

机器学习在肺炎死亡预测模型中的研究进展

申艳梅¹, 马普艳¹, 许均通¹, 郑鹏程^{2*}

¹大理大学药学院, 云南 大理

²云南省第一人民医院/昆明理工大学附属医院药学部, 云南 昆明

收稿日期: 2025年3月5日; 录用日期: 2025年4月3日; 发布日期: 2025年4月14日

摘要

肺炎(pneumonia)是指各种致病因素导致的肺实质及肺间质炎症, 是全球范围内常见的呼吸系统疾病之一, 而重症肺炎是在肺炎的基础上发展而来, 具有病死率高, 并发症多和预后差的特点, 并加重医疗经济负担, 同时对人类健康构成重大威胁。随着人工智能的迅速发展和医疗健康数据的急剧增长, 机器学习已经在肺炎死亡预测中得到了广泛应用, 帮助临床医生精准的进行个体化临床用药和提高治疗效果, 延长患者的生存时间, 提高其生存质量。

关键词

肺炎, 机器学习, 死亡, 风险预测

Research Progress of Machine Learning in Pneumonia Death Prediction Model

Yanmei Shen¹, Puyan Ma¹, Juntong Xu¹, Pengcheng Zheng^{2*}

¹College of Pharmacy, Dali University, Dali Yunnan

²Department of Pharmacy, The First People's Hospital of Yunnan Province/The Affiliated Hospital of Kunming University of Science and Technology, Kunming Yunnan

Received: Mar. 5th, 2025; accepted: Apr. 3rd, 2025; published: Apr. 14th, 2025

Abstract

Pneumonia refers to pulmonary parenchyma and interstitial inflammation caused by various pathogenic factors, which is one of the common respiratory diseases in the world, and severe pneumonia develops on the basis of pneumonia, with high mortality, multiple complications and poor prognosis, and increases the economic burden of medical care, and poses a major threat to human health.

*通讯作者。

文章引用: 申艳梅, 马普艳, 许均通, 郑鹏程. 机器学习在肺炎死亡预测模型中的研究进展[J]. 生物医学, 2025, 15(3): 453-458. DOI: 10.12677/hjbm.2025.153052

With the rapid development of artificial intelligence and the rapid growth of medical and health data, machine learning has been widely used in the prediction of pneumonia death, helping clinicians to accurately personalize clinical medication and improve treatment effects, extend the survival time of patients, and improve their quality of life.

Keywords

Pneumonia, Machine Learning, Death, Risk Prediction

Copyright © 2025 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

尽管自 21 世纪以来,人类疾病谱发生了变化,但是感染性疾病仍然是人类生命健康的主要威胁,肺炎正是其中之一[1]。肺炎(pneumonia)是一种常见的感染性呼吸系统疾病,多因感染、免疫损伤等因素引起肺实质炎症反应,肺组织炎症发展至一定程度则会进展至重症肺炎,可能继发脑水肿、心功能不全、中毒性脑病等并发症,引发各器官功能障碍,危及生命安全[2]。依据 2022 年中国卫生健康统计年鉴记载,我国 2021 年城市居民肺炎的病死率为 11.2/10 万人,农村居民肺炎的病死率为 7.06/10 万人。积极探索有效的干预措施,在缩短病程,降低肺炎治疗失败率及病死率方面具有重大意义[3]。对于优化医疗资源调控、保障重症患者得到及时有效的疗救以及降低整体死亡率极为关键,按照这一背景局面,传统的临床经验与以统计尺度的办法在大规模疫情时也许有局限,而机器学习(ML)技术借助依靠强大的数据分析能力、模式识别能力以及预测能力,持续成为肺炎死亡风险预测的关键手段。

2. 机器学习在肺炎临床结局预测中的研究现状

机器学习借助自身强大的数据分析与模式识别能力,在肺炎相关情况的风险评估、重症预测和死亡预测等任务中被大量采用,与传统统计学方法相比,机器学习可从复杂又呈现多维度的医疗数据中自动提取关键特征。识别潜在的隐患因素,并建立高效的预测模型,极大提高了疾病风险评估的精准度与效率[4]。既有相关研究主要集中于采用电子健康记录(EHR)、医学影像数据(如胸部 CT 和 X 光片)以及实验室检测指标(如血氧饱和度、炎症标志物和肝肾功能指标)等多源数据,构建肺炎患者的死亡风险预测模型[5][6]。这些模型不仅可以辅佐医务工作者迅速识别出高风险患者,恰当规划有限的医疗资源,还可为制定个性化治疗方案给出科学证据,降低重症患者死亡风险[7][8]。

