

机器学习在医学专业教育教学评价中应用的范围综述

岳芳岩, 李 静, 陈 欧, 贾 愚, 郭玉芳

山东大学护理与康复学院, 山东 济南

收稿日期: 2022年6月18日; 录用日期: 2022年7月15日; 发布日期: 2022年7月21日

摘 要

目的: 对在医学专业教育教学评价体系中应用机器学习的相关研究进行范围综述, 为医学专业教育教学评估体系发展提供新思路。方法: 依据范围综述的方法框架, 确立研究问题, 系统检索中国知网、万方、维普、PubMed、Web of Science 5个中英文库, 检索时间为建库至2022年3月15日。筛选符合纳入标准的文献, 并对文献进行分析讨论。结果: 共检索文献301篇, 纳入文献15篇。总结显示, 机器学习主要应用在学生能力评价、课程反馈信息处理、预测和识别高危学生等方面。应用机器学习形成的教育教学评价模型评价效果较好, 其在提高评价效率、节约评价成本、构建科学的评价指标等方面具有较大优势。结论: 目前将机器学习应用于医学专业教育教学评价的研究较少, 现有研究验证了机器学习在医学教育教学评价中应用的可行性, 但其准确性、泛化性和有效性等方面仍待进一步完善。

关键词

机器学习, 医学教育, 教育评价

A Review of the Application Scope of Machine Learning in Medical Education and Teaching Evaluation

Fangyan Yue, Jing Li, Ou Chen, Yu Jia, Yufang Guo

School of Nursing and Rehabilitation, Shandong University, Jinan Shandong

Received: Jun. 18th, 2022; accepted: Jul. 15th, 2022; published: Jul. 21st, 2022

Abstract

Objective: To review the research on the application of machine learning in the evaluation system of medical professional education and teaching, and to provide new ideas for the development of medical professional education and teaching system. **Methods:** According to the method framework of scope review, the research questions were identified and the five Chinese and English databases of CNKI, Wanfang, VIP, PubMed and Web of Science were searched systematically from the establishment of the database to March 15, 2022. The literatures that met the inclusion criteria were screened and analyzed. **Results:** A total of 301 papers were retrieved, 15 of which were included according to the inclusion criteria. Machine learning methods are mainly applied in evaluating students' abilities, processing course feedback information, predicting and identifying high-risk students, and so on. The results show that the evaluation effect of machine learning is better, and it has great advantages in improving evaluation efficiency, saving evaluation cost and constructing scientific evaluation indexes. **Conclusion:** Currently, there are few researches on the application of machine learning in the evaluation of medical education and teaching. Existing researches have verified the feasibility of the application of machine learning in the evaluation of medical education and teaching, but its accuracy, generalization and effectiveness still need to be further improved.

Keywords

Machine Learning, Medical Education, Teaching Evaluation

Copyright © 2022 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

教育教学评价是根据教学目标,通过搜集教学过程中的主要信息,准确地了解教学情况并进行科学分析,对学校教学质量做出评价,进而调整、优化教学的活动[1]。教育教学评价是高校进行教学质量监控与管理的核心内容,一套高效合理的评价体系可以有效地提高教学和管理水平。医学教育涉及目前中国发展改革中教育与医药卫生事业两大热点和难点,作为医学人才培养的基础和起始点,医学教育受到越来越多的重视[2]。目前大部分医学高校采用加权平均法、专家评价法、AHP层次分析法、模糊综合评价法、神经网络模型法等传统方法[3]进行教学评价,但传统教学评价体系存在以下问题:①评价指标的设定主要靠专家商定或教师经验,缺乏科学性和客观性;传统的统计方法确定评价指标权重过程繁琐,不能挖掘大量数据中可指导教学的有效信息[3]。②评价过程消耗大量的人力成本,评价效率较低。③评价算法不能动态添加大量数据,评价模型固定,缺乏自适应能力[3]。面对教学信息化快速发展以及医学教育模式向能力标准方向转变的趋势,传统的教学评价方法越来越无法满足医学教育发展的需求。

