

机器学习在新型冠状病毒感染诊疗中的应用研究进展

郭桦^{1,2}, 丁俊谕^{1,2}, 刘长鑫^{1,2}, 张侃², 马琳², 王博^{1,2}, 赵慧珺^{1,2}, 宋曼雅^{1,2}, 管希周^{2*}

¹解放军医学院, 北京 100853; ²解放军总医院第八医学中心呼吸与危重症医学部, 北京 100091

[中图分类号] R563.1 [文献标志码] A [DOI] 10.11855/j.issn.0577-7402.2022.1214

[声明] 本文所有作者声明无利益冲突

[引用本文] 郭桦, 丁俊谕, 刘长鑫, 等. 机器学习在新型冠状病毒感染诊疗中的应用研究进展[J]. 解放军医学杂志, 2023, 48(7): 863-870.

[收稿日期] 2021-12-21 [录用日期] 2022-08-12 [上线日期] 2022-12-14

[摘要] 新型冠状病毒感染(COVID-19)暴发以来, 机器学习被广泛应用于其流行趋势的预测、高风险人群的筛查与追踪、早期诊断与监护等, 显著提高了疫情期间信息的处理效率, 为临床医师提供了高效的决策支持, 但因用于开发模型的数据类型与规模、训练方法等不同, 执行诊断或预后任务的模型存在不同的局限性。本文从机器学习结合影像资料、实验室检查结果, 以及整合这两部分数据所训练的模型综述机器学习在COVID-19诊断与预后中的应用, 以期机器学习的训练与应用提供更贴合实践的思路。

[关键词] 新型冠状病毒感染; 诊断; 预后; 机器学习

Application progress of machine learning in diagnosis and treatment of the coronavirus disease 2019

Guo Hua^{1,2}, Ding Jun-Yu^{1,2}, Liu Chang-Xin^{1,2}, Zhang Kan², Ma Lin², Wang Bo^{1,2}, Zhao Hui-Jun^{1,2}, Song Man-Ya^{1,2}, Guan Xi-Zhou^{2*}

¹Medical School of Chinese PLA, Beijing 100853, China

²Department of Respiratory and Critical Care Medicine, the Eighth Medical Center of Chinese PLA General Hospital, Beijing 100091, China

*Corresponding author, E-mail: gxz301@126.com

[Abstract] Since the outbreak of the coronavirus disease 2019 (COVID-19), machine learning has been widely used in forecasting the epidemic trend of COVID-19, screening and tracking high-risk people, early diagnosis and monitoring of patients, etc., which has greatly improved the efficiency of information processing during the epidemic period and provided efficient decision support for clinicians. However, due to the different data types and scales and training methods used to develop models, machine learning that perform diagnosis or prognosis tasks also have different limitations. This review introduces the application of machine learning in the diagnosis and prognosis of COVID-19 from the aspects of machine learning combined with imaging data, laboratory results, and the model trained by integrating these two aspects, trying to provide more practical ways for machine learning in training and application.

[Key words] coronavirus disease 2019; diagnosis; prognosis; machine learning

自2019年12月中国湖北省武汉市报告首例新型冠状病毒感染(coronavirus disease 2019, COVID-19)以来, 疫情已在全球蔓延。至2022年8月, 新型冠状病毒(SARS-CoV-2)感染超过5亿人, 累计死亡病例超过600万例^[1]。研究发现, 多数COVID-19患者的潜伏期为3~7 d, 最常见的症状为发烧、咳嗽、嗅/味觉丧失及肌肉酸痛等^[2-3]。实验室检查结果显

示, 多数COVID-19患者具有淋巴细胞和白细胞减少以及C反应蛋白升高表现, COVID-19的临床表现与其他病毒性肺炎均非常相似, 部分患者最初只有轻微的非典型症状, 甚至没有症状, 临床和实验室特征难以将其与其他常见呼吸道病原体引起的肺炎区分, 显著增加了诊断难度, 给有限的医疗资源带来了巨大负担^[3-4]。而机器学习在COVID-19肺炎疫

[作者简介] 郭桦, 硕士研究生, 主要从事肺部感染与大数据利用方面的研究

[通信作者] 管希周, E-mail: gxz301@126.com

情中的应用,如流行趋势预测、高风险人群筛查与追踪、早期诊断与监护等,显著提高了疫情期间信息的处理效率,为临床医师提供了准确高效的决策支持。机器学习的常见类型包括无监督学习、半监督学习、监督学习及强化学习等,临床资料与影像资料是其常用的数据模态。构建机器学习模型常采用随机森林、决策树、支持向量机等方法,而卷积神经网络(convolutional neural networks, CNN)则被广泛用于对胸部CT影像资料或磁共振图像的学习^[5]。本文从机器学习结合影像资料和临床资料,以及整合这两部分数据所训练的模型介绍机器学习在COVID-19诊断与预后中的应用,以期对机器学习的临床应用实践提供新思路。

1 机器学习在COVID-19诊断中的应用

1.1 机器学习与影像资料结合 SARS-CoV-2具有快速传播性,早期准确诊断对于防治COVID-19具有重要意义^[6-7]。目前,核酸检测作为COVID-19检测的金标准,受样本采集部位、采集时间的影响,其敏感度仅为79%^[8-9],一定程度的假阴性意味着无法对COVID-19患者及密切接触者进行早期隔离和治疗,进一步增加了传播风险。Santosh^[10]指出,合理利用多模态数据可预测COVID-19在全球范围内的流行趋势。有研究发现,将美国普通人群的基线数据如年龄、性别、既往住院次数、Charlson合并症指数等结合,可预测普通人群罹患COVID-19的风险^[11]。

