

基于因果贝叶斯网络的急性肾损伤患者死亡风险预测

徐乃岳¹, 周亮², 刘坤¹, 周梦雨¹

(1.上海理工大学健康科学与工程学院, 上海 200093;

2.南京医科大学附属江宁医院医工融合实验室, 江苏 南京 211100)

✉ xunaiyue21@163.com; wenzhou6@sjtu.edu.cn; lkun1111@163.com; zhou_meng_yu_66@163.com



摘要:为了及早发现重症监护室中的急性肾损伤高危患者,为其提供适当的护理,实现医疗资源的合理利用,研究建立因果贝叶斯网络模型进行急性肾损伤高危患者死亡风险预测。从重症监护医学信息市场(Medical Information Mart for Intensive Care III, MIMIC-III)数据库中筛选了25个研究变量和3 870条患者数据,使用因果发现算法进行特征降维。通过NO TEARS算法构建因果图并建立因果贝叶斯网络进行实验,通过机器学习算法验证重要特征的合理性,并对网络结构进行因果效应估计,模型具有最高的受试者工作特征曲线下面积(Area Under the Receiver Operating Characteristic, AUROC)分数,为81.7%,优于逻辑回归(Logistic Regression, LR)、随机森林(Random Forest, RF)和极端梯度提升树(eXtreme Gradient Boosting, XGBoost)。此外,模型的重要特征预测能力在各种建模中都很稳健,构建的因果贝叶斯网络具有更好的预测效果并具备良好的解释能力。

关键词:急性肾损伤;因果贝叶斯网络;因果发现;死亡风险预测

中图分类号:TP391 **文献标志码:**A

Predicting Mortality Risk of AKI Patients Based on Causal Bayesian Network

XU Naiyue¹, ZHOU Liang², LIU Kun¹, ZHOU Mengyu¹

(1.School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China;

2.Biomedical Engineering Fusion Laboratory, Jiangning Hospital Affiliated to Nanjing Medical University, Nanjing 211100, China)

✉ xunaiyue21@163.com; wenzhou6@sjtu.edu.cn; lkun1111@163.com; zhou_meng_yu_66@163.com

Abstract: In order to promptly identify high-risk patients for Acute Kidney Injury (AKI) in the Intensive Care Unit (ICU), provide appropriate care, and achieve rational utilization of medical resources, this paper proposes to establish a causal Bayesian network model for predicting mortality risk in high-risk AKI patients. 25 study variables and 3 870 patient records are selected from the MIMIC-III (Medical Information Mart for Intensive Care III) database, and causal discovery algorithm is used for feature dimension reduction. The NO TEARS algorithm is employed to construct a causal graph and establish a causal Bayesian network for experimentation. Machine learning algorithm is utilized to validate the rationality of important features, and causal effect estimation is performed on the network structure. The model achieves the highest Area Under the Receiver Operating Characteristic (AUROC) score of 81.7%, which is superior to Logistic Regression (LR), Random Forest (RF), and eXtreme Gradient Boosting (XGBoost). Additionally, the predictive ability of important features in the model remains robust across various modeling scenarios. The proposed causal Bayesian network has better prediction performance and good interpretability.

Key words: AKI; causal Bayesian network; causal discovery; mortality risk prediction

0 引言(Introduction)

急性肾损伤(Acute Kidney Injury, AKI)是外科患者或危

重患者经常出现的一种复杂的全身综合征,具有高死亡风险^[1-2]。流行病学证据表明轻度、可逆的AKI也有严重的临床

后果^[3]。过去 20 年,AKI 患者的死亡率居高不下,其中住院患者的死亡率为 15%,重症监护室(Intensive Care Unit, ICU)中的患者死亡率为 50%^[4],全球每年约有 200 万人死于 AKI^[5-6]。

随着临床医学信息学的进步和电子健康记录完整性的提高,机器学习算法已被广泛地应用于构建精确的风险预测模型。复杂的机器学习模型具有高复杂度和不可解释的特征,无法在临床中获得医护人员的信任。复杂模型通常依赖完备的高维数据进行预测,在病人进入 ICU 的初期,因为时间较短,所以难以进行较多的实验室检测,导致模型无法发挥出早期预测的优势^[7]。构建适用于 ICU 早期的 AKI 高风险患者的预测模型具有一定的挑战性,但对于制定 AKI 治疗的新策略至关重要^[8]。

1 相关工作(Related work)

