

# 机器学习在肝包虫病诊疗中的研究进展

陈志恒, 周 瀛\*

青海大学附属医院肝胆胰外科, 青海 西宁

收稿日期: 2023年6月25日; 录用日期: 2023年7月19日; 发布日期: 2023年7月25日

## 摘 要

本综述主要介绍机器学习在肝包虫病诊断方面的应用, 包括数据集和预处理、特征提取与选择、分类器方法等方面的内容。先进的机器学习技术在协助肝包虫病的诊断、分型、临床手段、机制评价、治疗和预后评估等方面具有重要价值, 在这一领域的研究对肝包虫病应该受到更多的关注和支持。同时, 还将讨论当前机器学习在肝包虫病诊断中面临的挑战和限制。

## 关键词

肝包虫病, 机器学习, 预测模型, 精准医学

# Research Progress of Machine Learning in the Diagnosis and Treatment of Hepatic Echinococcosis

Zhiheng Chen, Ying Zhou\*

Department of Hepatobiliary and Pancreatic Surgery, Affiliated Hospital of Qinghai University, Xining Qinghai

Received: Jun. 25<sup>th</sup>, 2023; accepted: Jul. 19<sup>th</sup>, 2023; published: Jul. 25<sup>th</sup>, 2023

## Abstract

This review mainly introduces the application of machine learning in the diagnosis of hepatic echinococcosis, including data sets and preprocessing, feature extraction and selection, classifier methods, and so on. Advanced machine learning technology is of great value in assisting the diagnosis, classification, clinical methods, mechanism evaluation, treatment and prognosis evaluation of hepatic echinococcosis. More attention and support should be paid to the research in this field.

\*通讯作者。

**At the same time, the current challenges and limitations of machine learning in the diagnosis of hepatic echinococcosis will be discussed.**

## Keywords

**Hepatic Echinococcosis, Machine Learning, Prediction Model, Precision Medicine**

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

肝包虫病(hepatic echinococcosis)是人和动物都能患上的疾病,分为囊型包虫(Cystic echinococcosis, CE)和泡型包虫(Alveolar echinococcosis, AE),分别由细粒棘球蚴和多房棘球蚴感染引起[1]。其发病率和死亡率在全球范围内仍然较高[2]。在我国多见于西北和西南地区[3]。牧区医疗条件有限,多数牧区患者是在出现明显占位或包虫并发症引起不适后就诊[4],在此阶段多需要住院手术治疗,给牧区群众带来巨大经济负担[5][6]。传统的肝包虫病诊断方法往往需要依靠医生的临床经验,这导致了诊断结果的不确定性和误诊率的高发。因此,包虫的早诊断和治疗以及鉴别,影响着患者的预后及生存质量。

但随着人工智能的快速发展,人工智能技术越来越多地应用于医学方面的分析[7],其中机器学习作为人工智能的一部分,能够快速有效的处理分析复杂医疗数据,基于机器学习的肝包虫病诊断方法是近年来新兴的研究方向,受到了广泛的关注和研究。其中逻辑回归、神经网络、支持向量机、深度卷积神经网络等已被广泛用于构建预测模型辅助诊断[8],取得良好效果。本文将对机器学习在肝包虫诊治中应用价值的研究进展进行综述。

## 2. 机器学习概述

机器学习是运用计算机从给出的数据中学习发现规律。机器学习最早是被 Samuel 提出的用来训练跳棋[9],并对机器学习过程进行了较为详细的研究,证实机器学习训练能取得较好的效果,具有训练时长短的特点,通过设定规则实验验证的机器学习原理也适用于许多其他情况。机器学习是数学和计算机的结合科学,具有高效率快速处理数据的特点[10]。机器学习的一般过程是数据的收集、预处理、切分数据集、训练与预测等[11],运用机器学习通常来讲我们需要一定的数据量,临床收集的数据有很多种类,需要对数据进行处理。这些数据中有的数据包含一定的噪声,有些数据存在缺失值还有一些数据没有特征值等等这些都是需要进行清洗预处理的。机器学习时计算机需要使用统计学方法尝试从给定的数据中学习分析出因素关联。计算机处理大量的数据从而建立模型。

## 3. 机器学习的分类

计算机使用的学习类型可以简单的分为两类:有监督学习和无监督学习[12]。在临床应用中监督学习的重要步骤是将数据中的一部分作为训练集对特征进行处理提取,之后对训练集进行训练,然后使用训练集的预测模型对验证集数据进行预测,以验证最终的预测效能。对于无监督学习来讲无监督学习和有监督学习最大区别是有人为参与标记的数据集。在机器学习的运用中由于某些种类的数据在标记特征时产生了大量的无标签数据和少量有标签数据,研究者将无标签数据加入有便签数据中一起训练,由此

