

# 基于机器学习的原发性肺癌患者合并肌肉减少症 风险预测模型构建与验证

唐萍<sup>1</sup>, 周亮<sup>1</sup>, 李松达<sup>1</sup>, 韩金慧<sup>2</sup>, 李砥<sup>2</sup>, 姚雪梅<sup>1</sup>

1. 新疆医科大学公共卫生学院, 新疆 乌鲁木齐 830000; 2. 克拉玛依市中心医院临床医学研究中心

**摘要:** **目的** 基于原发性肺癌患者合并肌肉减少症的危险因素, 构建肺癌患者合并肌肉减少症的机器学习模型, 旨在提高对该病的早期识别和诊断能力, 提供个性化的营养和治疗建议, 改善患者的健康状态和预后。 **方法** 选取 2023 年 10 月—2024 年 7 月在克拉玛依市中心医院住院的 263 例原发性肺癌患者作为研究对象, 收集患者性别、年龄、BMI、IGF-1、GDF-8、白介素 6 (IL-6)、肿瘤坏死因子 (TNF- $\alpha$ ) 等资料。根据亚洲肌肉减少症工作组 (AWGSOP) 制定标准将患者分为肌肉减少症组及非肌肉减少症组, 通过单因素分析及 LASSO 回归分析筛选出影响原发性肺癌患者出现肌肉减少症的危险因素, 将筛选出的危险因素纳入 R 软件构建 K 最近邻算法模型 (KNN)、高斯朴素贝叶斯 (GNB) 与多因素 logistic 回归模型, 采用 Bootstrap 方法进行内部验证, 采用受试者工作特征 (ROC) 曲线、Hosmer-Lemeshow 校准曲线及临床决策曲线 (DCA) 评估机器学习模型的预测价值和临床实用性。 **结果** 共调查 263 名患者, 肌肉减少症组 137 例, 非肌肉减少症组 126 例; 肌肉减少症组的饮酒比例、IL-4、IL-6、IL-17、TNF- $\alpha$ 、GDF-8、尿素氮、低密度脂蛋白等显著高于非肌肉减少症组, 而 BMI、吸烟、PSQI 评分、IGF-1、血小板计数等显著低于非肌肉减少症组, 差异均具有统计学意义 ( $P < 0.05$ )。ROC 曲线显示, KNN 模型 C 指数为 0.936, GNB 模型 C 指数为 0.935, 显著优于多因素 logistic 回归模型 C 指数为 0.926, Hosmer-Lemeshow 拟合优度检验结果显示, KNN 模型与 GNB 模型与多因素 logistic 回归模型的预测值与实际值之间的平均误差分别为 0.026 9、0.018 8 与 0.028 7, GNB 模型显著优于 KNN 模型与多因素 logistic 回归模型, 预测结果和观察结果之间具有良好的 consistency。DCA 曲线显示, 使用 GNB 预测模型预测肺癌患者发生肌肉减少症的风险显著优于 KNN 模型与多因素 logistic 回归模型。 **结论** 高水平的 IL-6、TNF- $\alpha$ 、GDF-8 及低水平的 IGF-1、白蛋白和较差的睡眠质量、低 BMI 是肺癌患者合并肌肉减少症的独立危险因素, 本研究构建的 GNB 预测模型显著优于 KNN 模型及多因素 logistic 回归模型, 能够精准个体化预测肺癌患者合并肌肉减少症的风险, 可为临床提供个性化的营养和治疗建议, 从而改善患者的健康状态和预后。

**关键词:** 肌肉减少症; 肺癌; K 最近邻算法模型; LASSO; 高斯朴素贝叶斯 (GNB) 模型

中图分类号: R734.2; R685 文献标志码: A 文章编号: 1003-8507(2025)16-2900-08

DOI: 10.20043/j.cnki.MPM.202504089

## Construction and validation of a machine learning based risk prediction model for primary lung cancer patients with muscle atrophy

TANG Ping\*, ZHOU Liang, LI Song-da, HAN Jin-hui, LI Di, YAO Xue-mei

\* School of Public Health, Xinjiang Medical University, Urumqi, Xinjiang 830000, China

