

机器学习在肝脏疾病中的应用现状和未来趋势

李富源^{1,2} 汪占金^{1,2} 蔡俊杰^{1,2} 薛张佗^{1,2} 周瀛³

1. 青海大学 青海西宁 810000

2. 青海大学临床医学院 青海西宁 810000

3. 青海大学附属医院 青海西宁 810000

摘要: 随着计算能力的提升和数据量的增加,机器学习(ML)在医疗领域的应用越来越广泛,尤其是在肝脏疾病的早期诊断、风险评估和治疗决策中。本文综述了机器学习在肝脏疾病中的关键应用,包括寻找肝癌的潜在标志物、预测肝脏术后并发症、识别肝脏疾病的早期危险因素,以及结合影像组学等方面的应用现状与未来趋势。通过分析现有研究,探讨机器学习在肝脏疾病管理中的潜力与挑战。

关键词: 机器学习; 肝脏疾病; 风险评估; 肝包虫病

引言

肝脏疾病是一类严重影响全球健康的疾病,包括肝炎、肝硬化、肝癌,以及一些特殊类型如肝包虫病等。根据世界卫生组织的数据,肝病是导致全球致死率较高的主要疾病之一。由于这些疾病的早期症状往往不明显,很多患者在被诊断时已处于晚期。因此,寻求新技术来提高诊断和管理效率变得尤为重要。机器学习(Machine Learning),作为一种能够从大量数据中提取信息和制定预测模型的工具,已经在肝脏疾病的研究与治疗中展现出重要的应用价值。本文将系统探讨机器学习在肝脏疾病相关的研究领域中的应用现状及未来趋势。

1. 机器学习概述

机器学习是一门人工智能的科学,它使用计算机模拟或实现人类学习行为,指导计算机从数据中学习,并利用自身经验来改善性能。机器学习算法经过持续训练,可以从大型数据库中发现模式和相关性,然后根据数据分析结果做出最佳决策和预测。其应用范围非常广泛,常用于医学研究^[1-2]的多个领域,例如疾病诊断、个性化治疗、患者风险预测等方面。机器学习主要分为三大类:监督学习、无监督学习和强化学习。监督学习利用已标记的医疗数据集进行训练,常用于疾病预测、诊断和风险分层,例如通过分析医学影像来识别肿瘤。常用算法有线性回归、决策树、随机森林和支持向量机等。无监督学习则处理未标记的数据,主要用于发现潜在的患者亚群体或疾病模式,帮助研究人员更好地理解复

杂的疾病机制。常用算法包括聚类(如K-means)和主成分分析(PCA)。深度学习^[3]作为机器学习的一个子领域,利用多层神经网络架构在处理复杂数据时表现出色,尤其是在医学图像分析、自然语言处理和基因组学中。例如,卷积神经网络(CNN)在肿瘤检测中的应用显著提高了影像诊断的敏感性和特异性。此外,强化学习在动态治疗决策中崭露头角,能够在临床试验和个性化治疗方案中实时优化决策,它通过学习在环境中的行动-反馈关系,逐步提升决策效果。随着电子病历的普及、基因组学和生物信息学的快速发展,机器学习方法已经在药物发现、临床预测和个性化医疗等领域取得了显著进展,展现了其在现代医学研究中的重要性。

2. 机器学习寻找肝癌潜在标志物

肝癌,尤其是肝细胞癌^[4](HCC),是全球范围内致死率较高的癌症之一。传统的肝癌筛查方法,比如甲胎蛋白(AFP)检测和影像学检查,虽有一定的临床效果,但灵敏度和特异性不足。近年来,机器学习通过分析复杂的生物标志物数据,提高了早期发现肝癌的能力^[5-7]。

2.1 数据来源与特征选择

研究人员通常从血液样本、影像学数据和基因组数据中提取特征。对基因组数据的分析特别重要,因为某些基因突变与肝癌的发生密切相关。例如,HCC通常与TP53、CTNNB1等基因的突变有关。机器学习算法,如随机森林和支持向量机(SVM),可以用于特征选择和标志物的识别^[8]。使用这些算法,研究表明可以通过分析患者血液中的代谢

