

# 基于MIMIC-IV数据库的万古霉素血药谷浓度机器学习分类预测模型构建<sup>△</sup>

林小惠<sup>\*</sup>, 汪余嘉, 张玲玲, 许双临<sup>#</sup>(宁德师范学院附属宁德市医院临床药学室, 福建宁德 352100)

中图分类号 R969.3;TP181 文献标志码 A 文章编号 1001-0408(2025)19-2448-06

DOI 10.6039/j.issn.1001-0408.2025.19.16



**摘要 目的** 构建万古霉素血药谷浓度的分类预测模型,优化其精准用药策略。**方法** 从重症监护医学信息集市数据库中筛选符合条件的患者数据,经过数据清洗和预处理,最终纳入9 902例患者,结合相关性分析和Boruta特征选择算法进行特征选择,根据临床治疗窗标准离散化万古霉素血药谷浓度结果为低浓度( $<10 \mu\text{g}/\text{mL}$ )、中浓度( $10\sim20 \mu\text{g}/\text{mL}$ )和高浓度( $\geq20 \mu\text{g}/\text{mL}$ )。采用6种机器学习算法:表格先验数据拟合网络(TabPFN)、逻辑回归(LR)、随机森林(RF)、极端梯度提升(XGBoost)、支持向量机(SVM)、K近邻(KNN)构建分类模型,通过10折交叉验证(10-CV)评估模型性能,主要性能评估指标包括准确率、平衡准确率、宏平均精确率、宏平均召回率、宏平均F1、多类ROC曲线的曲线下面积(OvR-AUC)。采用沙普利加性解释(SHAP)分析不同特征对模型预测结果的影响方向与强度。**结果** RF和TabPFN模型表现最优(准确率为0.741 4和0.737 7,OvR-AUC为0.907 0和0.895 8),XGBoost模型表现中等,而LR、SVM和KNN模型的性能较差。混淆矩阵热力图显示,RF和TabPFN模型在高浓度类别上的预测准确率较高,但在低、中浓度类别上的表现略显不足。自举法结合10-CV评估显示,RF模型各项性能评价指标表现稳定(准确率0.741 4,平衡准确率0.740 3,宏平均精确率0.732 1,宏平均召回率0.736 0,宏平均F1 0.736 0,OvR-AUC 0.907 0),具备良好的分类性能与判断能力。SHAP法分析发现,肌酐、尿素氮及万古霉素日累计量和给药频率等关键特征对预测结果具有显著影响。**结论** RF和TabPFN模型在万古霉素血药谷浓度分类预测任务中表现出一定优势,在低、中浓度类别上的表现仍有改进空间。

**关键词** 机器学习;万古霉素;血药浓度;MIMIC-IV数据库;分类预测

## Construction of machine learning classification prediction model for vancomycin blood concentrations based on MIMIC-IV database

LIN Xiaohui, WANG Yujia, ZHANG Lingling, XU Shuanglin (Dept. of Clinical Pharmacy, Ningde Municipal Hospital Affiliated to Ningde Normal University, Fujian Ningde 352100, China)

**ABSTRACT OBJECTIVE** To construct a classification prediction model for vancomycin blood concentration, and to optimize its precision dosing strategies. **METHODS** Patient records meeting inclusion criteria were extracted from the Medical Information Mart for Intensive Care database. Following data cleaning and preprocessing, a final cohort of 9 902 patient was analyzed. Feature selection was performed through correlation analysis and the Boruta feature selection algorithm. Vancomycin blood concentrations were discretized into three categories based on clinical therapeutic windows: low ( $<10 \mu\text{g}/\text{mL}$ ), intermediate ( $10\sim20 \mu\text{g}/\text{mL}$ ), and high ( $\geq20 \mu\text{g}/\text{mL}$ ). Six machine learning algorithms were employed to construct classification models: tabular prior-data fitted network (TabPFN), logistic regression (LR), random forest (RF), extreme gradient boosting (XGBoost), support vector machine (SVM), K-nearest neighbors (KNN). Model performance was evaluated using 10-fold cross-validation (10-CV), with primary metrics including: accuracy, balanced accuracy, precision macro, recall macro, macro F1, area under the receiver operating characteristic curve (OvR-AUC). Shapley Additive Explanations (SHAP) was adopted to analyze the direction and magnitude of the impact that different features had on the model's predictive outcomes. **RESULTS** The results showed that the RF and TabPFN models performed the best (with accuracy of 0.741 4 and 0.737 7, and OvR-AUC of 0.907 0 and 0.895 8, respectively). XGBoost model exhibited moderate performance, while LR, SVM, and KNN models demonstrated relatively poor performance. Confusion matrix heatmap analysis revealed that both RF and TabPFN achieved higher accuracy in predicting high-concentration cases but exhibited slightly lower performance in the low and medium concentration categories. Bootstrap with 10-CV revealed that the RF model demonstrated stable performance across various evaluation metrics (accuracy: 0.741 4; balanced accuracy: 0.740 3; precision macro: 0.732 1; recall macro: 0.736 0; macro F1: 0.736 0; OvR-AUC: 0.907 0), indicating good classification performance and generalization ability. SHAP analysis revealed that creatinine, urea nitrogen, daily cumulative dose and administration frequency of vancomycin, which were key predictors, had a significant impact on the prediction results. **CONCLUSIONS** RF and TabPFN models demonstrate certain advantages in the classification prediction

