基于PSO-SVM的肺癌患者免疫治疗相关不良事件预测模型*

王明灏[1] 张 强[2] 毕 楠[1,3△] 李 光[1△]

【提 要】 **目的** 探讨改进支持向量机在小样本数据情况下的分类效果，并对辽宁省某三级甲等医院2019—2024年296例肺癌患者免疫治疗相关不良事件进行预测。**方法** 回顾性地收集了确诊肺癌且接受抗程序性死亡受体1（programmed death-1，PD-1）/抗程序性死亡配体1（programmed death ligand-1，PD-L1）治疗的296例患者临床数据。在对患者数据进行预处理后，构建了基于粒子群算法优化支持向量机（particle swarm optimization support vector machine，PSO-SVM）的肺癌患者免疫治疗相关不良事件预测模型，并与改进前的支持向量机（support vector machine，SVM）模型进行比较。**结果** 经过改进的SVM表现出了更好的预测和分类能力，PSO-SVM的准确率为82.0%，高于SVM算法的78.7%。**结论** 小样本数据情况下，PSO-SVM算法具有较高的准确率，为提前预知肺癌患者免疫治疗相关不良事件的发生和为肺癌患者的免疫治疗临床决策提供了新的方法尝试。

【关键词】 肺癌 免疫检查点抑制剂 免疫治疗相关不良事件 机器学习
【中图分类号】 R195.1 【文献标志码】 A **DOI** 10.11783/j.issn.1002-3674.2026.04.018

原发性支气管肺癌简称肺癌，在我国有较高的发病率和死亡率，严重威胁着居民的生命健康[1]。根据最新数据，2022年肺癌新发病例占中国的所有恶性肿瘤新发病例总数的18.06%，肺癌死亡人数占中国恶性肿瘤死亡总数的23.9%，均排名在第1位[2]。早期肺癌并不产生明显的症状，多数患者在出现症状时往往已经到了晚期，晚期肺癌患者整体的5年生存率只有20%左右[3]。近年来，免疫检查点抑制剂（immune checkpoint inhibitors，ICIs）已成为治疗恶性肿瘤的重要手段[4]。免疫检查点抑制剂通过调控机体免疫系统来实现抗肿瘤的作用，所以它被称为“广谱抗肿瘤药物”，为患者的长期生存提供了可能[5]。随着ICIs的广泛应用以及使用患者的不断增多，由于免疫治疗的独特作用机制，会不可避免地出现一系列免疫相关不良反应（immune-related adverse events，irAEs）[6]。因此，现阶段ICIs在治疗肺癌患者中的有效性和安全性仍存在争议，还需要进行大量的研究。

机器学习为一种新兴的数据计算与统计方法，可以从多维度进行医学数据的分析与挖掘，目前机器学习已经逐步应用于预防、治疗和监测等方面[7-9]。支持向量机作为机器学习算法的一种，其通过寻求结构化、风险最小，来提高学习泛化能力，这使其在统计样本数量较小的情况下，依然可以获得良好统计规律的目的[10]。本研究利用使用免疫检查点抑制剂的肺癌患者的临床资料，构建小样本数据下的机器学习模型对irAEs的发生进行预测，对ICIs在治疗肺癌患者中的临床风险进行有效的识别与评估，为今后临床上对irAEs的管理及肺癌患者的免疫治疗临床决策提供一些思路。

模型介绍

1. 支持向量机

支持向量机（support vector machine，SVM）是一种从统计学习理论中衍生出来的分类算法，已被广泛应用于模式识别。其基本原理是对于非线性可分的数据，通过非线性映射函数转换到高维空间实现分类。