**Abstract: Objective** To construct a machine learning model based on Insulin-like Growth Factor-1 (IGF-1) and Growth Differentiation Factor-8 (GDF-8, Myostatin) for predicting sarcopenia in lung cancer patients, with the aim of improving early detection and diagnostic capabilities, providing personalized nutrition and treatment recommendations, and enhancing patients' health status and prognosis. **Methods** A total of 263 primary lung cancer patients hospitalized at Karamay Central Hospital between October 2023 and July 2024 were selected as research subjects. Data on gender, age, BMI, IGF-1, GDF-8, interleukin-6 (IL-6), tumor necrosis factor-alpha (TNF- $\alpha$ ), and other variables were collected. Patients were divided into a sarcopenia group and a non-sarcopenia group according to the criteria established by the Asian Working Group for Sarcopenia (AWGSOP). Univariate analysis and LASSO regression analysis were used to identify risk factors for sarcopenia in primary lung cancer patients. The selected risk factors were incorporated into the K-nearest neighbors (KNN) algorithm

**基金项目:** 自治区自然科学基金项目 (2024D01C131); 新疆维吾尔自治区“十四五”高等学校特色学科-公共卫生与预防医学; 天山英才医药卫生高层次人才培养计划 (TSYC202301B150)

**作者简介:** 唐萍 (2001—), 女, 硕士在读, 研究方向: 流行病与统计学

**通信作者:** 姚雪梅, E-mail: yxm0421@126.com

model, Gaussian Naive Bayes (GNB), and binary logistic regression models, using the R software. Internal validation was performed using the Bootstrap method. **Results** A total of 263 patients were investigated, with 137 in the sarcopenia group and 126 in the non-sarcopenia group. The sarcopenia group had significantly higher proportions of alcohol consumption, IL-4, IL-6, IL-17, TNF- $\alpha$ , GDF-8, blood urea nitrogen, and low-density lipoprotein, while BMI, smoking, PSQI score, IGF-1, and platelet count were significantly lower than those in the non-sarcopenia group (all  $P < 0.05$ ). The ROC curve showed that the C-index of the KNN model was 0.936, the C-index of the GNB model was 0.935, both significantly better than the binary logistic regression model's C-index of 0.926. The Hosmer-Lemeshow goodness-of-fit test showed that the average prediction error between the predicted and actual values of the KNN model, GNB model, and binary logistic regression model were 0.026 9, 0.018 8, and 0.028 7, respectively, with the GNB model significantly outperforming the KNN model and the binary logistic regression model. The prediction results of the GNB model were highly consistent with the observed outcomes. DCA curves demonstrated that the GNB model outperformed both the KNN model and the binary logistic regression model in predicting sarcopenia risk in lung cancer patients. **Conclusion** High levels of IL-6, TNF- $\alpha$ , and GDF-8, low levels of IGF-1 and albumin, poor sleep quality, and low BMI are independent risk factors for sarcopenia in lung cancer patients. The GNB prediction model constructed in this study significantly outperforms both the KNN model and the binary logistic regression model, providing precise and individualized predictions for sarcopenia risk in lung cancer patients. This model can offer personalized nutrition and treatment recommendations for clinical practice, improving patients' health status and prognosis.

**Keywords:** Sarcopenia; Lung cancer; K-nearest neighbors algorithm model; LASSO; Gaussian Naive Bayes (GNB) model

## 1 前言

肺癌是全球发病率和死亡率最高的恶性肿瘤之一。根据全球癌症统计数据(GLOBOCAN)显示,2022年全球新发肺癌病例约为220万例,占所有癌症病例的11.4%;因肺癌死亡的人数约为180万,占所有癌症死亡的18%<sup>[1-2]</sup>。尽管近年来在早期诊断和治疗方面取得了一定进展,但肺癌的预后仍较差。肺癌患者中,肌肉减少症(Sarcopenia)是一种常见的并发症,其特征是骨骼肌量减少和肌肉功能下降。肌肉减少症不仅影响患者的生活质量,还与化疗毒性增加、住院时间延长及较高的死亡率相关<sup>[3]</sup>。

