

- with alteplase for patients with acute ischemic stroke: a systematic review and meta-analysis [J]. *Pharmazie*, 2021, 76 (2): 109–113.
- [12] 曹广辉, 邱胜利, 刘涛, 等. 依达拉奉右莰醇对颅脑损伤患者脑代谢及预后的影响 [J]. *临床合理用药杂志*, 2021, 14 (33): 62–64.
- [13] 杨晓洁. 依达拉奉右莰醇联合阿替普酶静脉溶栓对急性脑梗死患者出血性转化及神经功能的影响 [J]. *现代诊断与治疗*, 2022, 33 (8): 1153–1156.
- [14] 邹娜, 李聪灵, 常莉莎. 依达拉奉右莰醇对急性脑梗死的远期疗效及安全性研究进展 [J]. *慢性病学杂志*, 2022, 23 (1): 20–24.
- [15] QUINN TJ, RICHARD E, TEUSCHL Y, *et al.* European stroke organisation and European Academy of Neurology joint guidelines on post-stroke cognitive impairment [J]. *Eur J Neurol*, 2021, 28 (12): 3883–3920.

[文章编号] 1007-7669 (2024) 01-0044-07

[DOI号] 10.14109/j.cnki.xyylc.2024.01.09

## 真实世界拉帕替尼治疗乳腺癌预后预测研究

于泽<sup>1a</sup>, 叶璇<sup>2</sup>, 吕春明<sup>1b</sup>, 张津源<sup>3</sup>, 郝昕<sup>4</sup>, 王瑞文<sup>4</sup>, 翟青<sup>2</sup>, 高飞<sup>3</sup>

(1. 上海中医药大学 a. 交叉科学研究院, b. 科学技术实验中心, 上海 201203; 2. 复旦大学附属肿瘤医院, 上海 200032; 3. 北京诺道认知医学科技有限公司, 北京 100071; 4. 大连诺道认知医学技术有限公司, 辽宁 大连 116000)

[关键词] 拉帕替尼; 乳腺肿瘤; 机器学习; 真实世界研究

[摘要] 目的 基于真实世界数据和机器学习技术, 构建拉帕替尼治疗乳腺癌患者无进展生存期 (PFS) 的预测模型。方法 回顾性收集复旦大学附属肿瘤医院 2016 年 7 月至 2017 年 6 月收治的 150 例病例。预测模型的结局指标为患者 PFS 是否  $\leq 1$  年。使用序列前向选择算法进行特征选择, 并比较极限梯度提升 (XGBoost)、分类提升 (CatBoost)、随机森林 (RF)、光梯度提升机 (LightGBM)、梯度提升决策树 (GBDT)、逻辑回归 (LR)、支持向量机 (SVR)、人工神经网络 (ANN) 和 TabNet 算法的预测性能。结果 挖掘得到的重要变量包括给药方案、年龄、化疗次数、蒽环类药物、铂类药物、雌激素受体阳性、肿瘤分期、转移部位数量。XGBoost 模型预测性能最佳, 对 PFS  $\leq 1$  年的预测准确率为 93%, 召回率为 87%; 对 PFS  $> 1$  年的预测准确率为 71%, 召回率为 83%。结论 本研究构建的拉帕替尼治疗乳腺癌患者的预后预测模型性能和稳健性良好, 可为乳腺癌临床治疗提供更好的辅助决策依据。

[中图分类号] R737.9

[文献标志码] A

## Prediction of prognosis in breast cancer patients using lapatinib based on real-world evidence

YU Ze<sup>1a</sup>, YE Xuan<sup>2</sup>, LÜ Chun-ming<sup>1b</sup>, ZHANG Jin-yuan<sup>3</sup>, HAO Xin<sup>4</sup>, WANG Rui-wen<sup>4</sup>, ZHAI Qing<sup>2</sup>, GAO Fei<sup>3</sup>

(1. a. Institute of Interdisciplinary Integrative Medicine Research, b. Experiment Center for Science and Technology, Shanghai University of Traditional Chinese Medicine, SHANGHAI 201203, China; 2. Fudan University Shanghai Cancer Center,

[收稿日期] 2023-08-31 [接受日期] 2023-10-27

[基金项目] 国家重点研发计划项目 (2020YFC2005503)

[作者简介] 于泽, 男, 副研究员, 博士在读, 主要从事临床个体化用药与真实世界研究, E-mail: 15910865863@163.com。高飞, 男, 教授, 博士生导师, 主要从事人工智能医疗大数据研究, E-mail: gaofei9000@163.com

[责任作者] 高飞

SHANGHAI 200032, China; 3. Beijing Medicinovo Technology Co. Ltd., BEIJING 100071, China; 4. Dalian Medicinovo Technology Co. Ltd., Dalian LIAONING 116000, China)

[ **KEY WORDS** ] lapatinib; breast neoplasms; machine learning; real world study