机器学习是使用计算机作为工具,并致力于真实实时模拟人类学习方式,将现有内容进行知识结构划分来提高学习效率和大数据利用的技术[4]。利用机器学习技术可以有效地分析和提取大型数据集中隐藏的信息,在大量数据中发现模式和知识,预测结果或行为[3]。目前机器学习在国内外医学教育教学评价方面的应用研究还处于起步阶段,研究文献较少,且相关研究存在较大的异质性。因此本文以范围综述报告框架为指导,对运用机器学习算法进行医学教育教学评价的相关研究进行分析和梳理,旨在为今

后机器学习算法在医学教育教学评价中推广应用提供依据。

2. 资料与方法

2.1. 确定相关概念, 确定研究问题

范围综述(Scoping review)是基于循证的研究方法, 可用于探索某一类研究的范围、程度和本质, 从而总结和推广研究结果[5]。本范围综述提出的研究问题主要有: ① 机器学习在医学教育教学评价中应用的研究现状如何? 包括所用的机器学习算法有哪些? 评价模型应用于教学评价的哪些方面? ② 评价模型性能如何?

2.2. 文献检索

检索中国知网、万方数据、维普、Web of Science、PubMed 等 5 个中英文文献数据库, 使用主题词与自由词相结合检索自建库至 2022 年 3 月 15 日的相关文献。英文检索以 PubMed 为例, 检索式为: “machine learning” [Title/Abstract] AND (“medical student” [Title/Abstract] OR “medical education” [Title/Abstract] OR “nursing education” [Title/Abstract]) AND (“evaluation” [Title/Abstract] OR “predict” [Title/Abstract])。中文检索以万方数据为例, 检索式为: 主题: (机器学习 OR Machine Learning OR 决策树算法 OR 朴素贝叶斯算法 OR 支持向量机算法 OR 随机森林算法 OR 神经网络算法 OR Boosting 与 Bagging 算法 OR 关联规则算法 OR EM (期望最大化)算法 OR 深度学习) and 主题: (医学专业 OR 临床医学 OR 护理学 OR 药学 OR 基础医学 OR 中医学 OR medical major) and 主题: (教学质量评价 OR 教学评价 OR 教学评价指标 OR 教学评价指标体系)。

2.3. 文献纳入标准

纳入标准: ① 运用机器学习算法对医学教育中教学过程、学习活动、教学效果、教学管理等进行评估的相关研究。② 评价对象为学生、教师、课程、医学高校等。③ 可提取完整数据的文章。排除标准: ① 非中、英文文献。② 会议论文和综述类文章。③ 重复发表的文章。

2.4. 文献的筛选、数据提取

由两名有循证学习经历的研究者独立筛选文献, 如遇分歧, 咨询第三方。首先将检索到的文献导入 Endnote、Noteexpress 文献管理软件进行去重, 根据文献的纳入、排除标准进行文献的初筛, 之后两位研究者阅读文献的题目及摘要, 排除不符合要求的文献, 进一步阅读全文进行复筛, 最终对纳入的文献进行信息提取, 提取内容包括: ① 文献的基本信息, 即作者、发表年限、国家或地区; ② 研究类型、所用算法、评价模型功能、评估应用及评价性能等。

2.5. 数据总结、报告研究结果

对纳入的研究文献的范围、性质和特点进行综述, 以表格的形式进行具体呈现。

3. 结果

3.1. 文献筛选结果

共获得文献 301 篇, 其中中国知网 7 篇, 万方数据 141 篇, 维普 58 篇, PubMed 32 篇, Web of Science 63 篇。去除重复文献 37 篇; 经过阅读标题和摘要去除文献 226 篇; 阅读全文, 排除研究主题、对象、类型不符以及无法获取全文和内容重复的文献 23 篇, 最终纳入篇文献 15 篇, 包括万方数据 1 篇, 维普