<sup>△</sup>基金项目 福建省自然科学基金联合资助项目(No.2024J01942);  
宁德师范学院校级科研项目(No.2023ZX715)

\*第一作者 主管药师,硕士。研究方向:循证药学、药物经济学。  
E-mail:linxiaohui@ndnu.edu.cn

#通信作者 副主任药师。研究方向:临床药学、数据科学。  
E-mail:sunbird01@163.com

of vancomycin trough blood concentrations; however, their performance in the low to moderate concentration categories still requires improvement.

**KEYWORDS** machine learning; vancomycin; blood concentration; MIMIC-IV database; classification prediction

万古霉素作为糖肽类抗菌药物的代表,是治疗耐药革兰氏阳性菌感染的核心药物,但因其存在治疗窗窄及个体间药代动力学差异,临床应用面临挑战<sup>[1]</sup>。传统的治疗药物监测(therapeutic drug monitoring, TDM)依赖经验性剂量调整与滞后的监测反馈,存在局限性<sup>[2]</sup>,临床亟待探索新型预测模型以优化决策流程。机器学习可对药代动力学中的复杂非线性关系进行建模,捕捉药物浓度变化规律<sup>[3]</sup>。目前已有多机器学习算法被应用于TDM预测,包括逻辑回归(logistic regression, LR)、随机森林(random forest, RF)、极端梯度提升(extreme gradient boosting, XGBoost)、支持向量机(support vector machine, SVM)和K近邻(K-nearest neighbors, KNN)等,在预测精度与个体化给药支持方面表现优秀<sup>[4-5]</sup>。随着机器深度学习的发展,表格先验数据拟合网络(tabular prior-data fitted network, TabPFN)作为一种基于生成式Transformer的表格基础模型,在小样本场景下展现出较大的临床应用潜力<sup>[6]</sup>。

重症监护医学信息集市(Medical Information Mart for Intensive Care, MIMIC)是一个大型、开源数据库,MIMIC-IV是其最新版本,覆盖超65 000例重症医学病房(intensive care unit, ICU)患者的人口统计学特征、实验室指标、药物治疗参数等多维信息,为医疗数据分析提供了丰富资源,便于研究者提取关键特征,以探索临床数据关联,构建个性化预测模型,指导精准用药<sup>[7]</sup>。本研究通过机器学习方法,基于MIMIC-IV数据库的数据,探索万古霉素血药谷浓度与患者病理生理状态之间的关系,构建分类预测模型,旨在为个体化用药提供支持,并评估MIMIC-IV数据库在TDM中的应用价值。

## 1 资料与方法

### 1.1 研究对象

本研究为一项回顾性研究,所有数据来源于MIMIC-IV数据库,根据《健康保险携带和责任法案》进行去标识化处理。数据库的访问遵循PhysioNet认证的健康数据许可1.5.0的要求,需完成以下流程:(1)通过PhysioNet官方研究人员身份验证和专业资质审核;(2)完成CITI Program规定的“Data or Specimens Only Research”伦理培训课程(ID: 50610476),通过考核;(3)签署PhysioNet数据使用协议,承诺遵守数据安全规范,包括禁止重新识别患者信息、限制数据传播范围及安全存储要求。由于本研究使用已完全去标识化的回顾性数据,故豁免知情同意及额外的伦理审批。

本研究的纳入标准包括:(1)入住ICU且年龄≥18岁;(2)接受静脉输注万古霉素治疗;(3)进行过万古霉素血药谷浓度监测。本研究的排除标准包括:(1)接受透析治疗或肾脏替代治疗者;(2)体重<40 kg或>150

kg者;(3)数据记录缺失量>70%者。

### 1.2 方法

#### 1.2.1 数据预处理

基于MIMIC-IV数据库资料,采用R 4.3.3软件构建数据分析框架,通过Apache Parquet列式存储格式高效管理临床数据,利用arrow包实现快速数据读取和处理。本研究整合了患者人口统计学特征、实验室检查结果及药物治疗记录等多维信息,所有数据转换和处理均在tidyverse生态系统中完成。研究对象严格按照纳排标准筛选,以减少噪声干扰。数据提取过程通过dbplyr包实现。

本研究提取的核心变量包括:(1)人口统计学特征——年龄、性别、体重;(2)实验室指标——肝功能、肾功能、营养状态、血液学、代谢及凝血功能等;(3)药物治疗参数——万古霉素给药方案、谷浓度监测值(μg/mL)。