胸部CT是肺炎诊断的常规影像工具,多数COVID-19患者具有特征性的CT表现,如不同程度的磨玻璃影、浸润影,以及肺叶间隔增厚、胸膜增厚等^[12-14]。有研究利用贝叶斯模型分析病毒感染和临床症状开始出现后RT-PCR假阴性率随时间变化的情况,发现在典型症状出现前4 d,RT-PCR假阴性率逐天变化,且在阴性人群中观察到了COVID-19的影像特征^[15-17]。一项研究纳入1014例患者,评估胸部CT对COVID-19的诊断价值,发现在阳性人群中胸部CT诊断COVID-19的敏感度为97%,而75%的阴性受检者胸部CT也提示COVID-19,其中81%被认定为疑似病例^[18]。胸部CT具有较高的敏感度,可作为COVID-19的筛查方法^[19-20],同时补充核酸检测结果,是诊断COVID-19的重要证据。然而疑似患者激增给影像诊断带来了巨大压力,机器学习正是应对这一挑战的方法。近年来,具有高效分割分析能力的机器学习与影像资料结合,已经成为医疗决策的重要辅助工具,如脑出血的分型诊断、肺结核分类、乳腺癌的早期诊断等^[21-23]。

将胸部CT与机器学习算法结合构建模型,有助于发热门诊快速分诊疑似病例^[24],高效识别COVID-19的影像特征,定位病灶,在病原学证据之前提供临床诊断^[6,25]。随着疫情进展,用于开发模型的临床资料逐渐丰富,机器学习执行的功能已从最初的筛查发展到与其他肺炎或典型病毒性肺炎的鉴别诊断。有研究采用4352例患者的胸部CT影像资料用于训练模型COV-Net,该模型具有区分COVID-19与社区获得性肺炎的能力,其敏感度、特异度分别为90%、96%^[26]。利用多中心的胸部X线片和CT资料构建的诊断系统可鉴别诊断COVID-19、甲型/乙型流感、非病毒性社区获得性肺炎,且受试者工作特征曲线下面积(AUC)为97.81%^[27]。

在模型的训练过程中,合理选择与COVID-19相关的显著性特征对模型的准确率影响较大,特征的筛选方法多样,适当的筛选方法可降低冗余度,提高分类任务的性能。部分研究将磨玻璃区域作为区分普通肺炎与COVID-19的有效特征,例如,Song等^[28]提出DR-Net可专注于辨认评估磨玻璃混浊区域的特征;Liu等^[7]在分割磨玻璃区域的基础上采用ReliefF算法,该算法的分类准确率为94.16%;Kang等^[29]利用一种学习潜在表征的方法,以189维影像特征构建模型来鉴别诊断COVID-19与社区获得性肺炎;Wang等^[30]将优先注意机制与残差学习相结合以提高特征筛选的性能;也有研究将随机森林与决策树结合构建感染区域面积感知模型,该模型可将CT影像按照决策树分成不同病灶大小范围的组,然后再使用随机森林进行分类,减少了感染面积差异较大对特征提取造成的偏差^[31]。此外,双向对抗网络也被用来训练和提取COVID-19的放射组学特征,以增加机器学习算法的可解释性^[32]。

除了特征选择,在COVID-19的鉴别分类任务中,对感染区域的分割方法也很重要。CT中的肺部感染区域面积和边界是鉴别诊断的关键特征。Shan等^[33]对比了三维卷积神经网络(VB-NET)与人工勾画感染区域的结果,对两种方法的Dice相似系数(Dice similarity coefficient)以及所判定感染区域的体积差异、感染面积百分比进行比较,量化了基于机器学习分割出的感染区域。Xu等^[34]采用位置注意力机制对分割出的感染区域进行分类,再用Noisy-OR或贝叶斯函数计算分类的置信度确定主要的感染类型,其鉴别诊断COVID-19、甲型流感病毒肺炎和健康对照者的准确率为86.7%。Wu等^[35]采用阈值分割和形态学优化算法进行分割,所训练的基于ResNet-50的深度学习模型在测试集中获得的准确率、敏感度和特异度分别为0.760、0.811和0.615。

机器学习模型往往需要大量的临床数据才能减少过度拟合,而迁移学习算法可以用有限的数据实现更好的分类效果^[36],利用已有的COVID-19影像和对抗性生成算法可生成更多具有类似特征的CT图像^[37]。基于残差网络的迁移学习算法(ResNet-50)可鉴别诊断COVID-19、细菌性和病毒性肺炎、健康对照者^[38],同样利用ResNet-50的快速COVID-19分类网络的敏感度和特异度优于VGG16、Inception-V3或Xception算法^[39]。Gifani等^[40]比较了15种训练好的CNN架构,选择其中5种在迁移学习背景下构建分类集成器,其效能优于单独的学习模型。另有研究提出了一种基于剪枝高效网络模型的迁移学习方法用于COVID-19的诊断,在构建过程中对所训练的模型进行插值,可提高预测结果的可解释性^[41]。