1 篇, PubMed 3 篇, webofscience 10 篇。

3.2. 纳入文献基本特征

15 篇文献发表于 2013 至 2022 年, 分别在美国($n=7$)、英国($n=1$)、加拿大($n=3$)、中国($n=3$)、意大利($n=1$)进行。样本量 15 至 9853 名。纳入文献汇总信息见表 1。

Table 1. Basic information of the included literature

表 1. 纳入文献的基本信息

作者、发表年份	国家	评价对象	模型所用的机器学习算法	评价模型的功能	评价应用	性能评价	研究类型	样本量
朴杰等(2013) [6]	中国	医学生	支持向量机(SVM)算法	临床能力评估	教学效果	评价系统具有可行性	定量研究	346
张俊飞(2018) [7]	中国	教师	改进朴素贝叶斯算法	评教评语情感分析	教学过程	较传统朴素贝叶斯算法评价性能提高	定量研究	9854
Yukunchen, MS 等(2014) [8]	美国	医学生	朴素贝叶斯、逻辑回归和线性 SVM	老年医学能力评估	教学效果	评价模型在部分老年医学能力评价性能良好, 在部分能力评价上效果不佳	定量研究	119
Jun Yang 等(2020) [9]	中国	医学生	随机森林	自我调节学习水平评价	教学管理	评价模型具有良好的预测精度、辨别性和泛化性	多中心横断面研究	2052
LiHannaford 等(2021) [10]	美国	医学生	C5.0 (决策树)、随机森林、xgboost、神经网络、支持向量机、朴素贝叶斯、k-最近邻和逻辑回归及堆积集成方法	识别不能完成学业的高危学生和与毕业结果相关因素	教学管理	验证了预测有效性; 敏感性低, 缺乏泛化性	定量描述性研究	1018
Matthew S. Holden (2018) [11]	加拿大	学员	决策树算法	医学技术技能评价	教学效果	验证了评估方法有较好的准确性和反馈的有用性。	定量研究	19
Syed Latifi 等(2015) [12]	加拿大	医学生	决策树算法	论文自动评分评估临床能力	教学效果	评分系统结果可靠性, 具有可行性	定量研究	8007
Aditya Borakati (2021) [13]	英国	课程	潜在狄利克雷分配(LDA)算法	分析医科学生对电子学习课程的反馈信息	教学过程	评价方法可行	多中心队列研究	1396
NykanMirchi 等(2020) [14]	加拿大	医生、医学生	支持向量机	评估和训练医学精神运动技能	教学效果	评价方法可行	定量研究	50

Continued

Roberto Bertolini 等 (2021) [15]	美国	医学生	单项机器学习算法 1) 逻辑回归(LR)、 2) 支持向量机(SVM) 集成机器学习算法 1) 广义线性模型(GLMNET)、2) 随机森林(RF)和 3) 随机梯度增强(GBM)	早期识别高危风险学生, 减少 STEM 相关课程(护理学、预防医学等)的学生流失率。	教学管理	评价结果准确性较好, 敏感性及泛化性较差	定量描述性研究	3325
Shang Zhao 等(2021) [16]	美国	学员	神经网络	利用有限数据进行技术技能评估	教学效果	验证了评价模型利用小样本量进行评价的可行性	定量研究	478
Andrea Moglia 等 (2022) [17]	意大利	医学生	集成深度学习	评估外科技能习得的进展, 预测医学生的技术技能	学习过程 教学效果	集成深度学习模型的性能优于集成树的模型(随机森林和GBRT)和传统的非线性 ML 算法	定量研究	176
Mohamed S. Baloul, MBBS 等(2020) [18]	美国	学员	人工神经网络(ANN)	评价学员的非技术认知技能	教学效果	机器学习算法比传统方法具有更好的准确性。本研究验证了其可行性	定量研究	81
Yuanyuan Gao 等(2020) [19]	美国	医学生	核偏最小二乘(KPLS)多元监督机器学习算法	根据学习者的初始技能水平预测其到达熟练程度的训练次数及最终的表现水平, 描绘学习曲线	学习过程 教学效果	评价方法可行	定量研究	15
Yousi A. Oquendo 等 (2018) [20]	美国	学员	正则化最小二乘回归(lasso)和回归树的组合	自动评估学员在儿童腹腔镜缝合任务中的技术技能水平	教学效果	验证了评价模型的可行性	定量研究	63