#### 1.2.2 异常值处理

为提升数据质量并降低极端值对模型性能的潜在偏倚,本研究采用基于四分位距(interquartile range, IQR)的方法识别并处理关键变量中的异常值<sup>[8]</sup>。异常值定义为小于 $Q_1 - 1.5 \times IQR$ 或大于 $Q_3 + 1.5 \times IQR$ 的数据点,其中 $Q_1$ 和 $Q_3$ 分别为第一和第三四分位数,  $IQR = Q_3 - Q_1$ 。识别出的异常值被统一标记为缺失值。

#### 1.2.3 缺失值处理

计算各个特征的缺失率,删除缺失率>30%的变量,其余变量采用多重插补法进行缺失值插补,设置插补数据集数量为1( $m=1$ ),最大迭代次数为5( $max_{it}=5$ ),插补方法根据变量类型自动选择(method=default method)。

#### 1.2.4 特征工程

本研究采用两阶段特征选择方法,分别通过相关性分析和Boruta特征选择算法(Boruta feature selection algorithm, 简称Boruta算法)对特征进行筛选与优化。首先,采用Pearson相关系数( $r$ )评估各数值型特征间的线性关联强度,使用pheatmap包绘制特征相关性热力图。相关性强度界定:  $|r| < 0.3$  为弱相关,  $0.3 \leq |r| < 0.7$  为中等相关,  $|r| \geq 0.7$  为强相关,并记录 $r$ 的正负号以反映变量间的变化方向(正相关或负相关)。其次,采用Boruta算法进行特征选择,识别对模型预测具有显著贡献的特征。通过比较原始特征与随机生成的影子特征的重要性得分,判断其统计显著性:若某特征的重要性得分持续高于所有影子特征的最大值,则被标记为“confirmed important”;若低于所有影子特征的最小值,则标记为“unimportant”;其余为“tentative important”。最终,按重要性得分均值排序,筛选出对模型预测贡献最大的关键特征,构建精简且高效的建模特征集<sup>[9]</sup>。

### 1.2.5 目标变量离散化

为满足机器学习分类任务需求,将万古霉素血药谷浓度按临床治疗窗标准离散化为三类: <10  $\mu\text{g}/\text{mL}$  为低浓度, 10~20  $\mu\text{g}/\text{mL}$  为中浓度, >20  $\mu\text{g}/\text{mL}$  为高浓度。

### 1.2.6 模型构建与性能评估

采用 Python 技术栈构建实验环境, 具体配置如下: 以 Python 3.12.9 作为基础编程语言, Scikit-learn 1.6.1 提供传统机器学习算法支持, XGBoost 3.0.0 用于构建梯度提升树模型, TabPFN 2.0.8 作为基于 Transformer 的新型表格数据学习框架。数据集按 7:3 的比例随机划分为训练集与测试集。在训练集上进行模型训练、超参数调优。在测试集上进行模型预测, 计算性能指标, 评估模型的泛化能力。

(1) 算法选择与实现: 基于医学预测任务的特性和前期研究基础, 本研究选择以下具有代表性的分类器进行基准测试。①TabPFN: 基于 Transformer 的先验拟合网络, 适用于小样本医学数据; ②LR: 作为基础线性模型参照; ③RF: 通过集成学习降低过拟合风险; ④XGBoost: 优化后的梯度提升决策树; ⑤SVM: 寻找最大化间隔的高效分类器; ⑥KNN: 基于局部相似性的非参数分类。

(2) 数据平衡与训练策略: 采用自适应合成采样 (adaptive synthetic sampling, ADASYN) 对数据类别不平衡问题进行处理, 通过自适应机制过采样, 实现类别数量的平衡, 保留原始数据的临床分布特征<sup>[10]</sup>。所有模型均采用 10 折交叉验证 (10-fold cross-validation, 10-CV), 确保每折数据保持类别比例的一致性。

(3) 模型性能评估: 采用多层次评估框架衡量各模型性能, 包括量化指标对比、可视化分析、分类错误分布评估。①量化指标对比: 选取 6 项核心性能评估指标, 展示各模型在数据集上的表现。准确率 (accuracy): 正确预测样本占总样本的比例, 适用于类别分布均衡的场景; 平衡准确率 (balanced accuracy): 各类准确率的算术平均值, 可缓解类别不平衡对评估结果的影响; 宏平均精确率 (precision macro): 各类精确率的未加权平均, 强调模型的正例识别精度; 宏平均召回率 (recall macro): 各类召回率的未加权平均, 反映模型对各类别正例的覆盖能力; 宏平均 F1 (macro F1): 宏平均精确率与宏平均召回率的调和均数, 衡量模型在不平衡数据中的表现; 多类受试者工作特征曲线 (receiver operating characteristic curve, ROC 曲线) 的曲线下面积 (AUC): 采用“One-vs-Rest” (OvR) 策略计算每类与其他类之间的 AUC (以下简称“OvR-AUC”), 取平均值, 评估模型的整体判别能力。所有指标取值范围为 0~1, 数值越大表示模型性能越好<sup>[6, 11]</sup>。②可视化分析: 构建多类 ROC 曲线。将每一类别视为正类, 其余类别合并为负类, 分别绘制 ROC 曲线并计算 AUC。通过平均各类 AUC 得到整体多类 AUC, 以评估模型在不同分类阈值下的判别能力。③分类错误分布评估: 绘制混淆矩阵热力图展示模型的分类效果, 矩阵中对角线元素 (左上角到右下角) 表示各类别样本的正确预测比例, 非对角线元素反映误分类的比