1.2 机器学习与临床资料结合 部分研究中,基于机器学习的临床特征分析也执行了分类诊断功能,无需使用RT-PCR或CT资料。早期临床特征虽不能全面反映COVID-19的免疫过程,但使用机器学习算法处理常见的实验室指标可为医护人员提供快速准确的诊断意见,尤其是对于社区医院及全科医院可以更好地控制院内感染,降低传播风险。将一些非实验室数据如年龄、性别、血压、基础疾病以及是否与COVID-19确诊患者接触等用于构建模型,来确定门诊就诊者的分诊程序,有利于社区医院和发热门诊的医疗资源分配^[42-43]。

味觉异常、发热、呼吸困难、咳嗽、肌痛/关节痛和腹泻等临床症状在RT-PCR阳性与阴性人群中差异显著,Wagner等^[44]采用机器学习确定了RT-PCR检测前1周,咳嗽、发热与味觉异常可共同构成COVID-19的鉴别特征。有研究将COVID-19的4种流行病学特征和6种临床表现分别用于logistic回归、支持向量机、决策树、随机森林4种机器学习方法进行早期诊断,结果表明,logistic回归预测模型的特异度为0.95,AUC为0.971,最适合筛查早期COVID-19^[45]。Li等^[46]关注确诊患者的免疫相关指标如白细胞、中性粒细胞、淋巴细胞、C反应蛋白等,用SOM算法对具有不同免疫特征和临床表现的患者进行分组,构建XGBoost模型进行诊断分类。Meng等^[47]选择9个变量(年龄、活化部分凝血活酶时间、红细胞分布宽度SD、尿酸、甘油三酸酯、血清钾、白蛋白/球蛋白、3-羟丁酸、血钙)训练诊断模型,并对模型进行外部验证,其AUC为0.872。

此外,全面整合诊断元素(如临床表现、CT影像、实验室检查结果等)可提高模型诊断的准确率。有研究将胸部CT与临床症状、接触史和实验

室检查结果相结合,构建了一个综合临床信息的诊断模型,通过比较AUC发现,整合CT影像和相关临床信息的联合模型的性能优于仅利用CT影像或临床信息训练的模型^[48]。Xu等^[49]将参与者的临床信息、实验室检查结果和入院时CT作为3个输入特征模态开发了多模态的机器学习模型,并用其区分非严重COVID-19、严重COVID-19、健康个体和病毒性肺炎。Song等^[50]将影像特征、中性粒细胞/淋巴细胞比值、T细胞、性别纳入诊断模型,并添加年龄和有意义的呼吸道症状构建COVID-19预警评分以对重症高风险患者进行提前预警,该研究还获得了疾病严重程度预测指标如CT评分、CD8⁺T细胞计数、CD4⁺T细胞计数等的临界值。

2 机器学习在COVID-19预后中的应用

2.1 机器学习与影像资料结合 核酸检测作为定性指标,无法判断COVID-19患者的严重程度,难以提前预测患者是否应转入重症监护室。虽然大多数患者临床症状轻微,预后良好,但重症患者会出现急性呼吸窘迫综合征、感染性休克甚至死亡,因此在医疗资源有限的情况下对高危患者进行早期预警非常重要。疫情暴发初期,患者激增,及时准确地评估并发现预后较差的高危患者是控制和管理COVID-19、改善预后及降低病死率的关键。

研究发现,多次重复CT检查可准确反映肺炎进展、监测治疗效果,随访发现42%的患者胸部CT改善时间早于RT-PCR转阴时间,且以铺路石征为主的持续性进展是COVID-19死亡患者的主要表现^[51-52]。Tang等^[53]将全肺感染体积比、磨玻璃区域特征等63项定量指标纳入随机森林模型,评估176例COVID-19患者的严重程度,准确率为0.875,AUC为0.910。Li等^[54]利用Dice系数比较人工智能自动分割与放射科医师人工标记的准确性,以两个影像标志即感染面积和平均CT值评估疾病的严重程度及进展情况,发现人工智能系统与放射科医师在评估感染区域方面取得了很好的一致性。也有研究将2D与3D分析相结合,用子系统A对感染区域体积进行三维分析,子系统B检测并定位较大的弥漫性浸润,用CNN对感染区域进行学习分类后,提出了评分系统可衡量随着时间推移疾病的进展情况^[55]。

2.2 机器学习与临床资料结合 康复者与死亡者的流行病学、临床特征和实验室特征显著不同,高龄和住院时间长是COVID-19患者死亡的两个主要危险因素,呼吸困难、胸闷和意识障碍在死亡患者中更常见,且死亡者的丙氨酸氨基转移酶、天冬氨酸氨基转移酶、肌酐和D-二聚体水平明显高于康复者^[56-57]。机器学习与COVID-19患者的临床资

