

# 机器学习在心血管疾病中的临床应用进展

王苏淮<sup>1</sup> 李晶洁<sup>1,2</sup>

(1. 哈尔滨医科大学附属第一医院, 黑龙江 哈尔滨 150000; 2. 哈尔滨医科大学附属第一医院心内科, 黑龙江 哈尔滨 150000)

**【摘要】** 人工智能已从各个方面改变了人类的生活。机器学习是人工智能的子集, 其从大型数据库中通过提取模式来自动获取信息, 作为一种结合数据科学和统计技术的方法, 已越来越多地应用于医学界, 特别是心血管疾病领域。现阐述机器学习在心血管疾病不同方面的临床应用, 便于大家更好地理解其与现代医学的关联。

**【关键词】** 人工智能; 机器学习; 心血管疾病

**【DOI】** 10.16806/j.cnki.issn.1004-3934.2021.02.013

## Clinical Applications of Machine Learning in Cardiovascular Diseases

WANG Suhuai<sup>1</sup>, LI Jingjie<sup>1,2</sup>

(1. The First Affiliated Hospital of Harbin Medical University, Harbin 150000, Heilongjiang, China; 2. Department of Cardiology, The First Hospital of Harbin Medical University, Harbin 150000, Heilongjiang, China)

**【Abstract】** Artificial intelligence has transformed various aspects of human life. Machine learning, which is a subset of artificial intelligence that autonomously acquires information from large databases by extracting patterns. As a method combining data science and statistical techniques, it has been increasingly applied in the medical field, especially in the field of cardiovascular diseases. This review describes the clinical application of machine learning in different aspects of cardiovascular diseases for a better understanding of its relevance to modern medicine.

**【Key words】** Artificial intelligence; Machine learning; Cardiovascular diseases

### 1 机器学习概述

#### 1.1 简介

人工智能 (artificial intelligence, AI) 是研究计算机模拟人的思维过程和智能行为的学科, 已在多个领域得到了广泛应用。机器学习 (machine learning, ML) 作为 AI 的核心, 不仅涉及信息技术领域, 目前也逐渐成为医学领域的重要研究手段<sup>[1-6]</sup>。

#### 1.2 算法

##### 1.2.1 ML 算法分类

ML 主要分为监督学习和无监督学习。目前临床医学最常用的是监督学习算法, 通过学习一组包含输入对象 (通常为矢量) 和期望的输出值 (也称为监督信号) 的标签数据集, 产生具有推断功能的模型, 然后将模型应用于未知数据集, 达到分类或回归的目标 (当目标变量是分类变量时, 完成分类; 当目标变量是连续性数值时, 完成回归)。常用的监督学习算法包括 K-

近邻 (KNN)、支持向量机 (SVM)、决策树、随机森林 (RF)、逻辑性回归、提升法 (Boosting) 和人工神经网络 (ANN) 等<sup>[7-8]</sup>。无监督学习是在无标签的输入数据集中发现其自然关系和分组。常用的无监督学习算法包括 K 均值聚类、分层聚类和主成分分析等。

##### 1.2.2 ML 的一般流程

ML 的一般流程主要包括以下几个方面: (1) 数据采集; (2) 数据预处理; (3) 特征提取; (4) 特征选择; (5) 根据数据特征选择不同 ML 算法; (6) 构建 ML 模型并评估其性能<sup>[5]</sup>。输入的数据集通常被划分为训练、验证和测试集: 训练集包含大量变量, 用于模型的开发; 验证集用于微调模型的超参数; 测试集则输入到构建的 ML 模型中, 用于性能的评估<sup>[9]</sup>。

### 2 ML 在心血管疾病医学中的应用

在临床医学领域, ML 主要通过分析患者的复杂多维数据 (如电子病历记录<sup>[10-11]</sup>、实验室指标<sup>[12]</sup> 和辅

助检查成像<sup>[13-16]</sup>等),以非线性和高度交互的方式组合变量,对疾病进行诊断分类<sup>[11,17]</sup>、预测<sup>[10,18-19]</sup>和构建风险模型<sup>[9,20-21]</sup>以辅助医生决策<sup>[22-23]</sup>。

