

丙戊酸血药浓度预测的小样本多分类机器学习模型对比[△]

陈 曜^{1*},袁申奥²,袁海玲^{1#},赵 杰³,陈 鹏²,田春艳¹,苏 怡¹,张云松¹,张 玉¹(1. 西安国际医学中心医院药学部,西安 710100;2. 长安大学信息工程学院,西安 710064;3. 长安大学经济与管理学院,西安 710064)

中图分类号 R969.3;TP181 文献标志码 A 文章编号 1001-0408(2025)11-1399-06

DOI 10.6039/j.issn.1001-0408.2025.11.20



摘要 目的 构建用于预测丙戊酸(VPA)血药浓度的三分类(不足、正常、超限)和二分类(不足、正常)模型,并比较这2种模型的性能,为临床制定用药方案提供参考。方法 收集2022年11月—2024年9月在西安国际医学中心医院接受VPA治疗并进行血药浓度检测的480名患者的临床数据(共695份数据)。分别针对三分类和二分类模型的目标变量构建预测模型,利用XGBoost特征重要性评分进行特征排名和选取,采用12种机器学习算法进行训练和验证,并通过准确率、F1分数及受试者工作特征曲线下面积(AUC)3个指标对模型的性能进行评价。结果 在三分类模型中,合并肾病和合并电解质紊乱的XGBoost特征重要性评分排名较高;然而在二分类模型中,这些特征的重要性排名显著降低,提示其与VPA血药浓度超限之间存在紧密的关联。在三分类模型中,随机森林法表现最佳,但其测试集F1分数仅达到0.704 0,AUC仅为0.519 3;而在二分类模型中,CatBoost方法表现最佳,其测试集F1分数为0.785 7,AUC达到了0.819 5。结论 本研究构建的三分类模型具有预测VPA血药浓度超限的能力,但预测及模型泛化能力较差;构建的二分类模型仅能对血药浓度不足和正常情况进行分类预测,但模型预测性能较强。

关键词 丙戊酸;机器学习;血药浓度预测;小样本数据集;模型对比

Comparison of small-sample multi-class machine learning models for plasma concentration prediction of valproic acid

CHEN Xi¹, YUAN Shen'ao², YUAN Hailing¹, ZHAO Jie³, CHEN Peng², TIAN Chunyan¹, SU Yi¹, ZHANG Yunsong¹, ZHANG Yu¹(1. Dept. of Pharmacy, Xi'an International Medical Center Hospital, Xi'an 710100, China; 2. School of Information Engineering, Chang'an University, Xi'an 710064, China; 3. School of Economics and Management, Chang'an University, Xi'an 710064, China)

ABSTRACT OBJECTIVE To construct three-class (insufficient, normal, excessive) and two-class (insufficient, normal) models for predicting plasma concentration of valproic acid (VPA), and compare the performance of these two models, with the aim of providing a reference for formulating clinical medication strategies. **METHODS** The clinical data of 480 patients who received VPA treatment and underwent blood concentration test at the Xi'an International Medical Center Hospital were collected from November 2022 to September 2024 (a total of 695 sets of data). In this study, predictive models were constructed for target variables of three-class and two-class models. Feature ranking and selection were carried out using XGBoost scores. Twelve different machine learning algorithms were used for training and validation, and the performance of the models was evaluated using three indexes: accuracy, F1 score, and the area under the working characteristic curve of the subject (AUC). **RESULTS** XGBoost feature importance scores revealed that in the three-class model, the importance ranking of kidney disease and electrolyte disorders was higher. However, in the two-class model, the importance ranking of these features significantly decreased, suggesting a close association with the excessive blood concentration of VPA. In the three-class model, Random Forest method performed best, with F1 score of 0.704 0 and AUC of 0.519 3 on the test set; while in the two-class model, CatBoost method performed optimally, with F1 score of 0.785 7 and AUC of 0.819 5 on the test set. **CONCLUSIONS** The constructed three-class model has the ability to predict excessive VPA blood concentration, but its prediction and model generalization abilities are poor; the constructed two-class model can only perform classification prediction for insufficient and normal blood concentration cases, but its model performance is stronger.