料结合为多模态数据集的方法被多项研究用于预测COVID-19患者的预后。有研究深度学习入院时患者的信息(包括人口学资料、合并症、吸烟史和症状)来预测病情进展和死亡风险^[58-59]。1987例确诊为COVID-19且未入住ICU患者的生命体征、实验室数据和心电图资料被作为随机森林模型的训练变量,结果显示该模型可预测24 h内可能需要转入ICU的患者,其敏感度为72.8%,特异度为76.3%,准确率为76.2%^[60]。此外,地理位置、旅行史、健康和人口学数据也可用于预测病情严重程度。Iwendi等^[61]将AdaBoost算法与随机森林模型结合获得的准确率为94%,F1得分为0.86(F1值是对诊断分类模型精确率与召回率的一种调和平均)。有研究者使用深度学习分析SARS-CoV-2感染和COVID-19病情严重程度的生物组学,并开发了一种建立在注意力学习机制上的机器学习模型,利用监督学习提取特征,应用于患者支气管肺泡灌洗液样本与RNA序列,以识别不同临床分型患者的基因和细胞特点^[62]。

在预测病情进展的同时,机器学习模型也获得了与预后相关的预测因子。有学者利用人口学资料、合并症、生命体征、实验室检查等78个临床变量结合深度学习算法构建的风险分层系统可预测患者的死亡率,并证实了与死亡率最相关的5个预测因子为D-二聚体、氧合指数、中性粒细胞/淋巴细胞比值、C反应蛋白和乳酸脱氢酶^[63]。另有研究发现,利用乳酸脱氢酶、淋巴细胞和高敏C反应蛋白构建的模型预测患者死亡风险的准确率超过90%^[64]。Ikemura等^[65]收集患者在核酸阳性36 h内的47项生物标志物以及1个月的预后情况,

筛选出血压、年龄、呼吸频率、血氧饱和度、尿素氮、乳酸脱氢酶、D-二聚体、肌钙蛋白和血糖等与死亡率相关的重要生物标志物。Hu等^[66]分析183例重症COVID-19患者的人口学资料、临床特征和入院后首次实验室检查结果,使用5种机器学习方法(logistic回归、偏最小二乘回归、弹性网络、随机森林和柔性判别法)选择特征并预测预后,结果显示,5个模型均认为4个相关变量(年龄、超敏C反应蛋白、淋巴细胞计数和D-二聚体)与患者预后相关。

Ning等^[67]在健康人群和疑似患者的验证队列中,比较仅基于CT的严重程度预测模型、仅基于临床特征的预后预测模型,以及基于CT与临床特征结合的多模态模型的预测性能,发现影像资料与临床资料结合以及利用更多维度信息构建的模型在COVID-19患者的预后预测中可以提供更加客观的决策支持。Fang等^[68]将确诊患者分为症状轻微且住院期间无重症进展与住院期间进展至危重阶段两组,利用临床数据与影像资料构建了可自动识别导致恶性进展关键指标的模型。另有研究利用机器学习算法将患者入院时和入院第4天的双肺磨玻璃样阴影体积、半实变体积和实变体积、急性生理与慢性健康评分II(acute physiology and chronic health evaluation, APACHE II)评分、中性粒细胞/淋巴细胞比值及D-二聚体水平相结合,用logistic回归和Cox比例风险模型预测28 d内的重症风险,为COVID-19的临床治疗提供了一个可靠的预后指示^[69]。基于机器学习的COVID-19模型总结如表1所示。

表1 基于机器学习的COVID-19模型

Tab.1 Machine learning-based COVID-19 models

模型类型	临床获益	算法	预测因子类型	主要结果	是否验证
诊断(分类)模型					
Liu et al[7]	区分COVID-19与一般肺炎	DL	胸部CT	准确率为94.16%, 灵敏度为88.62%, 特异度为100%	是
Wang et al[24]	识别发热门诊疑似COVID-19患者	DL	胸部CT	AUC=0.953, 灵敏度为92.3%, 特异度为85.1%	是
Wang et al[25]	区分COVID-19与其他典型病毒性肺炎	DL	胸部CT	准确率为89.5%, 灵敏度为87%, 特异度为88%	是
Ni et al[6]	识别COVID-19	DL	胸部CT	F1=0.97	是
Jin et al[27]	识别COVID-19	DL	胸部CT	AUC=0.97	是
Song et al[28]	识别COVID-19	DL	胸部CT	AUC=0.99, 灵敏度为93%	是
Kang et al[29]	区分COVID-19与社区获得性肺炎	DL	胸部CT	准确率为95%, 灵敏度为96%, 特异度为93%	是
Wang et al[30]	区分COVID-19与其他病毒性肺炎	DL	胸部CT	AUC=0.97	是
Shi et al[31]	区分COVID-19与社区获得性肺炎	DL	胸部CT	准确率为87.9%, 灵敏度为90.7%, 特异度为83.3%	是
Wang et al[32]	识别COVID-19	DL	胸部CT	灵敏度>73%, 特异度>75%	是
Shan et al[33]	识别COVID-19	DL	胸部CT	与人工分割相比, Dice为91.6% ± 10.0%	是

(续 表)