[ **ABSTRACT** ] **AIM** Based on real world data and machine learning technology, a predictive model of progression free survival (PFS) of patients with breast cancer treated with lapatinib was constructed. **METHODS** A retrospective collection of 150 patients admitted to the Fudan University Shanghai Cancer Center from July 2016 to June 2017 was conducted. The outcome indicator of the prediction model was whether the patient's PFS was  $\leq 1$  year. Using sequential forward selection algorithms for feature selection, and comparing the predictive performance of 9 algorithms for building models, including extreme gradient boost (XGBoost), classification boost (CatBoost), random forest (RF), light gradient boost (LightGBM), gradient boost decision tree (GBDT), logistic regression (LR), support vector regression (SVR), artificial neural network (ANN), and TabNet. **RESULTS** Important variables included medication regimen, age, frequency of chemotherapy, anthracycline drugs, platinum drugs, estrogen receptor, disease stage, and number of metastatic sites. The XGBoost model had the best prediction performance, with a prediction accuracy of 93% and a recall rate of 87% for PFS  $\leq 1$  year. And a prediction accuracy was 71%, and a recall rate was 83% for PFS  $> 1$  year. **CONCLUSION** The performance and robustness of the prognosis prediction model for patients with breast cancer treated with lapatinib established are good, which can provide a better auxiliary decision-making basis for clinical treatment of breast cancer.

人体表皮生长因子受体 2 (human epidermal growth factor receptor 2, HER2) 基因是一种与预后不良相关的标志物, 其在大约 15% 的肿瘤患者中表达水平明显升高<sup>[1]</sup>。拉帕替尼 (lapatinib) 是一种酪氨酸激酶抑制剂, 可阻断 HER2 酪氨酸激酶活性, 从而抑制乳腺肿瘤细胞生长<sup>[2]</sup>。在我国, 拉帕替尼与卡培他滨联用治疗方案已被批准用于 HER2 阳性晚期/转移性乳腺癌患者<sup>[3,4]</sup>。真实世界数据 (real-world data, RWD) 是目前医学研究的重要数据来源<sup>[5]</sup>。与随机对照试验相比, 基于 RWD 的研究更接近临床实际情况, 可为个体化治疗提供更多的证据<sup>[6-8]</sup>。但真实世界研究所使用的数据多数为复杂高维数据, 因此通常采用一些新技术如机器学习、深度学习技术进行处理。目前, 对于拉帕替尼治疗乳腺癌患者预后预测的研究较少, 本研究基于 RWD 和机器学习技术, 构建拉帕替尼治疗乳腺癌患者预后的预测模型, 以期对乳腺癌临床治疗提供更好的辅助决策依据。

### 材料与方法

**研究人群** 回顾性收集复旦大学附属肿瘤医院 2016 年 7 月至 2017 年 6 月期间收治的 150 例被诊断为乳腺癌且接受拉帕替尼治疗的患者资料。

**纳入标准** 纳入标准: (1) 被诊断为乳腺癌; (2) 接受拉帕替尼治疗; (3) 初始治疗剂量为 1 000 mg 和 1 250 mg。排除标准: (1) 乳腺癌 II 期; (2) 资料缺失。

筛选流程见图 1。

**数据收集和处理** 数据来源于电子医疗记录, 包括人口统计信息 (如年龄、身高和体重)、药物组合 (如蒽环类药物、紫杉醇、铂类药物、氟尿嘧啶和曲妥珠单抗的使用情况)、生理和病理状况 (如高血压、糖尿病、心脏病、其他潜在疾病和绝经后情况), 以及治疗方案 [如既往化疗次数、既往内分泌治疗、Ki-67 指数、雌激素受体 (ER) 和孕激素受体 (PR) 的检测情况、疾病分期、手术情况、东部肿瘤协作组 (ECOG) 评分、转移部位数量、肺转移、肝转移、骨转移和脑转移发生率以及不同治疗方案]。其他潜在疾病包括癫痫、肝炎、甲状腺功能亢进、慢性肠炎、乙型肝炎和桥本甲状腺炎等情况。治疗方案包括方案 1 (拉帕替尼 + 卡培他滨)、方案 2 (紫杉醇 + 卡铂 + 曲妥珠单抗 + 拉帕替尼)、方案 3 (长春瑞滨 + 拉帕替尼) 和方案 4 (其他联合方案)。拉帕替尼的初始剂量方案有两种, 即 1 000 mg 和 1 250 mg, 将其转化为二元变量, 其中 1 250 mg 对应为 “1”, 1 000 mg 对应为 “0”。根据中位年龄对年龄数据集进行分层分析及独热编码。由于数据集中存在某些变量样本量极不平衡的情况, 因此删除了阳性或阴性样本数与样本总数比值小于 0.1 的变量。对于存在缺失值的数据, 利用随机森林 (RF) 算法学习类似患者的信息, 对变量进行插值处理。结局指标是接受拉帕替尼治疗的乳腺癌患者的无进展生存期 (PFS), PFS  $\leq 1$  年对应为 “0”,