△基金项目 陕西省自然科学基础研究计划(No.2022JQ-657);西安国际医学中心医院院级课题青年项目(No.2024QN11)

*第一作者 主管药师,硕士。研究方向:精准药学服务。E-mail: cxj9@foxmail.com

#通信作者 主任药师,硕士。研究方向:精准药学服务与药事管理。E-mail:aliceyuanhailing@163.com

丙戊酸(valproic acid, VPA)是一种广谱抗癫痫药,通过增加脑内 γ -氨基丁酸(γ -aminobutyric acid, GABA)水平等机制发挥抗癫痫和稳定情绪作用,可用于全面性癫痫、部分性癫痫及肌阵挛癫痫等多种类型的癫痫发作及双相情感障碍等精神疾病的治疗^[1-3]。但VPA体内代谢的个体差异较大,血药浓度与剂量呈非线性相关^[4]。此外,VPA的有效血药浓度范围较窄,当VPA血药浓度低于50 $\mu\text{g}/\text{mL}$ 时可能导致临床治疗失败,而超过100 $\mu\text{g}/\text{mL}$ 时患者出现不良反应的风险又会显著升高^[5]。因此,监测VPA的血药浓度对于提升临床用药的安全性和有效性至关重要。

VPA血药浓度通常通过血液生化检测,这需要对患者进行采血,并且其检测过程不仅耗时还需投入较高经济成本。因此,开发VPA血药浓度预测模型成为了近年来的研究焦点^[6]。传统的VPA药代动力学预测模型主要包括一室模型^[6]、二室模型^[7]、群体药代动力学模型^[8]以及基于生理的药代动力学模型^[9]4种。然而,由于药代动力学的高度非线性和多因素影响,使得药代动力学预测模型的构建颇具挑战性。随着机器学习技术的发展,VPA血药浓度预测建模成为可能,且国内外相关研究已经取得了一定成果,如有研究者通过构建模型来预测癫痫患者转氨酶异常升高情况^[10]、老年癫痫患者血药浓度^[11]、儿科抗癫痫血药浓度^[12]以及精神类疾病患者血药浓度^[13]等。鉴于此,本研究回顾性采集了西安国际医学中心医院的真实世界数据,构建了用于VPA血药浓度预测的三分类和二分类模型并比较了其性能,旨在辅助临床制定更精确的用药方案。

1 资料与方法

1.1 研究对象与样本收集

收集2022年11月—2024年9月在西安国际医学中心医院服用VPA并接受血药浓度检测的患者的临床资料。患者纳入标准如下:连续接受VPA治疗4~5个半衰期达到稳态血药浓度后,于下次给药前采血的患者。本研究共纳入554例门诊及住院患者(包括重症监护室患者),共计841份VPA血药浓度数据样本。排除VPA血药浓度异常(低于VPA血药浓度检测下限)的数据样本28份和VPA给药途径不详的数据样本118份,最终获得有效数据样本695份,涉及480例患者。根据文献报道及临床经验,本研究统计了患者临床资料的32个特征,共分为5大类,具体包括:基础信息(性别、年龄、身高、体重),给药相关信息(给药途径、日剂量),实验室生化指标[白细胞计数、血小板计数、总胆红素、间接胆红素、胱抑素C、丙氨酸转氨酶(alanine transaminase, ALT)、天冬氨酸转氨酶(aspartate transaminase, AST)、白蛋白(albumin, ALB)、肌酐(creatinine, Cr)、肌酐清除率(creatinine clearance rate, CCR)],合并症(高同型半胱氨酸

酸血症、糖尿病、高血压、电解质紊乱、肝病、肾病、感染、低蛋白血症)以及合并用药(使用碳青霉烯类抗生素、具有酶诱导作用的抗癫痫药、其他抗癫痫药、质子泵抑制剂、组胺H₂受体拮抗剂、蛋白酶抑制剂、护肝药、抗精神病药)。本研究经西安国际医学中心医院伦理委员会批准(批准号:GJYX-KY-2024-032),在研究过程中采用的数据已完全删除患者隐私信息。

1.2 数据预处理与规范化

(1)分类特征的数值化:将合并症和合并用药情况转换为二进制数值(“1”代表存在/使用,“0”代表不存在/不使用);同样地,将性别和给药途径也转换为二进制数值(“1”代表男性/静脉给药,“0”代表女性/口服给药)。

(2)临床资料特征的缺失值填充:695份数据样本中,仅有285份样本的数据完整,临床资料特征的缺失比例具体为:身高(29.74%),体重(28.59%),CCR(26.58%),Cr(15.8%),间接胆红素(19.54%),胱抑素C(16.09%),ALB(13.22%),ALT(11.21%),AST(11.93%),白细胞计数(11.64%),血小板计数(11.64%),总胆红素(10.78%),日剂量(3.59%),年龄(0.14%)。本研究按照缺失值的常规填充步骤,采用中位数填充法对缺失的临床资料特征进行填充。

