

基于机器学习算法的大于胎龄儿风险预测模型

白哲¹ 罗云云¹ 周智博¹ 苏明亮² 杨柳青² 陈适¹ 阳洪波¹ 朱惠娟¹ 潘慧¹

¹中国医学科学院/北京协和医学院/北京协和医院内分泌科/国家卫生健康委员会内分泌重点实验室/疑难重症及罕见病国家重点实验室, 北京 100730; ²东华医为科技有限公司, 北京 100190

通信作者: 潘慧, Email: panhui20111111@163.com

【摘要】目的 开发和验证基于机器学习算法的孕期大于胎龄儿(LGA)风险预测模型, 并比较其与传统逻辑回归方法建模的性能差异。**方法** 研究对象来自“中国免费孕前优生健康检查项目”, 于2010-2012年在全国31个省市的220个县开展, 覆盖全部农村计划妊娠夫妇, 本研究选取分娩新生儿胎龄在24-42周内, 单胎活产的所有育龄期夫妇及其新生儿为研究对象。应用10种机器学习算法分别建立LGA预测模型, 评估模型对LGA的预测性能。**结果** 最终纳入104 936名新生儿, 男婴54 856例(52.3%), 女婴50 080例(47.7%), LGA的发生率为11.7%(12 279例)。经过下采样数据平衡处理后, 机器学习方法建立模型的整体效能出现明显提高, 其中以CatBoost模型在预测LGA风险方面表现最佳, 模型的受试者工作特征曲线的曲线下面积(AUC)为0.932; 逻辑回归模型表现最差, AUC仅为0.555。**结论** 与传统的逻辑回归方法相比, 通过机器学习算法可建立更有效的孕期LGA风险预测模型, 具有潜在的应用价值。

【关键词】 机器学习; 大于胎龄儿; 风险预测模型

Development and evaluation of a machine learning prediction model for large for gestational age

Bai Xi¹, Luo Yunyun¹, Zhou Zhibo¹, Su Mingliang², Yang Liuqing², Chen Shi¹, Yang Hongbo¹, Zhu Huijuan¹, Pan Hui¹

¹Department of Endocrinology, Key Laboratory of Endocrinology of National Health Commission/State Key Laboratory of Complex Severe and Rare Diseases/Peking Union Medical College Hospital/Chinese Academy of Medical Science and Peking Union Medical College, Beijing 100730, China; ²DHC Mediway Technology Co., Ltd, Beijing 100190, China

Corresponding author: Pan Hui, Email: panhui20111111@163.com

【Abstract】 Objective To develop and validate a useful predictive model for large gestational age (LGA) in pregnancy using a machine learning (ML) algorithm and compare its performance with the traditional logistic regression model. **Methods** Data were obtained from the National Free Preconception Health Examination Project in China, carried out in 220 counties of 31 provinces from 2010 to 2012, covering all rural couples with a planned pregnancy. This study included all teams of childbearing age who delivered newborns within 24-42 weeks of gestational age and their newborns. Ten different ML algorithms were used to establish LGA prediction models, and the prediction performance of these models was evaluated. **Results** A total of 104 936 newborns were included, including 54 856 boys (52.3%) and 50 080 girls (47.7%). The incidence of LGA was 11.7% (12 279). The imbalance between the two groups was addressed by the under-sampling technique, after which the overall performance of the ML models was significantly

DOI: 10.3760/cma.j.cn112338-20210824-00677

收稿日期 2021-08-24 本文编辑 王岚

引用本文: 白哲, 罗云云, 周智博, 等. 基于机器学习算法的大于胎龄儿风险预测模型[J]. 中华流行病学杂志, 2021, 42(12): 2143-2148. DOI: 10.3760/cma.j.cn112338-20210824-00677.



improved. The CatBoost model achieved the highest area under the receiver-operating-characteristic curve (AUC) value of 0.932. The logistic regression model had the worst performance, with an AUC of 0.555. **Conclusions** In predicting the risk for LGA in pregnancy, the ML algorithms outperform the traditional logistic regression method. Compared to other ML algorithms, CatBoost could improve the performance, and it deserves further investigation.

