accuracy, and slow treatment due to limitations in manpower and time. With the in-depth research in artificial intelligence and its application in the medical field, machine learning models can not only predict and diagnose stroke more accurately but also identify risk factors and determine high-risk populations. This paper reviews the current research status of machine learning algorithms, the identification of stroke risk factors, common machine learning algorithms for stroke prediction, and the effectiveness of these algorithms in stroke risk prediction. Findings from this paper will help provide a scientific basis for the early identification of high-risk populations, the adoption of effective preventive measures, and the formulation of precise treatment plans.

【Key words】 Stroke; Machine learning; Forecasting; Risk factors; Decision trees; Support vector machine; Nerve Net; Diagnosis; Review

【Fund program】 Nursing Research Fund Project of Zhongda Hospital of Southeast University (KJZC-HL-202201)

脑卒中是一种脑血管的急性神经系统疾病, 可分为缺血性卒中(最常见, 约 85% 的病例发生)或出血性卒中^[1]。

根据世界卫生组织(WHO)最新报告, 全球每年约有 1 500 万人患脑卒中^[2], 并已成为全球第二大死因, 占全世界总死亡人

DOI: 10.3760/cma.j.cn341190-20231120-00462

收稿日期 2023-11-20 本文编辑 朱鹏

引用本文: 万红燕, 郝舒欣, 刘婕, 等. 机器学习在脑卒中风险预测中的应用进展[J]. 中国基层医药, 2024, 31(8): 1275-1280. DOI: 10.3760/cma.j.cn341190-20231120-00462.



数的 10%, 占有残疾调整寿命的 5%, 对人口健康构成严重威胁, 特别是在低收入或中等收入的发展中国家^[3]。在中国, 脑卒中发病率的年均增速已达到 8.7%, 对公众健康构成了极大的威胁, 并且是导致全国成年人死亡和伤残的主要原因^[4]。临床研究表明, 前期预防和干预, 早发现早处理能减少大脑受影响区域的进一步损伤和身体其他部位的并发症, 且越早干预效果越好, 可明显降低发病率、致残率和病死率。因此, 早期诊断对脑卒中的预后尤为重要^[5-6]。

随着大数据、人工智能等技术的发展和深入应用, 机器学习在医疗领域展现了无法比拟的优势, 可以提供快速准确的预测结果^[7], 基于机器学习方法构建脑卒中风险预测模型也成为近年来的研究热点。本综述通过文献检索汇总了机器学习在脑卒中风险预测中的相关研究, 为及早识别高危人群和危险因素, 采取有效的预防措施及制定精准的干预方法提供依据和参考。

1 机器学习算法的研究现状

机器学习是基于海量数据以及并行计算能力, 让机器自主模拟人类学习的过程, 通过不断从数据中学习和自我优化从而做出智能决策, 是人工智能领域中的一个分支^[8-9]。目前, 机器学习在许多领域广泛应用, 尤其是在医疗领域给医学实践带来了革命性的影响^[10]。通过文献检索, 脑卒中机器学习算法始于 2005 年^[11], 在此之前, 脑卒中风险预测主要以逻辑回归、LASSO 回归、Cox 比例风险回归模型^[12-13]等传统统计学模型为主, 但传统统计模型存在对数据质量要求较高、对缺失值和高维数据不友好; 海量数据导致模型效率低下; 无法有效探索数据中的复杂非线性关系以及无法在数据存在复杂相互作用下进行评估^[14]等问题。机器学习结合统计学和计算机科学的特点和优势, 以海量、高维数据驱动模型学习复杂结构, 通过多种算法进行预测^[7], 在数据特征提取及数据处理等方面明显优于传统模型^[15]。

机器学习主要包括监督学习、无监督学习、半监督学习、深度学习、强化学习及迁移学习。脑卒中风险预测常用决策树、支持向量机、朴素贝叶斯、K-相邻、随机森林、神经网络等^[2, 16]模型, 通过灵敏度、特异度、准确率、受试者工作特征曲线下面积等^[17]对模型进行评价。

1.1 支持向量机 支持向量机, 监督式模型, 在高维空间里建立最大间隔的超平面, 将线性算法与线性或非线性核函数相结合, 从而获得一个对分类结果泛化能力较强的线性方程^[18]。支持向量机能够很好地处理高维、非线性、小样本等复杂数据, 可避免过拟合风险。但训练大样本数据繁琐、耗资源、效率低, 分类较多时会不可分类或分类重叠^[14]。支持向量机对二分类、线性可分和非线性可分、小样本训练和高维数据具有较好的泛化性和鲁棒性。