(3)数值特征的归一化:对所有数值特征执行归一化处理。

1.3 目标数据集的构建

(1)三分类模型数据集:VPA的有效血药浓度为50~100 $\mu\text{g}/\text{mL}$ 。因此,本研究将VPA血药浓度作为目标变量结果,分为不足组(<50 $\mu\text{g}/\text{mL}$,n=278)、正常组(50~100 $\mu\text{g}/\text{mL}$,n=377)和超限组(>100 $\mu\text{g}/\text{mL}$,n=40),构成三分类模型数据集。分别将不足组、正常组和超限组内的样本取80%,重新组合为训练集(n=556),将各组剩余的20%样本重新组合为测试集(n=139)。为避免因超限组样本数量较少引起的样本不均衡问题,本研究采用SMOTE算法(参数为sampling_strategy='auto')对训练集数据进行了增强。经数据增强处理后,训练集内超限组和不足组的样本数量均扩增到训练集内正常组样本数量,即302例。

(2)二分类模型数据集:从695份VPA血药浓度样本中移除超限组的40份样本数据,将目标变量简化为不足组(<50 $\mu\text{g}/\text{mL}$,n=278)和正常组(50~100 $\mu\text{g}/\text{mL}$,n=377),构成二分类模型的数据集。与三分类模型数据集的划分相似,正常组和不足组内的所有样本数据也按照8:2的比例进行划分,并组合成训练集(n=524)和测试集(n=131)。需要指出的是,二分类模型与三分类模型的数据集构建过程是相互独立的,相关样本数据是随机选取的(随机种子不同)。

1.4 模型特征的选取

考虑到三分类模型和二分类模型所采用的数据集大小并不相同,本研究采用XGBoost算法(XGBoost特征筛选具有较高的稳健性)对特征重要性进行评分,以对这2种分类方法的模型特征变量进行排序和选取。在得到各模型的特征排序后,从排名前10位的特征开始训练模型,观测模型准确率,并依次递增特征数量,直至分类模型的准确率基本收敛到稳定值,将收敛时的特征选取为模型特征。

1.5 模型构建与性能评价

结合上述XGBoost方法选取出的特征,分别构建针对VPA血药浓度预测的三分类以及二分类模型。本研究采用随机森林(Random Forest)、梯度提升决策树(Gradient Boosting Decision Tree, GBDT)、自适应提升算法(Adaptive Boosting, AdaBoost)等12种常见的机器学习算法构建预测模型。为了更加公平地比较各种算法的性能,各算法均使用XGBoost算法选择的相同特征子集进行训练。训练过程的调优过程为:在过采样后的训练数据集合内,对各模型的参数进行超参数组合,并进行超参数空间的有限穷举网格搜索,最后利用5折交叉验证法筛选最优超参数组合。

本研究通过准确率、F1分数以及受试者工作特征曲线下面积(area under the working characteristic curve, AUC)3个指标对模型的性能进行比较,并挑选出表现最佳的模型进行深入分析。其中,准确率反映了模型正确分类的样本占总样本的比例,但当样本分布不均时,其可能会产生误导性结果;F1分数是精确率和召回率的调和平均值,能够更加全面地评价模型的性能^[13];AUC能够综合反映模型的整体分类能力,并具有区分正、负样本的能力以及模型的泛化能力。为了进一步提升模型的稳健性,本研究还采用机器学习领域常用的5倍交叉验证法来验证这12种机器学习方法的性能。此外,本研究对三分类和二分类模型的最优方法分别进行了敏感性分析,即分析模型对输入数据微小变动的稳健性。模型敏感性分析的过程为:通过在测试集上叠加高斯噪声(强度为特征标准差的0.005倍)生成扰动数据;重复100次迭代后分别用扰动数据预测模型,统计每次迭代的预测准确率;最后计算所有准确率的平均值和标准差以量化模型预测的稳定性。

1.6 统计学方法

采用Python进行数据分析。采用K-S检验对数据特征的连续变量进行正态性检验,发现数据均不符合正态分布,故采用 $M(P_{25}, P_{75})$ 表示,不同分类模型中2个数据集之间的差异采用秩和检验进行比较;计数资料用例数或百分比(%)表示,不同分类模型中2个数据集之间的差异采用卡方检验进行比较。

2 结果

2.1 2类模型的数据基线特征比较

2类模型采用的数据集的基线特征见表1。结果显示,在三分类模型中,仅给药途径(口服)在训练集和测试集中差异具有统计学意义($P<0.05$),其余指标差异均无统计学意义($P>0.05$);在二分类模型中,性别(男)、合并使用其他抗癫痫药、合并使用蛋白酶抑制剂在训练集和测试集中差异具有统计学意义($P<0.05$),其余指标差异均无统计学意义($P>0.05$)。

