

· 综述 ·

DOI: 10.12449/JCH250121

机器学习在慢性丙型肝炎诊疗中的应用

韩 华¹, 段钟平^{2a}, 王 扬^{2b}

1 华北电力大学医院内科, 北京 102206

2 首都医科大学附属北京佑安医院 a. 肝病四科, b. 肿瘤内科, 北京 100069

通信作者: 王扬, wangyangdoc@126.com (ORCID: 0000-0002-7631-1660)

摘要: 随着人工智能技术的发展,机器学习在医疗健康领域中展现出巨大的应用潜力。机器学习通过对患者的临床特征、血液检验、影像学检查等数据进行综合分析,建立相应的数学模型,以实现疾病的诊断、治疗及病情评估的预测,指导疾病的管理。本文结合最新的研究成果,综述了机器学习在慢性丙型肝炎中的应用情况及研究进展。

关键词: 丙型肝炎, 慢性; 机器学习; 诊断; 治疗学

基金项目: 首都医科大学附属北京佑安医院2022年度院内中青年人才孵育项目(BJYAYY-YN2022-08)

Application of machine learning in the diagnosis and treatment of chronic hepatitis C

HAN Hua¹, DUAN Zhongping^{2a}, WANG Yang^{2b}

1. Department of Internal Medicine, North China Electric Power University Hospital, Beijing 102206, China; 2. a. Fourth Department of Liver Diseases, b. Department of Medical Oncology, Beijing YouAn Hospital, Capital Medical University, Beijing 100069, China

Corresponding author: WANG Yang, wangyangdoc@126.com (ORCID: 0000-0002-7631-1660)

Abstract: With the development of artificial intelligence, machine learning has shown great potential in the field of medical health. Machine learning conducts a comprehensive analysis of patient data including clinical features, blood tests, and imaging examinations and establishes corresponding mathematical models to achieve the diagnosis and treatment of diseases and the prediction of disease conditions, thereby guiding disease management. With reference to the latest research findings, this article reviews the application of machine learning in chronic hepatitis C and related research advances.

Key words: Hepatitis C, Chronic; Machine Learning; Diagnosis; Therapeutics

Research funding: Scientific Research Project of Beijing YouAn Hospital, CCMU, 2022 (BJYAYY-YN2022-08)

目前,全球约有1.13亿HCV感染者,其中70%~80%的急性HCV感染者会进展为慢性携带者,10%~20%的慢性HCV感染者会出现肝硬化、肝功能衰竭或肝癌^[1]。HCV感染具有隐匿性,发展速度慢^[2],导致诊断率不足20%。传统的HCV诊断依赖于血清学检测和聚合酶链式反应技术,早期感染预警存在假阴性可能。尽管直接抗病毒药物(direct antiviral agent, DAA)在治疗上极大地提高了治愈率,但目前丙型肝炎的早期诊断率低,导致患者知晓、接受治疗的比例偏低。随着机器学

习对医学的赋能,利用大数据和高级算法,可以在疾病的早期诊断、疗效评估及预后等方面协助长期管理。本文旨在概述机器学习技术在慢性丙型肝炎诊断、治疗和预后中应用的研究进展。

1 机器学习概述

机器学习作为人工智能的一种形式,是使计算机具备学习能力的技术,通过数据学习和算法对未出现的情况作出预测^[3]。根据训练方法的模式,机器学习分为无

监督式学习、半监督式学习和强化学习^[4]。无监督式学习的一大优势在于其能够处理大量的非结构化数据,如文本、图像等,而且能够发现数据内的潜在结构和规律,这对于数据预处理、特征提取以及异常检测等具有重大意义,其常用的算法有:k-均值法、主成分分析等。监督式学习适合处理带注释的数据,通过带标签的训练数据来训练模型,然后用此模型来预测未知数据,多用于图像识别、语音识别等领域,其常用的算法包括:线性回归算法、逻辑回归算法、决策树、支持向量机、随机森林、朴素贝叶斯分类和梯度增强等^[3]。随机森林和梯度增强是两种基于决策树的集成统计方法,可以构建分类和回归预测模型。半监督式学习是介于监督式学习和无监督式学习之间的一种方法,其利用大量的未标记数据和少量的标记数据进行学习,降低了人工标注的成本,同时提高模型的泛化能力。不同于上述三种主要依赖于数据驱动的方法,强化学习侧重于智能体与环境的交互以学习最优的行为策略,无需大量标注数据,而是通过试错和接收环境反馈来逐渐优化决策过程。