3.3. 机器学习在医学教学评价中应用的研究现状

根据纳入文献可知, 在医学教育教学评价中应用最多的算法为支持向量机(SVM)算法($n = 5$), 其次是朴素贝叶斯算法($n = 3$)、随机森林算法($n = 3$)、决策树算法($n = 3$)、逻辑回归算法($n = 3$)、人工神经网络算法($n = 3$)、深度学习($n = 1$)、xgboost 算法($n = 1$)、k-最小邻算法($n = 1$)、广义线性模型($n = 1$)。机器学习算法在医学教育中主要应用于教学效果评价($n = 10$)、教学过程评价($n = 2$)、学习过程评价($n = 2$)及教学管理($n = 3$)等方面。其中在教学效果评价方面的研究数量最多, 纳入文献中有 11 项研究将机器学习应用于医学生的能力评估, 包括医学技术技能($n = 6$)、临床综合能力($n = 2$)、老年医学能力($n = 1$)、自

我调节水平($n = 1$)及医学非技术认知能力($n = 1$)的评估。在教学过程评价方面,目前的研究主要集中在应用机器学习算法进行课程反馈信息的分析;在教学管理方面,纳入的3篇文献均是关于应用机器学习早期识别高危风险学生及教学结果相关因素预测的研究。在学习过程评价方面,机器学习主要应用于学生学习过程的监控和跟踪,通过描绘学习曲线,对学生的学习情况进行即时反馈。

3.4. 机器学习评价模型的性能评估

纳入的15项研究均对机器学习算法模型的性能进行了验证和评估,其中大多数研究通过数据分割及交叉验证的方法进行性能评估,只有1项研究[9]通过外部验证的方法进行机器学习评价模型的性能评估。通过验证,在大多数研究中,机器学习评价模型表现良好,评价结果有较好的准确性,评价方法提高了评价效率和公平性,证明了机器学习评价模型在医学教育教学评价中应用的可行性。然而机器学习在性能方面仍有不足之处,如在朴杰等[6]的研究中,支持向量机算法对医学生临床能力的评价结果准确率不到90%,并未达到研究者的预期效果。Chen等[8]利用机器学习进行老年医学能力的评估中,姑息治疗(PC)及认知和行为障碍(CBD)的评估结果也并不理想。在Li等[10]及Bertolini等[15]所构建的早期预警(EWS)系统中,机器学习模型存在敏感度较低,泛化性较差等问题。

4. 讨论

4.1. 机器学习在医学教育教学评价领域应用研究现状分析

4.1.1. 机器学习在教学效果评价方面的应用

教学效果评价是教学评价体系中的重要组成部分,当前医学评价已由基于时间标准逐渐向基于能力标准转变[11],这种趋势增加了医学教育教学评价的复杂性和困难,急需纳入更为科学、有效的评估手段。机器学习在分类和预测领域的突出表现,使得学者们尝试机器学习应用于医学教育教学评价中。本研究纳入的文献中有11项研究将机器学习应用于医学生的能力评估如医学技术技能、临床综合能力、老年医学能力、自我调节水平、及医学非技术认知能力。通过分析文献发现,机器学习在医学专业教育教学效果评价方面的研究主要集中在两个方向:① 医学技能测试,通过将机器学习算法与盒子训练器或虚拟模拟器结合评估医学生的外科技能水平,评估模型可自动为学生在模拟训练器上的表现分配分数。② 非结构化文本自动评分,利用机器学习算法模型和自然语言处理技术进行文本挖掘,从而实现通过分析非结构化文本进行医学能力评估。如Chen等[8]通过分析临床笔记来评估学生的临床决策能力和熟练程度,该研究不仅是运用机器学习进行临床能力评估的尝试,更为医学教学评价体系的构建提供了新的视角。运用机器学习方法进行教学效果评估弥补了传统教学评价中主观性、评价效率低等不足,为实现自动、高效、准确的能力评估提供了可能。