1.2 随机森林 随机森林, 集成许多独立的决策树, 决策树之间没有关联, 以随机方式在决策树的基础上建立一个森林, 多棵决策树的结果决定随机森林的结果。随机森林无须降维就能够准确处理大量级、高维离散或连续数据引入随机性可避免过拟合; 训练速度快, 可对变量进行重要性排序。随机森林在处理大型数据集或包含大量数据特征的数据集, 计算成本可能会很高。同时随机森林模型中的每个决策树都是相互独立的, 它的解释性较差。

1.3 朴素贝叶斯 将特征词的关系视为可重复、可预测、可计算的机制, 推断出某个特定情况下的结果, 并根据结果计算出 X 的后验概率, X 的值越小, Y 值越高^[19]。计算速度快、对高维和小样本数据适应性较好。但应用模型需先了解先验概率, 模型与实际情况存在差异时准确性容易受影响。

1.4 神经网络 神经网络, 监督学习模型, 模拟人脑神经系统对复杂信息处理机制, 包含输入层、隐藏层和输出层, 每个输入值都有权重, 隐藏层的神经元先开始计算, 将输入的信息转换成有效的结果后将其传输到输出层。可处理大规模标注数据, 不断训练节点和隐含层, 改变算法特征, 从而使预测更加准确^[20]。但需要大样本数据训练模型; 大规模数据训练时需要较多的计算资源和时间; 内部决策过程较难解释和理解。

1.5 决策树 决策树, 根据某种准则(如信息增益、基尼指数等), 选择最佳特征作为树的根节点, 然后根据根节点特征取值, 将数据划分为多个子集, 每个子集逐步递归, 构建决策树的分支。模型结构直观、规则清晰明确, 易于理解和解释; 可处理离散和连续型数据, 也能处理多分类和回归问题; 异常和缺失值具有一定的鲁棒性。但容易生成复杂的模型, 造成过拟合; 数据微小变化可能导致完全不同的决策树模型; 不适于处理高维稀疏数据。

1.6 线性回归 线性回归, 经典的线性统计模型, 通过假设函数、损失函数、参数估计, 描述自变量与因变量之间的线性关系。应用广泛, 实现简单; 可解释性好, 预测结果较准确。但处理复杂交互特征数据和结果不准确; 非线性数据拟合效果较差。

1.7 逻辑回归 逻辑回归, 常用的分类算法模型, 将样本分为不同类别, 通过假设函数、损失函数、参数估计建立预测模型。模型简单, 容易理解和解释; 计算效率较高, 特别是处理大量级数据时表现良好; 适用广泛, 既可于二分类问题也可扩展到多分类问题。但非线性数据拟合能力有限; 异常值和离群点较敏感, 可能会影响模型的性能和鲁棒性。

1.8 K 近邻 K 近邻, 常用的分类和回归算法, 用于根据样本特征将其分为不同类别或预测其数值。将训练样本的特征和对应的类别或数值保存起来, 构建一个特征空间。对于一个新的样本, 计算其与训练样本特征之间的距离(如欧氏距离、曼哈顿距离等)。根据距离的大小, 选择与新样本最近的 K 个训练样本。分类问题, 根据 K 个最近邻的类别, 通过投票或权重计算确定新样本的类别。回归问题, 根据 K 个最近邻的数值, 通过平均或加权平均计算确定新样本的数值。无须显示训练过程, 只需保存训练样本的特征和类别或数值; 可以处理离散和连续型的数据, 可用于多分类和回归问题; 对异常值和噪声影响较小, 可较好地处理不规则数据。但不合适的 K 值可能导致模型性能下降; 要保存所有训练样本的特征和类别或数值, 对存储空间要求较高。

1.9 K 均值 K 均值, 聚类算法, 将样本数据划分为 K 个不同的簇(聚类)。随机选择 K 个数据样本作为初始的簇中

心,分配样本,对每个样本计算其与各个簇中心之间的距离,将样本分配到距离最近的簇中心所属的簇,更新簇中心,对于每个簇计算其所有样本的均值,并将该均值作为新簇中心,反复迭代执行分配样本和更新簇中心,直到簇中心不再发生明显变化或达到预定的迭代次数。能够处理较大规模数据,相对较快地收敛到局部最优解。但K均值对初始簇中心选择比较敏感;对离群点和噪声数据较为敏感。