物、脂质谱和蛋白质谱，发现新的生物标志物。

3. 机器学习预测肝脏术后并发症

肝癌、肝包虫病等患者通常需要进行肝切除或肝移植手术，这些手术的术后并发症对患者预后具有重要影响。常见的术后并发症^[9]包括感染、出血、肝功能不全、胆道并发症、血栓形成、代谢性并发症，以及免疫反应。这些并发症可能严重影响患者的康复和生存质量，因此围手术期的风险评估和管理变得至关重要，以提高患者的预后和生活质量。

3.1 风险评估模型的建立

机器学习可以通过多种患者数据（包括年龄、性别、病史、术前检查结果等）建立预测模型。在现有研究^[10-11]中，很多学者利用机器学习和深度学习方法来预测肝脏术后并发症的发生率及预后。例如，利用多种术前临床数据（如患者基本信息、实验室结果和影像学资料等）来训练一个深度学习或机器学习模型（如随机森林、支持向量机等），并对比了它们与传统方法（如Child-Pugh评分和MELD评分）的预测效果，结果显示，这些模型在预测术后并发症（出血、感染、肝功能不全等）方面的准确率显著高于传统的临床评分系统，在特定并发症预测中表现更佳。深度学习和机器学习在肝脏手术预后和并发症预测中的应用具有重要意义，能够显著提高预测准确性，帮助外科医生更好地评估患者风险，从而改善术后管理和患者预后。

4. 机器学习识别肝脏疾病早期危险因素

早期识别肝脏疾病的危险因素对于防止疾病的进展至关重要。许多肝脏疾病，如肝炎、肝硬化、肝癌和肝包虫病等，其早期阶段往往难以被发现，患者可能在无明显症状的情况下经历严重的肝脏损伤。通过主动识别和监测潜在危险因素，医疗专业人员可以采取预防措施，延缓或阻止疾病的发展。在这一背景下，机器学习技术显示出巨大的潜力。利用电子健康记录（EHR）中的丰富数据，机器学习算法可以有效提取和分析大量相关信息，包括患者的历史病历、实验室检查结果和生活方式因素。通过挖掘这些潜在的危险因素，机器学习不仅能帮助识别肝脏疾病的高危人群，还能提供个性化的干预策略，从而显著提高早期诊断的准确性和效率。

4.1 早期危险因素的挖掘

通过训练机器学习模型，研究人员能够识别与肝脏疾病相关的多种因素^[12-13]。这些潜在因素包括生活方式方面的因素，如饮酒、饮食习惯和身体活动水平；代谢指标，如

体重指数（BMI）、血糖水平和血脂异常；以及家庭病史，特别是与肝脏疾病相关的家族遗传因素。此外，其他因素如年龄、性别和种族也可能影响肝脏疾病的风险。近年来，多项研究采用了聚类分析和逻辑回归等机器学习算法，有效地确定个体的高危特征，从而实现早期识别和干预。例如，一项研究^[14]利用电子健康记录和机器学习模型，分析了大量患者的临床数据，成功识别出饮酒和高BMI与肝脏疾病发生之间的显著关联。这项研究强调了生活方式的调整对降低患病风险的重要性。诸如此类的研究有很多，采用机器学习的方法不仅提高了对肝脏疾病危险因素的认识，且为预防和治疗策略的制定提供了重要支持，推动了个性化医疗的发展。

5. 机器学习结合影像组学在肝脏疾病中的应用

影像组学结合机器学习为肝脏疾病的诊断和预后评估提供了新的视角和强有力的工具。影像组学^[15]是一种通过提取医学影像（如CT和MRI图像）中的定量特征来深入分析生物学信息的技术。这些定量特征不仅可以揭示肿瘤的形状、纹理和密度等重要信息，还能反映组织的微观环境及其生物学特性。当这些影像特征与机器学习算法相结合时，能够显著提高影像分析的效率和准确性。通过训练机器学习模型，研究人员能够识别出与肝脏疾病相关的复杂特征模式。这种方法不仅可以加速诊断过程，还能提高疾病预后的准确性，为个体化治疗决策提供支持。