近年来,随着机器学习技术的不断发展,医护人员对 ICU 中患者死亡风险的预测方法有了更多的选择^[9]。医护人员可以使用机器学习技术和计算密集型统计建模方案快速评估大量复杂的数据,用以确定患者的死亡风险^[10]。许多研究人员主张使用机器学习开发复杂临床场景的预测模型,由于临床特征和患者结局之间的关系多是非线性的,相较于传统的统计学方案和最普遍的逻辑回归(Logistic Regression, LR)算法,机器学习算法可以更好地捕获高维数据中的非线性关系,更有效地利用电子健康病历并挑选出高维数据中对预测贡献显著的数据字段,提高算法模型的预测性能^[11]。

LIN 等^[12]使用随机森林(Random Forest, RF)算法构建了一种死亡率预测模型,根据 AKI 患者的尿量、收缩压、年龄、血清碳酸氢盐和心率等变量预测其死亡率,避免延误高危患者的 AKI 治疗。AWAD 等^[13]使用 20 个变量,包括人口统计学、生命体征和实验室测试变量构建 RF 模型,用于 ICU 患者的早期死亡率预测,他们发现 RF 模型在死亡率预测方面表现出良好的性能。KIM 等^[14]使用人口统计学、慢性健康状况和生理学等相关变量构建用于 ICU 患者死亡率预测的支持向量机(Support Vector Machine, SVM)模型,并将其性能同急性生理与慢性健康评分(Acute Physiology and Chronic Health Evaluation III, APACHE III)的性能进行比较,发现 SVM 模型的性能优于 APACHE III。XU 等^[15]研究使用 ML 模型,即 LR、RF 和极端梯度提升树(eXtreme Gradient Boosting, XGBoost)预测 AKI 患者的死亡风险,并进行分层。蔺轲等^[16]使用 SVM 算法构建 ICU 急性肾损伤患者的死亡风险预测模型,同时将该模型与简化急性生理评分(Simplified Acute Physiology Score II, SAPS II)进行对比,表现出更好的模型性能。但是,上述方法存在以下问题:(1)它们大多针对特定模型,需要较多的预测特征,不具有普适性,需要一种精简并具有良好鲁棒性的方法,以适应日益复杂的医学场景;(2)关于如何预测 AKI 风险的内部流程无法彻底呈现,仅获得特征重要性评分无法解释特征与结局之间如何发挥作用;(3)以上方法只能提供全局解释,不能为每个案例提供个性化信息。

本研究基于因果层次结构^[17]理论,采用因果发现和贝叶斯网络方法构建因果贝叶斯网络进行 ICU 中 AKI 患者的死亡风险预测研究。

2 数据与方法(Data and methods)

2.1 数据

本研究使用的患者数据来自重症监护医学信息市场(Medical Information Mart for Intensive Care III, MIMIC-III)数据库^[18],该数据库整合了 2001—2012 年贝斯以色列女执事

医疗中心收治患者的综合临床数据^[19],仅限于 MIMIC-III 数据库的 MetaVision 信息系统中成年的患者(年龄为 18~89 岁)的首次 ICU 住院记录。首先从国际疾病分类(International Classification of Diseases, ICD)中查找急性肾损伤疾病定义的疾病代码;其次从 MIMIC-III 数据库找出收集的 ICU 中的急性肾损伤患者的全部信息;最后使用 Python 对信息进行筛选,只保留进入一次 ICU 的病人数据。本研究所做出的预测必须基于病人是第一次入院的前提,挑选出病人入院 24 h 内的生理数据并对搜索过程中产生的表格进行聚合,数据获取流程如图 1 所示。

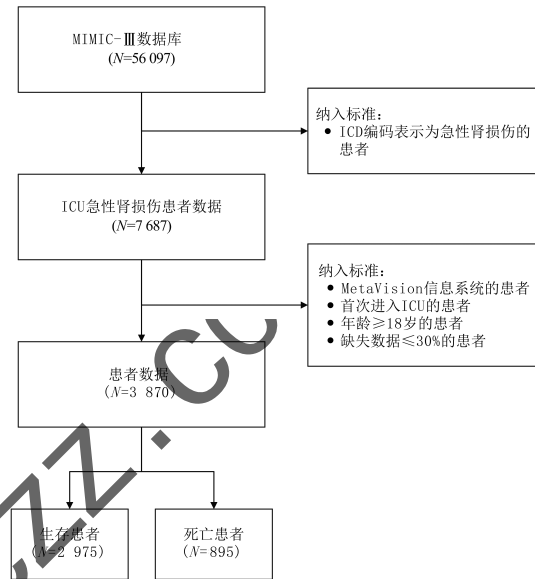


图 1 数据获取流程图

Fig. 1 Data acquisition process diagram

通过广泛的文献综述和向 ICU 专业人员的咨询,选择以下 25 个特征变量进行分析。

人口统计学(2 个):年龄(Age)、性别(Gender)。

生命体征(6 个):体温(Temperature, Temp)、心率(Heart Rate, HR)、呼吸频率(Respiratory Rate, Resp)、动脉血氧饱和度(SpO₂)、疼痛(Pain)、格拉斯昏迷量表(Glasgow Coma Scale, GCS)。