深度学习则是机器学习的一个子集,其通过深度神经网络来模拟人类大脑的认知功能,构建复杂的神经网络结构以实现更高级的学习任务。深度学习模型通常包含更多层次,能够自动提取更加抽象的特征,从而提供更准确的预测和分类。机器学习方法可以优化深度学习模型的性能,深度学习也扩展了机器学习的范畴,使其能够解决更为困难的问题,例如语音和图像识别。

不同于传统统计方法只能处理线性数据,机器学习可以处理线性和非线性信息,并识别大数据中变量和结果之间的隐藏关系^[5]。机器学习具有以下优势:(1)高度的数据适应性,通过机器学习模型可以捕捉到数据中的非线性关系和高维特征;(2)能够处理大规模数据集,便于作出精准的预测;(3)通过集成学习和深度学习等技术实现复杂任务的自动化,从而节省时间和资源。近年来,机器学习已逐渐用于临床的决策,包括诊断、治疗和预后的评估等。

2 机器学习在慢性丙型肝炎诊断中的应用

酶联免疫吸附测定法虽然具有较高的敏感性,但在感染初期可能会出现假阴性结果。利用机器学习技术,分析电子健康记录、问卷预测、临床实验室检测等数据,创建HCV的诊断模型,可以提高丙型肝炎的早期诊断率,并降低其传播风险,有助于缩短慢性感染未经治疗的时间,进而降低进展风险。

Doyle等^[6]通过分析美国约1 000万患者的处方数据和纵向医疗索赔等电子健康记录数据,开发了预测模型,来识别未确诊的HCV患者。该研究从患者的病史中提取了与HCV相关的人口统计学信息、危险因素、症状、治疗等特征,基于逻辑回归、随机森林、梯度增强和集成等算法,从284个特征中确定了46个HCV感染最重要的危险因素。该研究表明,被诊断为HCV的患者在诊断前平均2~3年表现出已知的HCV症状。当召回率<10%时,所有算法的准确率至少为95%;当召回率>50%时,集成算法表现最好,准确率为97%[受试者工作特征曲线下面积(AUC)为0.96],而梯度增强树的精度为87%,逻辑回归的精度仅为31%。Elshevey等^[7]建立了hyOPTGB模型,利用梯度增强模型和OPTUNA超参数调整来预测埃及的HCV相关疾病,同时使用前向选择法来识别数据集中的基本特征,采用Min-Max归一化预处理技术将数据集的值缩放到固定范围,准确率可达95.3%。Reiser等^[8]通过Kohonen人工神经网络分析了180万参保者的社会医学数据,建立了识别早期HCV感染的自组织图谱。该网络使用从2 544例确诊HCV患者亚组中获得的变量进行训练,排除了与HCV诊断直接相关的变量,并将训练结果三维可视化,从而在地图中探索聚类的分布与特征。这种人工神经网络方法可以更有效地筛查HCV感染,但预测模型还需要额外的研究进一步验证。

对于无症状人群,可以通过针对性的问卷数据识别HCV感染。Butaru等^[9]使用12项问卷建立了用来预测HCV感染的深度神经网络(deep artificial neural network, DANN),DANN预测器与第三代快速HCV抗体测试报告的特异度和阴性预测值分别为99.7% vs 99.7%和99.2% vs 99.99%,可以将其推广为资源有限的高流行环境中排除HCV感染的快速低成本工具。

基于常规血液检测数据,利用机器学习来预测丙型肝炎,可以提高患者的确诊率和接受治疗的比例。一项基于患者血液检查的HCV发病分类概率自动分类器研究^[10],通过结合随机森林和逻辑回归的级联两阶段混合算法,再利用人工蜂群算法确定滤波分离所需的最优阈值,建立了HCV检测模型,准确率最高达94.5%。Alizargar等^[11]利用美国国家卫生与营养检查调查数据库,比较了各种机器学习算法在预测丙型肝炎方面的性能,支持向量机和XGBoost技术具有较高的准确率(AUC>80%),可以作为基于血液检测数据预测丙型肝炎的有效工具。该研究还确定了丙氨酸转氨酶、白蛋白、碱性磷酸酶、天