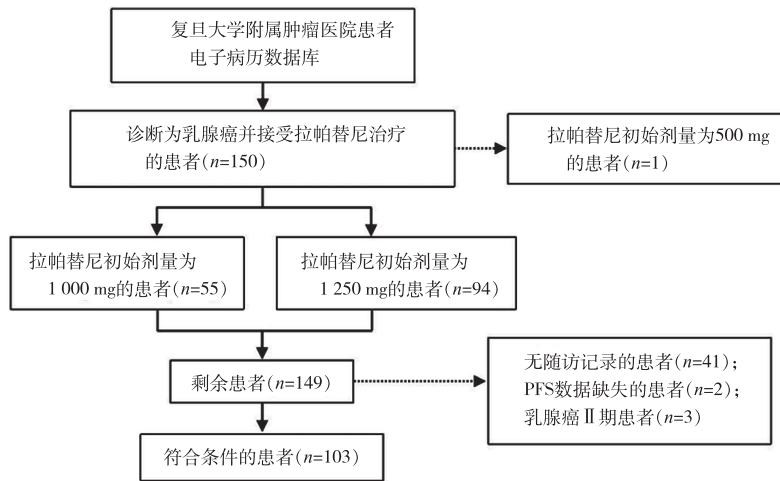


图 1 研究对象筛选流程图 PFS: 无进展生存期

PFS>1 年对应为“1”。

**特征选择和模型构建** 特征选择和模型构建的过程见图 2。对数据进行处理后, 按照 8:2 的比例划分训练集和测试集。采用 SMOTETomek 综合采样方法来处理训练队列中的数据<sup>[9]</sup>。综合采样的核心是经过采样扩大样本后, 再对处在胶着状态的样本用 Tomek Link 法进行删除。采用基于分类提升 (CatBoost) 的序列前向选择 (SFS) 算法进行特征选择<sup>[10]</sup>。选择过程是利用 SFS 算法每次向特征子集中添加一个特征, 迭代生成一个新模型, 并计算模型性能。

f1 值是精确率和召回率的综合评价指标, 用来衡量模型性能。精确率是指在所有被预测为正的样本中实际为正的样本的概率, 召回率是指在实际为正的样本中被预测为正的样本的概率。f1 值越高, 表示模型具有更强的稳健性 (robustness)。当使用某一特征子集构建的模型的 f1 值达到最优值时, 迭代停止。特征子集包含了筛选出来的重要变量, 要求规模最小, 性能最佳。在测试集中, 分别采用 9 种算法建立 PFS 预测模型, 并对它们的预测性能进行比较。算法包括极限梯度提升 (XGBoost)、CatBoost、RF、光梯度提升机 (LightGBM)、梯度提升决策树 (GBDT)、逻辑回归 (LR)、支持向量机 (SVR)、人工神经网络 (ANN) 和 TabNet。

本研究选择具有预测性能最佳的模型来预测患者的 PFS。使用具有最佳预测性能的算法计算选定变量的重要性得分, 即每个变量在提高整个模型预测能力方面的贡献程度。应用 Shapley 加性解释 (SHAP) 分析重要变量对模型输出的影响<sup>[11]</sup>。SHAP 值可以帮助解释变量如何影响模型对 PFS 的预测。最后, 使用混淆矩阵计算模型准确率和召回率, 并将预测模型在测试队列中的表现可视化。

**统计学分析** 为了展示预测性能, 本研究计算了评估指标, 包括准确率、召回率、f1 值、精确率、曲线下面积 (AUC)、灵敏度、特异度。精确率 = 真阳性例数 / (真阳性例数 + 假阳性例数) × 100%, 召回率 = 真阳性例数 / (真阳性例数 + 假阴性例数) × 100%, f1 值 = 2 × 真阳性例数 / (2 × 真阳性例数 + 假阳性例数 + 假阴性例数) × 100%, 准确率 = (真阳性例数 + 真阴性例数) / (真阳性例数 + 假阳性例数 + 真阴性例数 + 假阴性例数) × 100%, 灵敏度 = 真阳性例数 / (真阳性例数 + 假阴性例数) × 100%, 特异度 = 真阴