实验室测量值(17 个):碳酸氢盐(HCO₃)、血尿素氮(Blood Urea Nitrogen, BUN)、氯化物(Chloride, CL)、葡萄糖(Glucose, GLU)、血细胞比容(Hematocrit, HCT)、血红蛋白(Hemoglobin, HB)、国际标准化比值(International Normalized Ratio, INR)、血小板计数(Blood Platelet Count, PLT)、钾(Kalium, K)、凝血酶原时间(Prothrombin Time, PT)、部分凝血活酶时间(Partial Thromboplastin Time, PTT)、血清肌酐(Serum Creatinine, SCr)、钠(Natrium, Na)、白细胞计数(White Blood Cell Count, WBC)、总胆红素(Total Bilirubin, TBIL)、丙氨酸转氨酶(Alanine Transaminase, ALT)、天门冬氨酸氨基转移酶(Aspartate Aminotransferase, AST)。

从数据库中筛选出的实验数据存在部分缺失值,根据病人信息中的各项生理数据进行数据筛选,去除缺失率在 30% 以上的病人数据和异常值,并使用随机森林算法进行数据插补,确定了 25 个特征变量和 3 870 例患者。

根据《默沙东诊疗手册》,将连续数据进行离散化处理,方便模型预测与推理,变量离散信息如表 1 所示。

表1 变量离散信息
Tab.1 Variable discrete information

变量	分类	例数 (N=3 870)	范围	变量	分类	例数 (N=3 870)	范围
Gender	男	1 703(0.44)	M	GLU	低	81(0.02)	≥0 <69
	女	2 167(0.56)	F		中	927(0.24)	≥70 ≤105
CL	低	500(0.13)	≥0 <98		高	2 862(0.74)	>105
	中	1 644(0.42)	≥98 ≤106	HB	异常	3 207(0.83)	≥0 <12
	高	1 726(0.45)	>106		正常	663(0.17)	≥12 ≤17
HCT	异常	3 205(0.83)	≥0 <36	K	低	496(0.13)	≥0 <3.5
	正常	665(0.17)	≥36 ≤51		中	2 964(0.77)	≥3.5 ≤5
INR	低	1 783(0.46)	≥0.8 <1.2		高	410(0.10)	>5
	中	1 504(0.39)	≥1.2 <2	PLT	低	1 295(0.34)	≥0 <150
	高	583(0.15)	≥2		中	2 176(0.56)	≥150 ≤350
Na	低	1 101(0.29)	≥0 <136		高	399(0.10)	>350
	中	2 482(0.64)	≥136 ≤145	PTT	低	324(0.08)	≥0 <25
	高	287(0.07)	>145		中	1 811(0.47)	≥25 ≤35
PT	低	42(0.01)	≥0 <11		高	1 735(0.45)	>35
	中	1 175(0.30)	≥11 ≤13	SCr	低	290(0.07)	≥0 <0.7
	高	2 653(0.69)	>13		中	1 341(0.35)	≥0.7 ≤1.3
WBC	低	329(0.09)	≥0 <4.5		高	2 239(0.58)	>1.3
	中	1 780(0.46)	≥4.5 ≤11	HCO ₃	低	2 231(0.58)	≥0 <23
	高	1 761(0.45)	>11		中	1 190(0.31)	≥23 ≤28
GCS	低	1 284(0.33)	≥3 ≤8		高	449(0.11)	>28
	较低	2 586(0.67)	≥9 ≤12	Pain	低	1 988(0.51)	≥0 <1
HR	低	123(0.03)	≥0 <60		中	1 242(0.32)	≥1 <4
	中	2 869(0.74)	≥60 ≤100		高	640(0.17)	≥4
	高	878(0.23)	>100	SpO ₂	异常	26(0.01)	≥0 <90
Resp	低	45(0.01)	≥0 <12		正常	3 844(0.99)	≥90 ≤100
	中	2 305(0.60)	≥12 ≤20	ALT	正常	2 244(0.58)	≥0 ≤35
	高	1 520(0.39)	>20		异常	1 626(0.42)	>35
Temp	正常	3 293(0.85)	≤37.3	Age	1	58(0.01)	≥18 <28
	异常	577(0.15)	>37.3		2	135(0.03)	≥28 <38
TBIL	低	3 096(0.81)	<3.4		3	284(0.07)	≥38 <48
	中	620(0.16)	≥3.4 ≤17.1		4	621(0.17)	≥48 <58
	高	154(0.03)	>17.1		5	771(0.20)	≥58 <68
AST	正常	1 742(0.45)	≥0 ≤35		6	768(0.20)	≥68 <78
	异常	2 128(0.55)	>35		7	1 233(0.32)	≥78
BUN	异常	2 857(0.26)	≥0 <20				
	正常	1 013(0.74)	≥20				