近几年关于 ML 与心血管疾病的研究成为热点,以下将对心血管病学中应用 ML 的几个领域进行阐述。

## 2.1 心电图

心电图 (electrocardiography, ECG) 信号是评估心脏功能最常用的临床工具之一,其优点是无创且价格低廉。临床实践中,ECG 信号通常由专业的电生理学家来判读,耗时且高度依赖于个人经验。在新的技术医疗时代,ML 可对 ECG 信号进行有效量化和分析,最大程度地从心电图数据集中提取信息以帮助临床医生评估患者风险<sup>[24]</sup>。

ML 通过对 ECG 信号进行处理、分割、特征提取和学习分析,达到心律分类或预测疾病的目的,最常用的 ML 算法是 ANN 和 SVM。如 Li 等<sup>[25]</sup>分析了 MIT-BIH 公共数据库的 360 组 ECG 数据集,通过小波包分解统计方法提取了 48 个 ECG 特征,然后使用遗传算法结合反向传播神经网络的 ML 方法,将心律分为六类:正常搏动(N)、左束支传导阻滞(L)、右束支传导阻滞(R)、房性期前收缩(A)、起搏心律(P)和室性期前收缩(V)。分类准确率为 97.8%,敏感性为 97.9%,特异性为 99.5%,阳性预测值为 97.8%。Yildirim 等<sup>[26]</sup>也分析了 MIT-BIH 数据库中 45 例患者的 1 000 个 ECG 片段,把原始 ECG 信号输入到 16 层深度卷积神经网络算法中,将心律分为 17 种,分类精度为 91.3%,分类时间为 0.015 s。近年来的研究多采用离散小波变化的方法对 ECG 信号进行形态特征的提取,如 Qin 等<sup>[27]</sup>利用此方法提取了 MIT-BIH 数据库中 107 049 次 ECG 搏动的形态特征,结合 ECG 信号的时域特征,使用优化后的 SVM 算法,完成了同上所述的六种心律分类:N、L、R、A、P 和 V,分类敏感性为 99.1%,特异性为 99.8%。

此外,除了心律分类外,也有研究通过分析 ECG 信号进行疾病的预测。如 Mjahad 等<sup>[28]</sup>将 ECG 时频图像数据直接输入到四种 ML 分类器(逻辑性回归、ANN、SVM 和装袋法),以预测心室颤动和室性心动过速的发生,心室颤动预测的敏感性为 95.6%,特异性为 98.8%;室性心动过速预测的敏感性为 88.8%,特异性为 99.5%。Ebrahimzadeh 等<sup>[29]</sup>则从心房颤动预测数据库中选取了 53 对 ECG 记录的 106 个信号,对心率变异性信号进行了特征提取,结合四种 ML 算法预测心房颤动的发生,其中最优的混合专家分类法预测敏感性、特异性和准确性分别为 100%、95.6% 和 98.2%。同时,ECG 在冠心病方面的应用也逐渐增

加。如 Park 等<sup>[30]</sup>选取了欧洲 ST-T 数据库中 90 个 ECG 记录的 367 个缺血性 ST 段,提取了 3 个特征,使用核密度估计和 SVM 算法预测心肌缺血,其中最优性能 SVM 的敏感性和特异性分别为 94.1% 和 92.3%。另有 Simjanoska 等<sup>[31]</sup>首次提出应用 ECG 信号特征结合七种 ML 算法预测收缩压、舒张压和平均动脉压;Sengupta 等<sup>[32]</sup>利用信号处理过的心电图提取特征,输入至 RF 分类器中预测心肌异常松弛,预测敏感性和特异性分别为 80.0% 和 84.0%,曲线下面积(AUC)为 0.91(95% CI 0.86~0.95)。以上研究都说明了 ECG 结合 ML 算法在心血管疾病领域拥有着广泛的应用前景。