3. 患者死亡预测模型的应用研究

3.1. 预测模型构建流程

肺炎死亡预测模型的开发一般涉及数据收集与预处理、特征变量筛选、模型训练与优化、模型验证与测试以及模型部署与应用等关键举措,以维持模型的精准性与临床实用性[9]。在数据获取及初步处理阶段,务必从医院电子健康记录(EHR)、公开医疗数据库(如 WHO、CDC 数据集)以及影像数据(如 CT 扫描和 X 光片)等各类出处采集患者临床资料[10],并依照缺失值填补(如均值填充、多重插补、KNN 插补等)、数据标准化(Z-score 归一化、Min-Max 归一化等)和特征选择(LASSO 回归、Boruta 算法等)做法开展数据筛选,以增强数据水准及一致性[11]。在特征变量筛选阶段,纳入可能影响死亡率的关键变量,如人

口相关内容,以提升模型的运算效率及解释性,在模型执行训练及进行优化的阶段,选择恰当的机器学习算法(如逻辑回归、随机森林、支持向量机、XGBoost、深度神经网络等)开展训练,并凭借 K 折交叉验证对模型性能开展测评,同时借助超参数调节(如网格搜索、贝叶斯优化)强化模型的性能[12]。

3.2. 关键机器学习算法及其应用

在符合肺炎有关的死亡预测模型里面,一般采用的机器学习算法主要包含决策树及其衍生样式。决策树及其衍生形式鉴于可处理非线性关系,适宜应用在医疗数据的复杂特征交互,特别是在小样本与高维特征的情境当中表现优秀,随机森林凭借组接多个决策树强化预测稳定性与抗过拟合能力[13],而 XGBoost 在添加上运算效率的同时,能够更到位地完善预测结果[14]。支持向量机(SVM)在小样本、高维度数据分类任务实施阶段当中具备较强泛化能力,适用于医学数据中跟患者分组、生存分析相关的情境[15]。特别是深度学习技术的实施推进手段,在处理大规模数据时表现着实非凡,可以借助自身从高维数据(如医学影像、基因组数据)中提取深层特征,为医生提供更具洞察力的预判结论数值[16]。逻辑回归(LR)作为一般的统计学习途径,由于它显示出可解释性和稳定性,在医学研究分析阶段内依旧被广泛采用,尤其适宜估算各特征对患者死亡风险的影响[17]。

3.3. 影响患者死亡率的关键特征

肺炎患者的死亡风险受多种情形影响,其中牵扯到人口统计学特征、基础疾病、实验室检测指标、影像学特征以及治疗措施等关键方面,按照人口统计学特征情形,高龄患者的死亡牵扯风险明显高于年轻患者[18]。后续相关研究显示男性患者的死亡率一般高于女性,这也许是因为免疫系统和基础健康状况等生理差异。患者合并基础性疾病是影响预后与死亡情况的关键要素,如糖尿病、高血压、心血管疾病、慢性呼吸系统疾病等均会使病情进一步加重,增加死亡率,实验室检测指标同样是体现关键意义的参考凭据,血氧饱和度(SpO_2)下降一般显示患者也许存在严重的肺损伤或呼吸衰竭情形。而 C 反应蛋白(CRP)、D 二聚体、白细胞计数等指示炎症及凝血功能的指标异常升高,也被看作与病情加重及死亡风险高度关联[19][20]。影像学检查在死亡风险预测相关工作中同样发挥关键作用,肺部 CT 影像能体现病变范围、炎症程度及肺纤维化情形,严重的双肺弥漫性病变一般揭示更差的预后[21]。

3.4. 多模态数据融合

肺炎患者的临床数据含有多种类型,如结构化的电子健康记录(EHR)、非结构化的医学影像数据(CT 扫描、X 光片)以及基因组数据等,由于单一数据模态大概存在信息欠缺或局限[22],借助多模态数据融合技术,可以全面采纳不同种类的数据,增大模型的泛化能力及预测精准度,让它更贴合临床应用需求点。常见的数据融合手段有:特征阶段整合,即对不同形式的数据分别进行特征抽取,并把这些归并成统一的输入向量,让模型得以同步研习多种数据特征;模型阶段融聚,指分别开展针对不同数据格式的机器学习或深度学习模型的训练,并借助集成学习办法(如投票机制、加权平均、Stacking 等)做出最终推断,以强化模型的坚实度及精准度[23]。