【Key words】 Machine learning; Large for gestational age; Risk prediction model

大于胎龄儿 (large for gestational age, LGA) 是指出生体质量高于同种族同胎龄同性别新生儿出生体质量第 90 百分位的新生儿^[1]。LGA 不仅出现新生儿低血糖、肩难产等急性围产期并发症风险高于正常体重新生儿,其成年后患肥胖症、2 型糖尿病等慢性病的风险也会升高^[2-4],因此早期准确地预测 LGA 具有重要意义。目前,LGA 的检测手段仍以 B 型超声为主,但妊娠早期的超声检查结果对 LGA 的预测价值有限;既往虽有研究曾利用孕妇的社会人口学特征结合临床生化指标,建立 LGA 的预测模型^[5],但建模仅依赖于传统的统计学方法,存在固有的局限性,可能影响其在多变量、大样本量数据集中的应用和性能。相比之下,机器学习算法可以更准确地从复杂数据集中提取关键特征,已被越来越多地应用于不同疾病的诊断和预后预测中^[6-9]。迄今为止还没有研究应用机器学习算法建立适用于中国人口的 LGA 预测模型。本研究利用一项多中心、大样本数据集,开发和验证基于机器学习算法的孕期 LGA 预测模型,并比较其与传统逻辑回归方法建模的性能差异。

对象与方法

1. 研究对象:来自“中国免费孕前优生健康检查项目”。该项目于 2010 年 1 月至 2012 年 12 月,通过国家卫生健康委员会在全国 31 个省市的 220 个县开展检测,目标人群覆盖全部农村计划妊娠夫妇,包括流动人口。项目的总体设计及实施在既往文献中已报道^[10-11]。本研究从中选取分娩新生儿胎龄在 24~42 周内,单胎活产,且出生体重记录完整的所有育龄期夫妇及其新生儿为研究对象,共涉及 214 636 条记录。排除基线特征存在缺失值及异常值的记录后,最终纳入 104 936 条记录。该项目通过国家计划生育研究所伦理委员会批准,研究对象均签署知情同意书。

2. 研究方法:项目包括孕前检查以及产前和产后随访。由经过专业培训的工作人员通过面对面调查的方式,收集父母与新生儿的社会人口学特

征、孕前及孕期生活习惯、社会心理状态、饮食及营养状况、环境风险暴露因素、既往病史、家族史、用药情况、体格检查、实验室检查等方面的 223 个特征。所有数据均上传至全国电子数据采集系统,并由国家质量检测中心进行质量控制。

3. 指标定义:①LGA:基于 1995 年 WHO 的诊断标准^[12],将出生体质量高于相同胎龄体质量的第 90 百分位的新生儿定义为 LGA,具体标准参照《2015 年中国不同胎龄新生儿出生体重曲线研制》中的出生体重百分位数参考值^[13]。②母亲分娩时年龄:按照年龄数据将记录分为:<22、22~、30~、≥35 岁。③母亲身高:按照母亲身高数据将记录分为:<150、150~、160~、≥170 cm。④父亲身高:按照父亲身高数据将记录分为:<160、160~、170~、≥180 cm。⑤父母职业:将职业为农民和工人的记录归为“体力劳动”组,服务业归为“服务业”组,职员和经商归为“脑力劳动”组,家务归为“无业”组,其他归为“其他”组。⑥父母文化程度:将文盲、小学和初中者归为“初中及以下”组,高中者归为“高中”组,大学及研究生者归为“大专及以上”组。⑦孕期丈夫和妇女吸烟情况:将不吸烟和戒烟者归为“戒烟”组,吸烟量减少归为“减少”组,吸烟量不变和增加者归为“未戒烟”组。⑧父母生活、人际关系及经济压力:将“无”者归为“无”组,“很少”和“有一点”者归为“较小”组,“比较大”和“很大”者归为“较大”组。⑨人口密度:将西藏、新疆、青海、宁夏、甘肃、内蒙古 6 个省份划分为人口密度较低地区,其余省份为人口密度较高地区。