## 2.2 成像

图片成像作为心血管疾病诊断和指导治疗的辅助工具在临床上无处不在,ML 算法与检查成像的结合进一步提高了诊疗的准确性和及时性。

Narula 等<sup>[33]</sup>使用斑点超声心动图数据,结合三个 ML 算法(SVM、RF 和 ANN)构建模型以鉴别肥厚型心肌病和运动员生理性肥厚。研究纳入 139 例男性受试者,77 例生理性和 62 例肥厚型心肌病患者,使用信息增益进行特征选择,确定左心室容积,左心室中段纵向应变、平均纵向应变和左心室中段径向应变为最佳预测指标。Khamis 等<sup>[34]</sup>证明采用时空特征提取和监督字典学习方法可提高心脏超声图的分类精度。作者研究了 309 个由专家标记过的心脏超声图,使用 ML 算法对心尖二腔室图,心尖四腔室图和心尖长轴室图的分类准确率分别为 97.0%、91.0% 和 97.0%。

此外,Mannil 等<sup>[35]</sup>开发了基于心脏计算机断层摄影成像数据的 ML 模型以预测心肌梗死,使用纹理分析的方法对图像数据进行量化,采用 KNN 算法得到了最佳效能(敏感性 69.0%,特异性 85.0%,假阳性率 15.0%,AUC 值 0.78)。Arsanjani 等<sup>[36]</sup>则通过结合临床数据和心肌灌注断层显像的定量图像数据作为特征,输入到 Boosting 算法中预测疑似冠心病患者的血运重建,预测敏感性为 73.6%±4.3%,特异性为 74.7%±4.2%,AUC 为 0.81±0.02。

ML 与多种辅助检查成像的结合使用,可帮助初学者做出更精准的临床诊断并有望应用到医疗资源紧缺的地方。

## 2.3 风险预测模型

因 ML 使用的非线性和交叉学习模式弥补了传统分析线性和主观选择特征的不足,已被广泛应用于构建临床风险预测模型。在心血管疾病领域,ML 主要致力于预测患者的死亡率、再入院率和不良临床事件的发生率。如 Motwani 等<sup>[21]</sup>研究了 10 030 例疑似冠

心病患者的 25 个临床和 44 个冠状动脉 CT 血管成像参数,进行信息增益法则排名,选取重要特征,然后输入至 Boosting 集成算法预测患者 5 年全因死亡率,ML 预测模型的 AUC 为 0.79,优于传统的弗雷明汉风险评分等评估方法。Barrett 等<sup>[37]</sup>从重症监护医学信息中心(MIMIC-III)数据库中获取了 5 037 例由急性心肌梗死收入 ICU 的患者数据表,结合逻辑回归模型预测了患者一年死亡率,预测准确度为 85.1%,AUC 为 0.90。Frizzell 等<sup>[38]</sup>对 56 477 例心力衰竭患者数据进行变量选择后,使用了贝叶斯网络、RF、梯度增强、逻辑性回归和 LASSO 回归五种 ML 算法,预测心力衰竭患者 30 d 再入院率,预测的 C 统计量分别为 0.62、0.61、0.61、0.62 和 0.62,与传统的预测模型相比无明显优势。Wu 等<sup>[20]</sup>筛选了 508 例年轻高血压患者的 58 个变量,通过递归特征消除法选择 11 个重要变量,结合极端梯度增强法构建了预测临床不良后果事件的 ML 模型,并将模型的性能与传统 Cox 回归和弗雷明汉风险评分模型相比,C 统计量分别是 0.76(95% CI 0.66 ~ 0.85)、0.72(95% CI 0.64 ~ 0.81)和 0.53(95% CI 0.40 ~ 0.66)。

ML 通过分析大量丰富的临床变量来评估患者风险,往往能收获性能更优的预测模型,这对于临床医生决策和改善患者预后都具有重要意义。

### 3 ML 的局限性

尽管 ML 在医学上得到了广泛应用,但其仍有局限性。首先,ML 需大量数据,在医学中获取大量标签数据可能具有挑战性;其次,ML 算法使用人类收集和标记的数据进行训练,由于人为的主观选择性可能出现选择偏倚;最后,许多 ML 算法如 RF,工作时像一个“黑匣子”,计算过程难以被全面理解,导致医生难以完全相信基于 ML 推荐的意见,也很难发现其错误的推荐<sup>[39]</sup>。

### 4 ML 的展望与未来

合理高效地使用 ML 可实现高精度的自动诊断及预测疾病转归,以此辅助临床医生决策,在节省大量时间的同时可降低临床错诊率。因此,在医疗系统中对 ML 的需求势必会继续增长。