胰岛素样生长因子-1(Insulin-like Growth Factor-1, IGF-1)和生长分化因子-8(Growth Differentiation Factor-8, GDF-8, 肌肉生长抑制素)在调节肌肉代谢和肌肉量方面具有重要作用。IGF-1是一种多功能多肽,通过与受体结合,激活PI3K/Akt和MAPK信号通路,促进肌细胞增殖和分化,抑制细胞凋亡,从而维持肌肉质量和功能<sup>[4]</sup>。研究表明,IGF-1表达与肌肉重量高度正相关( $r = 0.803, P < 0.01$ ),补充外源性IGF-1可显著缓解肌肉萎缩,表明IGF-1可能是治疗肌肉减少症的潜在靶点<sup>[5]</sup>。相反,GDF-8通过与其受体(Activin受体)结合,激活Smad2和Smad3信号通路,抑制肌细胞增殖和分化,从而限制肌肉生长<sup>[6]</sup>。在癌症患者中,GDF-8抑制肌肉合成及卫星细胞活性,并与肌肉量减少有关<sup>[7]</sup>。GDF-8不仅通过直接抑制肌肉生长,还通过诱导炎症和脂肪代谢紊乱,进一步加剧肌肉减少症的发生和发展<sup>[8]</sup>。

近年来,随着机器学习技术的迅速发展,应用机器学习算法进行疾病的早期预测和风险评估已成为医学领域的一大趋势。相比于传统统计方法,机器学习算法具备处理大量复杂数据的能力,能够揭示变量间深层次的非线性关系,从而提供更为精确、可靠的预测结果。本研究旨在构建基于IGF-1和GDF-8水平的肺癌患者肌肉减少症预测模型。通过分析肺癌患者血清中IGF-1和GDF-8的浓度数据,并结合患者的临床特征和病史信息,采用机器学习算法对模型进行训练与验证,最终筛选出最优的预测模型。该模型不仅可以帮助临床医生早期识别高风险患者,还能为个性化治疗方案的制定提供重要参考。

通过本研究的开展,旨在为改善肺癌患者的预后及提升其生活质量提供新思路和方法,并为临床治疗决策的优化提供有力支持。

## 2 资料与方法

**2.1 一般资料** 选择2023年10月—2024年7月在克拉玛依市中心医院血液肿瘤科经组织病理学诊断为原发性肺癌得263例患者为研究对象,并根据亚洲肌肉减少症工作组(AWGSOP)制定标准将其分为肌肉减少症组(137例)及非肌肉减少症组(126例),所有患者在确诊后4个月内完成评估。纳入标准:①年龄:18~65岁;②经临床病理初次诊断为I~II期肺癌;③同意参与研究并签署知情同意书。排除标准:①合并有其他癌症。②其他原因引起的肌肉减少:慢性阻塞性肺疾病、运动损伤、严重营养不良等。③不能提供完整数据的患者。本研究已获得克拉玛依市中心医院伦理委员会的批准(YL-2024-85),所有研

究对象对研究内容等知情,并签署知情同意书。

**2.2 方法** 收集患者年龄、性别、吸烟情况、体质指数(body mass index, BMI)、肿瘤病理分型,肿瘤 TNM 分期等资料。采集受试者在清晨空腹状态下进行外周静脉血的抽取。测定白介素-6(IL-6)、白介素-2(IL-2)、C-反应蛋白(CRP)、肿瘤坏死因子 $\alpha$ (TNF- $\alpha$ )、球蛋白、白蛋白等。

**2.3 匹兹堡睡眠质量指数(Pittsburgh Sleep Quality Index, PSQI)** 由 Buysse 等<sup>[9]</sup>编制,用来测试近 1 个月的睡眠质量。该量表共 24 个条目,7 个维度,总分 0~21 分。得分越高,表示睡眠质量越差。<7 分为无睡眠障碍, $\geq 7$  分为存在睡眠障碍。本研究中该量表的 Cronbach S d 系数为 0.807。