模型类型	临床获益	算法	预测因子类型	主要结果	是否验证
Xu <i>et al</i> [34]	区分COVID-19与甲型流感病毒性肺炎	DL	胸部CT	准确率为86.7%	是
Wu <i>et al</i> [35]	识别COVID-19	DL	胸部CT	AUC=0.732, 准确率为70.0%, 灵敏度为73.0%, 特异度为61.5%	是
Pathak <i>et al</i> [36]	识别COVID-19	DL	胸部CT	测试集中准确率为93.0189%	是
Loey <i>et al</i> [37]	识别COVID-19	DL	胸部X线	准确率为99.9%	是
Kana <i>et al</i> [38]	识别COVID-19与健康人群、细菌与病毒性肺炎	DL	胸部X线	准确率为99%, 召回率为99.8%	是
Ko <i>et al</i> [39]	识别COVID-19、其他肺炎和健康人群	DL	胸部CT	准确率为99.87%, 灵敏度为99.58%, 特异度为100%	是
Gifani <i>et al</i> [40]	识别COVID-19	DL	胸部CT	准确率为85%, 召回率为85%, 精确度为85.7%	是
Jaiswal <i>et al</i> [41]	识别COVID-19	DL	胸部X线	准确率为90%	是
Sun <i>et al</i> [45]	识别COVID-19	DL	人口学特征、临床表现	AUC=0.971	是
Li <i>et al</i> [46]	区分COVID-19与流感	DL	实验室特征	灵敏度为92%, 特异度为97%	是
Meng <i>et al</i> [47]	识别COVID-19	DL	实验室特征	阳性预测值为86.35%, 阴性预测值为84.62%	是
Mei <i>et al</i> [48]	识别COVID-19	DL	胸部CT、临床表现、实验室结果	AUC=0.92	是
Xu <i>et al</i> [49]	区分非严重COVID-19、严重COVID-19、健康者和病毒性肺炎	DL	胸部CT、临床表现、实验室结果	准确率为95.4%~97.7%	是
Song <i>et al</i> [50]	识别COVID-19	DL	胸部CT、人口学特征、临床表现、实验室结果	AUC=0.956	是
预后模型					
Tang <i>et al</i> [53]	预测COVID-19患者的严重程度	RF	胸部CT	AUC=0.91	是
Gozes <i>et al</i> [55]	识别COVID-19, 衡量疾病随着时间的进展	CNN	胸部CT	AUC=0.996, 灵敏度为98.2%, 特异度为92.2%	是
Gao <i>et al</i> [59]	预测COVID-19患者的预后	LR、SVM、GBDT、NN	临床表现、实验室特征	AUC=0.9621	是
Cheng <i>et al</i> [60]	识别即将在24 h内转入ICU的COVID-19患者	RF	临床表现、实验室特征	AUC=0.799, 灵敏度为72.8%, 特异度为76.3%	是
Iwendi <i>et al</i> [61]	预测COVID-19患者的预后	RF	人口学特征	准确率为94%, F1=0.86	是
Sehanobish <i>et al</i> [62]	识别COVID-19, 预测患者严重程度	SSL	COVID-19支气管肺泡灌洗液样本的单细胞RNA测序	准确率为95.12%	是
Zhu <i>et al</i> [63]	预测COVID-19患者的死亡风险, 得到与死亡率独立相关的实验室特征	DL	人口学特征、实验室结果	AUC=0.968	是
Yan <i>et al</i> [64]	预测COVID-19重症患者的生存率	XGBoost	实验室特征	AUC>0.90	是
Ikemura <i>et al</i> [65]	预测COVID-19患者死亡率	DL	人口学特征、实验室结果	AUC=0.836	是
Hu <i>et al</i> [66]	预测COVID-19重症患者预后	LR	年龄、高敏C反应蛋白、淋巴细胞计数和D-二聚体	AUC=0.881	是
Ning <i>et al</i> [67]	预测COVID-19患者发病率和死亡率	DL	胸部CT、实验室结果	AUC=0.944、0.860、0.884	是
Fang <i>et al</i> [68]	预测COVID-19患者预后	DL	胸部CT、实验室结果	AUC=0.920(多中心); AUC=0.874(单中心)	是
Liu <i>et al</i> [69]	预测COVID-19患者预后	LR、COX比例风险模型	胸部CT、实验室结果	AUC=0.93	是

COVID-19. 新型冠状病毒感染; CNN. 卷积神经网络; RF. 随机森林; DL. 深度学习; SSL. 自监督学习; LR. Logistic回归; SVM. 支持向量机; NN. 神经网络

3 总结与展望

机器学习在COVID-19的早期筛查、诊断及预后预测中发挥了重要作用, 本文综述了机器学习

利用影像资料和临床资料在诊断与预后模型中的开发应用。尽管既往各项研究中机器学习模型获得了较高的准确率, 但由于数据规模有限、回顾性数据缺失程度不同、数据维度单一、验证不全面等原

因,存在过拟合及部署困难等问题,且大多数研究在对所构建的机器学习模型性能进行验证的过程中使用了内部验证方法,模型在多中心的适用性无法得到验证,故大多数模型在实际应用场景下仍面临诸多挑战。随着COVID-19传染力与致死率的不断变化,后续研究在纳入实验室检查结果与影像资料的基础上,可进一步收集患者的其他流行病学信息、影像数据、基因或蛋白组学数据,扩充机器学习模型的模态维度,以集成的观点建立更全面的信息处理与诊断模型;模型设计与构建应适当参考利用PROBAST(prediction model risk of bias assessment tool)评价工具对机器学习模型的偏倚与适用性进行评估诊断,从纳入人群、模型预测因子、模型表现性能与设计过程等4个方面进行判断^[70]。此外,随着SARS-CoV-2的变异,出现了越来越多的无症状感染者,前期利用临床资料与影像资料构建的大量深度学习模型在临床应用中表现出较大局限性,因此未来的研究可针对无症状感染者训练多模态机器学习模型,以在疫情防控中发挥更好的作用。