冬氨酸转氨酶、总胆红素、胆固醇、胆碱酯酶、 γ -谷氨酰转移酶、血肌酐、性别和年龄等变量在诊断中的重要性。

3 机器学习在慢性丙型肝炎治疗中的应用

DAA 的出现彻底改变了 HCV 感染的治疗,90% 以上的患者接受治疗 8 或 12 周后,可以实现持续病毒学应答(sustained virologic response, SVR),几乎没有副作用^[12]。Chirikov 等^[13]利用反映肾脏和肝脏功能的 15 个临床变量以及 HCV 特征(RNA 载量、基因型、耐药相关突变)建立了数学模型来预测 SVR,在接受 DAA 治疗的队列中,输入器显示外部马修斯相关系数(Matthews correlation coefficient, MCC)和 F1 评分分别为 0.98 和 0.999, MCC 和 F1 评分都是用于评估二元分类模型性能的常用指标。

DAA 的治疗带来了革命性的变化,固定时间的 DAA 治疗转变为个体化治疗,可以显著节约医疗资源,并为其他病毒感染的抗病毒治疗提供启示。HCV 动力学的数学模型为评估抗病毒治疗的有效性和估计治疗结果,提供了一种有效手段。Churkin 等^[14]使用模拟体内病毒动力学的双方程动力学模型为假设的 HCV 感染者创建了 HCV RNA 载量测量的训练数据集,再用该数据集训练了一个多层感知回归器,以预测 DAA 治疗后病毒转阴的确切时间。将临床研究的患者数据与来自计算机患者的数据集相结合,机器学习算法可以帮助处理关于优化治疗持续时间的过程,有助于准确地预测丙型肝炎的治愈时间,从而通过个体化抗病毒治疗的持续时间来改变丙型肝炎的治疗模式。

Park 等^[15]评估了 4 个机器学习模型的性能,以预测 DAA 治疗的无应答。经典的多变量逻辑回归只能检查 6 个临床和实验室可能的危险因素,而机器学习模型设法调查 179 个人口统计学和临床变量作为治疗失败的候选预测因素。其中,梯度增强是最突出的机器学习算法,但该模型的 MCC 和 F1 评分分别为 0.14 和 0.15。Haga 等^[16]将从 109 例 SVR 和 64 例非 SVR 患者中分离的 HCV 基因组的全长序列纳入 9 个机器学习模型,研究表明,支持向量机是最准确的机器学习预测器,其 MCC 和 F1 评分分别为 0.88 和 0.94,而且在 1 867 个基因变异序列中,支持向量机识别出 81 个导致 DAA 治疗失败的遗传变异。Janczewska 等^[17]也开发了 4 个机器学习模型,通过 14 012 例患者记录的 36 个变量,包括人口统计学特征和临床相关因素,来确定与 DAA 治疗失败相关的因素。除此之外,随机森林分类器达到了最佳的外部性能,准确率和 k 统计量约为 100%。根据该模型,控制治

疗失败的最重要变量是肝功能、病毒载量和基因型、年龄、身体质量指数、血红蛋白以及血清肌酐水平。相反,共感染 HBV 或艾滋病毒、肝外表现和肝细胞癌共存对治疗失败的影响最小。

4 机器学习在慢性丙型肝炎预后方面的应用

肝纤维化分期是丙型肝炎病情管理中评估患者预后的重要组成部分,机器学习技术可以作为预测慢性丙型肝炎肝纤维化风险的非侵入性方法,以减少肝活检。Shousha 等^[18]将数据挖掘策略和白细胞介素 28B 基因分型相结合,使用比天冬氨酸转氨酶与血小板比值和 FIB-4 指数性能更高的神经网络算法多层感知机(multilayer perceptron, MLP)来预测 HCV 患者的晚期纤维化。MLP 属于监督神经网络,是一种通过反向传播学习的多层前馈人工神经网络模型,对 427 例慢性丙型肝炎患者白细胞介素 28B 基因型和生化标志物的回顾性研究显示,MLP 作为最佳分类器来预测晚期纤维化,其敏感度为 0.825,特异度为 0.811, AUC 为 0.880。Hashem 等^[19]通过 39 567 例慢性丙型肝炎患者的血清生物标志物和临床信息建立了分类模型,开发了决策树、多元线性回归、粒子群优化和遗传算法,这些算法的准确率为 66.3%~84.4%, AUC 为 0.73~0.76,能够成功预测慢性丙型肝炎患者的晚期肝纤维化,而且预测中起重要作用的变量为年龄、血小板计数、天冬氨酸转氨酶和白蛋白。在长期抗病毒治疗肝硬化的慢性丙型肝炎患者队列中, Konevman 等^[20]利用纵向临床数据和使用逻辑回归、随机森林以及梯度增强构建的进展模型,预测纤维化进展(增加 ≥ 2 个 Ishak 分期)和肝脏相关死亡(肝失代偿、肝细胞癌、肝移植或 Child-Pugh 评分 ≥ 7),其 AUC 分别为 0.79、0.86 和 0.84。纳入纵向数据的预测模型可以捕捉慢性丙型肝炎的非线性疾病进展,在不影响风险预测准确性的情况下合并大量预测变量,从而克服传统统计分析形式的局限性。该模型能够准确区分高风险和低风险患者,敏感度为 74%,特异度为 78%,阴性预测值为 94%,并且进一步证实对模型预测能力贡献最重要的变量是晚期肝病的纵向实验室标志物,包括血小板计数、天冬氨酸转氨酶与血小板比值和白蛋白的变化。