表1中,BUN、SCr、GLU 的单位为 mg/dL, HB 的单位为 g/dL,CL、HCO₃ 的单位为 mmol/L,TBIL 的单位为 μmol/L, ALT、AST 的单位为 U/L,HCT、SpO₂ 的单位为%,K、Na 的单位为 mEq/L,WBC 的单位为 10³ cells/mcL,PLT 的单位为 10³/μL,PT、PTT 的单位为 s,Resp、HR 的单位为次/分钟,Temp 的单位为℃,INR、Pain、GCS、Age、Gender 无计量单位。

2.2 方法

许多学科的基本任务是发现、建模和理解自然现象背后的因果关系,数据科学正在从当前以数据为中心的范式转向以科学为中心的范式。通过分析纯粹的观察数据揭示因果信息,称为因果发现,它在近年来引起了很多研究人员的关注。从数据中学到的因果关系可以从感知和认知的角度进行测量、描述和评估,学习因果关系的一种传统且可行的方法是将贝叶斯网络(Bayesian Network, BN)扩展到因果贝叶斯网络,它可以通过生成因果定向无环图表示因果关系^[20]。

贝叶斯网络可以提供基于统计学原理的、可解释的预测结果,为获取和表示许多领域的知识提供了一种有效方法^[21]。为了将从数据中挖掘的因果知识更好地应用于 ICU 中 AKI 患者的死亡风险评估,本文基于因果层次结构理论在贝叶斯网络学习的框架内评估因果关系并构建因果贝叶斯网络进行预测。帮助医生重点关注具有高风险的患者,为 ICU 中医疗资源的合理分配提供帮助。

因果发现算法能够捕捉到变量之间的因果关系,因果关系算法有助于精准地进行风险预测,筛选出的特征可以更好地适用于复杂多变的医学环境。贝叶斯网络自身固有的有向无环图(Directed Acyclic Graph, DAG)特征和统计学优势,使因果发现算法和贝叶斯网络的结合可以方便地进行预测推理和模型解释。

采用 PC(Peter-Clark)、GES(Greedy Equivalence Search)、GRASP(Greedy Randomized Adaptive Search Procedure)三种因果发现算法进行投票加权处理,从所有的特征中发现有助于预测的真正风险因素进行特征降维,筛选出特征中投票数最高的部分特征作为重要特征。本文从 25 个特征变量中挑选出 PTT、TBIL、SCr、Resp、SpO₂,为了评估所提取特征的预测能力,使用机器学习方法进行建模和预测的准确性检查,重要特征获取流程如图 2 所示。

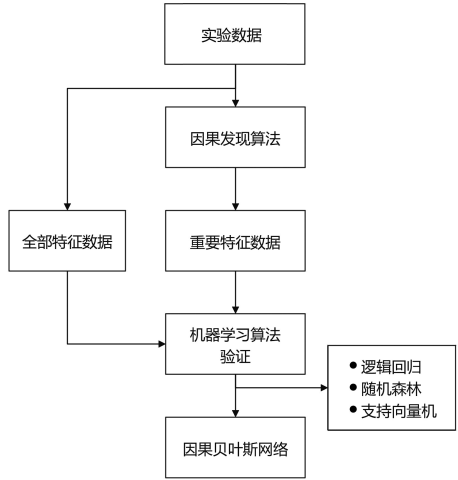


图2 重要特征获取流程图

Fig. 2 Important feature acquisition process diagram

确定有向无环图的结构非常具有挑战性,这是因为 DAG 的搜索空间是组合的,并且会随着节点数量的增加而进行超指数扩展。NO TEARS 算法将结构学习问题表述为一个在实矩阵上的连续优化问题,避免了这种组合约束,是通过对非循环的一种新颖的描述实现的,这种描述平滑且准确,可以更好地被用来发现数据中的因果结构^[22]。基于发现的因果图结构建立因果贝叶斯网络模型并进行分析,与建立在模式识别和相关性分析基础上的传统机器学习方法相比,利用贝叶斯网络能更直观地描述因果关系,因果贝叶斯网络模型结构如图 3 所示。