4. 机器学习在肺炎患者死亡预测情形中的挑战

4.1. 数据质量问题

在肺炎感染病例数据的收集与整理阶段中,经常面临诸如存在缺失值、样本分布不均衡等数据质量不高的问题。这些问题会影响机器学习模型训练效果的实际意义,降低预测的准确度。有的轻症患者只做了部分检查,引起电子健康档案(EHR)资料欠完整,而重症或逝去患者一般接受了全面的医疗监测,由此

数据愈发充裕,这种态势会引起数据分布不均衡,让模型更倾向于学习死亡患者的特征,难以精准鉴别高风险的轻症病患,且医疗数据有时因测量误差、输入错误而受影响,进而加大了数据偏差[24]。

4.2. 数据隐私问题

医疗数据因为关联就医体验对象隐私,一般需采取严格的保护行动,这造成数据共享面临诸多阻碍[25]。在多个医疗处所开展跨机构考察或利用大规模数据开展深度学习训练时,数据共享的管控往往是关键环节,为攻克这一难题,联邦学习作为一项创新手段顺势呈现。联邦学习的关键主张是准许不同机构在本地独自实施模型训练,而不是直接实行原始数据的互换行动,各机构可以在把牢数据隐私的前提下,依照共享模型参数来协同改进模型,实现跨机构的协同研讨[26]。差分隐私技术也在应对医疗数据共享时的隐私问题发挥了重要作用,通过在数据共享过程里面添加噪声,保证个体数据无法被辨识或倒推结论,从而进一步夯实了患者隐私保护[27]。

4.3. 数据标准化问题

不同医院采用不同的数据办法、编码样式和诊断规范,导致在推进跨机构数据整合及分析活动时,存在较大范围的数据不匹配情况[28],不仅加大了数据处理的繁杂程度,还可能对模型的泛化能力造成消极后果,引起模型在不同医院还是其他数据集上的表现差异突出,降低其通用性与精准度[29]。为攻克这一困境,可采用标准化的医疗数据矩阵,比如挑选 FHIR,该格式想要借助统一的标准达成不同医院之间的数据兼容性,确定采用 FHIR 可保证不同机构的医疗数据在结构、编码和语义方面的一致程度[30],进而促进跨机构的数据流通与模型迭代。数据映射与转换技术同样能展现出关键效能,特别是借助自然语言处理(NLP)技术来统一不同起始源头的医疗记录格式[31]。

5. 小结与展望

机器学习技术在肺炎患者死亡预测模型中展示出庞大潜力,可以辅助临床医生更精准地筛选高风险患者群体,优化疗愈路径,虽然当前依旧面临数据异质性、隐私保护和标准化等问题,但新兴技术范式(如联邦学习框架下的隐私保护计算、基于 Transformer 的多模态数据融合架构)正系统性突破这些瓶颈。未来需通过跨机构协作建立标准化医疗数据联盟,构建兼顾安全性与互通性的分布式数据平台,以此实现模型性能的持续优化与临床泛化能力的验证,最终为重症肺炎的精准防控提供循证医学决策支持,为公共卫生事业构建更坚实的防线。