4. 统计学分析:在 Python 3.8.5 软件的环境下进行模型训练及数据分析。通过单因素显著性分析、单因素相关性分析 (Spearman) 及单因素逻辑回归分析,综合以上 3 项 P 值结果,将 $P>0.1$ 的特征进行删除。此外,经过特征共线性分析,将与其余特征之间有较强共线性的特征进行删除。最终保留 115 个特征 (表 1)。对数据集使用 9:1 比例随机划分训练集与测试集。此外,数据集中的阳性结局特征 LGA 的记录占比较低 ($n=12\ 279, 11.7\%$),属于不平衡数据集。为解决数据集不平衡的问题,本研究

通过下采样(under-sampling)的方法处理训练集,使用随机算法对阴性样本进行删除,进而得到目标等比的训练集。采用逻辑回归、决策树、随机森林、朴素贝叶斯、GBDT、XGBoost、AdaBoost、CatBoost、LightGBM、KNN 10种算法进行机器学习模型训练,评价各种算法建立的模型对 LGA 的预测价值。此外,在 115 个特征中选取既往研究证实的少量比较明确的 LGA 危险因素作为预测因子^[14-17],通过传统的逻辑回归方法构建 LGA 风险预测模型,比较机器学习模型与传统统计学模型的预测效能差异。计量资料采用中位数(P_{25}, P_{75})表示,其中符合正态分布的特征通过 t 检验进行组间比较,非正态分布的特征通过 Mann-Whitney U 检验进行组间比较。计数资料采用例数(百分比)的形式描述,通过 χ^2 检验进行组间比较。所有分析均采用双侧统计检验, $P < 0.05$ 为差异有统计学意义。

结 果

1. 基本特征:共纳入 104 936 名研究对象。其

中,男婴 54 856 例(52.3%),女婴 50 080 例(47.7%),胎龄为 40.0(39.0, 40.0)周,出生体重为 3 350.0(3 025.0, 3 600.0)g, LGA 的发生率为 11.7% (12 279 例)。纳入对象的社会人口学特征的单因素分析结果显示,母亲分娩年龄、母亲文化程度、母亲生活/工作压力、母亲是否服用叶酸及时间、父亲文化程度、父亲生活/工作压力、父亲饮酒情况在 LGA 组与非 LGA 组间的差异有统计学意义($P < 0.01$)。见表 1。

2. 预测模型结果:本研究人群中存在阳性结局的 LGA 新生儿仅占总新生儿人数的 11.7%,说明数据集存在严重的不平衡现象。因此,本研究通过 under-sampling 方法对原始数据集进行处理后,分别利用 10 种算法对原始数据集和平衡后的数据集构建 LGA 预测模型,预测模型的曲线下面积(AUC)、灵敏度、特异度、阳性预测值、阴性预测值。见表 2。在不平衡数据集中使用机器学习方法建模后,测试集中预测模型的 AUC 和特异度均较低。相比之下,经过数据平衡处理后,机器学习方法建立模型的整体效能出现显著提高,尤其以随机森