**2.4 肌肉减少症诊断标准** 根据亚洲工作组(Asian Working Group for Sarcopenia, AWGS)2019 年共识,肌肉减少症的诊断标准为四肢骨骼肌质量减少,同时伴有肌肉力量或躯体功能的降低<sup>[10]</sup>。

**2.4.1 四肢骨骼肌肌肉量** 采用韩国双生物电阻抗分析仪(inbody770)进行测量,获得的四肢骨骼肌含量(appendicular skeletal muscle mass, ASM)是双上肢及双下肢骨骼肌肌肉含量之和并以身高(Ht)进行校正。校正公式为:ASM/Ht<sup>2</sup>。男性 ASM/Ht<sup>2</sup> < 7.0 kg/m<sup>2</sup>,女性 ASM/Ht<sup>2</sup> < 5.7 kg/m<sup>2</sup> 判定为低骨骼肌质量<sup>[10]</sup>。

**2.4.2 握力** 采用德国 JOTTMM 电子握力器来测量优势侧握力。测试前将握距调整在合适的范围内。测试时,受试者端坐,上臂与前臂成 90 度,测试臂稍外展,但不超过 30 度,用最大力握住仪器握柄,左右手各握两次,取最大值,两侧数据相差较小,以右侧进行统计分析。若男性握力 < 28 kg,女性握力 < 18 kg,则判定为肌力降低<sup>[10]</sup>。

**2.4.3 步速** 使用 6 米步速测定法。用彩色胶布在起点到终点标记出 12 米直线距离,并标记出起点、3 米点、9 米点和终点。受试者行至 3 米线时开始计

时,行至 9 米时计时结束。测试 3 次,取其中最快一次进行统计分析。若步速 < 1.0 m/s 则判定为躯体功能降低<sup>[10]</sup>。

**2.5 GDF-8 和 IGF-1 的测定** 分别于患者入院后次日清晨及治疗后次日清晨抽血,收集血标本 2 管(EDTA 抗凝),每管约 4 ml,离心后取出上层血清样本,立即放于冷藏冰箱中冷藏;采用酶联免疫吸附试验(ELISA)法检测血清 GDF-8 及 IGF-1 水平,试剂盒购自瑞迪生物科技(武汉)有限公司,操作步骤按照说明书进行。

**2.6 统计学方法** 使用 R studio(version 4.1.3)建立模型,使用 R 软件的“glmnet”包进行最小绝对收缩和选择算子(LASSO)回归,使用 R 软件的“caret”包与“glm”包及“e1071”包建立 KNN 与多因素 logistic 回归模型及高斯朴素贝叶斯模型。首先使用 Kolmogorov-Smirnow 对数据进行正态性检验,符合正态分布的计量资料采用( $\bar{x} \pm s$ )表示,进行独立样本 *t* 检验,非正态分布的计量资料采用 [ $M(P_{25}, P_{75})$ ] 表示,并使用 Mann-whitney 检验进行组间比较。使用卡方检验比较肌肉减少症组和非肌肉减少症组患者基线特征。采用单因素分析对变量进行初步筛选,使用 LASSO 回归模型对变量进行进一步筛选,将最终筛选出得变量纳入 K 最近邻算法模型(KNN)、多因素 logistic 回归模型及高斯朴素贝叶斯模型构建模型。最终绘制 ROC 曲线及校准曲线及临床预测曲线(DCA),采用 Bootstrap 随机重复抽样 1 000 次进行模型内部验证。

### 3 结果

**3.1 肺癌患者并发肌肉减少症的单因素分析** 单因素分析中,性别、年龄、糖尿病、肿瘤类型、IL-2、CRP、球蛋白在非肌肉减少症组和肌肉减少症组之间没有统计学意义,其他指标在非肌肉减少症组和肌肉减少症组之间均有统计学意义。

表 1 肺癌患者并发肌肉减少症的单因素分析( $\bar{x} \pm s$ ), [ $M(P_{25}, P_{75})$ ]