2.2 2类模型的特征评分及特征选取结果

采用XGBoost排序的特征进行模型训练,当特征数量递增到16个时,各分类模型的准确率基本收敛到稳定值,此时各分类模型前16个特征的重要性评分之和已经超过了97%。因此,本研究采用前16个特征进行模型训练,其详细信息见表2。根据表2可知:(1)在三分类和二分类模型的前16个特征中,相同特征有15个。(2)合并肾病和合并电解质紊乱对2个模型的影响差异较大。其中,合并肾病在三分类模型中的XGBoost评分排第1位,但在二分类模型中的评分仅排第15位;合并电解质紊乱在三分类模型中的XGBoost评分排第3位,但未出现在二分类模型中。(3)结合2类模型的数据基线特征比较结果和表2可以看出,尽管性别(男)这一特征在训练集与测试集间存在分布差异($P<0.05$),但其在XGBoost模型特征重要性排名中位列第4位(二分类模型中),表明该特征与VPA血药浓度关联性较强,仍具有潜在预测价值,故保留用于模型训练。

2.3 2类模型的性能分析结果

12种机器学习方法在不同分类模型中的性能分析结果见表3。

2.3.1 三分类模型中的结果

由表3可知,在三分类模型中,以Random Forest方法的表现最佳,其训练集中5倍交叉验证的性能指标为:准确率0.7163、F1分数0.7141;测试集中的性能指标为:准确率0.7050、F1分数为0.7040。值得注意的是,在三分类模型中,各机器学习方法的测试集性能普遍低于训练集,这暗示了数据分布的不一致性以及存在过拟合的可能性。图1进一步展示了Random Forest方法在三分类模型中的学习曲线。从图1可以看出,训练集的准确率始终接近于1,即训练集的拟合程度较高;然而,当训练集样本数<500时,其准确率基本不超过0.5;随着训练集样本数的增加,其交叉验证准确率最终稳定在0.72左右,与训练准确率差异较大。

此外,Random Forest方法在三分类模型中的AUC计算结果显示:(1)不足组 vs. (正常组+超限组), $AUC=0.24$,表明模型对“不足”类别与正常组+超限组