表 1 大于胎龄儿与非大于胎龄儿新生儿的社会人口学特征

特征指标	非大于胎龄儿(n=92 657)	大于胎龄儿(n=12 279)	P 值
男性(%)	46 374(50.1)	8 482(69.1)	<0.001
胎龄[周(P_{25}, P_{75})]	40.0(39.0, 40.0)	39.0(38.0, 40.0)	<0.001
出生体重[g(P_{25}, P_{75})]	3 250.0(3 000.0, 3 500.0)	4 000.0(3 900.0, 4 250.0)	<0.001
母亲分娩年龄组(岁, %)			0.002
<22	13 818(14.9)	1 691(13.8)	
22~	67 720(73.1)	9 020(73.4)	
30~	7 559(8.2)	1 076(8.8)	
≥35	3 560(3.8)	492(4.0)	
母亲文化程度(%)			<0.001
初中及以下	66 198(71.4)	8 572(69.8)	
高中	18 439(19.9)	2 493(20.3)	
大专及以上	8 020(8.7)	1 214(9.9)	
母亲生活/工作压力(%)			0.002
无	75 233(81.2)	10 082(82.1)	
较小	17 071(18.4)	2 134(17.4)	
较大	353(0.4)	63(0.5)	
母亲是否服用叶酸及时间(%)			0.005
停经前至少 3 个月	34 952(37.7)	4 570(37.2)	
停经前 1~2 个月	16 037(17.3)	2 284(18.6)	
停经后	21 860(23.6)	2 872(23.4)	
未服用	19 808(21.4)	2 553(20.8)	
父亲文化程度(%)			0.001
初中及以下	63 159(68.2)	8 245(67.1)	
高中	20 190(21.8)	2 672(21.8)	
大专及以上	9 308(10.0)	1 362(11.1)	
父亲生活/工作压力(%)			<0.001
无	71 338(77.0)	9 720(79.2)	
较小	20 678(22.3)	2 434(19.8)	
较大	641(0.7)	125(1.0)	
父亲饮酒(%)			<0.001
否	61 327(66.2)	7 732(63.0)	
是	31 330(33.8)	4 547(37.0)	

林、GBDT、LightGBM、XGBoost、CatBoost 模型最为显著。其中, CatBoost 模型的 AUC 从 0.692 增加到 0.932, 特异度从 93.1% 增加到 99.2%, 灵敏度从 20.2% 增加到 39.6%, 阳性预测值从 61.9% 增加至 95.2%, 阴性预测值从 67.8% 增加至 80.8%。

经过模型训练, 10 种算法建立的 LGA 预测模型中, CatBoost 模型对 LGA 的预测效能最佳, 在训练集中, 模型的受试者工作特征曲线(ROC)的 AUC 为 1.000; 在测试集中, 模型的 ROC 的 AUC 为 0.932。其次是 GBDT 模型, 在训练集中 ROC 的 AUC 为 0.977, 测试集中 AUC 为 0.866。在测试集中, 除逻辑回归模型在预测 LGA 风险方面表现较差, AUC 仅为 0.555 外, 其余 9 种机器学习算法建立的 LGA 预测模型的效能均较佳: XGBoost(AUC: 0.860), LightGBM(AUC: 0.841), 随机森林(AUC: 0.739), 决策树(AUC: 0.726), AdaBoost(AUC: 0.698), KNN(AUC: 0.676), 朴素贝叶斯(AUC: 0.602)。见图 1, 2。

从机器学习模型纳入的 115 个预测特征中, 选取已证实的少量比较明确的 LGA 危险因素作为预测因子, 包括父母孕前 BMI、母亲既往怀孕次数、母亲孕前血糖值、母亲年龄、父母身高、父母民族、胎儿性别、母亲妊娠第 1~6 月 PM2.5 暴露值, 通过传