【参考文献】

- [1] World Health Organization. WHO coronavirus (COVID-19) dashboard overview[EB/OL]. [2022-08-31]. <https://covid19.who.int/>.
- [2] Chen T, Wu D, Chen H, *et al*. Clinical characteristics of 113 deceased patients with coronavirus disease 2019: Retrospective study[J]. *BMJ*, 2020, 368: m1091.
- [3] Holshue ML, DeBolt C, Lindquist S, *et al*. First case of 2019 novel coronavirus in the United States[J]. *N Engl J Med*, 2020, 382(10): 929-936.
- [4] Rodriguez-Morales AJ, Cardona-Ospina JA, Gutiérrez-Ocampo E, *et al*. Clinical, laboratory and imaging features of COVID-19: A systematic review and meta-analysis[J]. *Travel Med Infect Dis*, 2020, 34: 1016-1023.
- [5] Deo RC. Machine learning in medicine[J]. *Circulation*, 2015, 132(20): 1920-1930.
- [6] Ni Q, Sun ZY, Qi L, *et al*. A deep learning approach to characterize 2019 coronavirus disease (COVID-19) pneumonia in chest CT images[J]. *Eur Radiol*, 2020, 30(12): 6517-6527.
- [7] Liu C, Wang X, Liu C, *et al*. Differentiating novel coronavirus pneumonia from general pneumonia based on machine learning[J]. *Biomed Eng Online*, 2020, 19(1): 66.
- [8] Chen D, Jiang X, Hong Y, *et al*. Can chest CT features distinguish patients with negative from those with positive initial RT-PCR results for coronavirus disease (COVID-19)?[J]. *AJR Am J Roentgenol*, 2021, 216(1): 66-70.
- [9] Yang Y, Yang M, Yuan J, *et al*. Laboratory diagnosis and monitoring the viral shedding of SARS-CoV-2 infection[J]. *Innovation (Camb)*, 2020, 1(3): 100061.
- [10] Santosh KC. AI-driven tools for coronavirus outbreak: Need of active learning and cross-population train/test models on multitudinal/multimodal data[J]. *J Med Syst*, 2020, 44(5): 93.
- [11] DeCaprio D, Gartner J, McCall CJ, *et al*. Building a COVID-19 vulnerability index[J]. *medRxiv*, 2020. doi: 10.1101/2020.03.16.20036723.
- [12] Xia W, Shao J, Guo Y, *et al*. Clinical and CT features in pediatric patients with COVID-19 infection: Different points from adults[J]. *Pediatr Pulmonol*, 2020, 55(5): 1169-1174.
- [13] Xu X, Yu C, Qu J, *et al*. Imaging and clinical features of patients with 2019 novel coronavirus SARS-CoV-2[J]. *Eur J Nucl Med Mol Imaging*, 2020, 47(5): 1275-1280.
- [14] Shi H, Han X, Jiang N, *et al*. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study[J]. *Lancet Infect Dis*, 2020, 20(4): 425-434.
- [15] Kucirka LM, Lauer SA, Laeyendecker O, *et al*. Variation in false-negative rate of reverse transcriptase polymerase chain reaction-based SARS-CoV-2 tests by time since exposure[J]. *Ann Intern Med*, 2020, 173(4): 262-267.
- [16] Xie X, Zhong Z, Zhao W, *et al*. Chest CT for typical coronavirus disease 2019 (COVID-19) pneumonia: Relationship to negative RT-PCR testing[J]. *Radiology*, 2020, 296(2): E41-E45.
- [17] Fang Y, Zhang H, Xie J, *et al*. Sensitivity of chest CT for COVID-19: Comparison to RT-PCR[J]. *Radiology*, 2020, 296(2): E115-E117.
- [18] Ai T, Yang Z, Hou H, *et al*. Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: A report of 1014 cases[J]. *Radiology*, 2020, 296(2): E32-E40.
- [19] Zhao W, Zhong Z, Xie X, *et al*. Relation between chest CT findings and clinical conditions of coronavirus disease (COVID-19) pneumonia: A multicenter study[J]. *AJR Am J Roentgenol*, 2020, 214(5): 1072-1077.
- [20] Li Y, Xia L. Coronavirus disease 2019 (COVID-19): Role of chest CT in diagnosis and management[J]. *AJR Am J Roentgenol*, 2020, 214(6): 1280-1286.
- [21] Ye H, Gao F, Yin Y, *et al*. Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network[J]. *Eur Radiol*, 2019, 29(11): 6191-6201.
- [22] Lakhani P, Sundaram B. Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks[J]. *Radiology*, 2017, 284(2): 574-582.
- [23] Depeursinge A, Chin AS, Leung AN, *et al*. Automated classification of usual interstitial pneumonia using regional volumetric texture analysis in high-resolution computed tomography[J]. *Invest Radiol*, 2015, 50(4): 261-267.
- [24] Wang M, Xia C, Huang L, *et al*. Deep learning-based triage and analysis of lesion burden for COVID-19: A retrospective study with external validation[J]. *Lancet Digit Health*, 2020, 2(10): e506-e515.
- [25] Wang S, Kang B, Ma J, *et al*. A deep learning algorithm using CT images to screen for corona virus disease (COVID-19)[J]. *Eur Radiol*, 2021, 31(8): 6096-6104.
- [26] Li L, Qin L, Xu Z, *et al*. Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: Evaluation of the diagnostic accuracy[J]. *Radiology*, 2020, 296(2): E65-E71.
- [27] Jin C, Chen W, Cao Y, *et al*. Development and evaluation of an artificial intelligence system for COVID-19 diagnosis[J]. *Nat Commun*, 2020, 11(1): 5088.
- [28] Song Y, Zheng S, Li L, *et al*. Deep learning enables accurate