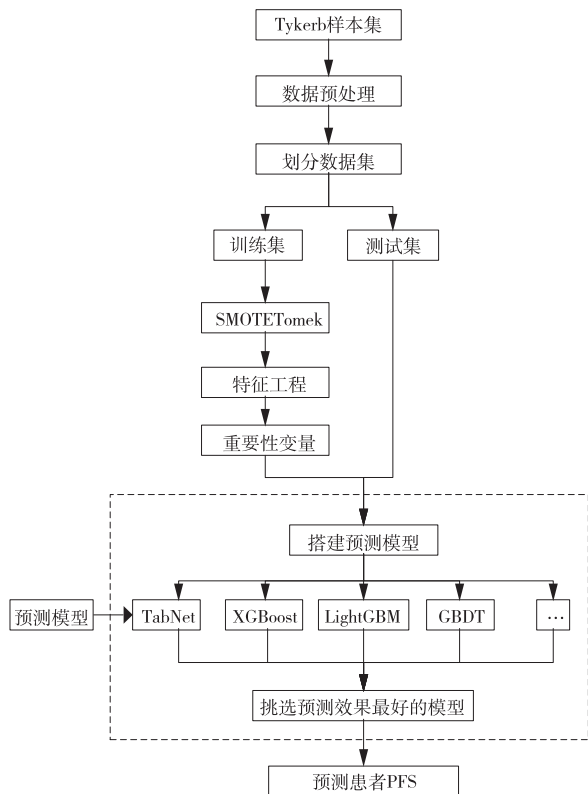


图 2 特征选择与模型构建的流程图 PFS: 无进展生存期

性例数 / (真阴性例数 + 假阳性例数) × 100%。

混淆矩阵由 Matplotlib 包实现。所有机器学习和深度学习算法的操作都在 Windows 10 上运行, CPU @ 2.90GHz 12cpu, 内存 512GB。使用 python 3.8.8 和 IBM SPSS Statistics 22 进行数据分析。

## 结 果

**患者资料** 根据纳入、排除标准筛选后, 共纳入 103 例经拉帕替尼治疗的乳腺癌患者。本研究目标变量为 PFS, 其中 82 例 (79.6%) 患者 PFS ≤ 1 年, 21 例 (20.4%) 患者 PFS > 1 年。31 例 (30.1%) 患者使用的拉帕替尼剂量为 1 000 mg, 72 例 (69.9%) 患者使用的拉帕替尼剂量为 1 250 mg。中位年龄为 52 岁 (43~57 岁), 52 岁以上患者有 57 例 (55.3%), 中位身高为 160.0 cm (158.0~162.0 cm), 体重为 57.0 kg (51.3~63.8)。中位 Ki-67 指数为 40.0 (20.0~60.0) (缺失率 7.77%)。46 例 (44.7%) 患者既往化疗次数 ≥ 3 次。42 例 (40.8%) 患者有既往内分泌治疗史。ER 阳性的患者 47 例 (45.6%), PR 阳性的患者 33 例 (32.0%)。

在药物治疗方面, 蒽环类药物、紫杉烷类药物、铂类药物、氟尿嘧啶和曲妥珠单抗的使用率分别为 75.7%、93.2%、42.7%、57.3% 和 92.2%。患者的生理、病理状态包括高血压、糖尿病、心脏病、其他基础疾病和绝经期, 分别占 11.7%、4.9%、3.9%、9.7% 和 67.0%。乳腺癌 IV 期患者有 88 例 (85.4%), III 期患者 15 例 (14.6%)。95 例 (92.2%) 患者接受了根治性乳房切除术 / 保乳手术, 未经手术治疗的患者 7 例 (6.8%), 接受姑息性手术的患者 1 例 (1.0%)。99 例 (96.1%) 患者 ECOG 评分为 1 分, 4 例 (3.9%) 患者评分为 2 分。未发生转移的患者 14 例 (13.6%), 有 1 个转移部位患者 47 例 (45.6%), 2 个转移部位的患者 24 例 (23.3%), 3 个转移部位的患者 12 例 (11.7%), 4 个转移部位的患者 6 例 (5.8%)。发生肺转移、肝转移、骨转移、脑转移和其他转移的患者分别有 53 例 (51.5%)、32 例 (31.1%)、35 例 (34.0%)、21 例 (20.4%) 和 24 例 (23.3%)。采用方案 1 的患者最多, 为 71 例 (68.9%), 采用方案 2 的患者 15 例 (14.6%), 采用方案 3 的患者 6 例 (5.8%), 采用方案 4 的患者 11 例 (10.7%)。

对训练集 (103 × 0.8 = 82 例) 中的阳性样本 (15 例) 采用 SMOTETomek 算法进行综合采样, 最终训练集得到 134 例样本, 阳性样本与阴性样本比例为 1:1。

**变量选择** 根据 SFS 方法, 从 22 个变量中选取了 8 个特征。通过使用由 1 到 22 个变量建立的 CatBoost 模型,

本研究得到了每个模型的 f1 值, 见图 3。随着纳入变量数量的增加, f1 值不断上升, 当模型纳入 8 个变量时达到最大值 (f1 值 = 0.86)。为了追求简洁准确的模型, 需要纳入最少变量, 使模型性能达到较好的效果。选择前 8 个变量建立预测模型, 包括给药方案、年龄、化疗次数、蒽环类药物、铂类药物、ER 阳性、疾病分期、转移部位数量。