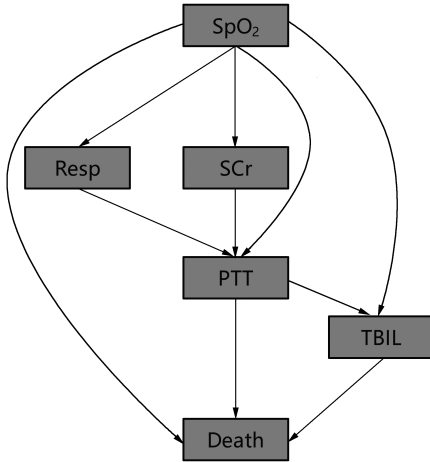


图 3 因果贝叶斯网络模型结构图

Fig. 3 Structure diagram of causal Bayesian network model

借助 CausalNex 软件包为模型引入干预操作,可以对数据中的任何节点应用干预更新其分布并分析干预措施的影响,如图 4 所示,当 SCr 和 TBIL 特征受到人为干预操作时,就会切断一切指向该特征的路径。此时,这些特征的值不再受到其他父级特征的影响,但是可以影响子级特征,同时对病人死亡风险的预测结果也会发生改变。干预操作可以帮助医护人员验证治疗方案的结果是否符合预期,为治疗方案的制订提供参考。

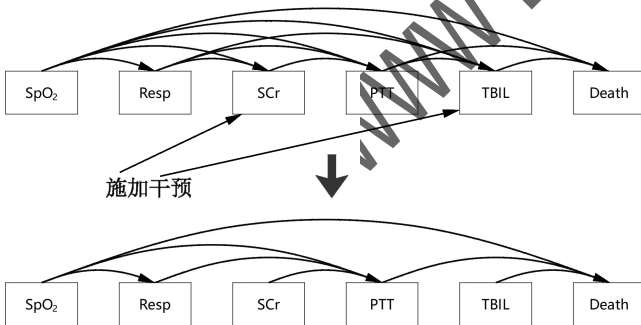


图 4 干预演示图

Fig. 4 Intervention demonstration diagram

3 结果(Result)

3.1 模型评估

本研究使用前人文献中使用的 LR、RF 和 XGBoost 进行验证。图 5 和图 6 分别展示了使用全部特征和重要特征建模时 3 种算法的十次交叉验证的准确率,图 7 对基于全部特征建模和重要特征建模的评估进行对比,可以明显地发现特征筛选前后准确率的差距都在 4% 以内,选定特征的预测能力在各种建模技术中都表现得相当稳健,表明仅依靠 5 个选定特征的模

型就可以有效地识别高危患者。

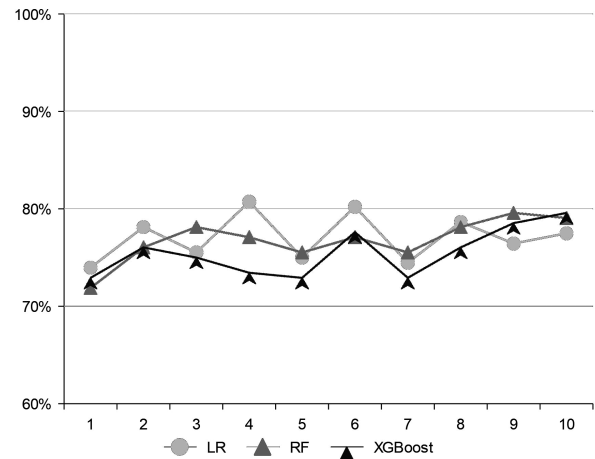


图 5 基于全部特征建模评估图

Fig. 5 Modeling evaluation chart based on all features

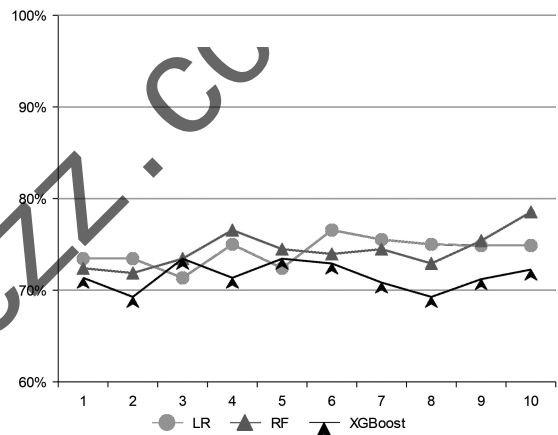


图 6 基于重要特征建模评估图

Fig. 6 Modeling evaluation chart based on important features

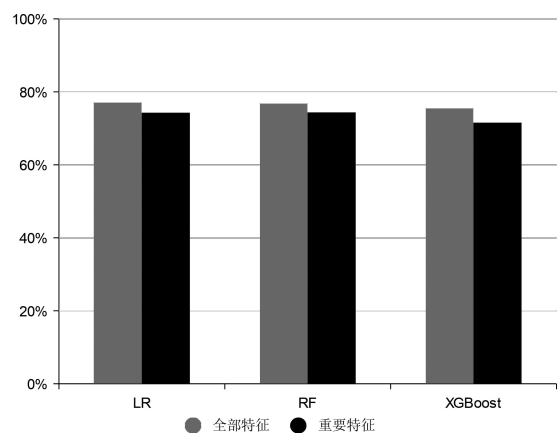


图 7 基于全部特征建模和重要特征建模的评估对比图

Fig. 7 Comparison diagram of modeling evaluation based on all features and important features