然而利用机器学习评估学生当前的水平并不是终点,形成性评价是医学教学评价中重要的评价方式,机器学习,尤其深度学习曾被比喻为“黑匣子”,外界对其评价过程全然不知。但透明度和信任是确保教育者和学习者之间成功联系的重要组成部分,近年来,Holden [11]、Mirchi [14]及Latifi [12]等学者开始进行透明化机器学习评价模型的研究,评价模型可基于学生的学习情况提供个性化的客观反馈,此方法将透明化的机器学习评价模型应用于医学教学形成性评价中,为教育者调整教学计划和提供依据,学生也可根据反馈信息进行自我指导和训练。

4.1.2. 机器学习在教学过程评价方面的应用

目前,机器学习在教学过程评价方面的应用研究较少,纳入研究中仅有2篇文献[7] [13]运用机器学习进行教学过程所产生的质性数据进行建模和分析。从两项研究中总结发现,运用无监督机器学习算法

对质性数据进行建模和自主学习, 尤其是通过情感分析从大量的质性数据中挖掘对教学有指导意义的信息, 可以帮助教育者建立和完善更加客观的评价指标, 设置更为合理科学的考核指标。机器学习在教学过程评价中的应用模型有潜力成为优化课程结构和提高教学水平的有效工具。

4.1.3. 机器学习在学习活动评价方面的应用

学习过程评价动态监控和调节学习者的学习行为, 促进学习者持续、有效地学习, 对教师动态掌握学习者的学习状况, 实施教学调控和工作反思具有极其重要的作用[21]。纳入研究中将机器学习方法应用于外科教育模拟器手术训练中, 根据学习者的初始技能水平预测其到达熟练程度的训练次数及最终的表现水平, 描绘学习曲线, 评估技能习得的进展, 使外科教育者在个人基础上定制技术技能的教学, 从而制定个性化的培训课程, 降低成本和提高效率。由此可见, 机器学习对学习过程的预测和评估有利于医学教育形成性评价的开展, 同时契合基于能力标准医学课程评价的需要, 将成为医学教育教学评价的一个重要组成部分。

4.1.4. 机器学习在教学管理中的应用

教学管理是教学评价体系中的重要环节, 决定教学评价的结果能否反馈于教学和学习活动中以提高教学质量。Yang [9]、Bertolini [15]和 Li 等[10]运用机器学习在教学管理方面进行了系列的研究, 运用机器学习的分类和预测功能构建早期预警系统(EWS), 用于高危学生早期识别和风险因素预测, 对此类学生根据风险因素及早采取干预措施, 提高毕业率或课程的合格率。机器学习在教学管理方面的应用研究不仅可以帮助提高教学质量, 还有利于提高学校教学管理者的决策能力和管理水平。

4.2. 机器学习评价模型的不足及改进方向

4.2.1. 机器学习评价模型性能的优化

本研究纳入的4篇报告[9] [17] [18] [19]中提及机器学习评价模型的性能受限于样本量的不足, 后续研究中可以采取多中心研究并增加样本量进行训练, 以提高评价模型的性能。在 Zhao 等[16]的研究中, 利用有限数据进行建模, 对原有样本中没被标记的数据信息进行分析, 同样提高了评价的准确性。因此, 建议在未来的研究中可尝试提高样本的利用率来优化评价模型的性能。此外, 在纳入文献中只有一项研究通过外部验证进行评价模型的性能评价, 其他研究均采用数据分隔和交叉验证进行评价, 在今后的研究中可以增加外部验证或者专家验证, 提高性能评价的准确性。