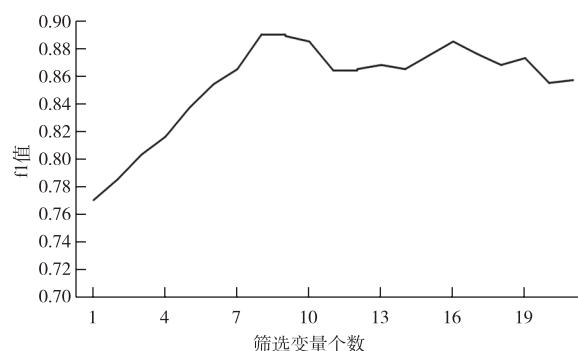


图 3 筛选变量个数对应的模型 f1 值图

**模型构建** 9 个模型的预测性能见表 1。其中 XGBoost 模型性能最好, 预测 PFS ≤ 1 年的准确率为 0.93, 召回率为 0.87, f1 值 = 0.90; 预测 PFS > 1 年的准确率为 0.71, 召回率为 0.83, f1 值 = 0.77。整个 XGBoost 模型的准确率为 0.86, AUC 为 0.83, 灵敏度为 0.83, 特异度为 0.87。利用 XGBoost 模型计算 8 个变量的重要性得分由高到低依次为: 转移部位数量 (0.310)、化疗次数 (0.142)、ER 阳性 (0.115)、年龄 (0.097)、铂类药物 (0.094)、疾病分期 (0.093)、剂量方案 (0.088) 和蒽环类药物 (0.060)。

**模型解释** SHAP 分析表明了 8 个变量分别对模型输出的影响, 见图 4。ER 阳性和剂量方案对患者 PFS 均有正向影响。化疗次数、转移部位数量、年龄、疾病分期和是否同服蒽环类药物对患者 PFS 有负向影响。是否同服铂类药物对患者 PFS 的影响方向尚不明确。

测试集共 21 例患者, 其中 15 例患者 PFS ≤ 1 年, 6 例患者 PFS > 1 年。构建基于 XGBoost 的混淆矩阵预测患者 PFS, 见图 5。该模型对 PFS ≤ 1 年的预测 14 例正确, 1 例不正确, 准确率为 93%, 召回率为 87%。该模型预测 PFS > 1 年的患者情况: 7 例正确, 2 例不正确, 准确率为 71%, 召回率为 83%。

## 讨 论

乳腺癌在肿瘤类别中是导致女性死亡的第二大病因<sup>[12]</sup>。拉帕替尼是 HER2 阳性晚期 / 转移性乳腺癌患者最常用的治疗药物。本研究以接受拉帕替尼治疗的乳腺癌患者为研究对象, 从这些患者的临床诊疗记

表 1 不同模型评价结果对比表

算法	标签	准确率	召回率	f1 值	精确率	AUC	灵敏度	特异度
XGBoost	0	0.93	0.87	0.90	0.86	0.83	0.83	0.87
	1	0.71	0.83	0.77				
LightGBM	0	0.85	0.73	0.79	0.71	0.78	0.67	0.73
	1	0.50	0.67	0.57				
CatBoost	0	0.76	0.87	0.81	0.71	0.76	0.33	0.87
	1	0.50	0.33	0.40				
RF	0	0.76	0.87	0.81	0.71	0.74	0.33	0.87
	1	0.50	0.33	0.40				
GBDT	0	0.75	0.80	0.77	0.67	0.78	0.33	0.80
	1	0.40	0.33	0.36				
SVR	0	0.80	0.80	0.80	0.71	0.68	0.50	0.80
	1	0.50	0.50	0.50				
LR	0	0.80	0.80	0.80	0.71	0.78	0.50	0.80
	1	0.50	0.50	0.50				
ANN	0	0.74	0.93	0.82	0.71	0.70	0.17	0.93
	1	0.50	0.17	0.25				
TabNet	0	0.81	0.87	0.84	0.76	0.77	0.50	0.87
	1	0.60	0.50	0.55				

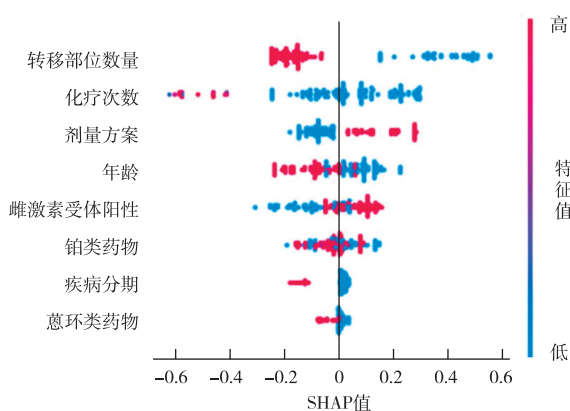


图 4 重要变量的 Shapley 加性解释 (SHAP) 值 每个圆点代表一个样本。圆点颜色越红, 代表 SHAP 值越大; 圆点颜色越蓝, 代表 SHAP 值越小。对于一个变量, 当大部分红点出现在 SHAP 值 >0 的区域时, 说明该变量对无进展生存期 (PFS) 有正向影响; 当大部分红点出现在 SHAP 值 <0 的区域时, 说明该变量对 PFS 有负向影响