使用 NO TEARS 因果发现算法分别基于全部特征和重要特征进行建模预测并进行准确率 (Accuracy)、精准率 (Precision)、召回率 (Recall) 和受试者工作特征曲线下面积

(AUROC)的对比,降维前后性能对比图如图 8 所示。

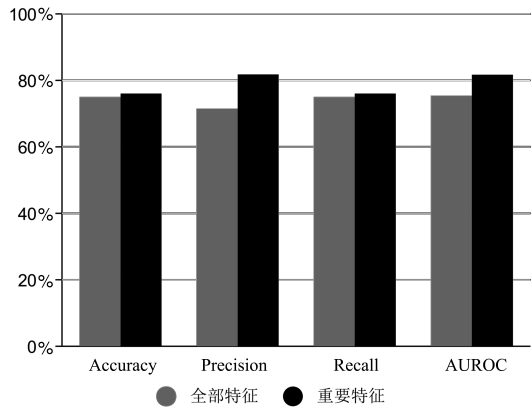


图 8 降维前后性能对比图

Fig. 8 Performance comparison chart before and after dimension reduction

对全部特征和重要特征都进行因果贝叶斯网络的建模,基于 Accuracy、Precision、Recall、AUROC 等评价指标进行分析。基于重要特征建模的预测性能比基于全部特征建模时的效果更好,重要特征排除了其他贡献较低的因素成为模型预测的主要依据。通过对特征降维前后的多种机器学习算法的建模进行对比发现,使用因果发现算法降维后的重要特征具有较强的预测能力,能够在多种不同的模型中发挥出良好的预测效果,有着优秀的泛化性能。

将因果贝叶斯网络模型与 LR、RF 和 XGBoost 三种算法筛选后的重要特征数据进行对比,基于 AUROC、Accuracy 和 Precision 三个评价指标展开分析,模型性能评估结果如表 2 所示。

表 2 模型性能评估结果

Tab.2 Results of model performance evaluation

模型	AUROC/%	Accuracy/%	Precision/%
因果贝叶斯网络	81.7	76.0	81.8
LR	72.3	75.1	63.4
RF	74.6	76.8	65.3
XGBoost	71.7	75.4	53.9

模型性能对比图如图 9 所示,Accuracy 反映模型总体的分类准确情况,结果显示因果贝叶斯网络的准确率略低于 LR 和 RF 的准确率,依然有着不错的准确度;AUROC 反映的是模型对于任意一例正负样本的区分能力,性能评估结果显示,因果贝叶斯网络模型的 AUROC 远高于 LR、RF 和 XGBoost,证明对于任意一例正负样本因果贝叶斯网络模型的正确区分结果的能力更好;Precision 着重评估在预测为 Positive 的所有数据中,真实 Positive 的数据占比,结果显示因果贝叶斯网络模型效果最好,表明模型可以准确地发现高危患者。

3.2 模型解释

模型可以从两个方面进行可解释分析。

3.2.1 模型的结构

因果贝叶斯网络模型的结构主要通过 NO TEARS 因果发现算法构建,可以很好地捕捉到变量之间的因果关系。模型的网络结构清楚地展现了模型预测过程中潜在的特征依赖关系,具有很高的透明度,这种可视化极大地简化了对因果关系的理解。