5.1 影像特征提取与分析

通过使用计算机视觉技术，可以从影像数据中提取数百个定量特征，如形状、纹理和强度等。这些特征可用于评估肝脏病变的性质和阶段。影像组学的应用不断扩展，在肝脏疾病的诊断和治疗中展现出了显著的潜力。例如，在一项国际合作研究^[16]中，研究人员利用多中心影像数据和机器学习算法，成功地预测了肝癌患者的生存期。该研究整合了来自不同医疗机构的CT和MRI影像数据，通过提取肝肿瘤的形状、纹理、密度等多种影像特征，建立了高效的预测模型。这些模型不仅分析了影像特征，还结合了患者的临床特征（如年龄、性别、肝功能指标等），从而为个体患者提供了更为准确的生存预后评估。这种方法的优势在于，其能够识别出复杂的特征模式，这些模式可能在传统影像分析中被忽视。通过机器学习和深度学习模型，研究者能够高效地处理大量数据，并识别出与预后相关的重要变量，从而为临床

医生提供可靠的决策支持。

6. 机器学习在肝包虫病中的应用

随着养殖业的扩张和人类与动物之间的密切接触，包虫病^[17]在全球一些地区（如南美、非洲、欧洲及亚洲某些地区）成为公共卫生问题。在我国西部地区如新疆、青海、西藏和甘肃等地，肝包虫病较为常见，主要与当地的畜牧业发展、特定的生态环境以及人们的生活习惯密切相关。该病早期诊断困难，且症状多样，轻者无明显症状，严重者可出现肝脏肿块、黄疸、腹痛等，严重时会导致肝脏破裂、感染等并发症^[18]。在一些高发地区，由于缺乏及时的诊断和治疗，以及医疗资源的相对匮乏，会导致较高的死亡率。近年来，机器学习技术的迅速发展为肝包虫病的早期诊断和管理提供了新的契机。

6.1 影像学特征分析

在很多研究中，研究人员利用机器学习算法分析了肝脏的超声、CT和MRI影像数据，以识别肝包虫病的特征。通过提取影像特征，如囊肿的大小、形状、边界平滑度等，使用支持向量机（SVM）和随机森林等算法，研究者能够有效区分活跃的肝包虫病和其他肝脏病变，如肝囊肿^[19-20]等。这些模型大大提高了诊断的准确性和效率，减少了误诊几率。

6.2 肝包虫病灶活性的评估

利用机器学习算法评估肝包虫病灶的活性^[21]也是近期研究的热点之一。例如，研究者们通过分析患者的影像学数据，采用卷积神经网络（CNN）来提取影像特征，与传统特征相比，这种方法能够自动识别出微小的结构变化，进而评估包虫病灶的活性。模型能够区分活跃病灶与钙化病灶，预测其可能的生长和并发症风险。此类研究为临床医生提供了依据，使其能够采取相应的干预措施，如定期监测或及时手术。

7. 多模态数据融合

现代医学研究的一个重要发展趋势是多源数据融合，尤其是在肝包虫病等复杂疾病的管理中。这种方法通过结合不同类型的数据，如影像数据、生化指标和临床信息，提供了更全面的临床预测和决策支持。以CT影像特征为例，研究人员可以提取各种量化数据，包括病灶的大小、形状、纹理和密度等，同时结合患者的生化指标（如肝功能测试、炎症指标等）和临床背景（如年龄、性别和既往病史），创建

更具信息含量的模型。

7.1 提升预测的准确性

通过数据整合，融合模型在疾病预测中的成功率显著提高。这种提高归因于多源数据可以捕捉到更复杂的生物标志物和潜在的临床特征，从而使模型在识别风险时更加敏感和特异。多源数据的融合不仅增强了模型的表现能力，还为医生提供了更为全面的洞察，帮助他们在患者管理和干预措施的制定上做出更为科学的决策。