参考文献

- [1] GBD 2017 Disease and Injury Incidence and Prevalence Collaborators (2018) Global, Regional, and National Incidence, Prevalence, and Years Lived with Disability for 354 Diseases and Injuries for 195 Countries and Territories, 1990-2017: A Systematic Analysis for the Global Burden of Disease Study 2017. *The Lancet*, **392**, 1789-1858.
- [2] Aliberti, S., Dela Cruz, C.S., Amati, F., Sotgiu, G. and Restrepo, M.I. (2021) Community-Acquired Pneumonia. *The Lancet*, **398**, 906-919. [https://doi.org/10.1016/s0140-6736\(21\)00630-9](https://doi.org/10.1016/s0140-6736(21)00630-9)
- [3] Dinh, A., Duran, C., Ropers, J., et al. (2021) Factors Associated with Treatment Failure in Moderately Severe Community-Acquired Pneumonia: A Secondary Analysis of a Randomized Clinical Trial. *JAMA Network Open*, **4**, e2129566.
- [4] 李重锦, 苏新星, 蒋丽君, 等. 重症肺炎患儿预后的危险因素及其列线图预测模型构建研究[J]. 实用心脑血管病杂志, 2021, 29(9): 47-52.
- [5] 殷菲. 影响重症肺炎患者预后的相关因素分析及构建 Nomogram 预测模型的价值研究[J]. 临床急诊杂志, 2020, 21(10): 819-825.
- [6] 刘佳琦, 曾炳亮, 陈会, 等. 基于 CT 影像组学建立肺孢子菌肺炎的预后模型[J]. 江西医药, 2023, 58(9): 1022-1025.
- [7] Collins, G.S., Moons, K.G.M., Dhiman, P., et al. (2024) TRIPOD+AI Statement: Updated Guidance for Reporting

- Clinical Prediction Models that Use Regression or Machine Learning Methods. *British Medical Journal*, **385**, e078378.
- [8] Wei, J., Cao, H., Peng, M., Zhang, Y., Li, S., Ma, W., *et al.* (2025) An Interpretable Machine Learning Model for Predicting In-Hospital Mortality in ICU Patients with Ventilator-Associated Pneumonia. *PLOS ONE*, **20**, e0316526. <https://doi.org/10.1371/journal.pone.0316526>
 - [9] 罗松平, 刘单霞, 韦兆吉, 等. 重症肺炎行有创机械通气患者 ICU 死亡的多因素分析及风险模型建立[J]. 中国急救医学, 2023, 43(4): 268-272.
 - [10] Riley, R.D., Ensor, J., Snell, K.I.E., Harrell, F.E., Martin, G.P., Reitsma, J.B., *et al.* (2020) Calculating the Sample Size Required for Developing a Clinical Prediction Model. *British Medical Journal*, **368**, m441. <https://doi.org/10.1136/bmj.m441>
 - [11] Chen, R., Chen, J., Yang, S., Luo, S., Xiao, Z., Lu, L., *et al.* (2023) Prediction of Prognosis in COVID-19 Patients Using Machine Learning: A Systematic Review and Meta-Analysis. *International Journal of Medical Informatics*, **177**, Article 105151. <https://doi.org/10.1016/j.ijmedinf.2023.105151>
 - [12] 金静怡, 王丽华, 季建红. ICU 呼吸机相关性肺炎风险预测模型的研究进展[J]. 当代护士(上旬刊), 2024, 31(8): 13-17.
 - [13] 苗若琪, 乔瑞萍. 基于数据挖掘技术的社区获得性肺炎患病风险预测模型的构建[J]. 郑州大学学报(医学版), 2023, 58(3): 306-310.
 - [14] Wang, B., Li, Y., Tian, Y., Ju, C., Xu, X. and Pei, S. (2023) Novel Pneumonia Score Based on a Machine Learning Model for Predicting Mortality in Pneumonia Patients on Admission to the Intensive Care Unit. *Respiratory Medicine*, **217**, Article 107363. <https://doi.org/10.1016/j.rmed.2023.107363>
 - [15] 蔡佩良, 刘超武, 朱振刚, 等. 基于人工智能构建非危重型新型冠状病毒肺炎的中医辅助决策模型[J]. 中国中医急症, 2024, 33(3): 399-404.
 - [16] 曾梦, 赵娜, 王显棋, 等. 基于胸部 CT 图像的肺炎深度学习分类预测模型[J]. 陆军军医大学学报, 2023, 45(21): 2266-2274.
 - [17] Xiang, B., Liu, Y., Jiao, S., Zhang, W., Wang, S. and Yi, M. (2024) Development and Validation of Interpretable Machine Learning Models for Postoperative Pneumonia Prediction. *Frontiers in Public Health*, **12**, Article 1468504. <https://doi.org/10.3389/fpubh.2024.1468504>
 - [18] 吴明琴, 王若雁, 张莉, 等. 肺超声、X 线联合实验室指标预测新型冠状病毒肺炎危重型患者预后的临床价值[J]. 临床超声医学杂志, 2024, 26(5): 370-374.
 - [19] 王彤, 季达峰. 老年重症肺炎的危险因素及诊断与治疗进展[J]. 现代医学与健康研究电子杂志, 2023, 7(24): 37-40.
 - [20] 李欣昱, 武铁群, 林连君. 社区获得性肺炎远期死亡风险的研究进展[J]. 国际老年医学杂志, 2023, 44(1): 99-101.
 - [21] 颜宇飞, 刘明利, 金一鸣, 等. 新型冠状病毒肺炎患者血液 NLR, DD 和 CRP 水平联合检测对死亡风险的评估研究[J]. 现代检验医学杂志, 2021, 36(1): 92-95+164.
 - [22] Misra, N.K., Das, S., Satpathy, S., Addula, S.R. and Trivedi, M.C. (2024) COVID-19 Pandemic: A Worldwide Critical Review with the Machine Learning Model-Based Prediction. *Journal of the Institution of Engineers (India): Series B*, **106**, 339-349. <https://doi.org/10.1007/s40031-024-01155-3>
 - [23] Rostami, A., Mousavi, F., Javadinia, S.A., Robotjazi, M. and Mehrpouyan, M. (2024) Predictive Value of Machine Learning Models in Mortality of Coronavirus Disease 2019 (COVID-19) Pneumonia. *International Journal of Computational Intelligence Systems*, **17**, Article No. 221. <https://doi.org/10.1007/s44196-024-00633-2>
 - [24] Alhassoon, K., Alhassoon, M.A., Alsunaydih, F., Alsaleem, F., Salim, O., Aly, S., *et al.* (2024) Machine Learning Predictive Modeling of the Persistence of Post-Covid19 Disorders: Loss of Smell and Taste as Case Studies. *Heliyon*, **10**, e35246. <https://doi.org/10.1016/j.heliyon.2024.e35246>
 - [25] Malin, B., Karp, D. and Scheuermann, R.H. (2010) Technical and Policy Approaches to Balancing Patient Privacy and Data Sharing in Clinical and Translational Research. *Journal of Investigative Medicine*, **58**, 11-18. <https://doi.org/10.2310/jim.0b013e3181c9b2ea>
 - [26] Brauneck, A., Schmalhorst, L., Kazemi Majdabadi, M.M., Bakhtiari, M., Völker, U., Baumbach, J., *et al.* (2023) Federated Machine Learning, Privacy-Enhancing Technologies, and Data Protection Laws in Medical Research: Scoping Review. *Journal of Medical Internet Research*, **25**, e41588. <https://doi.org/10.2196/41588>
 - [27] Moshawrab, M., Adda, M., Bouzouane, A., Ibrahim, H. and Raad, A. (2023) Reviewing Federated Machine Learning and Its Use in Diseases Prediction. *Sensors*, **23**, Article 2112. <https://doi.org/10.3390/s23042112>
 - [28] Wang, M., Li, W., Wang, H. and Song, P. (2024) Development and Validation of Machine Learning-Based Models for Predicting Healthcare-Associated Bacterial/Fungal Infections among COVID-19 Inpatients: A Retrospective Cohort