4.2.2. 机器学习评价模型的有效性评价

目前机器学习在医学教育教学评价体系中的应用处于初步探索阶段, 研究重点集中在评价模型的可用性, 而缺乏评价模型在教学评价体系中应用的用户或利益相关者使用效果的研究, 同时缺乏纵向研究。在后续的研究中, 建议进行纵向研究, 以发现机器学习在评价体系中应用的不足及对教学管理水平和教学质量的影响, 从而检验机器学习评价模型在医学教学评价体系中应用的有效性。

4.2.3. 机器学习评价模型的推广

评价模型是否得到推广使用取决于评价方法的性价比。在以上研究中机器学习方法在医学教学评价体系中应用的可行性已被验证, 并且其具有节约人力成本和提高效率的优势, 但评价模型构建成本却未被提及。Latifi 等[12]研究显示, 自动论文评分系统(AES)建立评分模型需要大量的样本量进行训练, 在小型评估环境中并不适用, 且每个评分任务都需要构建新的模型, 这将带来大量的开发成本。此外, 大多数评价模型处于开发阶段, 若想推广使用, 首先需要优化评价模型的性能, 而评价结果的准确性及可靠性很大程度上依赖于样本量, 构建完善的评价模型需要大量的时间和成本。而纳入研究并未对模型的成

本效益进行评价, 因此, 推荐后续的研究考虑评价模型的成本效益评价, 以推动机器学习教学评价模型的在实际医学教育教学评价体系中使用和推广。

5. 小结和展望

机器学习所具有的自主学习能力及在分类和预测领域的突出表现, 可以有效地克服传统医学教育教学评价方法中主观性、线性、评价模型固定及评价效率低下等问题, 其在医学教育教学评估方面有较大发展空间。但相关研究数量较少, 评价模型性能、评价结果有效性及泛化性等方面还存在不足。在今后的研究中, 可以扩大样本量, 改进机器学习算法模型性能, 增加评价模型应用效果的研究, 促进机器学习在医学教育教学评价中更加高效的应用和推广。