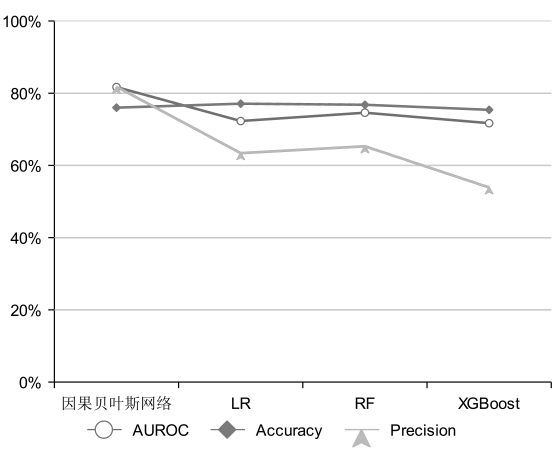


图 9 模型性能对比图

Fig. 9 Comparison chart of models performance

如图 10 所示,以 PTT 变量为例,使用微软 DoWhy 因果效应估计软件包对模型进行评估,a 列表示增加一个混杂因子前后平均效应如何变化,测试该结构的鲁棒性,差距越小则模型鲁棒性越好;b 列表示使用随机变量替代干预变量时平均效应如何变化,在理想状态下,新的平均效应应该为 0;c 列表示使用随机子集查看平均效应如何变化,说明模型具有很好的泛化能力。模型网络结构在 a、b、c 这 3 种情况下的因果效应估计都有着很好的表现,说明具有可信的因果效应。

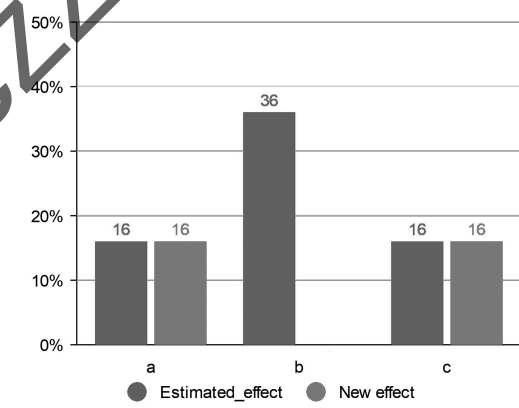


图 10 因果效应估计图

Fig. 10 Estimation chart of causal effect

3.2.2 预测的依据

因果贝叶斯网络模型的本质依然是一种概率图模型,在具有因果特性的图结构中支持使用概率推理的方式进行预测,各变量的不同取值概率可以推理出不同的预测结果,解释预测结果的推理依据。因果贝叶斯网络模型不仅可以基于各变量取值预测死亡风险概率,还可以通过逆向推理当死亡风险最大时各变量取值的概率,用来进行诊断推理,概率推理演示如图 11 所示。

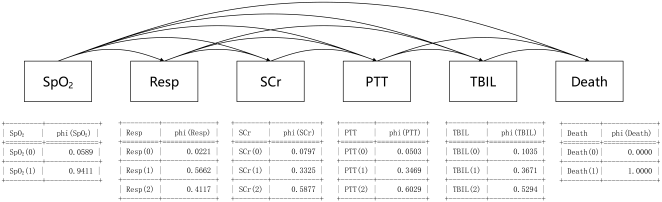


图 11 概率推理演示图

Fig. 11 Demonstration diagram of probabilistic reasoning

4 结论(Conclusion)

为了实现ICU中AKI高危患者的及早发现、及早治疗,本研究开发了一个患者死亡风险预测模型,通过患者进入ICU 24 h内的生命体征进行可解释的预测。使用因果发现算法寻找到的PTT、TBIL、SCr、Resp、SpO₂五个特征变量在LR、RF和XGBoost三种机器学习模型的验证下都表现出优异的预测性能,这五个特征变量可以作为ICU中AKI患者死亡风险预测研究的重要特征。基于重要特征构建的因果贝叶斯网络模型在微软的DoWhy因果效应估计中也表现出良好的因果效应和鲁棒性,因果贝叶斯网络模型结构透明的天然优势,使因果贝叶斯网络模型的预测流程易于理解,同时因果贝叶斯网络模型具备初步的干预能力,对于医学研究有着重要意义。通过学习因果发现算法选择的特征和因果关系,可以提高分类准确性。在研究中没有考虑患者的合并症和治疗措施,基于以上研究结果,有望进一步扩展NO TEARS等因果发现算法,从更广泛的医疗数据中挖掘影响患者结局的关键因素。未来的研究工作将着眼于探究影响ICU急性肾损伤患者的真正风险因素的同时,寻找合适的干预措施,降低患者最终死亡概率。此外在接下来的研究中还需要通过对因果理论的深入研究赋予模型更为准确的解释方法和推理手段,以更有效地助力医学决策。

参考文献(References)