Table 1 Univariate analysis of sarcopenia in lung cancer patients( $\bar{x} \pm s$ ), [ $M(P_{25}, P_{75})$ ]

项目	非肌肉减少症组 (126)	肌肉减少症组(137)	<i>t</i> / $\chi^2/Z$	<i>P</i>
男性[n(%)]	56(44.8)	69(55.2)	0.923	0.337
女性[n(%)]	70(50.7)	68(49.3)		
年龄( $\bar{x} \pm s$ , 岁)	65.4 $\pm$ 12.7	66.2 $\pm$ 11.9	-0.521	0.603
BMI[ $M(P_{25}, P_{75})$ , kg/m <sup>2</sup> ]	23.8(22.6, 27.6)	22.8(20.5, 24.4)	-5.968	<0.001
吸烟[n(%)]	62(40.0)	93(60.0)	9.460	0.002
不吸烟[n(%)]	64(59.3)	44(40.7)		
饮酒[n(%)]	25(36.8)	43(63.2)	4.564	0.033
不饮酒[n(%)]	101(51.8)	94(48.2)		
糖尿病[n(%)]	10(37.0)	17(63.0)	1.425	0.233
无糖尿病[n(%)]	116(49.2)	120(50.8)		

(续表)

项目	非肌肉减少症组 (126)	肌肉减少症组(137)	$\sqrt{\chi^2/Z}$	P
高血压[n(%)]	34(59.6)	23(40.4)	4.019	0.045
无高血压[n(%)]	92(44.7)	114(55.3)		
肿瘤类型[n(%)]				
腺癌	40(48.8)	42(51.2)	0.095	0.954
鳞癌	41(46.6)	47(53.4)		
其他	45(48.4)	48(51.6)		
PSQI 评分[M(P <sub>25</sub> ,P <sub>75</sub> ),分]	4.5(2.0,8.0)	5.0(3.0,11.0)	-2.183	0.029
IGF-1[M(P <sub>25</sub> ,P <sub>75</sub> ),ng/ml]	86.37(81.46,91.26)	77.36(70.96,85.01)	-6.107	<0.001
GDF-8[M(P <sub>25</sub> ,P <sub>75</sub> ),ng/ml]	15.39(12.78,17.62)	16.58(14.75,19.55)	-3.760	<0.001
IL-2[M(P <sub>25</sub> ,P <sub>75</sub> ),pmol/L]	2.28(1.83,2.28)	2.23(1.86,2.75)	-1.612	0.107
IL-4[M(P <sub>25</sub> ,P <sub>75</sub> ),pmol/L]	2.28(0.96,3.35)	3.48(2.79,4.99)	-6.428	<0.001
IL-6[M(P <sub>25</sub> ,P <sub>75</sub> ),pmol/L]	13.25(6.47,18.77)	25.5(7.74,68.01)	-4.914	<0.001
IL-17[M(P <sub>25</sub> ,P <sub>75</sub> ),pmol/L]	3.36(2.77,4.65)	3.65(3.10,5.13)	-2.489	0.013
TNF- $\alpha$ [M(P <sub>25</sub> ,P <sub>75</sub> ),pmol/L]	3.00(1.94,4.13)	4.13(2.18,22.24)	-3.369	0.001
CRP[M(P <sub>25</sub> ,P <sub>75</sub> ),mg/L]	19.60(10.39,30.79)	20.95(11.00,38.77)	-0.928	0.353
血小板计数[M(P <sub>25</sub> ,P <sub>75</sub> ), $\times 10^9/L$ ]	243.60(197,279)	219.20(179,247)	-3.118	0.002
直接胆红素[M(P <sub>25</sub> ,P <sub>75</sub> ), $\mu\text{mol/L}$ ]	3.38(2.62,3.83)	2.56(2.12,2.89)	-6.240	<0.001
间接胆红素[M(P <sub>25</sub> ,P <sub>75</sub> ), $\mu\text{mol/L}$ ]	11.91(10.60,12.99)	9.96(8.23,10.57)	-6.090	<0.001
总胆红素[M(P <sub>25</sub> ,P <sub>75</sub> ), $\mu\text{mol/L}$ ]	15.68(14.18,16.89)	13.05(11.04,14.18)	-6.519	<0.001
白蛋白[M(P <sub>25</sub> ,P <sub>75</sub> ),g/L]	42.09(40.40,44.23)	41.07(40.00,42.15)	-3.964	<0.001
球蛋白[M(P <sub>25</sub> ,P <sub>75</sub> ),g/L]	27.88(26.00,29.25)	28.13(26.95,28.95)	-1.529	0.126
肌酐[M(P <sub>25</sub> ,P <sub>75</sub> ), $\mu\text{mol/L}$ ]	62.86(55.45,68.85)	77.41(66.81,84.55)	-7.857	<0.001
尿素氮[M(P <sub>25</sub> ,P <sub>75</sub> ),mmol/L]	4.41(3.54,4.79)	5.58(4.66,6.43)	-7.891	<0.001
高密度脂蛋白[M(P <sub>25</sub> ,P <sub>75</sub> ),mmol/L]	1.23(1.03,1.46)	1.13(0.97,1.33)	-2.233	0.026
低密度脂蛋白[M(P <sub>25</sub> ,P <sub>75</sub> ),mmol/L]	2.66(2.22,3.09)	2.88(2.28,3.56)	-1.992	0.046
碱性磷酸酶[M(P <sub>25</sub> ,P <sub>75</sub> ),U/L]	66.15(52.30,82.13)	73.90(64.40,91.70)	-3.514	<0.001
癌胚抗原(CEA)[M(P <sub>25</sub> ,P <sub>75</sub> ),U/ml]	2.51(1.45,2.89)	3.15(2.42,4.53)	-4.146	<0.001