参考文献

- [1] 何锡涛. 构建本科教育教学质量评价体系的设想[J]. 上海电机学院学报, 2005, 8(4): 25-27+35. <https://doi.org/10.3969/j.issn.2095-0020.2005.04.007>
- [2] 朱雪波, 吕帆, 刘燕楠. 健康中国目标与医学教育改革[J]. 医学与哲学, 2019, 40(13): 64-67. <https://doi.org/10.12014/j.issn.1002-0772.2019.13.15>
- [3] 胡延雪. 基于机器学习的教学评价模型的研究与实现[D]: [硕士学位论文]. 吉林: 延边大学, 2019.
- [4] 李昊朋. 基于机器学习方法的智能机器人探究[J]. 通讯世界, 2019, 26(4): 241-242. <https://doi.org/10.3969/j.issn.1006-4222.2019.04.157>
- [5] Davis, K., Drey, N. and Gould, D. (2009) What Are Scoping Studies? A Review of the Nursing Literature. *International Journal of Nursing Studies*, **46**, 1386-1400. <https://doi.org/10.1016/j.ijnurstu.2009.02.010>
- [6] 朴杰, 李勇, 杨琳丽, 赵士斌. 运用机器学习法构建临床能力评价系统的研究[J]. 中国高等医学教育, 2013(3): 103-104+137.
- [7] 张俊飞. 基于改进朴素贝叶斯算法实现评教评语情感分析[J]. 现代计算机, 2018(32): 3-6. <https://doi.org/10.3969/j.issn.1007-1423.2018.32.001>
- [8] Chen, Y., Wrenn, J., Xu, H., Spickard, A., Habermann, R., Powers, J., et al. (2014) Automated Assessment of Medical Students' Clinical Exposures According to AAMC Geriatric Competencies. *AMIA Annual Symposium Proceedings*, 2014, 375-384.
- [9] Yang, J., Zhang, G., Huang, R., Yan, P., Hu, P., Huang, L., et al. (2020) Nomograms Predicting Self-Regulated Learning Levels in Chinese Undergraduate Medical Students. *Frontiers in Psychology*, **10**, Article No. 2858. <https://doi.org/10.3389/fpsyg.2019.02858>
- [10] Hannaford, L., Cheng, X. and Kunes-Connell, M. (2021) Predicting Nursing Baccalaureate Program Graduates Using Machine Learning Models: A Quantitative Research Study. *Nurse Education Today*, **99**, Article ID: 104784. <https://doi.org/10.1016/j.nedt.2021.104784>
- [11] Holden, M.S., Xia, S., Lia, H., Keri, Z., Bell, C., Patterson, L., et al. (2019) Machine Learning Methods for Automated Technical Skills Assessment with Instructional Feedback in Ultrasound-Guided Interventions. *International Journal of Computer Assisted Radiology and Surgery*, **14**, 1993-2003. <https://doi.org/10.1007/s11548-019-01977-3>
- [12] Latifi, S., Gierl, M.J., Boulais, A.P. and De Champlain, A.F. (2016) Using Automated Scoring to Evaluate Written Responses in English and French on a High-Stakes Clinical Competency Examination. *Evaluation & the Health Professions*, **39**, 100-113. <https://doi.org/10.1177/0163278715605358>
- [13] Borakati, A. (2021) Evaluation of an International Medical E-Learning Course with Natural Language Processing and Machine Learning. *BMC Medical Education*, **21**, Article No. 181. <https://doi.org/10.1186/s12909-021-02609-8>
- [14] Mirchi, N., Bissonnette, V., Yilmaz, R., Ledwos, N., Winkler-Schwartz, A. and Del Maestro, R.F. (2020) The Virtual Operative Assistant: An Explainable Artificial Intelligence Tool for Simulation-Based Training in Surgery and Medicine. *PLOS ONE*, **15**, Article ID: e0229596. <https://doi.org/10.1371/journal.pone.0229596>
- [15] Bertolini, R., Finch, S.J. and Nehm, R.H. (2021) Testing the Impact of Novel Assessment Sources and Machine Learning Methods on Predictive Outcome Modeling in Undergraduate Biology. *Journal of Science Education and Technology*, **30**, 193-209. <https://doi.org/10.1007/s10956-020-09888-8>
- [16] Zhao, S., Zhang, X., Jin, F. and Hahn, J. (2021) An Auxiliary Tasks Based Framework for Automated Medical Skill Assessment with Limited Data. 2021 43rd Annual International Conference of the IEEE Engineering in Medicine &

-
- Biology Society*, Mexico, 1-5 November 2021, 1613-1617. <https://doi.org/10.1109/EMBC46164.2021.9630498>
- [17] Moglia, A., Morelli, L., D'Ischia, R., Fatucchi, L.M., Pucci, V., Berchiolli, R., *et al.* (2022) Ensemble Deep Learning for the Prediction of Proficiency at a Virtual Simulator for Robot-Assisted Surgery. *Surgical Endoscopy*, 1-7. <https://doi.org/10.1007/s00464-021-08999-6>
- [18] Baloul, M.S., Yeh, J.H., Mukhtar, F., Traynor Jr., M.D., Shaikh, N., Rivera, M., *et al.* (2020) Video Commentary & Machine Learning: Tell Me What You See, I Tell You Who You Are. *Journal of Surgical Education*, In Press. <https://doi.org/10.1016/j.jsurg.2020.09.022>
- [19] Ygm, A., Ukea, B., Xia, B., Schwaitzberg, S. and De, S. (2020) A Machine Learning Approach to Predict Surgical Learning Curves. *Surgery*, **167**, 321-327. <https://doi.org/10.1016/j.surg.2019.10.008>
- [20] Oquendo, Y.A., Riddle, E.W., Hiller, D., Blinman, T.A. and Kuchenbecker, K.J. (2018) Automatically Rating Trainee Skill at a Pediatric Laparoscopic Suturing Task. *Surgical Endoscopy*, **32**, 1840-1857. <https://doi.org/10.1007/s00464-017-5873-6>
- [21] 郑志高, 张立国, 张春荣. xMOOC 的学习评价方法调查研究[J]. 中国电化教育, 2014(11): 44-49.