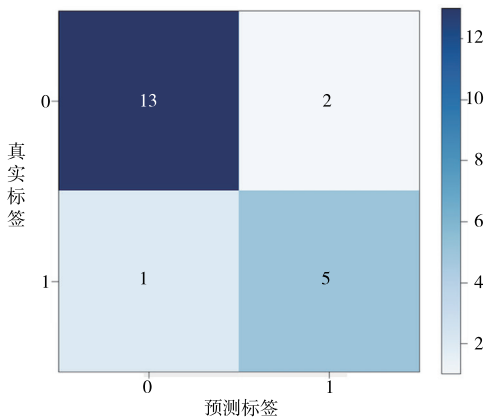


图 5 XGBoost 混淆矩阵 预测乳腺癌患者无进展生存期

录中探索挖掘影响患者 PFS 的重要变量, 最终应用机器学习技术构建拉帕替尼治疗乳腺癌患者 PFS 预测模型。本研究发现, 影响乳腺癌患者 PFS 的重要变量包括剂量方案、年龄、化疗次数、蒽环类药物、铂类药物、ER 阳性、疾病分期和转移部位数量。其中, 依据重要性得分, 转移部位数量为影响乳腺癌患者使用拉帕替尼治疗后 PFS 的最重要变量, 这也与既往研究结果及临床经验相符合, 尤其是腋窝淋巴结转移数目对乳腺癌的预后尤为重要<sup>[13]</sup>。相对于其他的重要变量, 蒽环类药物的使用对患者 PFS 的影响相对较小。应用 XGBoost 构建的拉帕替尼治疗乳腺癌患者 PFS 预测模型具有良好的性能和稳健性 (精确率 = 0.86, AUC=0.83, 灵敏度 = 0.83, 特异度 = 0.87)。该模型预测乳腺癌患者接受拉帕替尼治疗后 PFS ≤ 1 年的准确率为 93%, 预测乳腺癌患者接受拉帕替尼治疗后 PFS > 1 年的准确率为 71%。本研究比较了 9 种机器学习和深度学习算法, 最终选择 XGBoost 进行建模来预测乳腺癌患者接受拉帕替尼治疗后是否 PFS > 1 年。XGBoost 是一种可扩展且高度准确的机器学习库, 它突破了梯度提升树算法的计算能力极限<sup>[9]</sup>。在真实世界研究中, 大多数变量具有非线性关系, 且可能存在缺失值。XGBoost 可以快速有效地挖掘处理真实世界研究中的复杂高维数据, 填补缺失值及防止模型过拟合<sup>[9]</sup>。在真实世界研究中, 与传统模型相比, 机器学习和深度学习模型具有更强的泛化能力和更好的稳健性<sup>[14-16]</sup>。随着输入数据的增加, 模型性能会变得更

加稳定,这是机器学习和深度学习模型的一个特点。

本研究结果显示,多个变量均可对乳腺癌患者接受拉帕替尼治疗后 PFS 有显著影响。根据 SHAP 分析结果,ER 阳性、剂量方案与患者 PFS 呈正相关,表明 ER 阳性和 / 或接受大剂量拉帕替尼治疗的乳腺癌患者 PFS 更长。化疗次数、转移部位数量、年龄、疾病分期和蒽环类药物与患者 PFS 呈负相关,这意味着年龄较大、疾病分期为 IV 期、转移部位数量  $\geq 2$ 、接受 3 次以上化疗以及接受过蒽环类药物治疗的患者 PFS 较短。与以往的研究和临床经验基本一致或类似,FINN 等<sup>[17]</sup>发现拉帕替尼改善了 ER 或 PR 阳性的转移性乳腺癌患者的中位无事件生存期(5.7 个月 vs. 4.5 个月,  $P=0.351$ )。YE 等<sup>[3]</sup>的研究证明,与乳腺癌 II / III 期患者相比,IV 期患者的生存率更低(中位 PFS 为 15.8 个月 vs. 6.2 个月,  $P<0.001$ )。