例, 以此反映模型的分类性能。

(4) 最优模型稳健性评估: 为评估最优模型的稳健性和泛化能力, 采用自举法对训练集数据进行 100 次有放回抽样, 并在每次样本上执行 10-CV, 最终展示模型的 6 项核心性能评估指标的均值及 95% 置信区间 (CI)。

(5) 模型可解释性分析: 采用沙普利加性解释 (SHapley Additive exPlanations, SHAP) 对表现最优的模型进行系统性解释分析。基于训练集计算各特征的 SHAP 值, 绘制所有样本的 SHAP 摘要图, 以可视化不同特征对模型预测结果的影响方向与强度。

## 2 结果

### 2.1 一般资料

从 MIMIC-IV 数据库中筛选符合条件的患者数据, 经数据清洗与预处理后, 计算各特征缺失率发现, 原始数据包括多数关键临床指标, 如年龄、体重、万古霉素剂量及给药频率、血清肌酐、尿素氮和白蛋白等变量的缺失率均低于 5%, 数据完整性较好。部分实验室指标存在较高比例的缺失, 其中总蛋白、C 反应蛋白、N-末端脑钠利肽前体及尿酸的缺失率超过 70%; 氧饱和度、肌酸激酶、皮质醇、铁结合能力、转铁蛋白、功能性纤维蛋白原、尿肌酐、肌钙蛋白、直接胆红素、酮体、糖化血红蛋白等变量的缺失率也在 30% 以上。根据预设标准予以删除后进行缺失值插补, 最终纳入 9 902 例患者, 其中男性 5 990 例 (占 60.5%)、女性 3 912 例 (占 39.5%), 年龄  $(61.92 \pm 15.65)$  岁, 体重  $(82.56 \pm 20.72)$  kg; 万古霉素血药谷浓度分布范围为 1.20~36.40  $\mu\text{g}/\text{mL}$ , 均值为 15.21  $\mu\text{g}/\text{mL}$ ; 按照临床治疗窗标准进行离散化处理, 万古霉素血药谷浓度为低浓度的患者有 2 692 例 (占 27.2%)、中浓度有 4 907 例 (占 49.6%)、高浓度有 2 303 例 (占 23.3%)。数据集按照 7:3 的比例随机划分为训练集和测试集, 其中训练集包含 6 931 条记录 (70.0%), 测试集包含 2 971 条记录 (30.0%)。

### 2.2 特征工程

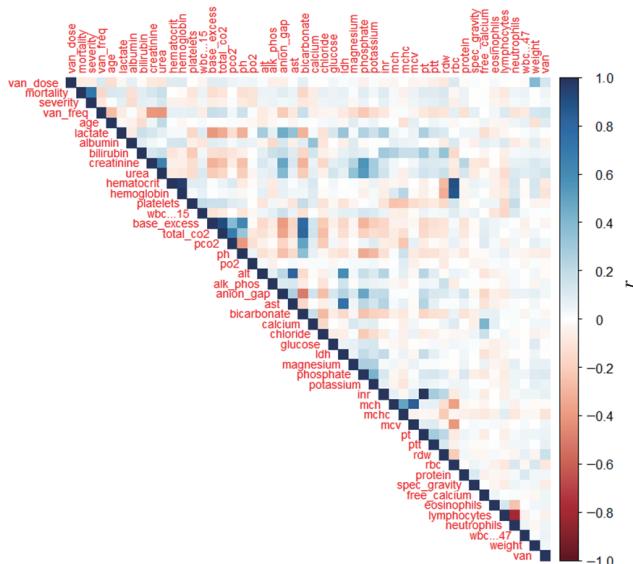
#### 2.2.1 相关性分析

特征相关性热力图分析结果表明: 血液学指标红细胞比容 (hematocrit) 与血红蛋白 (hemoglobin) 呈强正相关; 代谢指标乳酸 (lactate) 与胆红素 (bilirubin) 呈中等正相关; 血液学指标和电解质代谢指标呈现明显的特征分组, 这些簇反映了生理机制的内在联系。结果见图 1。