7.2 集成机器学习模型

在这一背景下，集成学习^[22-23]作为一种新兴的机器学习方法，也表现出了良好的效果。集成学习通过结合多个模型的输出，可以更有效地提高预测的准确性和鲁棒性。它通过将单一模型的弱点最小化，从而弥补单模型可能存在的过拟合或偏差问题。这种方法可以将不同算法的优势结合在一起，形成综合的预测结果。在实际应用中，集成学习模型能够通过加权平均或投票机制结合多种模型的预测结果，从而提升整体性能。例如，在肝包虫病的研究中，集成模型可以同时调用影像组学特征和临床数据的分析结果，使其在复杂的临床环境中更具适应性。而且，对于数据集中的噪声和异常值，集成学习模型也显得更为稳健，能够有效提升模型的泛化能力。

8. 机器学习面临的挑战和局限性

尽管机器学习在医学研究中展现出强大的应用潜力，推动了疾病诊断、治疗方案优化和个性化医疗的发展，但其实际应用依然面临诸多挑战和局限性。首先，机器学习模型，特别是深度学习模型，常被视为“黑箱”，缺乏可解释性^[24]。这一问题在医学领域尤为重要，医生和患者需理解模型的预测结果，以做出知情决策。其次，临床转化存在障碍，实验室表现良好的模型在复杂多变的临床环境中可能面临性能下降的问题。此外，数据质量与可获得性也制约了机器学习模型的表现，医学数据常存在缺失、噪声及不一致性，且隐私保护法规使数据共享困难。最后，机器学习在医学应用成功的关键在于跨学科合作，包括医学、统计学、计算机科学和伦理学的专家共同参与，以解决复杂问题并促进技术的有效转化。

结语

机器学习在肝脏疾病及其他疾病领域的应用展现出强大的潜力与广泛适用性，正在改变传统的医疗实践与研究模

式。通过利用患者多种数据和影像学特征，机器学习能够建立精准的预测模型，帮助医生在早期识别疾病风险、评估术后并发症及制定个性化治疗方案，显著提高诊断效率和患者预后。在肝脏疾病研究中，机器学习与影像组学的结合为临床诊断提供了新的视角，能够有效提取和分析医学影像数据。此外，多模态数据融合的兴起，使不同类型的数据得以整合，形成更全面的临床预测模型，进一步增强了风险识别的敏感性和特异性。随着数据挖掘技术和计算能力的提升，机器学习在医疗领域将扮演越来越重要的角色，尤其是在复杂病症管理、疾病预防和个性化医疗方面。通过集成学习等先进方法，未来医疗研究将不断提高预测模型的准确性与稳健性，推动医学创新，造福更多患者。