存在器官转移的乳腺癌患者生存率通常较差,如肝转移和脑转移是影响预后的负面因素<sup>[18-20]</sup>。在临床治疗中,拉帕替尼通常在患者已接受一种或多种抗 HER2 药物(如蒽环类药物)治疗后使用,但此类患者的生存率较差<sup>[21-23]</sup>。EGF104900 和 EGF100151 临床试验的生存分析结果显示,既往接受抗 HER2 治疗次数较少的患者可获得更长的 PFS<sup>[24]</sup>。在本研究中,约 50% 患者接受了 3 次以上的化疗且生存率较差,这与 YE 等<sup>[3]</sup>的研究结果一致。随着样本量的增大,在 SHAP 分析中这些变量对 PFS 的影响会更加准确。预测模型的结局指标是乳腺癌患者接受拉帕替尼治疗后 PFS 是否  $\leq 1$  年。患者 PFS  $>1$  年表明目前的治疗方案可能会获得相对满意的预后,但如果患者的 PFS  $\leq 1$  年,就表明预后相对较差,临床医生可以考虑适当调整当前的治疗方案,如拉帕替尼与其他化疗药物联合治疗或增加拉帕替尼的剂量。XIE 等<sup>[23]</sup>研究表明,其他化疗药物与拉帕替尼联合治疗可以改善肿瘤患者的 PFS,在 EGF104900 III 期临床试验中,接受拉帕替尼与曲妥珠单抗联合治疗的患者 PFS 明显长于接受拉帕替尼单药治疗的患者(3 个月 vs. 2 个月,  $P=0.008$ )。此外,接受拉帕替尼、曲妥珠单抗、化疗联合治疗的患者中位 PFS 显著长于接受拉帕替尼和卡培他滨联合治疗的患者(8.8 个月 vs. 5.0 个月,  $P<0.000\ 001$ )。EGF109491 临床试验结果显示,拉帕替尼与卡培他滨联合治疗获得中位 PFS 为 6.34 个月,而拉帕替尼与紫杉醇联合治疗获得的中位 PFS 为 8.9 个月。因此预测 PFS  $\leq 1$  年的患者可以选择加入曲妥珠单抗治疗或其他联合治疗方案,以提高治疗效果。然而,有一项研究比较拉帕替尼单药与拉帕

替尼联合培唑帕尼治疗 HER2 阳性乳腺癌患者的有效性,结果发现拉帕替尼组和联合组的中位 PFS 分别为 16.1 周和 14.3 周,且药物毒性增加<sup>[25]</sup>。因此,预测 PFS  $\leq 1$  年的患者应谨慎使用拉帕替尼与培唑帕尼联合治疗<sup>[25]</sup>。本研究建立的乳腺癌患者接受拉帕替尼治疗后 PFS 预测模型可以帮助临床医生准确预测患者的预后情况,以便尽早制定治疗方案。未来可以通过纳入更多的数据来进一步完善该模型。

本研究具有多个优势:首先,应用机器学习技术对 RWD 中影响 PFS 的重要变量进行探索挖掘,并最终确定具有最佳预测性能、最小规模的变量集,因此该模型更加简洁高效。其次,为了充分利用患者诊疗数据,本研究采用了 SMOTETomek 采样方法来处理分布不均衡的数据集,提高模型的准确率。本研究的新颖之处在于利用机器学习技术充分挖掘了 RWD,构建了首个接受拉帕替尼治疗的乳腺癌患者预后预测模型,筛选出了影响拉帕替尼治疗后 PFS 的关键变量。机器学习与个体化用药的结合,提高了临床实践中精准医疗的水平,并为辅助乳腺癌临床治疗提供了决策依据。本研究也存在局限性:其一,研究样本量相对较小,需要更大规模的 RWD 集来优化模型性能;其二,为了验证模型的稳定性和泛化性,未来可以使用来自多中心的数据进行验证并进行增量学习。

综上,本研究构建的拉帕替尼治疗乳腺癌患者预后预测模型性能和稳健性良好,可以为乳腺癌患者的治疗方案选择提供更多的参考依据。

#### [参考文献]

- [1] FIGUEROA-MAGALHAES MC, JELOVAC D, CONNOLLY R, *et al.* Treatment of HER2-positive breast cancer [J]. *Breast*, 2014, 23 (2): 128-136.
- [2] VOIGTLAENDER M, SCHNEIDER-MERCK T, TREPEL M. Lapatinib [J]. *Recent Results Cancer Res*, 2018, 211 (4): 19-44.
- [3] YE X, LUO X, DU Q, *et al.* Effectiveness and safety of lapatinib in Chinese breast cancer patients: a real-world study [J]. *Ann Transl Med*, 2020, 8 (5): 240.
- [4] ESTEVEZ LG, SUAREZ-GAUTHIER A, GARCIA E, *et al.* Molecular effects of lapatinib in patients with HER2 positive ductal carcinoma in situ [J]. *Breast Cancer Res*, 2014, 16 (4): R76.
- [5] BLONDE L, KHUNTI K, HARRIS SB, *et al.* Interpretation and impact of real-world clinical data for the practicing clinician [J]. *Adv Ther*, 2018, 35 (11): 1763-1774.
- [6] PALACIOS M. The quality of research with real-world evidence [J]. *Colomb Med (Cali)*, 2019, 50 (3): 140-141.
- [7] SCHAD F, THRONICKE A. Real-world evidence-current