假设患者临床信息的数据集 $T=\{(x_i,y_i),i=1,2,\cdots,n\}$，$x_i$为患者临床信息的特征向量集合，$y_i$为免疫相关不良事件的标签，其分类超平面可表示为：

$$f(x)=\omega^T x+b \quad (1)$$

式（1）中：ω为权系数向量；x为样本空间内任意一点，b为偏置。

在SVM分类器中引入松弛变量$\xi_i\geqslant 0$，以容忍异常数据，此时超平面求解问题变为约束条件下的凸函数优化问题，可表示为：

$$\min_{\omega,b,\xi_i}\frac{1}{2}\|\omega\|^2+c\sum_{i=1}^{m}\xi_i \quad (2)$$

式（2）中：c为惩罚系数，用于控制不满足最小区间的数据样本贡献，当样本分类错误时，c越大，代表惩罚程度越重。

采用高斯核函数将x映射到高维空间$\varphi(x)$上，使样本空间线性可分，其中φ为映射函数。高斯核函数k

* 基金项目：中国医学科学院医学与健康科技创新工程（2023-I2M-C&T-B-073）
1. 中国医科大学附属第一医院放疗科（110001）
2. 中国医科大学附属第一医院全科医学科
3. 国家癌症中心/国家肿瘤临床医学研究中心/中国医学科学院北京协和医学院肿瘤医院放疗科
△ 通信作者：毕楠，E-mail：binan_email@163.com；李光，E-mail：13804058616@163.com

(x_i, x_j)的表达式为：

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) = \exp\left(-g\|x_i - x_j\|^2\right) \quad (3)$$

式(3)中：g为核函数参数，$g = \frac{1}{2\sigma^2}$，决定高维空间中支持向量个数且影响训练速度；σ为高斯核的带宽参数；$\|x_i - x_j\|^2$为样本点x_i和样本点x_j之间的距离。

因此，可通过对参数c和g进行优化以提高SVM分类器的分类精度。

2. 粒子群优化算法

粒子群算法(particle swarm optimization, PSO)是一种能够在全局范围内进行参数寻优的群体智能优化算法。PSO算法将待优化问题的每种解转化为D维空间中包含N个粒子的粒子群，在每次迭代期间，粒子的速度和位置保持不变，在下一次迭代中，每个粒子根据当前个体最优位置与群体最优位置运动，最终获得最优解。粒子的速度和位置分别由式(4)、式(5)确定。

$$v_{iD}^{k+1} = \omega v_{iD}^{k} + c_1 r_1 (p_{best_{iD}}^{k} - x_{iD}^{k}) + c_2 r_2 (g_{best_{iD}}^{k} - x_{iD}^{k}) \quad (4)$$

$$x_{iD}^{k+1} = v_{iD}^{k+1} + x_{iD}^{k} \quad (5)$$

式中：i=0, 1, …, 2^N−1为每次迭代中的粒子数；k为当前迭代次数；ω为惯性系数；v_{iD}^{k+1}、x_{iD}^{k}分别为第k次迭代中粒子i的速度和位置；r_1、r_2为(0, 1)的随机数；c_1、c_2为加速系数，用于调整个体最优位置$p_{best_{iD}}^{k}$和群体最优位置$g_{best_{iD}}^{k}$。

研究对象

1. 资料来源

选取2019年4月至2024年4月在辽宁省某三级甲等医院就诊的患者作为研究对象，本研究共纳入了296例患者，利用其临床信息开展irAEs预测模型构建和相关预测分析。患者纳入标准：①年龄≥18岁；②患者病理报告诊断为恶性肺癌；③首次在我院使用抗程序性死亡受体1(programmed death−1, PD−1)/抗程序性死亡配体1(programmed death ligand−1, PD−L1)进行治疗，同时可接受辅助化疗或放疗；④患者在使用PD−1/PD−L1治疗前经过了全面系统的评估，在用药治疗前不存在严重的内分泌系统异常。排除标准：①年龄<18岁；②接受PD−1/PD−L1治疗的非肺癌患者；③缺少恶性肿瘤病理诊断报告或原发部位不明；④来我院就诊前接受了其他医院的PD−1/PD−L1治疗。

2. 患者基本情况

通过电子病历系统查阅患者的临床资料，收集接受PD−1/PD−L1治疗的肺癌患者的临床数据(表1)，包括患者的基本人口特征(性别、年龄、BMI指数)、家族肿瘤史、肺癌的病理分类、PD−1/PD−L1的用药剂量、治疗周期、治疗方案(采用单药或联合化疗)、患者的基础疾病信息等。对于发生irAEs的患者，同时需要收集其irAEs的分级、采取的治疗方案、发生时间和持续时间等信息。

表1　本研究患者基本情况(n=296)