#### 参考文献

- [1] Schwalbe N, Wahl B. Artificial intelligence and the future of global health[J]. *Lancet*, 2020, 395(10236):1579-1586.
- [2] Freedman DH. Hunting for new drugs with AI[J]. *Nature*, 2019, 576(7787):S49-S53.
- [3] Rajkomar A, Dean J, Kohane I. Machine learning in medicine[J]. *N Engl J Med*, 2019, 380(14):1347-1358.
- [4] Beam AL, Kohane IS. Big data and machine learning in health care[J]. *JAMA*, 2018, 319(13):1317-1318.
- [5] Alfaref SJ, Anchouche K, Singh G, et al. Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging[J]. *Eur Heart J*, 2019, 40(24):1975-1986.
- [6] Sevakula RK, Au-Yeung WM, Singh JP, et al. State-of-the-art machine learning techniques aiming to improve patient outcomes pertaining to the cardiovascular system[J]. *J Am Heart Assoc*, 2020, 9(4):e013924.
- [7] Bzdok D, Krzywinski M, Altman N. Machine learning: supervised methods[J]. *Nat Methods*, 2018, 15(1):5-6.
- [8] Feldmann J, Youngblood N, Wright CD, et al. All-optical spiking neurosynaptic networks with self-learning capabilities[J]. *Nature*, 2019, 569(7755):208-214.
- [9] Chen PC, Liu Y, Peng L. How to develop machine learning models for healthcare[J]. *Nat Mater*, 2019, 18(5):410-414.
- [10] Mahmoudi E, Kamdar N, Kim N, et al. Use of electronic medical records in development and validation of risk prediction models of hospital readmission: systematic review[J]. *BMJ*, 2020, 369:m958.
- [11] Liang H, Tsui BY, Ni H, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence[J]. *Nat Med*, 2019, 25(3):433-438.
- [12] Goldstein BA, Navar AM, Carter RE. Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges[J]. *Eur Heart J*, 2017, 38(23):1805-1814.
- [13] 渠海贤, 李涛, 程流泉. 人工智能在心脏磁共振成像中的应用进展[J]. *心血管病学进展*, 2019, 40(5):659-662.
- [14] Ting DSW, Cheung CY, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes[J]. *JAMA*, 2017, 318(22):2211-2223.
- [15] Vernooij MW. A data-driven update of arterial perfusion territories[J]. *Nat Rev Neurol*, 2019, 15(11):624-625.
- [16] Hinton G. Deep learning—A technology with the potential to transform health care[J]. *JAMA*, 2018, 320(11):1101-1102.
- [17] Lynch CJ, Liston C. New machine-learning technologies for computer-aided diagnosis[J]. *Nat Med*, 2018, 24(9):1304-1305.
- [18] Hyland SL, Faltys M, Hüser M, et al. Early prediction of circulatory failure in the intensive care unit using machine learning[J]. *Nat Med*, 2020, 26(3):364-373.
- [19] Chen JH, Asch SM. Machine learning and prediction in medicine—beyond the peak of inflated expectations[J]. *N Engl J Med*, 2017, 376(26):2507-2509.
- [20] Wu X, Yuan X, Wang W, et al. Value of a machine learning approach for predicting clinical outcomes in young patients with hypertension[J]. *Hypertension*, 2020, 75(5):1271-1278.
- [21] Motwani M, Dey D, Berman DS, et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis[J]. *Eur Heart J*, 2017, 38(7):500-507.
- [22] Obermeyer Z, Lee TH. Lost in thought—The limits of the human mind and the future of medicine[J]. *N Engl J Med*, 2017, 377(13):1209-1211.
- [23] Asselbergs FW, Meijboom FJ. Big data analytics in adult congenital heart disease: why coding matters[J]. *Eur Heart J*, 2019, 40(13):1078-1080.
- [24] Mincholé A, Camps J, Lyon A, et al. Machine learning in the electrocardiogram[J]. *J Electrocardiol*, 2019, 57S:S61-S64.
- [25] Li H, Yuan D, Ma X, et al. Genetic algorithm for the optimization of features and neural networks in ECG signals classification[J]. *Sci Rep*, 2017, 7:41011.
- [26] Yildirim Ö, Plawiak P, Tan RS, et al. Arrhythmia detection using deep convolutional neural network with long duration ECG signals[J]. *Comput Biol Med*, 2018, 102:411-420.
- [27] Qin Q, Li J, Zhang L, et al. Combining low-dimensional wavelet features and support vector machine for arrhythmia beat classification[J]. *Sci Rep*, 2017, 7(1):6067.