1.10 极端梯度提升 极端梯度提升,基于梯度提升树的集成学习算法,广泛应用于分类和回归问题。设置初始预测值,然后迭代生成决策树,通过优化目标函数(如平方损失和对数损失)求解每个叶子节点的最佳分数,更新模型预测值,最后将每棵树的预测值与初始预测值相加,得到新的模型预测值,反复迭代生成决策树,直到达到预定的迭代次数或满足停止条件,最后将所有决策树的预测值相加,得到最终预测结果。能够准确地进行分类和回归预测;能够较好地处理大规模数据;有较强的可解释性;支持多种损失函数。但参数调优较为繁琐,容易过于拟合;大规模数据训练时间较长。

1.11 集成学习 集成学习,构建一个组合多个基学习器的集成模型,解决单一机器学习模型的问题以提高整体性能,包括 Hardvoting、soft-voting 和 Stacking 等^[21-22]。从原始数据集中构建多个基学习器,分别对每个基学习器进行独立的训练,将多个基学习器的预测结果进行组合,最后生成最终的集成模型。能够有效减少过拟合,提高模型的泛化性能;能够减少模型因为数据变动而产生的波动提高模型的稳定性;减少个体学习器的错误;对于复杂的机器学习问题取得较好的效果。但需要训练和组合多个基学习器会增加训练复杂度和计算成本;对原始数据集质量要求较高。

2 机器学习在脑卒中风险预测中的应用

2.1 脑卒中相关危险因素的识别 脑卒中病例特征复杂,影响因素众多,发病的病因涉及遗传因素和环境暴露、生活方式等影响因素^[3,23-26]。国内外相关研究已证实脑卒中与多种暴露因素有关。孙勇等^[20]研究中随机森林模型及SHAP预测值提示,高龄、收缩压、糖尿病、脂质核心体积、有无糖尿病治疗及壁径比值是新发缺血性脑卒中最重要预测因素。张晓林等^[27]研究中随机森林模型显示,载脂蛋白A、天门冬氨酸氨基转移酶、白蛋白、红细胞压积、糖尿病、乳酸脱氢酶为脑卒中排名前6位的复发影响因素。Emon等^[28]研究中证实性别、年龄、高血压、心脏病、平均血糖水平、BMI和吸烟状态等因素与中风的相关性。Pradeepa等^[29]研究中运用APRIORI算法显示,脑卒中的危险因素从高往低依次为流产和更年期提前等女性生殖因素、肥胖、高血压、肾脏疾病、胆固醇、生活方式、遗传史、高钠的摄入、饮酒、牙龈疾病、艾滋病及炎症。Wu等^[30]研究中运用生存条件推理树方法显示,控制不佳的血压和极低的LDL-C浓度(≤ 40 mg/dL)在脑卒中15个预测因子中排序最高。吴菊华等^[31]研究中运用多层感知器神经网络方法和邵泽国等^[32]研究中运用优化决策树方法均显示高血压、糖尿病、胆固醇、吸烟和饮酒等是脑卒中高危风险因素。Li等^[33]研究中运用决策树Boosting模型显示,脑卒中与饮酒史、家族史、心脏病

史、心脏杂音等具有相关性。Cheon等^[34]使用深度神经网络的方法,利用医疗服务使用和健康行为数据来检测中风并识别卒中危险因素。此外相关研究表明^[35-38],脑卒中也与空气污染、气象因素等多种暴露因素有关。