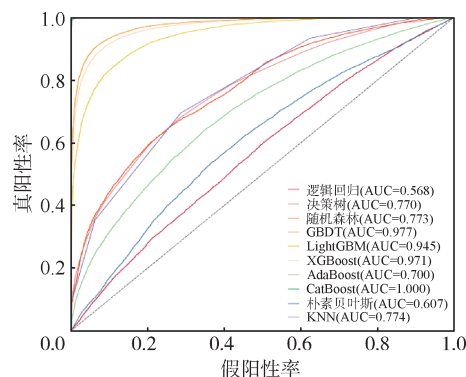


图 1 基于机器学习模型的训练集工作特征曲线

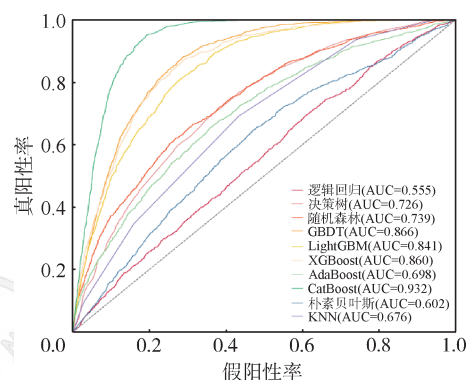


图 2 基于机器学习模型的测试集工作特征曲线

统的逻辑回归方法构建 LGA 风险预测模型。传统逻辑回归模型在测试集中的 AUC 仅为 0.584, 见图

表 2 大于胎龄儿预测模型的预测效能指标

模型	AUC(训练集)	AUC(测试集)	特异度	灵敏度	阳性预测值	阴性预测值
原始数据集						
逻辑回归	0.559	0.542	0.904	0.131	0.673	0.408
决策树	0.712	0.659	0.924	0.187	0.589	0.662
随机森林	0.748	0.688	0.935	0.185	0.684	0.603
GBDT	0.726	0.699	0.930	0.213	0.598	0.708
LightGBM	0.730	0.700	0.934	0.201	0.646	0.661
XGBoost	0.725	0.699	0.932	0.200	0.636	0.665
AdaBoost	0.691	0.692	0.942	0.181	0.737	0.560
CatBoost	0.870	0.692	0.931	0.202	0.619	0.678
朴素贝叶斯	0.604	0.600	0.924	0.148	0.718	0.455
KNN	0.789	0.554	0.896	0.144	0.385	0.698
under-sampling 数据集						
逻辑回归	0.568	0.555	0.906	0.131	0.696	0.389
决策树	0.770	0.726	0.949	0.189	0.771	0.563
随机森林	0.773	0.739	0.939	0.220	0.664	0.688
GBDT	0.977	0.866	0.979	0.288	0.883	0.711
LightGBM	0.945	0.841	0.966	0.277	0.810	0.721
XGBoost	0.971	0.860	0.977	0.285	0.872	0.711
AdaBoost	0.700	0.698	0.934	0.195	0.656	0.641
CatBoost	1.000	0.932	0.992	0.396	0.952	0.808
朴素贝叶斯	0.607	0.602	0.920	0.153	0.658	0.518
KNN	0.774	0.676	0.933	0.174	0.692	0.566

注: AUC: 曲线下面积

3. 其特异度、灵敏度、阳性预测值和阴性预测值分别为 91.7%、14.8%、66.3% 和 49.4%，均明显低于机器学习模型中表现最佳的 CatBoost 模型。

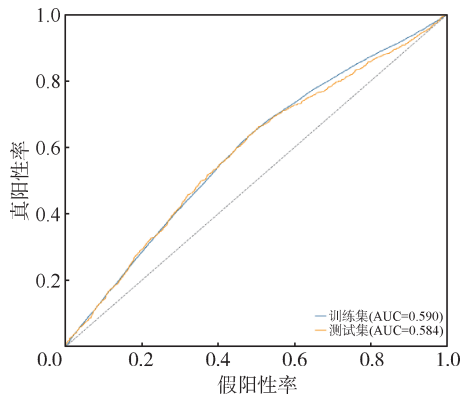


图3 基于传统逻辑回归模型的工作特征曲线

讨 论

本研究利用多中心、大样本数据集建立基于机器学习算法的 LGA 预测模型。结果表明,与传统的逻辑回归方法相比,利用机器学习算法可产生更有效的 LGA 预测模型,其中 CatBoost 模型对 LGA 的预测效能最佳,AUC 为 0.932。此外,本研究所采用的数据集属于不平衡数据集,LGA 新生儿($n=12\ 279$)和非 LGA($n=92\ 657$)新生儿的数量存在明显差异,基于此类阳性结局占比较少的预测模型往往表现不佳,通常因过度拟合低估了低危患者的事件发生率,同时高估了高危患者的事件发生风险^[18]。与预期一致,直接利用不平衡数据集建立的 LGA 预测模型的性能较差,而经过数据平衡处理后,机器学习预测模型的综合性能出现明显提高,说明通过数据平衡方法,可以有效地改善数据集不平衡的问题,对准确预测来说至关重要。