- [1] KELLUM J A, PROWLE J R. Paradigms of acute kidney injury in the intensive care setting[J]. *Nature reviews nephrology*, 2018, 14: 217-230.
- [2] BANG J Y, LEE J B, YOON Y, et al. Acute kidney injury after infrarenal abdominal aortic aneurysm surgery: a comparison of AKIN and RIFLE criteria for risk prediction[J]. *British journal of anaesthesia*, 2014, 113(6): 993-1000.
- [3] KELLUM J A, LAMEIRE N, for the KDIGO AKI Guideline Work Group. Diagnosis, evaluation, and management of acute kidney injury: a KDIGO summary (Part 1)[J]. *Critical care*, 2013, 17: 1-15.
- [4] WALD R, MCARTHUR E, ADHIKARI N K J, et al. Changing incidence and outcomes following dialysis-requiring acute kidney injury among critically ill adults: a population-based cohort study[J]. *American journal of kidney diseases*, 2015, 65(6): 870-877.
- [5] WANG Y, BELLOMO R. Cardiac surgery-associated acute kidney injury: risk factors, pathophysiology and treatment[J]. *Nature reviews nephrology*, 2017, 13: 697-711.
- [6] HOSTE E A J, KELLUM J A, SELBY N M, et al. Global epidemiology and outcomes of acute kidney injury[J]. *Nature reviews nephrology*, 2018, 14: 607-625.
- [7] BHATRAJU P K, ZELNICK L R, KATZ R, et al. A prediction model for severe AKI in critically ill adults that incorporates clinical and biomarker data[J]. *Clinical journal of the American society of nephrology*, 2019, 14(4): 506-514.
- [8] MEERSCH M, SCHMIDT C, HOFFMEIER A, et al. Prevention of cardiac surgery-associated AKI by implementing the KDIGO guidelines in high risk patients identified by biomarkers: the PrevAKI randomized controlled trial[J]. *Intensive care medicine*, 2017, 43: 1551-1561.
- [9] LIU V X, PRESCOTT H C. Precision delivery in critical care: balancing prediction and personalization[C]//VINCENT J L. Annual Update in Intensive Care and Emergency Medicine 2019. Cham: Springer, 2019: 15-27.
- [10] NEMATI S, HOLDER A, RAZMI F, et al. An interpretable machine learning model for accurate prediction of sepsis in the ICU[J]. *Critical care medicine*, 2018, 46(4): 547-553.
- [11] POUCKE S V, ZHANG Z, SCHMITZ M, et al. Scalable predictive analysis in critically ill patients using a visual open data analysis platform[J]. *PLOS ONE*, 2016, 11(1): e0145791.
- [12] LIN K, HU Y H, KONG G L. Predicting in-hospital mortality of patients with acute kidney injury in the ICU using random forest model[J]. *International journal of medical informatics*, 2019, 125: 55-61.
- [13] AWAD A, BADER-EL-DEN M, MCNICHOLAS J, et al. Early hospital mortality prediction of intensive care unit patients using an ensemble learning approach[J]. *International journal of medical informatics*, 2017, 108: 185-195.
- [14] KIM S, KIM W, PARK R W. A comparison of intensive care unit mortality prediction models through the use of data mining techniques[J]. *Healthcare informatics research*, 2011, 17(4): 232-243.
- [15] XU Z X, LUO Y, ADEKKANATTU P, et al. Stratified mortality prediction of patients with acute kidney injury in critical care[J]. *Studies in health technology and informatics*, 2019, 264: 462-466.
- [16] 简柯, 谢俊卿, 胡永华, 等. 支持向量机在ICU急性肾损伤患者住院死亡风险预测中的应用[J]. *北京大学学报(医学版)*, 2018, 50(2): 239-244.
- [17] SHPITSER I, PEARL J. Complete identification methods for the causal hierarchy[J]. *Journal of machine learning research*, 2008, 9: 1941-1979.
- [18] JOHNSON A E W, POLLARD T J, SHEN L, et al. MIMIC-III, a freely accessible critical care database[J]. *Scientific data*, 2016, 3(1): 1-9.
- [19] 徐良辰, 郭崇慧. 基于时间序列特征表示与信息融合的ICU患者死亡风险预测[J]. *系统工程理论与实践*, 2022, 42(10): 2815-2828.
- [20] KONG H, SHI X H, WANG L M, et al. Averaged tree-augmented one-dependence estimators[J]. *Applied intelligence*, 2021, 51: 4270-4286.
- [21] JIANG L X, ZHANG L G, YU L J, et al. Class-specific attribute weighted naive Bayes[J]. *Pattern recognition*, 2019, 88: 321-330.
- [22] ZHENG X, ARAGAM B, RAVIKUMAR P, et al. DAGs with NO TEARS: continuous optimization for structure learning[J]. *Advances in neural information processing systems*, 2018, 31: 9472-9483.

作者简介:

徐乃岳(1999-),男,硕士生。研究领域:医疗数据分析。

周亮(1981-),男,博士,副教授。研究领域:数据建模与优化计算,医疗大数据分析与决策支持。本文通信作者。

刘坤(1998-),男,硕士生。研究领域:医疗数据分析。

周梦雨(1999-),女,硕士生。研究领域:医学图像分割。