- diagnosis of novel coronavirus (COVID-19) with CT images[J]. *IEEE/ACM Trans Comput Biol Bioinform*, 2021, 18(6): 2775-2780.
- [29] Kang H, Xia L, Yan F, *et al.* Diagnosis of coronavirus disease 2019 (COVID-19) with structured latent multi-view representation learning[J]. *IEEE Trans Med Imaging*, 2020, 39(8): 2606-2614.
- [30] Wang J, Bao Y, Wen Y, *et al.* Prior-attention residual learning for more discriminative COVID-19 screening in CT images[J]. *IEEE Trans Med Imaging*, 2020, 39(8): 2572-2583.
- [31] Shi F, Xia L, Shan F, *et al.* Large-scale screening to distinguish between COVID-19 and community-acquired pneumonia using infection size-aware classification[J]. *Phys Med Biol*, 2021, 66(6): 065031.
- [32] Wang H, Wang L, Lee EH, *et al.* Decoding COVID-19 pneumonia: Comparison of deep learning and radiomics CT image signatures[J]. *Eur J Nucl Med Mol Imaging*, 2021, 48(5): 1478-1486.
- [33] Shan F, Gao Y, Wang J, *et al.* Abnormal lung quantification in chest CT images of COVID-19 patients with deep learning and its application to severity prediction[J]. *Med Phys*, 2021, 48(4): 1633-1645.
- [34] Xu X, Jiang X, Ma C, *et al.* A deep learning system to screen novel coronavirus disease 2019 pneumonia[J]. *Engineering (Beijing)*, 2020, 6(10): 1122-1129.
- [35] Wu X, Hui H, Niu M, *et al.* Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: A multicentre study[J]. *Eur J Radiol*, 2020, 128: 109041.
- [36] Pathak Y, Shukla PK, Tiwari A, *et al.* Deep transfer learning based classification model for COVID-19 disease[J]. *Ing Rech Biomed*, 2022, 43(2): 87-92.
- [37] Loey M, Smarandache FM, Khalifa NEM. Within the lack of chest COVID-19 X-ray dataset: A novel detection model based on GAN and deep transfer learning[J]. *Symmetry*, 2020, 12(4): 651.
- [38] Kana EB, Kana MG, Kana AF, *et al.* A web-based diagnostic tool for COVID-19 using machine learning on chest radiographs (CXR)[J]. *medRxiv*, 2020. doi: 10.1101/2020.04.21.20063263.
- [39] Ko H, Chung H, Kang WS, *et al.* COVID-19 pneumonia diagnosis using a simple 2D deep learning framework with a single chest CT image: Model development and validation[J]. *J Med Internet Res*, 2020, 22(6): e19569.
- [40] Gifani P, Shalhaf A, Vafaezadeh M. Automated detection of COVID-19 using ensemble of transfer learning with deep convolutional neural network based on CT scans[J]. *Int J Comput Assist Radiol Surg*, 2021, 16(1): 115-123.
- [41] Jaiswal AK, Tiwari P, Rathi VK, *et al.* Covidpen: A novel COVID-19 detection model using chest X-rays and CT scans[J]. *medRxiv*, 2020. doi: 10.1101/2020.07.08.20149161.
- [42] Mcrae MP, Dapkins IP, Sharif I, *et al.* Managing COVID-19 with a clinical decision support tool in a community health network: Algorithm development and validation[J]. *J Med Internet Res*, 2020, 22(8): e22033.
- [43] Zoabi Y, Deri-Rozov S, Shomron N. Machine learning-based prediction of COVID-19 diagnosis based on symptoms[J]. *NPJ Digit Med*, 2021, 4(1): 3.
- [44] Wagner T, Shweta F, Murugadoss K, *et al.* Augmented curation of clinical notes from a massive EHR system reveals symptoms of impending COVID-19 diagnosis[J]. *Elife*, 2020, 9: e58227.
- [45] Sun NN, Yang Y, Tang LL, *et al.* A prediction model based on machine learning for diagnosing the early COVID-19 patients[J]. *medRxiv*, 2020. doi: 10.1101/2020.06.03.20120881.
- [46] Li WT, Ma J, Shende N, *et al.* Using machine learning of clinical data to diagnose COVID-19: A systematic review and meta-analysis[J]. *BMC Med Inform Decis Mak*, 2020, 20(1): 247.
- [47] Meng Z, Wang M, Song H, *et al.* Development and utilization of an intelligent application for aiding COVID-19 diagnosis[J]. *medRxiv*, 2020. doi: 10.1101/2020.03.18.20035816.
- [48] Mei X, Lee HC, Diao KY, *et al.* Artificial intelligence-enabled rapid diagnosis of patients with COVID-19[J]. *Nat Med*, 2020, 26(8): 1224-1228.
- [49] Xu M, Ouyang L, Han L, *et al.* Accurately differentiating COVID-19, other viral infection, and healthy individuals using multimodal features *via* late fusion learning[J]. *J Med Internet Res*, 2021, 23(1): e25535.
- [50] Song CY, Xu J, He JQ, *et al.* COVID-19 early warning score: A multi-parameter screening tool to identify highly suspected patients[J]. *medRxiv*, 2020. doi: 10.1101/2020.03.05.20031906.
- [51] Yuan M, Yin W, Tao Z, *et al.* Association of radiologic findings with mortality of patients infected with 2019 novel coronavirus in Wuhan, China[J]. *PLoS One*, 2020, 15(3): e0230548.
- [52] Pan F, Zheng C, Ye T, *et al.* Different computed tomography patterns of coronavirus disease 2019 (COVID-19) between survivors and non-survivors[J]. *Sci Rep*, 2020, 10(1): 11336.
- [53] Tang Z, Zhao W, Xie X, *et al.* Severity assessment of coronavirus disease 2019 (COVID-19) using quantitative features from chest CT images[J]. *arXiv*, 2020. doi: 10.48550/arxiv.2003.11988.
- [54] Li Z, Zhong Z, Li Y, *et al.* From community-acquired pneumonia to COVID-19: A deep learning-based method for quantitative analysis of COVID-19 on thick-section CT scans[J]. *Eur Radiol*, 2020, 30(12): 6828-6837.
- [55] Gozes O, Frid-Adar M, Greenspan H, *et al.* Rapid AI development cycle for the coronavirus (COVID-19) pandemic: Initial results for automated detection & patient monitoring using deep learning CT image analysis[J]. *arXiv*, 2020. doi: 10.48550/arxiv.2003.05037.
- [56] Yin R, Yang ZQ, Wei YX, *et al.* Clinical characteristics of 106 patients with neurological diseases and co-morbid coronavirus disease 2019: A retrospective study[J]. *Cold Spring Harb Lab Press*, 2020. doi: 10.1101/2020.04.29.20085415.
- [57] CDC COVID-19 Response Team. Severe outcomes among patients with coronavirus disease 2019 (COVID-19) - United States, February 12-March 16, 2020[J]. *MMWR Morb Mortal Wkly Rep*, 2020, 69(12): 343-346.
- [58] Abdulaal A, Patel A, Charani E, *et al.* Prognostic modeling of COVID-19 using artificial intelligence in the United Kingdom: model development and validation[J]. *J Med Internet Res*, 2020, 22(8): e20259.
- [59] Gao Y, Cai GY, Fang W, *et al.* Machine learning based early warning system enables accurate mortality risk prediction for COVID-19[J]. *Nat Commun*, 2020, 11(1): 5033.
- [60] Cheng FY, Joshi H, Tandon P, *et al.* Using machine learning to predict ICU transfer in hospitalized COVID-19 patients[J]. *J Clin Med*, 2020, 9(6): 1668.
- [61] Iwendi C, Bashir AK, Peshkar A, *et al.* COVID-19 patient health prediction using boosted random forest algorithm[J]. *Front Public Health*, 2020, 8: 357.