既往研究表明,孕妇具有糖尿病家族史、妊娠期糖尿病史、孕前 BMI 较高、孕前及孕期体重增加过多、LGA 分娩史、父亲 BMI 较高、胎次、男性新生儿等均为 LGA 的风险因素^[14-17]。如何通过综合这些 LGA 的风险因素建立日常可行的 LGA 孕期预测模型,是目前仍待解决的问题。虽然曾有研究尝试利用孕妇的社会人口学特征结合临床生化指标建立 LGA 的预测模型^[5],但建模仅依赖于传统的统计学方法,无法避免其存在的固有局限性。传统的统计学方法是模型驱动的,即从一个模型开始,并检查数据是否符合建立的模型,其潜在的假设是,数据由随机模型提供,验证是基于拟合优度检验,即

R^2 和 χ^2 检验^[19-20]。因此,其固有的局限性包括^[19-23]: 首先,传统的统计学方法通常假设数据符合正态分布,变量具有独立性和线性关联。然而,真实的数据是存在噪声的,并不符合这样的预先假设。其次,得到的结论是数据拟合模型,而非模型拟合数据,强制对数据行为进行严格的假设。此外,传统的统计学方法对大样本量数据集的处理能力是有限的。相比之下,可解释多种因素的机器学习算法的发展,可帮助建立稳健和准确的预测模型。

机器学习是人工智能的一种形式,它的理论基础为,系统从数据中学习、识别模式,并在无需明确编程的情况下做出决策,可通过使用从数据中迭代学习的算法预测复杂数据集的结果^[24]。机器学习算法不考虑对数据行为和变量预选的固定假设,而是允许数据通过检测或学习潜在模式来创建模型^[25]。既往已有研究提出,当阳性结果较少时,机器学习方法可以建立更优的预测模型^[26]。本研究采用了 10 种算法进行机器学习模型训练,其中逻辑回归、决策树、朴素贝叶斯和 KNN 属于简单模型,随机森林、GBDT、XGBoost、CatBoost、AdaBoost 和 LightGBM 属于集成模型,以上算法均适用于本研究中的二分类问题。

本研究首次利用多中心、大样本数据集建立了基于机器学习算法的孕早期 LGA 预测模型。结果表明,与传统的逻辑回归方法相比,利用机器学习算法可产生更有效的 LGA 预测模型。此外,在预测模型结果评估时,AUC 为 0.7~0.8 通常被解释为较好,0.8~0.9 为良好,0.9~1.0 为优秀^[27]。在本研究中,CatBoost 模型对 LGA 的预测效能最佳,在测试集中的 AUC 可达到 0.932,其特异度、灵敏度、阳性预测值和阴性预测值在 10 种算法模型中均为最高。CatBoost 算法是在 GBDT 算法框架下的一种改进实现,是一种基于对称决策树实现的参数较少、支持分类型变量和具有高准确性的算法,可以处理分类型、数值型变量,尤其对分类型变量表现优异,还解决了梯度偏差及预测偏移的问题,从而有效减少过拟合的发生,进而提高算法的准确性和泛化能力。这种算法采用特殊的方式处理分类型变量,极大减少了处理的工作量,还可以利用变量之间的联系,极大地丰富特征维度,为优化模型效果提供了可能性。然而,本研究存在一定的局限性。虽然模型的特异度、阳性预测值、阴性预测值分别高达 99.2%、95.2%、80.8%,但测试集中的灵敏度仅为 39.6%。原因可能与数据集中 LGA 的发生率不高

有关,尽管存在一些已知的危险因素,但每一种危险因素仅会导致 LGA 的发生风险小幅度增加^[28-30]。此外,当前模型中缺乏孕期超声结果等重要指标,也可能造成模型的灵敏度降低。如何使机器学习模型在测试集中仍维持较高的灵敏度已成为许多研究面临的问题^[24,28,31],值得进一步研究。

本研究首次利用多中心、大样本数据集建立了基于多种机器学习算法的孕期 LGA 风险预测模型。与传统的逻辑回归方法相比,利用机器学习算法可建立更有效的 LGA 预测模型,其中以 CatBoost 模型的预测效能最佳,AUC 可高达 0.932。虽然模型的误诊率极低,但灵敏度偏低,仍需进一步纳入更多重要特征以优化当前的风险预测模型。