表1 2类模型采用的数据集基线特征

基线资料	三分类模型				二分类模型			
	训练集(n=556)	测试集(n=139)	Z χ^2	P	训练集(n=524)	测试集(n=131)	Z χ^2	P
VPA血药浓度[M(P_{25}, P_{75})]/(μg/mL)	56.84(41.19, 76.00)	59.21(35.43, 74.75)	0.398	0.691	54.61(39.75, 72.73)	59.21(35.43, 74.75)	-0.099	0.922
性别(男)/例(%)	365(65.65)	93(66.91)	0.078	0.779	360(68.70)	78(59.54)	3.969	0.046
年龄[M(P_{25}, P_{75})]/岁	50.00(31.00, 64.00)	51.00(26.00, 64.00)	0.044	0.965	51.00(30.00, 65.00)	49.00(30.50, 61.00)	0.803	0.422
身高[M(P_{25}, P_{75})]/cm	170.00(162.00, 174.00)	170.00(162.00, 174.75)	-0.329	0.742	170.00(162.00, 175.00)	170.00(164.00, 173.00)	-0.213	0.831
体重[M(P_{25}, P_{75})]/kg	65.00(55.00, 74.00)	65.00(56.75, 75.00)	-0.688	0.491	65.00(55.00, 74.62)	65.00(57.00, 75.00)	-1.193	0.232
给药途径(口服)/例(%)	419(75.36)	117(84.17)	4.895	0.027	407(77.67)	102(77.86)	0.002	0.963
日剂量[M(P_{25}, P_{75})]/g	1.00(0.80, 1.20)	1.00(0.80, 1.00)	0.775	0.426	1.00(0.80, 1.20)	1.00(0.80, 1.20)	-1.410	0.148
ALT[M(P_{25}, P_{75})]/(U/L)	17.00(10.70, 26.34)	17.32(11.29, 27.50)	-0.406	0.685	17.00(11.24, 26.93)	18.19(11.19, 27.70)	-0.305	0.761
AST[M(P_{25}, P_{75})]/(U/L)	19.00(15.00, 27.00)	19.00(14.50, 26.00)	0.313	0.754	19.45(15.00, 27.08)	18.10(14.70, 25.93)	0.963	0.336
ALB[M(P_{25}, P_{75})]/(g/L)	37.20(33.19, 41.52)	36.59(33.24, 41.69)	0.186	0.853	37.24(33.29, 41.56)	36.55(32.35, 41.16)	0.716	0.398
Cr[M(P_{25}, P_{75})]/(μmol/L)	55.00(41.70, 68.90)	52.65(43.58, 68.03)	0.015	0.989	54.95(41.48, 70.50)	53.80(43.30, 65.20)	0.455	0.650
CCR[M(P_{25}, P_{75})]/(mL/min)	111.34(96.31, 126.72)	111.95(97.79, 126.06)	-0.134	0.894	110.47(95.52, 126.54)	114.25(103.62, 126.26)	-1.321	0.187
胱抑素C[M(P_{25}, P_{75})]/(μg/mL)	0.93(0.81, 1.12)	0.94(0.80, 1.05)	0.312	0.755	0.94(0.82, 1.12)	0.91(0.80, 1.08)	1.101	0.271
白细胞计数[M(P_{25}, P_{75})]/(×10 ⁹ L ⁻¹)	6.36(5.00, 8.44)	6.33(5.10, 8.49)	-0.075	0.940	6.36(5.04, 8.45)	6.52(5.04, 8.38)	0.309	0.758
血小板计数[M(P_{25}, P_{75})]/(×10 ⁹ L ⁻¹)	183.00(138.00, 233.00)	188.00(146.00, 244.00)	-1.110	0.267	187.00(140.00, 237.00)	191.00(145.00, 238.00)	-0.333	0.739
总胆红素[M(P_{25}, P_{75})]/(μmol/L)	7.72(5.40, 11.09)	7.34(5.27, 9.93)	1.079	0.281	7.67(5.36, 11.24)	7.22(5.21, 10.14)	0.963	0.336
间接胆红素[M(P_{25}, P_{75})]/(μmol/L)	4.50(3.09, 6.92)	4.65(2.86, 6.70)	-0.073	0.942	4.75(3.09, 7.01)	4.20(3.05, 6.38)	1.081	0.280
合并高同型半胱氨酸血症/例(%)	13(2.34)	4(2.88)	0.136	0.713	12(2.29)	3(2.29)	<0.001	1.000
合并高血压/例(%)	154(27.70)	32(23.02)	1.241	0.265	138(26.34)	38(29.01)	0.381	0.537
合并糖尿病/例(%)	67(12.05)	17(12.23)	0.003	0.954	64(12.21)	16(12.21)	<0.001	1.000
合并低蛋白血症/例(%)	72(12.95)	23(16.55)	1.219	0.269	81(15.46)	13(9.92)	2.611	0.106
合并电解质紊乱/例(%)	52(9.35)	12(8.63)	0.069	0.793	55(10.50)	8(6.11)	2.323	0.128
合并肝病/例(%)	36(6.47)	4(2.88)	2.653	0.103	28(5.34)	8(6.11)	0.118	0.732
合并肾病/例(%)	19(3.42)	5(3.60)	0.011	0.917	20(3.82)	3(2.29)	0.721	0.396
合并感染/例(%)	193(34.71)	45(32.37)	3.106	0.078	185(35.31)	45(34.35)	0.042	0.838
合并使用质子泵抑制剂/例(%)	75(13.49)	17(12.23)	0.153	0.695	71(13.55)	17(12.98)	0.030	0.864
合并使用组胺H ₂ 受体拮抗剂/例(%)	32(5.76)	7(5.04)	0.109	0.742	24(4.58)	9(6.87)	1.149	0.284
合并使用具有酶诱导作用的抗癫痫药/例(%)	30(5.40)	9(6.47)	0.244	0.621	28(5.34)	11(8.40)	1.745	0.187
合并使用其他抗癫痫药/例(%)	168(30.22)	46(33.09)	0.432	0.511	151(28.82)	50(38.17)	4.308	0.038
合并使用碳青霉烯类抗生素/例(%)	30(5.40)	6(4.32)	0.264	0.608	28(5.34)	8(6.11)	0.118	0.732
合并使用蛋白酶抑制剂/例(%)	2(0.36)	1(0.72)	0.335	0.563	1(0.19)	2(1.53)	4.102	0.043
合并使用护肝药/例(%)	44(7.91)	12(8.63)	0.078	0.780	41(7.82)	13(9.92)	0.611	0.435
合并使用抗精神病药/例(%)	85(15.29)	30(21.58)	3.191	0.074	88(16.79)	21(16.03)	0.044	0.834

表2 不同分类模型前16个特征的XGBoost重要性评分

排名	三分类模型		二分类模型	
	特征	XGBoost重要性评分/%	特征	XGBoost重要性评分/%
1 合并肾病	21.92	合并使用碳青霉烯类抗生素	21.01	
2 合并使用碳青霉烯类抗生素	12.02	合并使用具有酶诱导作用的抗癫痫药	8.06	
3 合并电解质紊乱	6.45	日剂量	6.76	
4 合并使用护肝药	6.43	性别(男)	5.82	
5 日剂量	5.20	间接胆红素	5.81	
6 血小板计数	4.71	AST	5.70	
7 合并使用具有酶诱导作用的抗癫痫药	4.55	ALT	5.48	
8 性别(男)	4.30	胱抑素C	5.30	
9 合并感染	4.23	给药途径(口服)	5.14	
10 胱抑素C	4.20	血小板计数	5.02	
11 AST	4.17	合并使用护肝药	4.84	
12 ALT	4.16	总胆红素	4.69	
13 总胆红素	4.04	体重	4.44	
14 间接胆红素	3.96	白细胞计数	4.34	
15 白细胞计数	3.51	合并肾病	3.83	
16 体重	3.16	合并感染	3.74	