2.2 脑卒中风险预测 脑卒中诊疗研究领域,机器学习被多项研究证实具有非常大的应用价值^[14,39-40]。常怀文和姚音^[41]研究中运用朴素贝叶斯模型及支持向量机模型构建江西地区缺血性脑卒中风险预测模型,结果表明支持向量机表现最优(AUC:1.0)且具有较高可信度。郭志恒等^[18]研究中运用支持向量机、随机森林和logistics回归等模型对脑卒中患者进行分类研究和脑卒中早期预测,随机森林模型的准确率、精确度、ROC值都优于支持向量机和逻辑回归模型。Jeena等^[42]研究中运用支持向量机预测脑卒中准确率达到91%。Bento等^[43]研究运用支持向量机识别早期颈动脉粥样硬化患者预防缺血性中风的发生,效果显著(AUC:0.91)。Yang等^[44]研究运用支持向量机预测脑卒中的发生,效果显著(AUC:0.91)。Dritsas等^[17]研究中运用了朴素贝叶斯、随机森林、logistics回归、K-相邻、决策树、随机梯度下降等机器学习方法对脑卒中进行预测,结果表明随机森林模型最优且具有较高的预测能力和有效性。Lee等^[45]研究中运用logistic回归、支持向量机、随机森林三种机器学习算法对脑卒中中进行预测,显示随机森林模型最佳(AUC:0.851)。吴菊华等^[31]研究中构建脑卒中神经网络预测模型,在完善脑卒中风险因素识别的同时也取得了较好的预测结果。Alanazi等^[46]研究中运用四种机器学习方法预测脑卒中,结果显示随机森林模型最佳(准确性:0.96、灵敏度:0.97、特异性:0.96、阳性预测值:0.75、阴性预测值:0.99、曲线下面积:0.97)。Colak等^[47]研究中使用知识发现过程方法、支持向量机及人工神经网络模型来预测卒中,结果显示人工神经网络具有更高的预测性能,所提出的神经网络模型在脑卒中临床决策时更具有意义。Cheon等^[34]研究中运用深度学习算法预测脑卒中,效果显著(AUC:0.8以上)。叶伟等^[48]研究利用Stacking集成学习方法将三种强监督学习器进行有效结合,提升了模型准确度和结果的准确性。Hwangbo等^[49]利用集成学习方法预测缺血性脑卒中患者6个月的死亡率(AUC:0.789)。许多研究表明,不同机器学习算法预测模型的预测性能均明显优于传统预测模型^[17,44,50]。使用机器学习方法进行脑卒中风险预测是一个极具前景的研究领域,与传统模型相比,具有诸多优势,如它能够自动、快速地整合和处理多维度大量级数据,并能够快速识别风险因素且能够提示风险因素的重要性程度,最重要的是机器学习方法能够基于更全面、更细致的数据进行分析,使得风险预测和风险评估具有较高的准确性,从而可以进一步帮助医生和研究人员更为精准地预测患者发生脑卒中的可能性,以便帮助医护人员更好地调整防控策略和制定更有效的预防措施。

3 讨论与展望

机器学习能够很好地协助医生做出更好的临床决策,能让脑卒中患者感体验全新的诊治模式,助力患者赢得更好的生活质量和预期寿命^[51]。随着机器学习的不断优化,

在脑卒中的风险管理和研究中,机器学习方法的应用日益广泛,不仅可以进行脑卒中的风险预测,还可以识别和分析脑卒中 CT、MRI、血管造影等大量的影像数据,辅助脑卒中的诊断。此外基于脑卒中患者特定情况的数据,机器学习还可以帮助医生选择最合适的治疗方案,并预测特定治疗方法的效果;在脑卒中的康复方面,机器学习还可以通过分析过往的康复数据,预测患者未来的康复潜力和需要的康复时间。因此,机器学习方法不仅能够为临床医生带来了良好的执业体验,也能为脑卒中患者带来更多的福音。虽然机器学习方法在脑卒中领域有诸多潜在的应用场景,但也要正视其在发展和应用中的局限性。首先,机器学习模型需要大量的数据进行不断训练和优化,数据量大小会影响模型的性能,目前脑卒中病例数据比较分散和封闭,很难实现高效共享,对模型训练造成一定困难;其次,数据质量制约了脑卒中机器学习模型研究的性能和准确性。除了常见的缺失、异常、逻辑错误、重复等数据外,不平衡数据对机器学习模型预测的准确率、精确度、ROC 值影响较大;再次,机器学习模型以数据驱动,模型内部运行方式很难被理解和解释,与传统的经典统计模型相比,可解释性是机器学习领域的一个重要问题。

机器学习属于交叉学科领域,学习曲线比较陡峭,需要医学、统计学、计算机、大数据等跨学科合作,才能取得更好的效果。但从长远角度看,机器学习相比传统统计方法还是具有显著的优势,也逐步成为医学研究的重要组成部分,将来会在脑卒中危险因素识别、早期诊断、预测和防治领域发挥重要作用。