利益冲突 所有作者均声明不存在利益冲突

参 考 文 献

- Esakoff TF, Cheng YW, Sparks TN, et al. The association between birthweight 4 000 g or greater and perinatal outcomes in patients with and without gestational diabetes mellitus[J]. *Am J Obstet Gynecol*, 2009, 200(6): 672.e1-672.e4. DOI:10.1016/j.ajog.2009.02.035.
- McGrath RT, Glastras SJ, Hocking SL, et al. Large-for-gestational-age neonates in type 1 diabetes and pregnancy: contribution of factors beyond hyperglycemia [J]. *Diabetes Care*, 2018, 41(8): 1821-1828. DOI:10.2337/dc18-0551.
- Oken E, Gillman MW. Fetal origins of obesity[J]. *Obes Res*, 2003, 11(4):496-506. DOI:10.1038/oby.2003.69.
- Yamamoto JM, Kallas-Koeman MM, Butalia S, et al. Large-for-gestational-age (LGA) neonate predicts a 2.5-fold increased odds of neonatal hypoglycaemia in women with type 1 diabetes[J]. *Diabetes Metab Res Rev*, 2017, 33(1):e2824. DOI:10.1002/dmrr.2824.
- Åmark H, Westgren M, Persson M. Prediction of large-for-gestational-age infants in pregnancies complicated by obesity: A population-based cohort study [J]. *Acta Obstet Gynecol Scand*, 2019, 98(6):769-776. DOI: 10.1111/aogs.13546.
- Deo RC. Machine learning in medicine[J]. *Circulation*, 2015, 132(20):1920-1930. DOI:10.1161/CIRCULATIONAHA.115.001593.
- Kourou K, Exarchos TP, Exarchos KP, et al. Machine learning applications in cancer prognosis and prediction [J]. *Comput Struct Biotechnol J*, 2015, 13: 8-17. DOI: 10.1016/j.csbj.2014.11.005.
- Pan LY, Liu GJ, Lin FQ, et al. Machine learning applications for prediction of relapse in childhood acute lymphoblastic leukemia[J]. *Sci Rep*, 2017, 7(1): 7402. DOI: 10.1038/s41598-017-07408-0.
- Shouval R, Bondi O, Mishan H, et al. Application of machine learning algorithms for clinical predictive modeling:a data-mining approach in SCT[J]. *Bone Marrow Transplant*, 2014, 49(3): 332-337. DOI: 10.1038/bmt.2013.146.
- 张世琨,王巧梅,沈海屏.中国免费孕前优生健康检查项目的设计、实施及意义[J].*中华医学杂志*, 2015, 95(3):162-165. DOI:10.3760/cma.j.issn.0376-2491.2015.03.002.
- Zhang SK, Wang QM, Shen HP. Design of the national free pre-conception health examination project in China[J]. *Natl Med J China*, 2015, 95(3):162-165. DOI:10.3760/cma.j.issn.0376-2491.2015.03.002.
- Wang YY, Li Q, Guo YM, et al. Association of long-term exposure to airborne particulate matter of 1 μm or less with preterm birth in China[J]. *JAMA Pediatr*, 2018, 172(3):e174872. DOI:10.1001/jamapediatrics.2017.4872.
- World Health Organization. Physical status: the use and interpretation of anthropometry[R]. Report of a WHO Expert Committee. Technical Report Series No. 854. Geneva:WHO, 1995.
- 朱丽,张蓉,张淑莲,等.中国不同胎龄新生儿出生体重曲线研制[J].*中华儿科杂志*, 2015, 53(2): 97-103. DOI: 10.3760/cma.j.issn.0578-1310.2015.02.007.
- Zhu L, Zhang R, Zhang SL, et al. Chinese neonatal birth weight curve for different gestational age[J]. *Chin J Pediatr*, 2015, 53(2): 97-103. DOI: 10.3760/cma.j.issn.0578-1310.2015.02.007.
- Júnior EA, Peixoto AB, Zamarian ACP, et al. Macrosomia[J]. *Best Pract Res Clin Obstet Gynaecol*, 2017, 38:83-96. DOI: 10.1016/j.bpobgyn.2016.08.003.