组别	项目	n(%)
性别	男性	227(76.7)
	女性	69(23.3)
年龄(岁)	≤39	17(5.8)
	40~49	34(11.5)
	50~59	59(19.9)
	≥60	186(62.8)
BMI(kg/m^2)	18~22.5	142(48.0)
	22.6~27	154(52.0)
长期抽烟	是	124(41.9)
	否	172(58.1)
长期饮酒	是	92(31.1)
	否	204(68.9)
基础疾病	有	116(39.2)
	没有	180(60.8)
肺癌类型	非小细胞肺癌	252(85.1)
	小细胞肺癌	44(14.9)
家族肿瘤史	有	68(23.0)
	没有	228(77.0)
是否转移	是	178(60.1)
	否	118(39.9)
治疗方案	联合治疗	193(65.2)
	单药治疗	103(34.8)
ICIs类型	PD−1	279(94.3)
	PD−L1	17(5.7)

irAEs的判断以中国临床肿瘤学会发布的《免疫检查点抑制剂相关的毒性管理指南(2023)》[11]为主，分为5个级别：1级为轻度毒性、2级为中度毒性、3级为重度毒性、4级为危及生命的毒性、5级为与毒性相关的死亡。为了模型研究方便，将未出现irAEs的患者定义为0级，本次收集到296例患者的临床数据，其中出现irAEs的患者为184例(均为1~4级)，未出现irAEs的患者为112例。

模型构建

本文采用Anaconda平台来搭建Python3.7的环境，在PyCharm中使用Python编程语言来实现肺癌患者免疫治疗相关不良事件预测模型，模型设置输入数据为患者的临床数据，对于非数值型的临床特征，对其进行标签化处理，输出数据为irAEs事件分类标签。惩罚因子c和核函数参数g对SVM分类器的性能起着决定性的作用，利用粒子群算法对SVM的参数c和g的选择进行优化。

基于PSO−SVM的肺癌患者免疫治疗相关不良事

件预测主要步骤为：①患者临床数据预处理，对于非数值型的临床特征进行数据标签化。②SVM 分类器模型构建。选择高斯核函数作为 SVM 分类器的核函数，初始化惩罚因子 c 和核函数参数 g。③参数优化。初始化粒子群后对参数 c 和 g 编码，利用数据集对 c 和 g 的选择进行优化，构建最优 SVM 分类器模型。④irAEs 事件预测，利用优化后的 SVM 分类器对 irAEs 事件进行分类预测。

预测结果

把患者的临床数据作为 2 种 irAEs 事件分类器 SVM、PSO—SVM 的输入，采用的数据集共包含 296 个样本，将其中的 70%（207 个样本）划分为训练样本，另外的 30%（89 个样本）划分为测试样本，采用 2 种分类算法进行分类，将分类过程重复多次以获得平均结果，分类结果如图 1 所示。

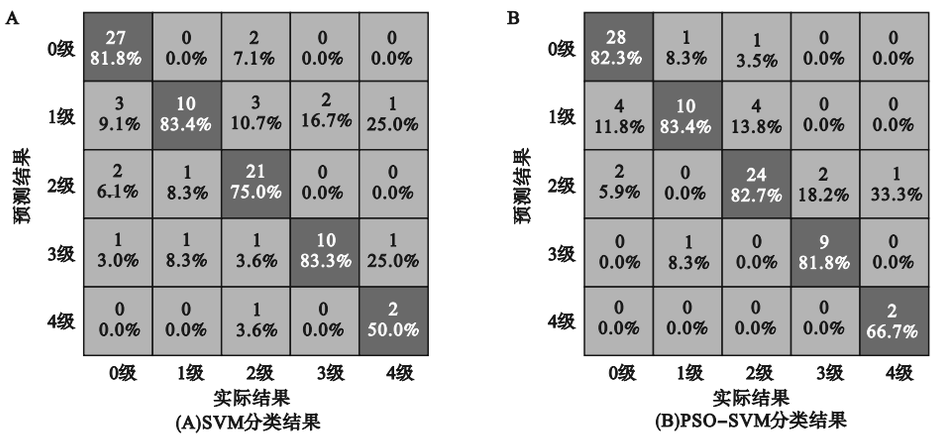


图 1 irAEs 预测结果

从图 1 中可以看出，本文所使用的经过改进的支持向量机对 5 种事件的分类预测中有 4 种事件的分类准确率超过了 80%，4 级 irAEs 事件的准确率为 66.7%，可能与其事件数量较少，造成了算法的训练不足。整体来看，PSO-SVM 的准确率为 82.0%，高于 SVM 算法的 78.7%，表明了经过改进的 SVM 表现出了更好的预测和分类能力。