的区分效果差;(2)正常组 vs. (不足组+超限组), AUC=0.21,表明模型对“正常”类别与不足组+超限组的区分效果差;(3)超限组 vs. (不足组+正常组), AUC=0.89,表明模型对“超限”类别与不足组+正常组的区分效果较好。

将上述3个AUC结果融合得到Random Forest方法在三分类模型中的平均AUC为0.5193,综合学习曲线,提示三分类模型存在过拟合问题,表明超限组与其他组(不足组+正常组)的样本分布存在显著差异,三分类模型的可靠性较低,泛化能力较差。

2.3.2 二分类模型中的结果

在二分类模型中CatBoost方法为最优,其训练集5倍交叉验证的性能指标为:准确率0.6688、F1分数为0.6685;测试集的性能指标为:准确率0.7863、F1分数0.7857、AUC 0.8195,对不足组和正常组的区分效果较好。由于二分类模型数据样本分布较为均衡,其测试集准确率和F1分数均能有效反映模型性能。此外,各机

器学习方法在二分类模型测试集的性能指标超过了训练集的性能,这表明模型在训练过程中已经充分掌握了数据的特征和规律,达到了较好的性能。

表3 不同分类模型中各机器学习方法的性能指标

机器学习方法	三分类模型			二分类模型		
	准确率(训练集/测试集)	F1分数(训练集/测试集)	AUC	准确率(训练集/测试集)	F1分数(训练集/测试集)	AUC
Random Forest	0.7163/0.7050	0.7141/0.7040	0.5193	0.6689/0.7405	0.6689/0.7390	0.8227
GBDT	0.6975/0.6115	0.6950/0.6173	0.5504	0.5993/0.7328	0.5991/0.7325	0.8071
AdaBoost	0.6446/0.6403	0.6408/0.6449	0.6774	0.6258/0.7176	0.6256/0.7187	0.7611
Support Vector Machine	0.4791/0.3885	0.4339/0.4003	0.4462	0.5910/0.6107	0.5824/0.6058	0.7545
Logistic Regression	0.5253/0.4748	0.4945/0.5107	0.4716	0.5960/0.6641	0.5934/0.6659	0.7481
Ridge Classifier	0.5309/0.3885	0.4802/0.3919	0.4277	0.6010/0.6947	0.5983/0.6957	0.7669
K-Nearest Neighbors	0.6523/0.5468	0.6342/0.5713	0.4789	0.6489/0.6107	0.6482/0.6126	0.6790
Decision Tree	0.6534/0.5755	0.6533/0.5806	0.5139	0.6324/0.6412	0.6321/0.6430	0.6414
Extra Trees	0.7340/0.6835	0.7296/0.6780	0.4995	0.6937/0.7405	0.6935/0.7410	0.8119
XGBoost	0.7096/0.6547	0.7069/0.6512	0.5371	0.6440/0.7023	0.6438/0.7019	0.7786
CatBoost	0.7306/0.6906	0.7266/0.6836	0.5243	0.6688/0.7863	0.6685/0.7857	0.8195
LightGBM	0.7008/0.6475	0.6993/0.6450	0.5505	0.6275/0.7252	0.6271/0.7257	0.7848

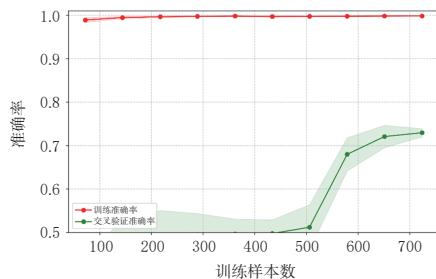


图1 三分类模型中的 Random Forest 方法的学习曲线

2.4 模型敏感性分析结果

图2和图3分别展示了2类模型中最优机器学习方法的敏感性分析的准确率分布直方图,具体结果为:三分类模型中 Random Forest 方法敏感性分析的准确率平均值为 0.6804、标准差为 0.0165,二分类模型中 CatBoost 方法敏感性分析的准确率平均值为 0.7618、标准差为 0.0101。