#### 2.2.2 Boruta 算法特征选择

Boruta 算法特征重要性分析结果如图 2 所示, 万古霉素给药频率 (van\_freq)、肌酐 (creatinine)、尿素氮 (urea)、红细胞分布宽度 (rdw)、年龄 (age) 和万古霉素剂量 (van\_dose) 等特征的重要性得分显著高于影子特征的最大值, 提示这些变量在模型预测中具有核心作用; 白细胞计数 (wbc) 和白蛋白 (protein) 的重要性得分较低, 表明其在模型中的预测能力有限。

基于 Boruta 算法结果, 本研究选择重要性得分较高的 38 个临床特征作为最终的建模特征子集, 涵盖多个具有临床意义的类别。(1) 药物治疗参数: 万古霉素日累计



注:红色表示正相关,蓝色表示负相关。

图1 特征相关性热力图

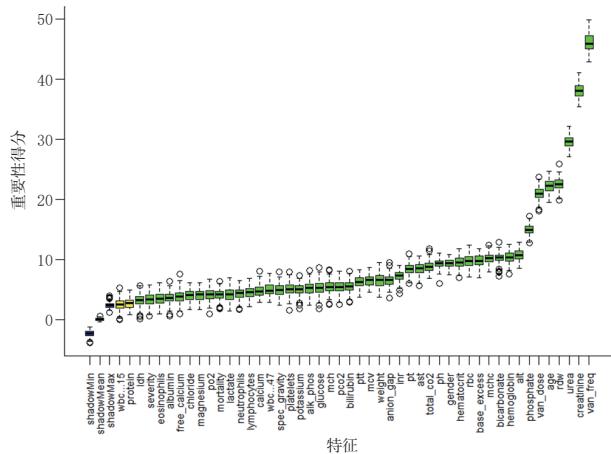


图2 Boruta 算法特征重要性分析结果

量、单次剂量及给药频率;(2)肾功能指标:肌酐、尿素氮;(3)血液学参数:红细胞计数、红细胞分布宽度、红细胞比容、血红蛋白、平均红细胞体积、平均红细胞血红蛋白含量、血小板计数、白细胞计数、中性粒细胞计数、淋巴细胞计数、嗜酸性粒细胞计数;(4)代谢、凝血及酸碱平衡指标:丙氨酸转氨酶、天冬氨酸转氨酶、血糖、乳酸、总胆红素、白蛋白、凝血酶原时间、碳酸氢盐、碱剩余、总二氧化碳、pH值、二氧化碳分压、氧分压、氧饱和度;(5)电解质相关指标:钾离子、钠离子、游离钙离子、镁离子、磷酸盐;(6)人口统计学特征:年龄、体重、性别。

### 2.3 模型性能比较

#### 2.3.1 量化指标对比

如表1所示,RF和TabPFN在各项指标中表现出色,准确率分别为0.7414和0.7377,平衡准确率为0.7403和0.7380,宏平均F1为0.7360和0.7393,OvR-AUC分别为0.9070和0.8958,表明其在分类精度和区分能力上显著优于其他模型。XGBoost模型表现中等,各项指标介于RF和LR之间;而LR、SVM和KNN模型的性能

较差,准确率、平衡准确率和OvR-AUC均接近随机猜测水平,不推荐用于万古霉素谷浓度的预测。

表1 各模型6项核心性能评估结果

模型	准确率/%	平衡准确率/%	宏平均精确率/%	宏平均召回率/%	宏平均F1/%	OvR-AUC
TabPFN	0.7377	0.7380	0.7458	0.7380	0.7393	0.8958
LR	0.5339	0.5309	0.5166	0.5309	0.5162	0.6987
RF	0.7414	0.7403	0.7372	0.7403	0.7360	0.9070
XGBoost	0.7001	0.6994	0.6990	0.6994	0.6985	0.8713
SVM	0.5354	0.5322	0.5171	0.5322	0.5107	0.6990
KNN	0.5493	0.5448	0.5453	0.5448	0.4739	0.7413

#### 2.3.2 可视化分析

ROC曲线分析结果(图3)显示,RF具有最优异的判别能力(AUC=0.91),显著优于其他模型;TabPFN、XGBoost也较好(AUC分别为0.89、0.87),而传统方法LR、SVM、KNN的表现相对较弱(AUC在0.70~0.74之间),表明集成学习方法在非线性分类任务中更具优势。