-
- Study. *Antimicrobial Resistance & Infection Control*, **13**, Article No. 42. <https://doi.org/10.1186/s13756-024-01392-7>
- [29] Wang, T., Zhao, Z., Li, W., Wu, J., Ye, Q. and Xie, H. (2023) Machine Learning Predictive Modeling for the Identification of Moderate Coronavirus Disease 2019 during the Pandemic: A Retrospective Study. *Cureus*, **15**, e50619e50619. <https://doi.org/10.7759/cureus.50619>
- [30] Ohno, Y., Aoki, T., Endo, M., Koyama, H., Moriya, H., Okada, F., *et al.* (2023) Machine Learning-Based Computer-Aided Simple Triage (CAST) for COVID-19 Pneumonia as Compared with Triage by Board-Certified Chest Radiologists. *Japanese Journal of Radiology*, **42**, 276-290. <https://doi.org/10.1007/s11604-023-01495-y>
- [31] Ismaeel, N.Q., Mohammed, H.J., Chalooob, I.Z., Kwekha-Rashid, A.S., Alhayani, B., Alkhayyat, A., *et al.* (2023) Application of Healthcare Management Technologies for COVID-19 Pandemic Using Internet of Things and Machine Learning Algorithms. *Wireless Personal Communications*, **122**. <https://doi.org/10.1007/s11277-023-10663-2>