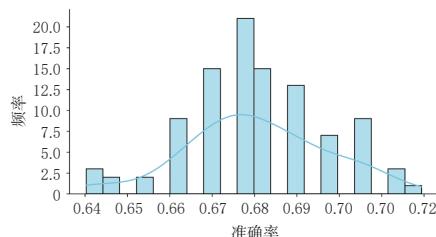


图2 三分类模型 Random Forest 方法敏感性分析的准确率分布

3 讨论

本研究采用综合性医院的真实世界数据,运用12种机器学习方法构建了用于预测VPA血药浓度的三分类和二分类模型。在三分类模型中,Random Forest方法表

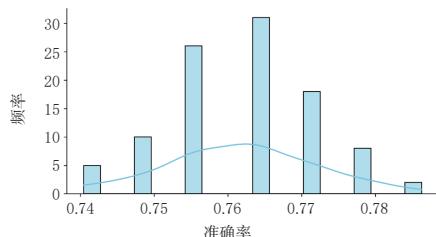


图3 二分类模型 CatBoost 方法敏感性分析的准确率分布

现最佳,其测试集 F1 分数为 0.7040;而在二分类模型中,CatBoost 方法最优,其测试集 F1 分数为 0.7857。通过 XGBoost 重要性评分发现,三分类和二分类模型的 16 个特征中有 15 个相同,仅合并电解质紊乱和给药途径(口服)2 个特征在 2 个模型中有所不同。这些特征组成及导致其在不同模型中排名高低的潜在原因包括:(1)患者的年龄、性别、体重及日剂量是影响 VPA 清除率的关键协变量,进而影响了 VPA 的血药浓度^[7]。(2)合并使用碳青霉烯类抗生素(如美罗培南、亚胺培南)后 VPA 血药浓度显著降低,会增加癫痫发作风险^[5,14];再者,合并使用具有酶诱导作用的抗癫痫药(如卡马西平、苯巴比妥、苯妥英钠)会降低 VPA 血药浓度,可能诱发惊厥或癫痫发作^[15]。(3)肝、肾功能指标如 AST、ALT、胱抑素 C 等异常会影响 VPA 的代谢与排泄,导致 VPA 血药浓度升高^[16-18]。(4)VPA 血药浓度与血液毒性相关,VPA 血药浓度升高可能会导致血小板和白细胞计数异常^[19-20]。(5)VPA 的超限可能会引起代谢性酸中毒,继而造成电解质紊乱^[21]。

相较于其他研究,本研究展现了 3 个显著特点:(1)样本源自综合性医院,覆盖多科室、多病种、多患者来源,还包括 VPA 血药浓度超限的样本,数据集多样性强,在三分类模型中测试集的 F1 分数达 0.7040。(2)本研究是在剔除超限组样本数据后完成的二分类建模,即便样本量相对较小,但二分类模型中测试集的 F1 分数仍达到 0.7857,接近大样本模型性能^[13]。(3)本研究通过比较三分类和二分类模型的 XGBoost 特征排名,发现合并肾病和合并电解质紊乱与 VPA 血药浓度超限关联性强;此外,本研究通过对比三分类和二分类模型的 AUC 值,发现 2 类模型性能存在较大差异的主要原因为超限组样本量少(仅占样本总数的 5.76%),超限组与其他 2 组数据分布差异较大。

尽管本研究中的模型展现出了较好的分类效果,但仍存在若干局限性:(1)本研究采用的数据样本量较小,原因在于所选医院运行时间不足 5 年,且纳入的数据不足 2 年,导致实际有效数据积累时间较短,这对模型性能的提升产生了不利影响。(2)由于数据集源自综合性医院的实际诊疗流程,VPA 血药浓度超限的样本在总体数

据集中占比较低,尽管本研究采取了技术手段对样本不均衡问题进行了缓解,但这种样本的不均衡性仍对三分类模型的训练造成了负面影响。(3)本研究所采用的训练和测试数据均仅来源于单一中心,这可能导致模型在外部的通用性不高。

4 结语

本研究利用西安国际医学中心医院的真实世界数据构建了VPA血药浓度的三分类和二分类模型。通过对2种分类模型的XGBoost特征排名,发现合并肾病和合并电解质紊乱这2个特征与VPA血药浓度的超限具有较强的相关性。三分类模型具有VPA血药浓度超限的预测能力,但因其复杂的模型构建和样本数据分布情况,表现最优的Random Forest方法的F1分数仅达到0.7040,AUC仅为0.5193,预测能力和模型泛化能力较差;二分类模型剔除了VPA血药浓度超限部分的数据,侧重于VPA血药浓度是否达到治疗要求,虽仅能对血药浓度不足和正常情况进行分类预测,但具有更高的准确性,其F1分数达到了0.7857,AUC达到了0.8195,预测及模型泛化能力较强。