中产生了半监督学习[13]。

#### 4. 机器学习的临床数据集及其预处理

数据集的质量和大小是机器学习研究成功的重要因素[14]。肝包虫病的数据集存在着困难与挑战,因为这是一种区域性分布的地方性疾病,其数据采集包含多种形态学和病理学异质性,同时医学图像和数据需要针对患者姓名和身份进行匿名化处理[15]。提高模型泛化能力至关重要[16],已有的肝包虫病数据集为了提高机器学习方法在肝包虫病诊断中的泛化能力,还需要更好地提取包虫病数据的生物学特征。这可以通过对肝包虫病图像数据的预处理来实现。

在数据预处理方面,肝包虫病的数据通常需要进行标准化、增强和配准等处理,以保证数据质量和可用性。例如,基于超声数据的肝包虫病诊断需要对超声图像进行构建和预处理[17],基于CT或MRI数据的诊断则需要对数据进行配准和分割处理[18]。另外研究者在临床数据的收集处理过程中,应承担相应的伦理责任[19]。

#### 5. 机器学习在肝包虫病诊治中的应用及现状

机器学习方法可以通过充分利用已有的临床病例数据集,提取其中的特征信息,构建分类器或者回归模型,从而提高诊断疾病的准确率和鉴别性[20],在机器学习应用于肝包虫病诊疗中同样适用,另外,更重要的是,在通过对肝包虫病数据的分析和处理过程中,人工智能可以帮助医生更好审视自己的研究领域[21],同样机器学习方法能够更好的帮助医生理解肝包虫病的发病机理、规律以及临床表现,最终发现新的治疗方法和手段。

机器学习现已在疾病诊断治疗,风险评估,并发症预测等方面取得了一定的发展[22][23]:肝包虫病灶有无活性[24],影响着包虫的治疗与用药,关系患者的预后,比如同为包虫治疗的药物阿苯达唑脂质体和阿苯达唑片剂,对包虫治疗的效果是否有显著区别?逻辑回归往往用于生成特征数少的二分类结果,可以生成简单线性的模型。有研究者运用逻辑回归结合影像学分析对包虫用药进行了临床疗效分析评价,结果显示阿苯达唑脂质体疗效好于阿苯达唑片剂[25]。

在机器学习的发展下,计算机医学图像分析也取得了长足的发展,特征的提取是机器学习应用于医学图像工作的一环,原本需要医学影像经验丰富的医生去完成特征的提取工作,如今在机器学习的辅助下变得更加便捷。因此有不少研究者将机器学习技术运用在影像学方面,做了一定的研究,其中成果较多的为预测模型方面的运用。

医工交叉以及人工智能和大数据的发展,越来越体现在个性化精准诊断方面,例如近年来影像组学利用生物影像分析技术进行疾病预测、诊断和治疗的研究就是很有力的证据。影像组学通过对大量病例的影像数据进行分析,建立与疾病相关的模型,从而提高疾病诊断和治疗的准确性和效率。现如今影像组学在肿瘤学、神经科学、心血管疾病等领域得到广泛应用,成为了医学领域的一个重要的研究方向[26]。比如超声具有便捷无辐射的特点,在肝包虫疾病方面的运用十分广泛,有不少研究者在机器学习与超声结合方面做了探索有研究者就运用超声影像组学来构建肝棘球蚴病分型模型取得良好的效果,在肝棘球蚴病精准超声诊断方面做了一定的贡献。有研究者以随机森林(random forest, RF)、极限梯度增强树(Extreme Gradient Boosting, XGBoost)和逻辑回归(logistic regression, LR)的方法,运用MRI影像组学来预测肝包虫病血管侵犯的机器学习模型,受试者工作特征曲线表明,模型取得了良好的效能[27]。影像组学在医学影像学 and 个性化精准医疗之间搭建了一座桥梁。

肝包虫的治疗有多种方式,其中手术为主要的治疗方式,有研究者对肝包虫术后并发症进行预测,使用机器学习筛选出独立危险因素,在此基础上构建了列线图风险预测模型并验证取得良好预测效果[28]。

贝叶斯是运用概率统计来分类的算法,可以用于大型数据库中,具有算法简单,分类准确快速的特点[29]。在医学领域有的研究者使用贝叶斯分析肝棘球蚴病的流行情况,从而得到肝棘球蚴病的发病因素及关联,可以以疾病特点对相关患者做好疾病的预防和诊疗[30]。支持向量机(Support Vector Machines, SVM),是基于数据集,对特定特征数据在决策面进行划分,一般用于数据的二分类,随着发展也可用于多分类。有研究者收集数据后,使用 SVM 等机器学习方法鉴别脑包虫与脑转移瘤,结果表明 SVM 分类方法可提高疾病病灶鉴别的准确率[31]。决策树是一树状结构,始于根节点,对数据样本进行测试,其数据往往是实例集组成[32]。它可以从无次序、无规则的数据中推理出分类规则,并用决策树形式表示。有研究中利用决策树(Decision Tree)模型对收集到的 1070 名包虫病患者得住院费用影响因素进行分析为医疗监管提供依据[33]。卷积神经网络(Convolutional Neural Networks)是一种机器学习具有多层感知器,一般用来进行图形分析。正因如此有学者以此方法对不同病灶类型的肝包虫 CT 图像进行分类,准确率在 78% 以上[34]。