利益冲突 所有作者声明无利益冲突

参 考 文 献

- [1] Bustamante A, Penalba A, Orset C, et al. Blood biomarkers to differentiate ischemic and hemorrhagic strokes[J]. *Neurology*, 2021, 96(15): e1928-e1939. DOI: 10.1212/WNL.00000000000011742.
- [2] Pitchai R, Dappuri B, Pramila PV, et al. An artificial intelligence-based bio-medical stroke prediction and analytical system using a machine learning approach[J]. *Comput Intell Neurosci*, 2022, 2022: 5489084. DOI: 10.1155/2022/5489084.
- [3] Wu SM, Wu B, Liu M, et al. Stroke in China: advances and challenges in epidemiology, prevention, and management[J]. *Lancet Neurol*, 2019, 18(4): 394-405. DOI: 10.1016/S1474-4422(18)30500-3.
- [4] «中国脑卒中防治报告»编写组. «中国脑卒中防治报告 2020»概要[J]. *中国脑血管病杂志*, 2022, 19(2): 136-144. DOI: 10.3969/j.issn.1672-5921.2022.02.011.
The compilation team of the report on stroke prevention and treatment in China. Brief report on stroke prevention and treatment in China, 2020[J]. *Chinese Journal of Cerebrovascular Diseases*, 2022, 19(2): 136-144. DOI: 10.3969/j.issn.1672-5921.2022.02.011.
- [5] «中国脑卒中防治报告»编写组. 我国脑卒中防治仍面临巨大挑战——«中国脑卒中防治报告 2018»概要[J]. *中国循环杂志*, 2019, 34(2): 105-119. DOI: 10.3969/j.issn.1000-3614.2019.02.001.
- [6] 巢宝华, 刘建民, 王伊龙, 等. 中国脑卒中防治: 成就、挑战和应对[J]. *中国循环杂志*, 2019, 34(7): 625-631. DOI: 10.3969/j.issn.1000-3614.2019.07.001.
Chao BH, Liu JM, Wang YL, et al. Stroke prevention and control in China: achievements, challenges and responses[J]. *Chinese Circulation Journal*, 2019, 34(7): 625-631. DOI: 10.3969/j.issn.1000-3614.2019.07.001.
- [7] 李洁洁, 张雁儒, 李昊, 等. 机器学习在脑卒中预测中的研究进展[J]. *河南医学研究*, 2022, 31(20): 3832-3835.
Li JJ, Zhang YR, Li H, et al. Research advances in machine learning for stroke prediction[J]. *Henan Medical Research*, 2022, 31(20): 3832-3835.
- [8] 赵晨阳. 机器学习综述[J]. *数字通信世界*, 2018(1): 109, 112. DOI: 10.3969/J.ISSN.1672-7274.2018.01.081.
Zhao CY. Overview of machine learning[J]. *Digital Communication World*, 2018(1): 109, 112. DOI: 10.3969/J.ISSN.1672-7274.2018.01.081.
- [9] Sirsat MS, Fermé E, Câmara J. Machine learning for brain stroke: a review[J]. *J Stroke Cerebrovasc Dis*, 2020, 29(10): 105162. DOI: 10.1016/j.jstrokecerebrovasdis.2020.105162.
- [10] 林振, 秦宇辰, 秦逸, 等. 深度学习在脑卒中诊疗中的应用[J]. *医疗卫生装备*, 2022, 43(2): 88-92. DOI: 10.19745/j.1003-8868.2022041.
Lin Z, Qin YC, Qin YY, et al. Application of deep learning in diagnosis and treatment of stroke[J]. *Chinese Medical Equipment Journal*, 2022, 43(2): 88-92. DOI: 10.19745/j.1003-8868.2022041.
- [11] Gottrup C, Thomsen K, Lochter P, et al. Applying instance-based techniques to prediction of final outcome in acute stroke[J]. *Artif Intell Med*, 2005, 33(3): 223-236. DOI: 10.1016/j.artmed.2004.06.003.
- [12] 赵楠楠, 郑印, 黄穹琼, 等. 急性缺血性脑卒中早期预后不良的危险因素分析及预测模型构建[J]. *中国临床研究*, 2022, 35(4): 456-461. DOI: 10.13429/j.cnki.cjcr.2022.04.003.
Zhao NN, Zheng Y, Huang QQ, et al. Risk factors analysis and predictive model construction of poor early prognosis in acute ischemic stroke[J]. *Chinese Journal of Clinical Research*, 2022, 35(4): 456-461. DOI: 10.13429/j.cnki.cjcr.2022.04.003.
- [13] 刘幼华, 郭红, 弓少华, 等. 急性缺血性脑卒中复发风险预测模型的构建[J]. *中国老年学杂志*, 2022, 42(11): 2612-2618. DOI: 10.3969/j.issn.1005-9202.2022.11.004.
Liu YH, Guo H, Gong SH, et al. Construction of a predictive model for the risk of recurrence in acute ischemic stroke[J]. *Chinese Journal of Gerontology*, 2022, 42(11): 2612-2618. DOI: 10.3969/j.issn.1005-9202.2022.11.004.
- [14] 武胜勇, 何倩, 郭铁斌, 等. 机器学习算法在脑卒中诊治中的应用现状及展望[J]. *中国卫生统计*, 2021, 38(3): 464-467. DOI: 10.3969/j.issn.1002-3674.2021.03.040.
Wu SY, He Q, Guo YB, et al. Application status and prospect of machine learning algorithm in diagnosis and treatment of stroke[J]. *Chinese Journal of Health Statistics*, 2021, 38(3): 464-467. DOI: 10.3969/j.issn.1002-3674.2021.03.040.
- [15] Abedi V, Goyal N, Tsvigoulis G, et al. Novel screening tool for stroke using artificial neural network[J]. *Stroke*, 2017,