为了进一步分析分类器性能，表 2 给出 PSO-SVM 模型对 5 种事件进行分类预测的精确率、召回率和 F_1 值。

表 2 PSO-SVM 算法评价指标（%）

算法评价指标	不良事件分类				
	0级	1级	2级	3级	4级
精确率	93.3	55.6	85.7	90.0	100.0
召回率	82.3	83.4	82.7	81.8	66.7
F_1	87.5	66.7	84.2	85.7	80.0

讨 论

本研究中 296 例接受 PD-1/PD-L1 治疗的肺癌患者男女比例为 3.29∶1，平均年龄 66.49±6.43 岁，中老年男性为肺癌的好发人群，这与其流行病学特征相一致[12]。一项中国人群肿瘤患者中替雷利珠单抗血药浓度与免疫相关不良事件的相关性研究显示[6]，纳入分析的 50 例患者中，共有 34 例发生了 irAEs，发生率为 68%，包括甲状腺功能减退、皮疹、肝功能损害等。一项中国人群 ICIs 相关不良事件的研究显示[13]，1063 例患者中 irAEs 发生率为 43.3%。本研究共有 62.16% 的患者出现 1~4 级的 irAEs，符合真实世界中 irAEs 的发生率。

ICIs 为肿瘤治疗领域带来革新性的改变，已广泛应用于晚期肺癌的治疗中，ICIs 通过阻断肿瘤细胞利用的免疫检查点来重新激活 T 细胞的抗肿瘤效应，从而增强免疫杀伤作用。目前国际上上市的肿瘤免疫治疗的 ICIs 药物主要包括抗细胞毒 T 淋巴细胞相关抗原 4 单抗、抗 PD-1 单抗、抗 PD-L1 单抗。在我国抗 PD-1/PD-L1 单抗已广泛应用于晚期非小细胞肺癌（non small cell lung cancer，NSCLC）与小细胞肺癌（small cell lung cancer，SCLC）患者治疗中[14-15]。然而，随着患者数量的增多，在接受 ICIs 治疗过程中也逐渐发现了全新的 irAEs，与传统的放疗、化疗或靶向治疗的药物相关毒性不同，irAEs 的发生往往由药物的间接作用引起，由于药物的作用，导致了自身免疫或自身炎症性的发生。临床预测和早期管理 irAEs 并进一步完善肿瘤免疫疗法和提高其治疗的有效性具有重要意义。目前，已经有学者通过探究器官特异性 irAEs 的生物标记物，来建立 irAEs 预测模型，这有助于临床提前预测和筛选出可能发生 irAEs 的高危人群，从而谨慎权衡免疫治疗方案的获益和产生的毒性，或者对接受免疫治疗患者进行更细致和严密的监测[16]。与许多传统的统计方法相比，随着与医疗保健相关的大型数据集越来越多，机器学习方法有可能大大改善经验分析的各个方面。

本研究比较了SVM、PSO-SVM两种模型对肺癌患者免疫治疗相关不良事件分类预测的能力，研究结果表明PSO-SVM算法具有较高的准确率，达到了82.0%，在小样本数据情况下，为提前预知irAEs的发生和肺癌患者的免疫治疗临床决策提供新的方法尝试。在其他研究中，Yang等[17]利用美国食品药品监督管理局不良事件报告系统的公开数据，构建SVM模型预测了接受PD-1/PD-L1治疗患者的高血糖不良反应，表明了SVM模型对药物不良反应的预测具有有效性。Yu等[18]利用接受ICIs治疗患者的临床数据，构建了8种机器学习模型来预测患者的急性肾损伤不良反应，包括：支持向量机算法（基于radial kernel核函数、sigmoid kernel核函数、polynomial kernel核函数）、决策树算法、随机森林算法、逻辑回归算法、神经网络算法、自适应增强算法、极限梯度提升算法、朴素贝叶斯算法，结果表明基于radial kernel核函数的支持向量机模型的性能最优，其预测准确率为75.68%。