参考文献

- [1] MARSON A, BURNSIDE G, APPLETON R, et al. The SANAD II study of the effectiveness and cost-effectiveness of valproate versus levetiracetam for newly diagnosed generalised and unclassifiable epilepsy: an open-label, non-inferiority, multicentre, phase 4, randomised controlled trial[J]. Lancet, 2021, 397(10282): 1375-1386.
- [2] MARSON A G, BURNSIDE G, APPLETON R, et al. Lamotrigine versus levetiracetam or zonisamide for focal epilepsy and valproate versus levetiracetam for generalised and unclassified epilepsy: two SANAD II non-inferiority RCTs[J]. Health Technol Assess, 2021, 25(75): 1-134.
- [3] 陈策,郑丽丹,谢作良,等.双相情感障碍住院患者的临床特征及用药情况的调查研究[J].中国全科医学,2020,23(2):245-250.
- [4] HIEMKE C, BERGEMANN N, CLEMENT H W, et al. 神经精神药理学治疗药物监测共识指南:2017版[J].实用药物与临床,2022,25(2):97-118.
- [5] LI Z, GAO W, LIU G, et al. Interaction between valproic acid and carbapenems: decreased plasma concentration of valproic acid and liver injury[J]. Ann Palliat Med, 2021, 10(5): 5417-5424.
- [6] 钱金,陈顺,谢新芳,等.建立丙戊酸钠在中国癫痫患者中的群体药代动力学模型[J].第二军医大学学报,2017,38(11):1449-1453.
- [7] METHANEETHORN J. A systematic review of population pharmacokinetics of valproic acid[J]. Br J Clin Pharmacol, 2018, 84(5): 816-834.
- [8] 张仲斌,吴晔,季双敏,等.不同年龄组中国癫痫患儿丙戊酸群体药代动力学模型的建立[J].中华实用儿科临床杂志,2014,29(9):698-703.
- [9] OGUNGBENRO K, AARONS L, CRESIM & EPI-CRESIM Project Groups. A physiologically based pharmacokinetic model for valproic acid in adults and children[J]. Eur J Pharm Sci, 2014, 63: 45-52.
- [10] MA H Y, HUANG S H, LI F X, et al. Development and validation of an automatic machine learning model to predict abnormal increase of transaminase in valproic acid-treated epilepsy[J]. Arch Toxicol, 2024, 98(9): 3049-3061.
- [11] MA P, SHANG S L, HUANG Y F, et al. Joint use of population pharmacokinetics and machine learning for prediction of valproic acid plasma concentration in elderly epileptic patients[J]. Eur J Pharm Sci, 2024, 201: 106876.
- [12] DAMNjanović I, Tsyplakova N, Stefanović N, et al. Joint use of population pharmacokinetics and machine learning for optimizing antiepileptic treatment in pediatric population[J]. Ther Adv Drug Saf, 2023, 14: 20420986231181337.
- [13] HSU C W, LAI E C, CHEN Y B, et al. Valproic acid monitoring: serum prediction using a machine learning framework from multicenter real-world data[J]. J Affect Disord, 2024, 347: 85-91.
- [14] 杜欣,李欣焱,谢晓慧,等.碳青霉烯类药物对丙戊酸血药浓度影响及两药联用后癫痫发作影响因素:文献Meta分析[J].中国医院药学杂志,2021,41(4):395-399.
- [15] 林小燕,林燕玲,庄俊鹏.癫痫患者丙戊酸钠血药浓度的影响因素[J].临床合理用药,2024,17(36):59-62.
- [16] 张文滨,林燕,蔡优生,等.丙戊酸钠血药浓度影响因素及不良反应分析[J].临床合理用药,2024,17(10):33-36.
- [17] 董露露,邹素兰,凌静.肝功能异常患者丙戊酸钠血药浓度的影响因素分析[J].淮海医药,2022,40(4):346-349.
- [18] 王利媛,段自峰,李晶,等.基于游离丙戊酸血药浓度开展丙戊酸钠剂量调整的药学实践1例[J].中国医院药学杂志,2023,43(1):117-120.
- [19] ACHARYA S, BUSSEL J B. Hematologic toxicity of sodium valproate[J]. J Pediatr Hematol Oncol, 2000, 22(1): 62-65.
- [20] NASREDDINE W, ATWEH S F, BEYDOUN A A, et al. Predicting the occurrence of thrombocytopenia from free valproate levels: a prospective study[J]. Seizure, 2022, 94: 33-38.
- [21] 王之舟,程红勤,白向荣,等.丙戊酸钠致高乳酸血症伴代谢性酸中毒[J].药物不良反应杂志,2020,22(6): 379-380.

(收稿日期:2025-01-22 修回日期:2025-05-06)

(编辑:林 静)