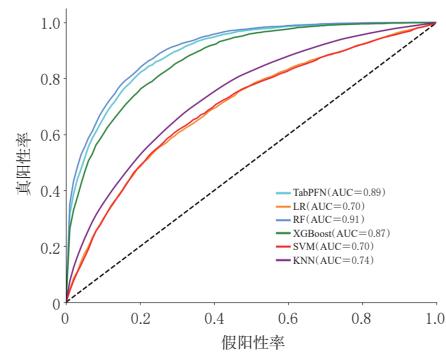


图3 各模型ROC曲线比较

#### 2.3.3 分类错误分布评估

混淆矩阵热力图结果(图4)显示,RF和TabPFN模型在混淆矩阵中预测正确的比例高于其他模型,TabPFN、RF和XGBoost模型在高浓度(类别2)的预测表现相对较好,准确率分别为0.75、0.60和0.59,但在低浓度(类别0)和中浓度(类别1)的表现略显不足。LR、SVM和KNN模型在中浓度的预测中表现较好,准确率为0.66、0.66和0.68。所有模型在某一特定类别上均表现出较高的误分类率,无法实现对各分类的完美识别,说明当前模型在处理该类别时可能存在固有局限性,例如数据分布不均、特征表达不足或类别间边界模糊等。

#### 2.3.4 最优模型稳健性评估

如表2所示,RF模型在自举法结合10-CV下的各项性能指标表现稳定:准确率为0.7414,平衡准确率为0.7403,宏平均F1为0.7360,OvR-AUC为0.9070,表明模型具备良好的分类性能与判别能力。

#### 2.3.5 模型可解释性分析

SHAP摘要图结果(图5)显示,肌酐、万古霉素日累计量、尿素氮、万古霉素给药频率是模型中影响最大的4个特征,符合万古霉素的药代动力学特性。红细胞分布宽度在模型中表现出一定的贡献度,可能与红细胞体积变异影响万古霉素药物分布容积或蛋白结合率相关。其他重要特征包括磷酸盐、丙氨酸转氨酶、年龄、游离钙