- [28] Mjahad A, Rosado-Muñoz A, Bataller-Mompeán M, et al. Ventricular fibrillation and tachycardia detection from surface ECG using time-frequency representation images as input dataset for machine learning[J]. *Comput Methods Programs Biomed*, 2017, 141:119-127.
- [29] Ebrahimzadeh E, Kalantari M, Joulani M, et al. Prediction of paroxysmal atrial fibrillation: a machine learning based approach using combined feature vector and mixture of expert classification on HRV signal[J]. *Comput Methods Programs Biomed*, 2018, 165:53-67.
- [30] Park J, Pedrycz W, Jeon M. Ischemia episode detection in ECG using kernel density estimation, support vector machine and feature selection[J]. *Biomed Eng Online*, 2012, 11:30.
- [31] Simjanoska M, Gjoreski M, Gams M, et al. Non-invasive blood pressure estimation from ECG using machine learning techniques[J]. *Sensors (Basel)*, 2018, 18(4):1160.
- [32] Sengupta PP, Kulkarni H, Narula J. Prediction of abnormal myocardial relaxation from signal processed surface ECG[J]. *J Am Coll Cardiol*, 2018, 71(15):1650-1660.
- [33] Narula S, Shameer K, Salem Omar AM, et al. Machine-learning algorithms to automate morphological and functional assessments in 2D echocardiography[J]. *J Am Coll Cardiol*, 2016, 68(21):2287-2295.
- [34] Khamis H, Zurakhov G, Azar V, et al. Automatic apical view classification of echocardiograms using a discriminative learning dictionary[J]. *Med Image Anal*, 2017, 36:15-21.
- [35] Mannil M, von Spiczak J, Manka R, et al. Texture analysis and machine learning for detecting myocardial infarction in noncontrast low-dose computed tomography: unveiling the invisible[J]. *Invest Radiol*, 2018, 53(6):338-343.
- [36] Arsanjani R, Dey D, Khachatryan T, et al. Prediction of revascularization after myocardial perfusion SPECT by machine learning in a large population[J]. *J Nucl Cardiol*, 2015, 22(5):877-884.
- [37] Barrett LA, Payrovnaziri SN, Bian J, et al. Building computational models to predict one-year mortality in ICU patients with acute myocardial infarction and post myocardial infarction syndrome[J]. *AMIA Jt Summits Transl Sci Proc*, 2019, 2019:407-416.
- [38] Frizzell JD, Liang L, Schulte PJ, et al. Prediction of 30-day all-cause readmissions in patients hospitalized for heart failure: comparison of machine learning and other statistical approaches[J]. *JAMA Cardiol*, 2017, 2(2):204-209.
- [39] Watson DS, Krutzinna J, Bruce IN, et al. Clinical applications of machine learning algorithms: beyond the black box[J]. *BMJ*, 2019, 364:l886.