参考文献：

- [1] Hou N, Li M, He L, Xie B, Wang L, Zhang R, Yu Y, Sun X, Pan Z, Wang K. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost. *J Transl Med.* 2020 Dec 7;18(1):462. doi: 10.1186/s12967-020-02620-5.
- [2] Du M, Haag DG, Lynch JW, Mittinty MN. Comparison of the Tree-Based Machine Learning Algorithms to Cox Regression in Predicting the Survival of Oral and Pharyngeal Cancers: Analyses Based on SEER Database. *Cancers (Basel).* 2020 Sep 29;12(10):2802. doi: 10.3390/cancers12102802.
- [3] van der Velden BHM, Kuijf HJ, Gilhuijs KGA, Viergever MA. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Med Image Anal.* 2022 Jul;79:102470. doi: 10.1016/j.media.2022.102470. Epub 2022 May 4. PMID: 35576821.
- [4] Forner A, Reig M, Bruix J. Hepatocellular carcinoma. *Lancet.* 2018 Mar 31;391(10127):1301-1314. doi: 10.1016/S0140-6736(18)30010-2. Epub 2018 Jan 5. PMID: 29307467.
- [5] Lv JH, Hou AJ, Zhang SH, Dong JJ, Kuang HX, Yang L, Jiang H. WGCNA combined with machine learning to find potential biomarkers of liver cancer. *Medicine (Baltimore).* 2023 Dec 15;102(50):e36536. doi: 10.1097/MD.00000000000036536. PMID: 38115320; PMCID: PMC10727608.
- [6] Gupta R, Kleinjans J, Caiment F. Identifying novel transcript biomarkers for hepatocellular carcinoma (HCC) using RNA-Seq datasets and machine learning. *BMC Cancer.* 2021 Aug 27;21(1):962. doi: 10.1186/s12885-021-08704-9. PMID: 34445986; PMCID: PMC8394105.
- [7] Matboli M, Diab GI, Saad M, Khaled A, Roushdy M, Ali M, ELsawi HA, Aboughaleb IH. Machine-Learning-Based Identification of Key Feature RNA-Signature Linked to Diagnosis of Hepatocellular Carcinoma. *J Clin Exp Hepatol.* 2024 Nov-Dec;14(6):101456. doi: 10.1016/j.jceh.2024.101456. Epub 2024 Jun 14. PMID: 39055616; PMCID: PMC11268357.
- [8] Tu DY, Cao J, Zhou J, Su BB, Wang SY, Jiang GQ, Jin SJ, Zhang C, Peng R, Bai DS. Identification of the mitophagy-related diagnostic biomarkers in hepatocellular carcinoma based on machine learning algorithm and construction of prognostic model. *Front Oncol.* 2023 Mar 1;13:1132559. doi: 10.3389/fonc.2023.1132559. PMID: 36937391; PMCID: PMC10014545.
- [9] Burchard PR, Dave YA, Loria AP, Parikh NB, Pineda-Solis K, Ruffolo LI, Strawderman M, Schoeniger LO, Galka E, Tomiyama K, Orloff MS, Carpizo DR, Linehan DC, Hernandez-Alejandro R. Early postoperative ERAS compliance predicts decreased length of stay and complications following liver resection. *HPB (Oxford).* 2022 Sep;24(9):1425-1432. doi: 10.1016/j.hpb.2022.01.008. Epub 2022 Jan 25. PMID: 35135723.
- [10] Liu Z, Liu Y, Zhang W, Hong Y, Meng J, Wang J, Zheng S, Xu X. Deep learning for prediction of hepatocellular carcinoma recurrence after resection or liver transplantation: a discovery and validation study. *Hepatol Int.* 2022 Jun;16(3):577-589. doi: 10.1007/s12072-022-10321-y. Epub 2022 Mar 29. PMID: 35352293; PMCID: PMC9174321.
- [11] Serenari M, Han KH, Ravaioli F, Kim SU, Cucchetti A, Han DH, Odaldi F, Ravaioli M, Festi D, Pinna AD, Cescon M. A nomogram based on liver stiffness predicts postoperative complications in patients with hepatocellular carcinoma. *J Hepatol.* 2020 Oct;73(4):855-862. doi: 10.1016/j.jhep.2020.04.032. Epub 2020 Apr 30. PMID: 32360997.
- [12] Su á rez M, Gil-Rojas S, Mart í nez-Blanco P, Torres AM, Ram ó n A, Blasco-Segura P, Torralba M, Mateo J. Machine Learning-Based Assessment of Survival and Risk Factors in Non-Alcoholic Fatty Liver Disease-Related Hepatocellular Carcinoma

for Optimized Patient Management. *Cancers (Basel)*. 2024 Mar 10;16(6):1114. doi: 10.3390/cancers16061114. PMID: 38539449; PMCID: PMC10969326.

[13] Peng HY, Duan SJ, Pan L, Wang MY, Chen JL, Wang YC, Yao SK. Development and validation of machine learning models for nonalcoholic fatty liver disease. *Hepatobiliary Pancreat Dis Int*. 2023 Dec;22(6):615–621. doi: 10.1016/j.hbpd.2023.03.009. Epub 2023 Mar 25. PMID: 37005147.

[14] Ji W, Xue M, Zhang Y, Yao H, Wang Y. A Machine Learning Based Framework to Identify and Classify Non-alcoholic Fatty Liver Disease in a Large-Scale Population. *Front Public Health*. 2022 Apr 4;10:846118. doi: 10.3389/fpubh.2022.846118. PMID: 35444985; PMCID: PMC9013842.