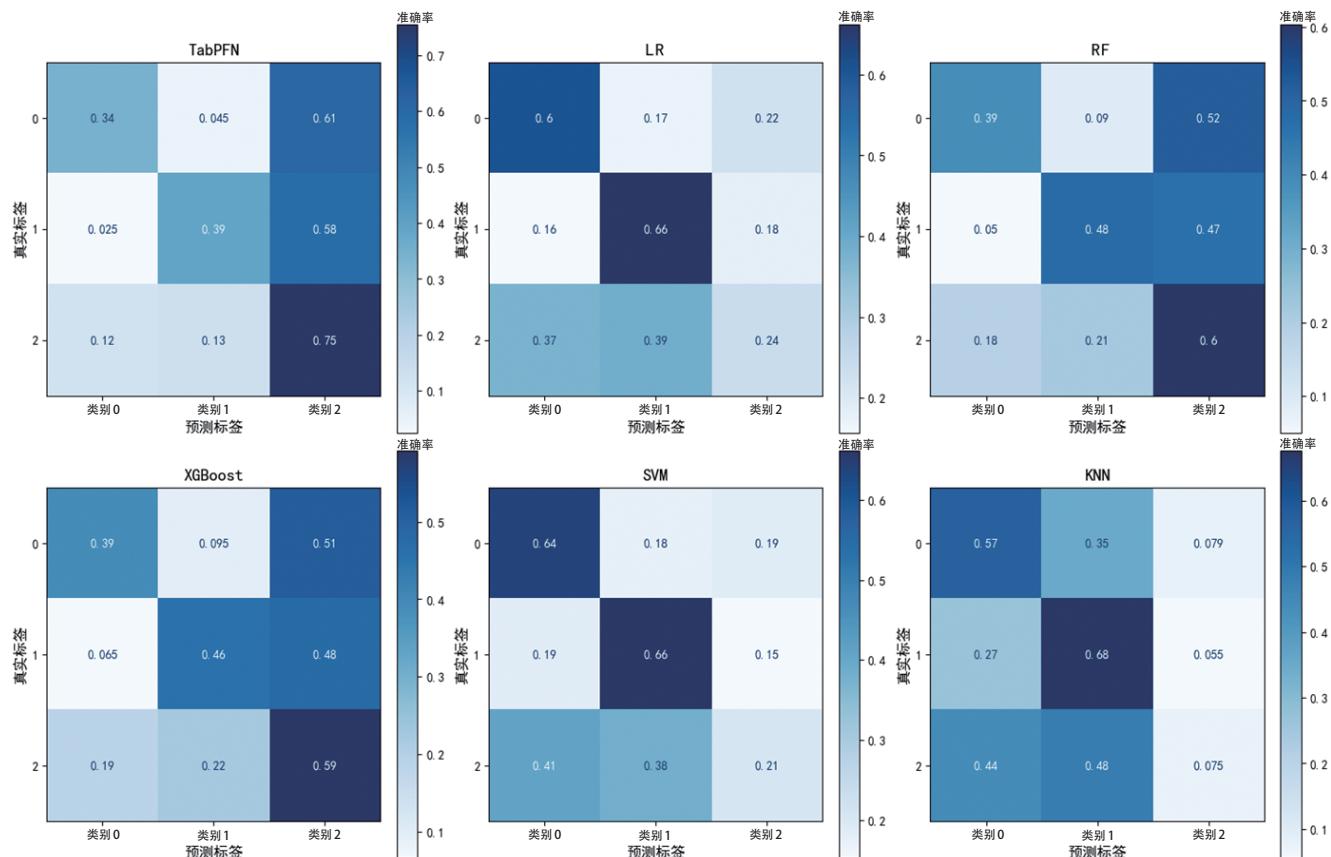


图4 各模型混淆矩阵热力图对比

表2 RF模型自举法结合10-CV的核心性能评估结果

指标	均值	95%CI
准确率	0.7414	0.7380~0.7450
平衡准确率	0.7403	0.7370~0.7440
宏平均精确率	0.7321	0.7280~0.7360
宏平均召回率	0.7360	0.7320~0.7400
宏平均F1	0.7360	0.7320~0.7400
OvR-AUC	0.9070	0.9030~0.9110

和万古霉素剂量,这些特征在一定程度上影响了模型的预测能力。以上结果表明,RF模型能够有效捕捉到影响万古霉素血药浓度的关键临床指标。

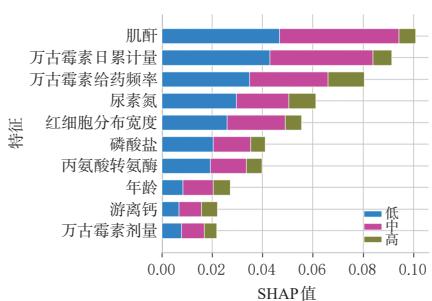


图5 SHAP摘要图

### 3 讨论

#### 3.1 本研究的主要发现

本研究通过评估6种机器学习模型在多分类任务中的表现,发现RF和TabPFN模型在分类精度(准确率>0.73)和区分能力(OvR-AUC>0.89)上优于其他模型,

但在低浓度与中浓度类别上的识别准确率相对较低,提示其分类判别能力仍有限,考虑可能与数据分布不均衡,关键特征在低、中浓度区间区分度不足,血药浓度本身存在生物学连续性导致类别边界模糊有关。已有研究表明,RF模型在临床决策支持任务中展现出一定的优势<sup>[12]</sup>,本研究结果进一步支持其在个体化用药场景中的应用潜力。TabPFN模型同样展现出较优异的性能,且该模型在数据集较小且存在缺失值的情况下也能保持高准确性,可能更适用于资源有限或数据质量欠佳的现实医疗环境<sup>[13]</sup>。这一发现与临床机器学习领域的发展趋势相契合<sup>[12]</sup>:不同算法因其内在假设各异,对数据的适应性也不同,应通过系统性评估筛选最优模型。

#### 3.2 本研究的临床意义

机器学习能有效处理药代动力学中的复杂非线性关系,其预测性能已表现出优于传统统计方法的潜力<sup>[14]</sup>。本研究基于MIMIC-IV数据库,整合患者病理生理状态与万古霉素药代动力学特性,构建并比较多种机器学习模型的分类预测性能。结果显示,万古霉素给药剂量和频率、患者肌酐、尿素氮等肾功能指标变量是影响血药浓度的关键预测因子,其中RF和TabPFN模型优于其他模型。本研究构建的机器学习模型具备一定的临床应用潜力,可为TDM提供辅助决策支持。未来可通过封装为轻量级应用程序编程接口嵌入电子病历系统,实现临床集成;并开展多中心外部验证,评估机器学习模型泛化能力与应用价值,推动精准医学发展。

### 3.3 本研究的局限性

本研究在多分类任务评估中存在一定局限性:(1)研究采用了ADASYN过采样技术,其效果依赖于少数类的原始分布,若少数类样本过少或特征重叠严重,可能生成与多数类重叠的“虚假样本”,导致分类器混淆,降低少数类的分类精度<sup>[15]</sup>。当类别之间存在重叠时,OvR-AUC评估方法的独立性假设可能不成立,从而导致AUC虚高,无法以AUC反映类别特定的误分类代价差异<sup>[16]</sup>。(2)特征工程未能充分表征浓度梯度变化规律,导致低、中浓度区间区分能力受限。针对这一问题,可通过构建时序相关特征(如剂量-效应矩阵),引入药物动力学相关的时序特征来提升模型对动态浓度变化的捕捉能力<sup>[17]</sup>。(3)一些潜在的混杂因素(如遗传多态性、合并用药、合并症或其他疾病表型等)可能对预测结果产生影响<sup>[18]</sup>,但这些因素均未被本研究纳入,可能造成系统偏差。(4)单一模型可能难以全面捕捉数据中的复杂模式,尤其在面对高维、非线性及多模态数据时表现受限,深度学习模型、图神经网络等先进架构有望弥补这一短板<sup>[19]</sup>。(5)模型的验证目前仅基于内部数据和交叉验证,尚未在外部独立数据集中进行测试,其泛化能力仍有待进一步验证。(6)尽管本研究中的RF和TabPFN模型在分类性能上表现出色,但其“黑箱”特性限制了对预测机制的深入理解。