本研究的不足之处：第一，受到患者临床资料收集困难的影响，在预测模型的变量中没有考虑用药前的检验学数据，如血常规数据、肝功能数据等等，限制了模型的学习能力；第二，缺乏对原始数据的深度预处理，数据的质量很大程度上决定了模型的质量，在未来的研究中可以尝试结合主成分分析等数据预处理方法来对原始数据进行深度预处理，进一步提高模型的预测精度。

参 考 文 献

[1] 中华医学会肿瘤学分会,中华医学会杂志社. 中华医学会肺癌临床诊疗指南(2023版)[J]. 中华肿瘤杂志,2023,45(7):539-574.

[2] Xia C, Dong X, Li H, et al. Cancer statistics in China and United States, 2022: profiles, trends, and determinants [J]. Chinese Medical Journal, 2022, 135(5): 584-590.

[3] Allemani C, Matsuda T, Carlo VD, et al. Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries [J]. Lancet (London, England), 2018, 391(10125): 1023-1075.

[4] 黄莹,许可,王炜,等. 免疫检查点抑制剂在非小细胞肺癌临床诊疗中的研究进展[J]. 中华肿瘤杂志,2020,42(9):713-717.

[5] 任晓蕾,詹轶秋,刘刚,等. 替雷利珠单抗在非小细胞肺癌患者中的安全性：一项68例单中心回顾性研究[J]. 中国药学杂志,2024,59(13):1262-1266.

[6] 诸慧,原永芳,徐影. 肿瘤患者中替雷利珠单抗血药浓度与免疫相关不良事件的相关性研究[J]. 中国临床药理学杂志,2024,40(8):1217-1221.

[7] 蔡偌欣,马亚楠,闻德亮. 机器学习在非酒精性脂肪肝预测中的应用[J]. 中国卫生统计,2024,41(2):316-318.

[8] 孙欣,魏东升,许允发,等. 基于影像组学构建2型糖尿病脑梗死后再发脑梗死的风险预测模型[J]. 中国卫生统计,2024,41(2):261-265.

[9] 姜超,冯哲,王均琴,等. 三种机器学习方法在脑出血中医辨证分类中应用的比较研究[J]. 中国卫生统计,2023,40(6):921-924.

[10] 张学工. 关于统计学习理论与支持向量机[J]. 自动化学报,2000(1):36-46.

[11] 中国临床肿瘤学会. 免疫检查点抑制剂相关的毒性管理指南(2023)[M]. 北京：人民卫生出版社,2023.

[12] 胡明明,阳苑,张同梅,等. 7843例原发肺恶性肿瘤流行病学特征分析[J]. 现代肿瘤医学,2021,29(4):682-685.

[13] Li L, Li G, Rao B, et al. Landscape of immune checkpoint inhibitor-related adverse events in Chinese population [J]. Scientific Reports, 2020, 10(1): 15567.

[14] 黄岩,张力. 2020 CSCO非小细胞肺癌诊疗指南更新要点解读[J]. 临床内科杂志,2020,37(8):603-605.

[15] 张新.《2020年CSCO小细胞肺癌诊疗指南》解读[J]. 临床内科杂志,2020,37(11):820-822.

[16] 姚卓然. 肺癌免疫治疗相关不良反应的多组学生物标记物探究与预测模型建立[D]. 北京：中国医学科学院,2021.

[17] Yang J, Li N, Lin W, et al. Machine Learning for Predicting Hyperglycemic Cases Induced by PD-1/PD-L1 Inhibitors [J]. Journal of Healthcare Engineering, 2022, 2022(1): 6278854.

[18] Yu X, Wu R, Ji Y, et al. Identifying patients at risk of acute kidney injury among patients receiving immune checkpoint inhibitors: a machine learning approach [J]. Diagnostics, 2022, 12(12): 3157.

（责任编辑：郭海强）