## 6. 机器学习在肝包虫诊疗中的不足

机器学习在医疗诊断中的应用本就有诸多方面的问题和挑战,数据获得方面,肝包虫为地方性疾病,多为牧民,随访困难后续数据获得不易,这就造成数据质量差和数据量不足等问题,机器学习算法依赖于模型具有一定局限性,数据偏倚直接影响模型性能,另外机器学习的算法相对复杂对医学工作者来说有一定门槛。另外机器学习的成果繁多缺乏整合,这就造成机器学习成果的应用不够。相信随着机器学习的算法的发展和医生思维的转变这些问题在不久的将来都能得到一定的改善。

## 7. 小结

人工智能的发展使医疗行业能更好的利用机器学习来辅助医疗,能够在诊断和预测疾病更加量化的分析,更早的对疾病进行诊断进而干预,目前传统的影像诊断主要依赖影像医生的个人经验,部分病例诊断信息有限,医学发展越来越向精准医疗方面靠拢,机器学习可以辅助提升疾病的诊断,有利于精准医学的发展[35]。精准医疗可以让患者获益更多。当然目前机器学习在医学领域的运用还是远远不够的,相信在以后随着科学技术的发展和科研工作者的探索和努力,机器学习将能更好的辅助医生进行疾病的诊治和决策。

## 基金项目

研究任务来源:青海省科技计划项目。项目全程:基于人工智能的肝囊型包虫和泡型包虫筛查模型及应用研究。项目编号:2021-ZJ-963Q。

## 参考文献

- [1] Wen, H., Vuitton, L., Tuxun, T., *et al.* (2019) Echinococcosis: Advances in the 21st Century. *Clinical Microbiology Reviews*, **32**. <https://doi.org/10.1128/CMR.00075-18>
- [2] Budke, C.M., Deplazes, P. and Torgerson, P.R. (2006) Global Socioeconomic Impact of Cystic Echinococcosis. *Emerging Infectious Diseases*, **12**, 296-303. <https://doi.org/10.3201/eid1202.050499>
- [3] Mihmanli, M., Idiz, U.O., Kaya, C., *et al.* (2016) Current Status of Diagnosis and Treatment of Hepatic Echinococcosis. *World Journal of Hepatology*, **8**, 1169-1181. <https://doi.org/10.4254/wjh.v8.i28.1169>
- [4] 于鹏, 段绍斌. 术前 MELD 评分对肝包虫病患者手术后并发症和远期预后的评估价值[J]. 肝脏, 2023, 28(2): 229-232, 237.
- [5] Chen, J.Z., Yu, D.G., Lan, X.R., *et al.* (2020) Multiple Disciplinary Team Discussion on Complex Hepatic Alveolar Echinococcosis. *Journal of Hepatopancreatobiliary Surgery*, **32**, 152-156.
- [6] 王正荣, 马勋, 张艳艳, 等. Th1/Th2 型免疫反应相关基因在巨噬细胞 RAW 264.7 应对细粒棘球绦虫原头蚴刺激时的表达谱研究[J]. 畜牧兽医学报, 2022, 53(1): 250-262.