收稿日期:2020-06-09

## (上接第 137 页)

- [12] Flett AS, Hayward MP, Ashworth MT, et al. Equilibrium contrast cardiovascular magnetic resonance for the measurement of diffuse myocardial fibrosis: preliminary validation in humans[J]. *Circulation*, 2010, 122(2):138-144.
- [13] Pagano JJ, Chow K, Khan A, et al. Reduced right ventricular native myocardial T1 in Anderson-Fabry disease: comparison to pulmonary hypertension and healthy controls[J]. *PLoS One*, 2016, 11(6):e0157565.
- [14] Patel RB, Li E, Benefield BC, et al. Diffuse right ventricular fibrosis in heart failure with preserved ejection fraction and pulmonary hypertension[J]. *ESC Heart Fail*, 2020, 7(1):253-263.
- [15] Reiter U, Reiter G, Kovacs G, et al. Native myocardial T1 mapping in pulmonary hypertension: correlations with cardiac function and hemodynamics[J]. *Eur Radiol*, 2017, 27(1):157-166.
- [16] Swift AJ, Rajaram S, Condliffe R, et al. Diagnostic accuracy of cardiovascular magnetic resonance imaging of right ventricular morphology and function in the assessment of suspected pulmonary hypertension results from the ASPIRE registry[J]. *J Cardiovasc Magn Reson*, 2012, 14(1):40.
- [17] Saunders LC, Johns CS, Stewart NJ, et al. Diagnostic and prognostic significance of cardiovascular magnetic resonance native myocardial T1 mapping in patients with pulmonary hypertension[J]. *J Cardiovasc Magn Reson*, 2018, 20(1):78.
- [18] Sanz J, Dellegrottaglie S, Kariisa M, et al. Prevalence and correlates of septal delayed contrast enhancement in patients with pulmonary hypertension[J]. *Am J Cardiol*, 2007, 100(4):731-735.
- [19] Roller FC, Wiedenroth C, Breithecker A, et al. Native T1 mapping and extracellular volume fraction measurement for assessment of right ventricular insertion point and septal fibrosis in chronic thromboembolic pulmonary hypertension[J]. *Eur Radiol*, 2017, 27(5):1980-1991.
- [20] Wang J, Zhao H, Wang Y, et al. Native T1 and T2 mapping by cardiovascular magnetic resonance imaging in pressure overloaded left and right heart diseases[J]. *J Thorac Dis*, 2018, 10(5):2968-2975.
- [21] Shehata ML, Lossnitzer D, Skrok J, et al. Myocardial delayed enhancement in pulmonary hypertension: pulmonary hemodynamics, right ventricular function, and remodeling[J]. *AJR Am J Roentgenol*, 2011, 196(1):87-94.
- [22] Spruijt OA, Vissers L, Bogaard HJ, et al. Increased native T1-values at the interventricular insertion regions in precapillary pulmonary hypertension[J]. *Int J Cardiovasc Imaging*, 2016, 32(3):451-459.
- [23] Chen YY, Yun H, Jin H, et al. Association of native T1 times with biventricular function and hemodynamics in precapillary pulmonary hypertension[J]. *Int J Cardiovasc Imaging*, 2017, 33(8):1179-1189.
- [24] Shiina Y, Taniguchi K, Nagao M, et al. The relationship between extracellular volume fraction in symptomatic adults with tetralogy of Fallot and adverse cardiac events[J]. *J Cardiol*, 2020, 75(4):424-431.
- [25] Habert P, Capron T, Hubert S, et al. Quantification of right ventricular extracellular volume in pulmonary hypertension using cardiac magnetic resonance imaging[J]. *Diagn Interv Imaging*, 2020, 101(5):311-320.
- [26] Rain S, Handoko ML, Trip P, et al. Right ventricular diastolic impairment in patients with pulmonary arterial hypertension[J]. *Circulation*, 2013, 128(18):2016-2025, 1-10.
- [27] Tello K, Dalmer A, Axmann J, et al. Reserve of right ventricular-arterial coupling in the setting of chronic overload[J]. *Circ Heart Fail*, 2019, 12(1):e005512.
- [28] Freed BH, Gomberg-Maitland M, Chandra S, et al. Late gadolinium enhancement cardiovascular magnetic resonance predicts clinical worsening in patients with pulmonary hypertension[J]. *J Cardiovasc Magn Reson*, 2012, 14(1):11.
- [29] Swift AJ, Rajaram S, Capener D, et al. LGE patterns in pulmonary hypertension do not impact overall mortality[J]. *JACC Cardiovasc Imaging*, 2014, 7(12):1209-1217.
- [30] Sato T, Tsujino I, Ohira H, et al. Paradoxical interventricular septal motion as a major determinant of late gadolinium enhancement in ventricular insertion points in pulmonary hypertension[J]. *PLoS One*, 2013, 8(6):e66724.
- [31] Roller FC, Kriebbaum S, Breithecker A, et al. Correlation of native T1 mapping with right ventricular function and pulmonary haemodynamics in patients with chronic thromboembolic pulmonary hypertension before and after balloon pulmonary angioplasty[J]. *Eur Radiol*, 2019, 29(3):1565-1573.

收稿日期:2020-07-11