- [62] Sehanobish A, Ravindra N, van Dijk D. Gaining insight into SARS-CoV-2 infection and COVID-19 severity using self-supervised edge features and graph neural networks[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(6): 4864-4873. doi: 10.1609/aaai.v35i6.16619.
- [63] Zhu JS, Ge P, Jiang C, *et al.* Deep - learning artificial intelligence analysis of clinical variables predicts mortality in COVID - 19 patients[J]. J Am Coll Emerg Physicians Open, 2020, 1(6): 1364-1373.
- [64] Yan L, Zhang HT, Xiao Y, *et al.* Prediction of survival for severe COVID-19 patients with three clinical features: development of a machine learning-based prognostic model with clinical data in Wuhan[J]. medRxiv, 2020. doi: 10.1101/2020.02.27.20028027.
- [65] Ikemura K, Goldstein D, Szymanski JM, *et al.* Using automated-machine learning to predict COVID-19 patient survival: Identify influential biomarkers[J]. medRxiv, 2020. doi: 10.1101/2020.10.12.20211086.
- [66] Hu C, Liu Z, Jiang Y, *et al.* Early prediction of mortality risk among patients with severe COVID-19, using machine learning[J]. Int J Epidemiol, 2020, 49(6): 1918-1929.
- [67] Ning W, Lei S, Yang J, *et al.* Open resource of clinical data from patients with pneumonia for the prediction of COVID-19 outcomes *via* deep learning[J]. Nat Biomed Eng, 2020, 4(12): 1197-1207.
- [68] Fang C, Bai S, Chen Q, *et al.* Deep learning for predicting COVID-19 malignant progression[J]. Med Image Anal, 2021, 72: 102096.
- [69] Liu F, Zhang Q, Huang C, *et al.* CT quantification of pneumonia lesions in early days predicts progression to severe illness in a cohort of COVID-19 patients[J]. Theranostics, 2020, 10(12): 5613-5622.
- [70] Wolff RF, Moons KGM, Riley RD, *et al.* PROBAST: A tool to assess the risk of bias and applicability of prediction model studies[J]. Ann Intern Med, 2019, 170(1): 51-58.

(责任编辑: 纪方方)