- developments and perspectives [J]. *Int J Environ Res Public Health*, 2022, 19 (16): 10159.
- [ 8 ] YU Z, YE X, LIU H, *et al.* Predicting lapatinib dose regimen using machine learning and deep learning techniques based on a real-world study [J]. *Front Oncol*, 2022, 12: 893966.
- [ 9 ] KIBRIA HB, NAHIDUZZAMAN M, GONI MOF, *et al.* An ensemble approach for the prediction of diabetes mellitus using a soft voting classifier with an explainable AI [J]. *Sensors (Basel)*, 2022, 22 (19): 7268.
- [ 10 ] HATAMIKIA S, MAGHOOLI K, NASRABADI AM. The emotion recognition system based on autoregressive model and sequential forward feature selection of electroencephalogram signals [J]. *J Med Signals Sens*, 2014, 4 (3): 194–201.
- [ 11 ] YI F, YANG H, CHEN D, *et al.* XGBoost-SHAP-based interpretable diagnostic framework for Alzheimer's disease [J]. *BMC Med Inform Decis Mak*, 2023, 23 (1): 137.
- [ 12 ] BONDE GV, YADAV SK, CHAUHAN S, *et al.* Lapatinib nano-delivery systems: a promising future for breast cancer treatment [J]. *Expert Opin Drug Deliv*, 2018, 15 (5): 495–507.
- [ 13 ] 赵建国, 袁 媛, 张应天. III c (pN3a) 期乳腺癌淋巴结转移数目或转移率对判断预后的价值 [J]. *临床外科杂志*, 2013, 21 (7): 507–509. ZHAO J, YUAN Y, ZHANG Y. The value of the number or rate of lymph node metastasis in predicting prognosis of stage III C (pN3a) breast cancer [J]. *J Clin Surg*, 2013, 21 (7): 507–509.
- [ 14 ] MO X, CHEN X, LI H, *et al.* Early and accurate prediction of clinical response to methotrexate treatment in juvenile idiopathic arthritis using machine learning [J]. *Front Pharmacol*, 2019, 10 (1): 1155.
- [ 15 ] LEE HC, YOON SB, YANG SM, *et al.* Prediction of acute kidney injury after liver transplantation: machine learning approaches vs. logistic regression model [J]. *J Clin Med*, 2018, 7 (11): 428.
- [ 16 ] KRUPPA J, ZIEGLER A, KONIG IR. Risk estimation and risk prediction using machine-learning methods [J]. *Hum Genet*, 2012, 131 (10): 1639–1654.
- [ 17 ] FINN RS, PRESS MF, DERING J, *et al.* Estrogen receptor, progesterone receptor, human epidermal growth factor receptor 2 (HER2), and epidermal growth factor receptor expression and benefit from lapatinib in a randomized trial of paclitaxel with lapatinib or placebo as first-line treatment in HER2-negative or unknown metastatic breast cancer [J]. *J Clin Oncol*, 2009, 27 (24): 3908–3915.
- [ 18 ] CAMERON D, CASEY M, OLIVA C, *et al.* Lapatinib plus capecitabine in women with HER2-positive advanced breast cancer: final survival analysis of a phase III randomized trial [J]. *Oncologist*, 2010, 15 (9): 924–934.
- [ 19 ] GOKSU SS, BOZCUK H, KORAL L, *et al.* Factors predicting lapatinib effectiveness in HER2+ metastatic breast carcinoma: does it work better in different histologic subtypes? [J]. *Indian J Cancer*, 2015, 52 (4): 517–519.
- [ 20 ] TOMASELLO G, BEDARD PL, de AZAMBUJA E, *et al.* Brain metastases in HER2-positive breast cancer: the evolving role of lapatinib [J]. *Crit Rev Oncol Hematol*, 2010, 75 (2): 110–121.
- [ 21 ] MA F, OUYANG Q, LI W, *et al.* Pyrotinib or lapatinib combined with capecitabine in HER2-positive metastatic breast cancer with prior taxanes, anthracyclines, and/or trastuzumab: a randomized, phase II study [J]. *J Clin Oncol*, 2019, 37 (29): 2610–2619.
- [ 22 ] JU UH, TSAI ML, LIN CS, *et al.* Lapatinib and platinum-based chemotherapy ameliorate breast cancer with choroidal metastasis and restore visual acuity [J]. *Tumori*, 2014, 100 (3): e67–e69.
- [ 23 ] XIE Y, GE R, SANG D, *et al.* Real-world data of lapatinib and treatment after lapatinib in patients with previously treated HER2-positive metastatic breast cancer: a multicenter, retrospective study [J]. *Cancer Med*, 2020, 9 (9): 2981–2988.
- [ 24 ] WANG X, WANG L, YU Q, *et al.* The effectiveness of lapatinib in HER2-positive metastatic breast cancer patients pretreated with multiline anti-HER2 treatment: a retrospective study in China [J]. *Technol Cancer Res Treat*, 2021, 20 (1): 15330338211037812.
- [ 25 ] CRISTOFANILLI M, JOHNSTON SR, MANIKHAS A, *et al.* A randomized phase II study of lapatinib + pazopanib versus lapatinib in patients with HER2+ inflammatory breast cancer [J]. *Breast Cancer Res Treat*, 2013, 137 (2): 471–482.