**3.2 肺癌患者肌肉减少症的 Lasso 回归分析** 以肺癌患者是否发生肌肉减少症为因变量,共纳入 11 个自变量。肺癌患者肌肉减少症的超参数  $\ln\lambda$  与均方误差 (Mean Squared Error) 的关系见图 1。由图 1 可知,第二条虚线对应的横轴中  $\ln\lambda \approx -3.25$ ,则  $\lambda \approx$

0.038 5,最终筛选出 IGF-1、GDF-8、IL-6、TNF- $\alpha$ 、PSQI、BMI、白蛋白、吸烟与低密度脂蛋白这 9 个预测肺癌患者肌肉减少症的自变量。

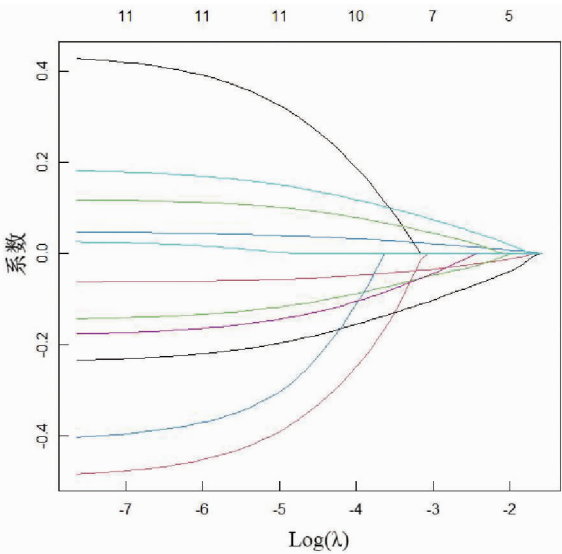


图 1 纳入 LASSO 回归的 11 个临床特征的系数曲线

Fig. 1 Coefficient curves of 11 clinical characteristics included in the LASSO regression