- 48(6): 1678-1681. DOI: 10.1161/STROKEAHA. 117.017033.
- [16] 杨晓静, 张福东, 胡长斌. 机器学习综述[J]. 科技经济市场, 2021(10): 40-42.
Yang XJ, Zhang FD, Hu CB. Overview of machine learning [J]. Science & Technology Economy Market, 2021(10): 40-42.
- [17] Dritsas E, Trigka M. Stroke risk prediction with machine learning techniques[J]. Sensors, 2022, 22(13): 4670. DOI: 10.3390/s22134670.
- [18] 郭志恒, 刘青萍, 刘芳, 等. 基于机器学习算法的脑卒中疾病早期预测模型研究[J]. 计算机与数字工程, 2021, 49(11): 2180-2183, 2247. DOI: 10.3969/j.issn.1672-9722.2021.11.002.
Guo ZH, Liu QP, Liu F, et al. Early prediction model of stroke based on machine learning algorithm[J]. Computer & Digital Engineering, 2021, 49(11): 2180-2183, 2247. DOI: 10.3969/j.issn.1672-9722.2021.11.002.
- [19] Saber H, Somai M, Rajah GB, et al. Predictive analytics and machine learning in stroke and neurovascular medicine [J]. Neurol Res, 2019, 41(8): 681-690. DOI: 10.1080/01616412.2019.1609159.
- [20] 孙勇, 王立强, 王芬, 等. 可解释的机器学习模型用于预测远期脑缺血事件[J]. 心脑血管病防治, 2022, 22(2): 53-56, 60. DOI: 10.3969/j.issn.1009-816x.2022.02.014.
Sun Y, Wang LQ, Wang F, et al. Interpretable machine learning model to predict long-term cerebral ischemic events[J]. Cardio-cerebrovascular Disease Prevention and Treatment, 2022, 22(2): 53-56, 60. DOI: 10.3969/j.issn.1009-816x.2022.02.014.
- [21] Morgan-Benita JA, Galván-Tejada CE, Cruz M, et al. Hard voting ensemble approach for the detection of type 2 diabetes in Mexican population with non-glucose related features[J]. Healthcare, 2022, 10(8): 1362. DOI: 10.3390/healthcare10081362.
- [22] Ali S, Hussain A, Aich S, et al. A soft voting ensemble-based model for the early prediction of idiopathic pulmonary fibrosis (IPF) disease severity in lungs disease patients[J]. Life, 2021, 11(10): 1092. DOI: 10.3390/life11101092.
- [23] 刘洁, 陈璐, 黄小雨, 等. 中性粒细胞/淋巴细胞与脑梗死复发的关系[J]. 医学研究杂志, 2020, 49(6): 61-65. DOI: 10.11969/j.issn.1673-548X.2020.06.014.
Liu J, Chen L, Huang XY, et al. Relevant research between neutrophils/lymphocytes and relapse of cerebral infarction[J]. Journal of Medical Research, 2020, 49(6): 61-65. DOI: 10.11969/j.issn.1673-548X.2020.06.014.
- [24] 郝天依, 韩素芹, 姚青, 等. 天津市脑卒中高住院风险的气象原因探析[J]. 气象学报, 2022, 80(3): 392-402. DOI: 10.11676/qxxb2022.041.
Hao TY, Han SQ, Yao Q, et al. Meteorological reasons for high hospitalization risk of stroke in Tianjin[J]. Acta Meteorologica Sinica, 2022, 80(3): 392-402. DOI: 10.11676/qxxb2022.041.
- [25] Verhoeven JJ, Allach Y, Vaartjes ICH, et al. Ambient air pollution and the risk of ischaemic and haemorrhagic stroke[J]. Lancet Planet Health, 2021, 5(8): e542-e552. DOI: 10.1016/S2542-5196(21)00145-5.
- [26] 刘婷, 朱琴, 徐琳, 等. 基于机器学习的原发性高血压并发脑梗死的风险预测模型[J]. 海军军医大学学报, 2022, 43(3): 258-265. DOI: 10.16781/j.cn31-2187/R.20211053.
Liu T, Zhu Q, Xu L, et al. Risk prediction models of essential hypertension complicated with cerebral infarction based on machine learning algorithm[J]. Academic Journal of Naval Medical University, 2022, 43(3): 258-265. DOI: 10.16781/j.cn31-2187/R.20211053.