- [7] Briganti, G. and Le Moine, O. (2020) Artificial Intelligence in Medicine: Today and Tomorrow. *Frontiers in Medicine*, 7, Article 27. <https://doi.org/10.3389/fmed.2020.00027>
- [8] 何新, 陈慧, 冯炜炜. 机器学习算法在辅助超声诊断附件肿块良恶性中的应用研究进展[J]. 诊断学理论与实践, 2022, 21(4): 541-546.
- [9] Samuel, A.L. (1959) Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, 3, 210-229. <https://doi.org/10.1147/rd.33.0210>
- [10] Hastie, T., Tibshirani, R. and Friedman, J. (2009) Overview of Supervised Learning. In: Hastie, T., Tibshirani, R. and Friedman, J. Eds., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, New York, 9-41. [https://doi.org/10.1007/978-0-387-84858-7\\_2](https://doi.org/10.1007/978-0-387-84858-7_2)
- [11] Tran, K.A., Kondrashova, O., Bradley, A., et al. (2021) Deep Learning in Cancer Diagnosis, Prognosis and Treatment Selection. *Genome Medicine*, 13, Article No. 152. <https://doi.org/10.1186/s13073-021-00968-x>
- [12] Rajkomar, A., Dean, J. and Kohane, I. (2019) Machine Learning in Medicine. *The New England Journal of Medicine*, 380, 1347-1358. <https://doi.org/10.1056/NEJMr1814259>
- [13] Sabbir, A., Jimeno-Yepes, A. and Kavuluru, R. (2017) Knowledge-Based Biomedical Word Sense Disambiguation with Neural Concept Embeddings. 2017 *IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE)*, Washington DC, 23-25 October 2017, 163-170. <https://doi.org/10.1109/BIBE.2017.00-61>
- [14] 何清, 李宁, 罗文娟, 等. 大数据下的机器学习算法综述[J]. 模式识别与人工智能, 2014(4): 327-336.
- [15] 赵镇东, 常晓林, 王逸翔. 机器学习中的隐私保护综述[J]. 信息安全学报, 2019, 4(5): 1-13.
- [16] 邹斌, 董雪梅, 付丽华. 基于算法稳定的分类机器学习泛化能力的研究[J]. 模式识别与人工智能, 2004, 17(4): 430-433.
- [17] 王海霞. 基于肝脏超声图像的包虫病智能辅助诊断方法研究[D]: [硕士学位论文]. 深圳: 深圳大学, 2020.
- [18] 木拉提·哈密提, 周晶晶, 严传波, 等. 新疆地方性肝包虫 CT 图像的灰度直方图特征提取与分析[J]. 科技导报, 2012, 30(6): 66-70.
- [19] 卢艺, 崔中良. 中国人工智能伦理研究进展[J]. 科技导报, 2022, 40(18): 69-78.
- [20] 陈旭, 刘鹏鹤, 孙毓忠, 等. 面向不平衡医学数据集的疾病预测模型研究[J]. 计算机学报, 2019, 42(3): 596-609.
- [21] 鞠强. 机器学习引领革新[J]. 科技导报, 2017, 35(22): 7.
- [22] 兰欣, 卫荣, 蔡宏伟, 等. 机器学习算法在医疗领域中的应用[J]. 医疗卫生装备, 2019, 40(3): 93-97.
- [23] 任珍, 李姝, 赵静静, 等. 机器学习在急诊医学中应用的研究进展及展望[J]. 中国急救医学, 2021, 41(3): 261-265.
- [24] 侯娇, 温浩, 王明坤, 等. 肝细粒棘球蚴病手术患者病灶活性状态的影响因素分析[J]. 中国寄生虫学与寄生虫病杂志, 2022, 40(3): 309-315.
- [25] 李海涛, 宋涛, 邵英梅, 等. Logistic 回归方程对阿苯达唑两种剂型药物治疗囊性包虫病的临床疗效评价[J]. 新疆医学, 2015(3): 281-285.
- [26] 王斯麒, 冉张申. 基于超声的影像组学在临床疾病诊断中的应用进展[J]. 中国医学创新, 2022, 19(18): 180-185.
- [27] 樊霞, 王健, 夏雨薇, 等. 基于 MRI 影像组学预测肝泡型包虫病边缘微血管侵犯[J]. 中国医学影像技术, 2021, 37(12): 1849-1853.
- [28] 郭兵, 庞明泉, 许晓磊, 等. 预测两型肝包虫病肝切除术后并发症列线图的建立及评价[J]. 临床肝胆病杂志, 2021, 37(11): 2626-2631.
- [29] 罗可, 林睦纲, 郗东妹. 数据挖掘中分类算法综述[J]. 计算机工程, 2005, 31(1): 3-5, 11.
- [30] 鄢靖欣, 邓满军, 樊羿辰, 等. 中国西部地区近 20 年肝棘球蚴病的贝叶斯单臂 meta 分析[J]. 中国病原生物学杂志, 2022, 17(10): 1164-1169, 1174.
- [31] 努尔比耶姆·阿布力克木, 杨静, 刘珺迪, 等. 基于磁共振灌注加权成像和支持向量机鉴别脑泡型包虫病与脑转移瘤[J]. 磁共振成像, 2022, 13(4): 26-31.
- [32] 杨学兵, 张俊. 决策树算法及其核心技术[J]. 计算机技术与发展, 2007, 17(1): 43-45.
- [33] 马明, 陈婷, 杨圆圆. 利用决策树模型的包虫病患者住院费用及病例组合分析[J]. 中国病案, 2022, 23(8): 33-37.
- [34] 茹仙古丽·艾尔西丁, 裴世宇, 严传波, 等. 基于深度学习的肝包虫疾病图像分类[J]. 中国数字医学, 2020, 15(4): 2-4, 21.
- [35] 周泊阳, 石一磊, 郭乐杭, 等. 人工智能技术赋能超声影像用于肝脏疾病精准诊疗[J]. 中国血吸虫病防治杂志, 2022, 34(5): 458-464.