[15] Xia T, Zhao B, Li B, Lei Y, Song Y, Wang Y, Tang T, Ju S. MRI-Based Radiomics and Deep Learning in Biological Characteristics and Prognosis of Hepatocellular Carcinoma: Opportunities and Challenges. *J Magn Reson Imaging*. 2024 Mar;59(3):767–783. doi: 10.1002/jmri.28982. Epub 2023 Aug 30. PMID: 37647155.

[16] Bo Z, Chen B, Zhao Z, He Q, Mao Y, Yang Y, Yao F, Yang Y, Chen Z, Yang J, Yu H, Ma J, Wu L, Chen K, Wang L, Wang M, Shi Z, Yao X, Dong Y, Shi X, Shan Y, Yu Z, Wang Y, Chen G. Prediction of Response to Lenvatinib Monotherapy for Unresectable Hepatocellular Carcinoma by Machine Learning Radiomics: A Multicenter Cohort Study. *Clin Cancer Res*. 2023 May 1;29(9):1730–1740. doi: 10.1158/1078-0432.CCR-22-2784. PMID: 36787379.

[17] Bresson-Hadni S, Spahr L, Chappuis F. Hepatic Alveolar Echinococcosis. *Semin Liver Dis*. 2021 Aug;41(3):393–408. doi: 10.1055/s-0041-1730925. Epub 2021 Jun 23. PMID: 34161992.

[18] Nunnari G, Pinzone MR, Gruttadauria S, Celesia BM, Madeddu G, Malaguarnera G, Pavone P, Cappellani A, Cacopardo B. Hepatic echinococcosis: clinical and therapeutic aspects. *World*

J Gastroenterol. 2012 Apr 7;18(13):1448–58. doi: 10.3748/wjg.v18.i13.1448. PMID: 22509076; PMCID: PMC3319940.

[19] Wang Z, Bian H, Li J, Xu J, Fan H, Wu X, Cao Y, Guo B, Xu X, Wang H, Zhang L, Zhou H, Fan J, Ren Y, Geng Y, Feng X, Li L, Wei L, Zhang X. Detection and subtyping of hepatic echinococcosis from plain CT images with deep learning: a retrospective, multicentre study. *Lancet Digit Health*. 2023 Nov;5(11):e754–e762. doi: 10.1016/S2589-7500(23)00136-X. Epub 2023 Sep 26. PMID: 37770335.

[20] 汪占金, 曹云太, 乔木, 等. 基于机器学习的CT平扫影像组学鉴别CE1型肝囊型包虫病与肝囊肿[J]. 临床放射学杂志, 2024,43(09):1512–1518. DOI:10.13437/j.cnki.jcr.2024.09.014.

[21] 任波. 基于影像组学评价肝泡型包虫病组织特征和生物活性的研究[D]. 新疆医科大学, 2021. DOI:10.27433/d.cnki.gxyku.2021.000001.

[22] Naimi AI, Balzer LB. Stacked generalization: an introduction to super learning. *Eur J Epidemiol*. 2018 May;33(5):459–464. doi: 10.1007/s10654-018-0390-z. Epub 2018 Apr 10.

[23] Sun L, Yang L, Liu X, Tang L, Zeng Q, Gao Y, Chen Q, Liu Z, Peng B. Optimization of Cervical Cancer Screening: A Stacking-Integrated Machine Learning Algorithm Based on Demographic, Behavioral, and Clinical Factors. *Front Oncol*. 2022 Feb 15;12:821453. doi: 10.3389/fonc.2022.821453.

[24] Barrag ́n-Montero A, Bibal A, Dastarac MH, Draguet C, Vald ́s G, Nguyen D, Willems S, Vandewinckele L, Holmstr ́m M, L ́ofman F, Souris K, Sterpin E, Lee JA. Towards a safe and efficient clinical implementation of machine learning in radiation oncology by exploring model interpretability, explainability and data-model dependency. *Phys Med Biol*. 2022 May 27;67(11):10.1088/1361-6560/ac678a. doi: 10.1088/1361-6560/ac678a. PMID: 35421855; PMCID: PMC9870296.