### 4 结语

机器学习模型RF和TabPFN在万古霉素血药谷浓度分类预测中表现出一定优势,但在低、中浓度类别上的表现仍有改进空间。未来研究应优化试验设计与特征工程,拓展多中心数据,验证模型的适用性,提升模型性能和可解释性,推动机器学习在TDM中的实际应用。

### 参考文献

- [1] WILLIAMS P G, TABAH A, COTTA M O, et al. International survey of antibiotic dosing and monitoring in adult intensive care units[J]. Crit Care, 2023, 27(1):241.
- [2] LIM A S, FOO S H W, BENJAMIN SENG J J, et al. Area-under-curve-guided versus trough-guided monitoring of vancomycin and its impact on nephrotoxicity: a systematic review and meta-analysis[J]. Ther Drug Monit, 2023, 45(4):519-532.
- [3] ANGELOPOULOS A N, BATES S, FANNJIANG C, et al. Prediction-powered inference[J]. Science, 2023, 382(6671):669-674.
- [4] HUANG X H, YU Z, BU S H, et al. An ensemble model for prediction of vancomycin trough concentrations in pediatric patients[J]. Drug Des Devel Ther, 2021, 15:1549-1559.
- [5] OTA R, YAMASHITA F. Application of machine learning techniques to the analysis and prediction of drug pharmacokinetics[J]. J Control Release, 2022, 352:961-969.
- [6] HOLLMANN N, MÜLLER S, PURUCKER L, et al. Accurate predictions on small data with a tabular foundation model[J]. Nature, 2025, 637(8045):319-326.
- [7] JOHNSON A E W, BULGARELLI L, SHEN L, et al. MIMIC-IV, a freely accessible electronic health record da-
- taset[J]. Sci Data, 2023, 10(1):1.
- [8] MRAMBA L K, LIU X, LYNCH K F, et al. Detecting potential outliers in longitudinal data with time-dependent covariates[J]. Eur J Clin Nutr, 2024, 78(4):344-350.
- [9] YAN F J, CHEN X H, QUAN X Q, et al. Association between the stress hyperglycemia ratio and 28-day all-cause mortality in critically ill patients with sepsis: a retrospective cohort study and predictive model establishment based on machine learning[J]. Cardiovasc Diabetol, 2024, 23(1):163.
- [10] PRIYADHARSHINI M, BANU A F, SHARMA B, et al. Hybrid multi-label classification model for medical applications based on adaptive synthetic data and ensemble learning[J]. Sensors, 2023, 23(15):6836.
- [11] CHICCO D, JURMAN G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation[J]. BMC Genomics, 2020, 21(1):6.
- [12] SWANSON K, WU E, ZHANG A, et al. From patterns to patients: advances in clinical machine learning for cancer diagnosis, prognosis, and treatment[J]. Cell, 2023, 186(8):1772-1791.
- [13] TRAN V Q, BYEON H. Predicting dementia in Parkinson's disease on a small tabular dataset using hybrid LightGBM-TabPFN and SHAP[J]. Digit Health, 2024, 10:1-15.
- [14] HUI K H M, LUI C Y G, WU K L A, et al. Multi-center prospective population pharmacokinetic study and the performance of web-based individual dose optimization application of intravenous vancomycin for adults in Hong Kong: a study protocol[J]. PLoS One, 2022, 17(5):e0267894.
- [15] PEREIRA T, ABBASI M, OLIVEIRA J L, et al. Optimizing blood-brain barrier permeation through deep reinforcement learning for de novo drug design[J]. Bioinformatics, 2021, 37(Suppl. 1):i84-i92.
- [16] MBIZVO G K, LARNER A J. Receiver operating characteristic plot and area under the curve with binary classifiers: pragmatic analysis of cognitive screening instruments[J]. Neurodegener Dis Manag, 2021, 11(5):353-360.
- [17] HUANG X H, YU Z, WEI X, et al. Prediction of vancomycin dose on high-dimensional data using machine learning techniques[J]. Expert Rev Clin Pharmacol, 2021, 14(6):761-771.
- [18] YIN M H, JIANG Y L, YUAN Y W, et al. Optimizing vancomycin dosing in pediatrics: a machine learning approach to predict trough concentrations in children under four years of age[J]. Int J Clin Pharm, 2024, 46(5):1134-1142.
- [19] ABBAS Q, DAADAA Y, RASHID U, et al. HDR-EfficientNet: a classification of hypertensive and diabetic retinopathy using optimize EfficientNet architecture[J]. Diagnostics, 2023, 13(20):3236.

(收稿日期:2025-04-21 修回日期:2025-08-01)

(编辑:刘明伟)