深度学习是一系列使用人工神经网络的机器学习模型,其使用的循环机制可以帮助捕获数据中的动态信息,还可以处理具有不同随访长度的纵向数据。Ioannou 等^[21]通过使用电子健康记录中提取的原始纵向数据,建立了深度学习递归神经网络模型,而且该模型在预测

HCV 相关肝硬化患者发生肝细胞癌的风险方面优于传统的逻辑回归算法模型。该研究表明,通过神经网络模型计算的肝细胞癌风险评分前 51% 的样本中,包含 80% 的未来 3 年将发展为肝细胞癌的患者,而目标前 66% 的样本包含 90% 将发展为肝细胞癌的患者。

5 展望

早期发现和治疗 HCV 感染对于成功治疗和预防长期并发症至关重要。机器学习技术能够帮助识别 HCV 感染的高风险患者并给予个体化的治疗方案。机器学习在丙型肝炎的应用中具有巨大的潜力,然而这些方法也存在一些局限性,比如对大数据集的依赖、算法偏见、模型解释性的限制,以及临床实践中技术实施的挑战。未来的研究应致力于解决这些问题,并进一步验证这些算法在多样化人群中的有效性和可扩展性。机器学习在丙型肝炎中的应用仍然是一个新兴领域,需更多的跨学科合作,通过不断的方法优化,最终实现更高效的健康服务。

利益冲突声明: 本文不存在任何利益冲突。

作者贡献声明: 王扬负责课题设计,资料分析,拟定写作思路;韩华负责查阅文献,撰写论文,修改论文;段钟平负责指导文章撰写并最后定稿。

参考文献:

- [1] FARRAG AN, KAMEL AM, EL-BARAKY IA. Opportunities and challenges for the application of artificial intelligence paradigms into the management of endemic viral infections: The example of chronic hepatitis C virus[J]. *Rev Med Virol*, 2024, 34(2): e2514. DOI: 10.1002/rmv.2514.
- [2] BRUNNER N, BRUGGMANN P. Trends of the global hepatitis C disease burden: Strategies to achieve elimination[J]. *J Prev Med Public Health*, 2021, 54(4): 251-258. DOI: 10.3961/jpmph.21.151.
- [3] LE BERRE C, SANDBORN WJ, ARIDHI S, et al. Application of artificial intelligence to gastroenterology and hepatology[J]. *Gastroenterology*, 2020, 158(1): 76-94. DOI: 10.1053/j.gastro.2019.08.058.
- [4] NADIF M, ROLE F. Unsupervised and self-supervised deep learning approaches for biomedical text mining[J]. *Brief Bioinform*, 2021, 22(2): 1592-1603. DOI: 10.1093/bib/bbab016.
- [5] HASSABIS D, KUMARAN D, SUMMERFIELD C, et al. Neuroscience-inspired artificial intelligence[J]. *Neuron*, 2017, 95(2): 245-258. DOI: 10.1016/j.neuron.2017.06.011.
- [6] DOYLE OM, LEAVITT N, RIGG JA. Finding undiagnosed patients with hepatitis C infection: An application of artificial intelligence to patient claims data[J]. *Sci Rep*, 2020, 10(1): 10521. DOI: 10.1038/s41598-020-67013-6.
- [7] ELSHEWEY AM, SHAMS MY, TAWFEEK SM, et al. Optimizing HCV disease prediction in Egypt: The hyOPTGB framework[J]. *Diagnostics*, 2023, 13(22): 3439. DOI: 10.3390/diagnostics13223439.
- [8] REISER M, WIEBNER B, HIRSCH J, et al. Neural-network analysis of socio-medical data to identify predictors of undiagnosed hepatitis C virus infections in Germany (DETECT)[J]. *J Transl Med*, 2019, 17(1): 94. DOI: 10.1186/s12967-019-1832-4.
- [9] BUTARU AE, MĂMULEANU M, STREBA CT, et al. Resource management through artificial intelligence in screening programs-key for the successful elimination of hepatitis C[J]. *Diagnostics*, 2022, 12(2): 346. DOI: 10.3390/diagnostics12020346.
- [10] LI TH S, CHIU HJ, KUO PH. Hepatitis C virus detection model by using random forest, logistic-regression and ABC algorithm[J]. *IEEE Access*, 2022, 10: 91045-91058. DOI: 10.1109/ACCESS.2022.3202295.
- [11] ALIZARGAR A, CHANG YL, TAN TH. Performance comparison of machine learning approaches on hepatitis C prediction employing data mining techniques[J]. *Bioengineering (Basel)*, 2023, 10(4): 481. DOI: 10.3390/bioengineering10040481.
- [12] FLAMM S, LAWITZ E, BORG B, et al. Efficacy and safety of sofosbuvir/velpatasvir plus ribavirin in patients with hepatitis C virus-related decompensated cirrhosis[J]. *Viruses*, 2023, 15(10): 2026. DOI: 10.3390/v15102026.
- [13] CHIRIKOV VV, MARX SE, MANTHENA SR, et al. Development of a comprehensive dataset of hepatitis C patients and examination of disease epidemiology in the United States, 2013-2016[J]. *Adv Ther*, 2018, 35(7): 1087-1102. DOI: 10.1007/s12325-018-0721-1.
- [14] CHURKIN A, KRIS S, UZIEL A, et al. Machine learning for mathematical models of HCV kinetics during antiviral therapy[J]. *Math Biosci*, 2022, 343: 108756. DOI: 10.1016/j.mbs.2021.108756.
- [15] PARK H, LO-CIGANIC WH, HUANG J, et al. Machine learning algorithms for predicting direct-acting antiviral treatment failure in chronic hepatitis C: An HCV-TARGET analysis[J]. *Hepatology*, 2022, 76(2): 483-491. DOI: 10.1002/hep.32347.
- [16] HAGA H, SATO H, KOSEKI A, et al. A machine learning-based treatment prediction model using whole genome variants of hepatitis C virus[J]. *PLoS One*, 2020, 15(11): e0242028. DOI: 10.1371/journal.pone.0242028.
- [17] JANCZEWSKA E, KOLEK MF, LORENC B, et al. Factors influencing the failure of interferon-free therapy for chronic hepatitis C: Data from the Polish EpiTer-2 cohort study[J]. *World J Gastroenterol*, 2021, 27(18): 2177-2192. DOI: 10.3748/wjg.v27.i18.2177.
- [18] SHOUSHA HI, AWAD AH, OMRAN DA, et al. Data mining and machine learning algorithms using IL28B genotype and biochemical markers best predicted advanced liver fibrosis in chronic hepatitis C[J]. *Jpn J Infect Dis*, 2018, 71(1): 51-57. DOI: 10.7883/yoken.JJID.2017.089.
- [19] HASHEM S, ESMAT G, ELAKEL W, et al. Comparison of machine learning approaches for prediction of advanced liver fibrosis in chronic hepatitis C patients[J]. *IEEE/ACM Trans Comput Biol Bioinform*, 2018, 15(3): 861-868. DOI: 10.1109/TCBB.2017.2690848.
- [20] KONERMAN MA, ZHANG YW, ZHU J, et al. Improvement of predictive models of risk of disease progression in chronic hepatitis C by incorporating longitudinal data[J]. *Hepatology*, 2015, 61(6): 1832-1841. DOI: 10.1002/hep.27750.
- [21] IOANNOU GN, TANG WJ, BESTE LA, et al. Assessment of a deep learning model to predict hepatocellular carcinoma in patients with hepatitis C cirrhosis[J]. *JAMA Netw Open*, 2020, 3(9): e2015626. DOI: 10.1001/jamanetworkopen.2020.15626.

收稿日期: 2024-05-22; 录用日期: 2024-07-05

本文编辑: 刘晓虹

引证本文: HAN H, DUAN ZP, WANG Y. Application of machine learning in the diagnosis and treatment of chronic hepatitis C[J]. *J Clin Hepatol*, 2025, 41(1): 141-144.
韩华, 段钟平, 王扬. 机器学习在慢性丙型肝炎诊疗中的应用[J]. *临床肝胆病杂志*, 2025, 41(1): 141-144.