- Curti A, Zanello M, de Maggio I, et al. Multivariable evaluation of term birth weight: a comparison between ultrasound biometry and symphysis-fundal height[J]. *J Matern Fetal Neonatal Med*, 2014, 27(13): 1328-1332. DOI:10.3109/14767058.2013.858241.
- Levy A, Wiznitzer A, Holcberg G, et al. Family history of diabetes mellitus as an independent risk factor for macrosomia and cesarean delivery[J]. *J Matern Fetal Neonatal Med*, 2010, 23(2): 148-152. DOI: 10.3109/14767050903156650.
- He XJ, Qin FY, Hu CL, et al. Is gestational diabetes mellitus an independent risk factor for macrosomia: a meta-analysis? [J]. *Arch Gynecol Obstet*, 2015, 291(4): 729-735. DOI:10.1007/s00404-014-3545-5.
- Pavlou M, Ambler G, Seaman SR, et al. How to develop a more accurate risk prediction model when there are few events[J]. *BMJ*, 2015, 351: h3868. DOI: 10.1136/bmj.h3868.
- Breiman L. Statistical modeling: The two cultures (with comments and a rejoinder by the author) [J]. *Stat Sci*, 2001, 16(3):199-231. DOI:10.1214/ss/1009213726.
- Hand DJ. Data mining: statistics and more? [J]. *Am Stat*, 1998, 52(2):112-118.
- Bagley SC, White H, Golomb BA. Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain[J]. *J Clin Epidemiol*, 2001, 54(10): 979-985. DOI: 10.1016/s0895-4356(01)00372-9.
- Sun GW, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis[J]. *J Clin Epidemiol*, 1996, 49(8): 907-916. DOI:10.1016/0895-4356(96)00025-X.
- Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes[J]. *J Clin Epidemiol*, 1996, 49(11): 1225-1231. DOI:10.1016/s0895-4356(96)00002-9.
- O'Neill AC, Yang DY, Roy M, et al. Development and evaluation of a machine learning prediction model for flap failure in microvascular breast reconstruction[J]. *Ann Surg Oncol*, 2020, 27(9): 3466-3475. DOI: 10.1245/s10434-020-08307-x.
- Shouval R, Hadanny A, Shlomo N, et al. Machine learning for prediction of 30-day mortality after ST elevation myocardial infarction: an acute coronary syndrome Israeli survey data mining study[J]. *Int J Cardiol*, 2017, 246:7-13. DOI:10.1016/j.ijcard.2017.05.067.
- Kruppa J, Ziegler A, König IR. Risk estimation and risk prediction using machine-learning methods[J]. *Hum Genet*, 2012, 131(10): 1639-1654. DOI: 10.1007/s00439-012-1194-y.
- Kleinbaum DG, Klein M. Logistic regression[M]. New York, NY:Springer, 2010. DOI:10.1007/978-1-4419-1742-3.
- Cao Y, Fang X, Ottosson J, et al. A comparative study of machine learning algorithms in predicting severe complications after bariatric surgery[J]. *J Clin Med*, 2019, 8(5):668. DOI:10.3390/jcm8050668.
- Stenberg E, Szabo E, Ågren G, et al. Early complications after laparoscopic gastric bypass surgery: results from the Scandinavian Obesity Surgery Registry[J]. *Ann Surg*, 2014, 260(6): 1040-1047. DOI: 10.1097/SLA.0000000000000431.
- Finks JF, Kole KL, Yenumula PR, et al. Predicting risk for serious complications with bariatric surgery: results from the Michigan Bariatric Surgery Collaborative[J]. *Ann Surg*, 2011, 254(4): 633-640. DOI: 10.1097/SLA.0b013e318230058c.
- Miller R, Tumin D, Cooper J, et al. Prediction of mortality following pediatric heart transplant using machine learning algorithms[J]. *Pediatr Transplant*, 2019, 23(3): e13360. DOI:10.1111/ptr.13360.