- [27] 张晓林, 彭晨, 殷淑娟, 等. 基于随机森林算法对首发缺血性脑卒中患者出院 90 d 复发的预测作用评价[J]. 当代医学, 2021, 27(14): 1-4. DOI: 10.3969/j.issn.1009-4393.2021.14.001.
Zhang XL, Peng C, Yin SJ, et al. Evaluation of the predictive effect of randomized forest algorithm on the recurrence of first ischemic stroke at 90days after discharge from hospital[J]. Contemporary Medicine, 2021, 27(14): 1-4. DOI: 10.3969/j.issn.1009-4393.2021.14.001.
- [28] Emon MU, Keya MS, Meghla TI, et al. Performance analysis of machine learning approaches in stroke prediction[C]// 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA). November 5-7, 2020. Coimbatore, India. IEEE, 2020-01-01. DOI: 10.1109/iceca49313.2020.9297525.
- [29] Pradeepa S, Manjula KR, Vimal S, et al. DRFS: detecting risk factor of stroke disease from social media using machine learning techniques[J]. Neural Process Lett, 2023, 55(4): 3843-3861. DOI: 10.1007/s11063-020-10279-8.
- [30] Wu ZJ, Huang Z, Lichtenstein AH, et al. The risk of ischemic stroke and hemorrhagic stroke in Chinese adults with low-density lipoprotein cholesterol concentrations < 70 mg/dL[J]. BMC Med, 2021, 19(1): 142. DOI: 10.1186/s12916-021-02014-4.
- [31] 吴菊华, 张烁, 陶雷, 等. 基于神经网络的脑卒中风险预测模型研究[J]. 数据分析与知识发现, 2019, 3(12): 70-75. DOI: 10.11925/infotech.2096-3467.2019.0691.
Wu JH, Zhang S, Tao L, et al. Predicting stroke risks with neural network[J]. Data Analysis and Knowledge Discovery, 2019, 3(12): 70-75. DOI: 10.11925/infotech.2096-3467.2019.0691.
- [32] 邵泽国, 陈晨, 陈炜. 基于优化决策树的脑卒中日常生活习惯风险因素分析[J]. 现代预防医学, 2018, 45(15): 2689-2693.
Shao ZG, Chen C, Chen W. Risk factors in the daily life of stroke patients based on an optimized decision tree[J]. Modern Preventive Medicine, 2018, 45(15): 2689-2693.
- [33] Li XM, Bian D, Yu JH, et al. Using machine learning models to improve stroke risk level classification methods of China national stroke screening[J]. BMC Med Inform Decis Mak, 2019, 19(1): 261. DOI: 10.1186/s12911-019-0998-2.
- [34] Cheon S, Kim J, Lim J. The use of deep learning to predict stroke patient mortality[J]. Int J Environ Res Public Health, 2019, 16(11): 1876. DOI: 10.3390/ijerph16111876.
- [35] Tang CZ, Chen Y, Song QL, et al. Short-term exposure to air pollution and occurrence of emergency stroke in Chongqing, China[J]. Int Arch Occup Environ Health, 2021, 94(1): 69-76. DOI: 10.1007/s00420-020-01557-y.
- [36] Chu SY, Cox M, Fonarow GC, et al. Temperature and precipitation associate with ischemic stroke outcomes in the United States[J]. J Am Heart Assoc, 2018, 7(22): e010020. DOI: 10.1161/JAHA.118.010020.
- [37] Kurtz P, Bastos LS, Aguilar S, et al. Effect of seasonal and temperature variation on hospitalizations for stroke over a 10-year period in Brazil[J]. Int J Stroke, 2021, 16(4): 406-410. DOI: 10.1177/1747493020947333.
- [38] 陈念, 王春茹. 气象环境因素与脑卒中发病的相关性研究[J]. 医学信息, 2019, 32(16): 34-37. DOI: 10.3969/j.issn.1006-1959.2019.16.011.
Chen N, Wang CR. Research of correlation between meteorological environmental factors and the onset of stroke[J]. Medical Information, 2019, 32(16): 34-37. DOI:



- 10.3969/j.issn.1006-1959.2019.16.011.
- [39] 李昇霖, 林晓强, 邓娟, 等. 影像机器学习方法在急性缺血性脑卒中的研究进展[J]. 临床放射学杂志, 2021, 40(4): 822-824. DOI: 10.13437/j.cnki.jcr.2021.04.044.
- Li SL, Lin XQ, Deng J, et al. Research progress of image machine learning method in acute ischemic stroke[J]. Journal of Clinical Radiology, 2021, 40(4): 822-824. DOI: 10.13437/j.cnki.jcr.2021.04.044.
- [40] 傅璠, 卢洁. 人工智能在缺血性脑卒中影像的研究进展[J]. 中华老年心脑血管病杂志, 2020, 22(10): 1110-1112. DOI: 10.3969/j.issn.1009-0126.2020.10.028.
- Fu F, Lu J. Research progress of artificial intelligence in ischemic stroke imaging[J]. Chinese Journal of Geriatric Heart Brain and Vessel Diseases, 2020, 22(10): 1110-1112. DOI: 10.3969/j.issn.1009-0126.2020.10.028.
- [41] 常怀文, 姚音. 基于机器学习构建江西地区缺血性脑卒中风险预测模型[J]. 西部医学, 2022, 34(8): 1182-1186. DOI: 10.3969/j.issn.1672-3511.2022.08.017.
- Chang HW, Yao Y. Constructing risk prediction model of ischemic stroke in Jiangxi based on machine learning[J]. Medical Journal of West China, 2022, 34(8): 1182-1186. DOI: 10.3969/j.issn.1672-3511.2022.08.017.
- [42] Jeena R, Kumar S. Stroke prediction using SVM[C]. 2016. IEEE, 2016-01-01. DOI: 10.1109/iccicct.2016.7988020.
- [43] Bento M, Souza R, Salluzzi M, et al. Automatic identification of atherosclerosis subjects in a heterogeneous MR brain imaging data set[J]. Magn Reson Imaging, 2019, 62: 18-27. DOI: 10.1016/j.mri. 2019. 06.007.
- [44] Yang YJ, Zheng J, Du ZZ, et al. Accurate prediction of stroke for hypertensive patients based on medical big data and machine learning algorithms: retrospective study[J]. JMIR Med Inform, 2021, 9(11): e30277. DOI: 10.2196/30277.
- [45] Lee H, Lee EJ, Ham S, et al. Machine learning approach to identify stroke within 4.5 hours[J]. Stroke, 2020, 51(3): 860-866. DOI: 10.1161/strokeaha.119.027611.
- [46] Alanazi EM, Abdou A, Luo J. Predicting risk of stroke from lab tests using machine learning algorithms: development and evaluation of prediction models[J]. JMIR Form Res, 2021, 5(12): e23440. DOI: 10.2196/23440.
- [47] Colak C, Karaman E, Turtay MG. Application of knowledge discovery process on the prediction of stroke[J]. Comput Methods Programs Biomed, 2015, 119(3): 181-185. DOI: 10.1016/j.cmpb.2015.03.002.
- [48] 叶伟, 陶永军, 陈锡程, 等. 脑卒中多分类预后预测的深度集成优化方法[J]. 计算机工程与应用, 2023, 59(5): 95-105. DOI: 10.3778/j.issn.1002-8331.2211-0077.
- Ye W, Tao YJ, Chen XC, et al. Deep ensemble evolutionary multi-classification method for predicting prognosis of stroke[J]. Computer Engineering and Applications, 2023, 59(5): 95-105. DOI: 10.3778/j. issn. 1002-8331. 2211-0077.
- [49] Hwangbo L, Kang YJ, Kwon H, et al. Stacking ensemble learning model to predict 6-month mortality in ischemic stroke patients[J]. Sci Rep, 2022, 12(1): 17389. DOI: 10.1038/s41598-022-22323-9.
- [50] Chun M, Clarke R, Cairns BJ, et al. Stroke risk prediction using machine learning: a prospective cohort study of 0.5 million Chinese adults[J]. J Am Med Inform Assoc, 2021, 28(8): 1719-1727. DOI: 10.1093/jamia/ocab068.
- [51] Ho KC, Speier W, Zhang HY, et al. A machine learning approach for classifying ischemic stroke onset time from imaging[J]. IEEE Trans Med Imaging, 2019, 38(7): 1666-1676. DOI: 10.1109/TMI.2019.2901445.

中華醫學會