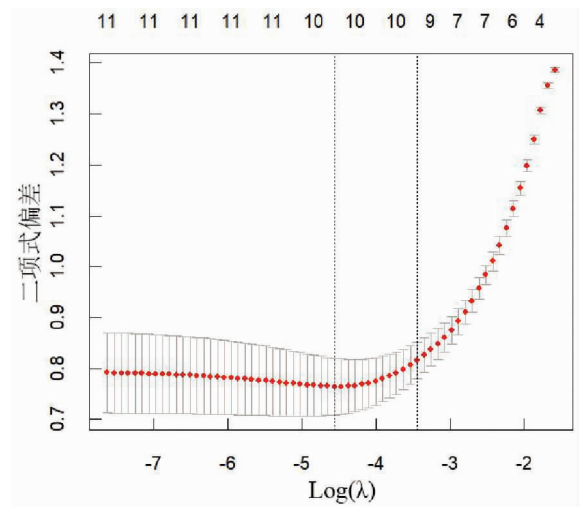


图 2 经 LASSO 回归、十折交叉验证法选择最合适的临床特征

Fig. 2 Selection of the most appropriate clinical characteristics through LASSO regression and ten-fold cross-validation method

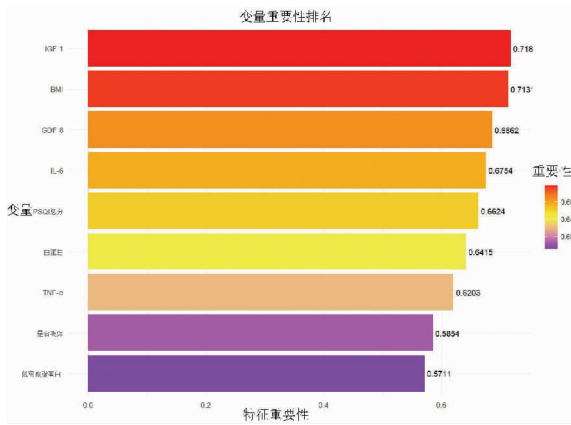


图 3 KNN 模型特征重要性图

Fig. 3 Feature importance graph of the KNN model

**3.3 肺癌患者并发肌肉减少症的危险因素特征重要性图** KNN 模型中导致肌肉减少症患病风险增加的影响因素的重要性排序依次是 IGF - 1、BMI、GDF - 8、IL - 6、PSQI 总分、白蛋白、TNF - α、是否吸烟与低密度脂蛋白见图 3, 高斯朴素贝叶斯模型中导致肌肉减少症患病风险增加的影响因素的重要性排序依次是 IL - 6、TNF - α、IGF - 1、PSQI 总分、GDF - 8、BMI、白蛋白、低密度脂蛋白与是否吸烟见图 4。

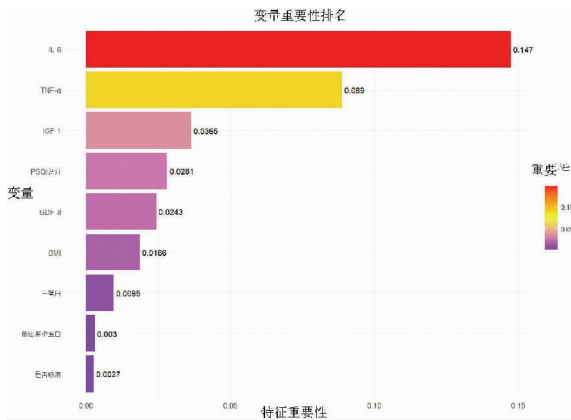


图 4 GNB 模型特征重要性图

Fig. 4 Feature importance graph of the GNB model

**3.4 KNN、多因素 logistic 回归和 GNB 模型对肺癌患者肌肉减少症的预测能力** 通过 C - index ROC 曲线对模型的区分度进行评价。ROC 曲线显示:KNN、多因素 logistic 回归和 GNB 模型的 ROC 曲线见图 5, KNN 模型曲线下面积 (AUC) = 0.936, 多因素 logistic 回归模型曲线下面积 (AUC) = 0.926, GNB 模型曲线下面积 (AUC) = 0.935, 区分度高, 模型预测性能良好; 三者均表明模型具有良好的预测价值。

**3.5 肺癌患者并发肌肉减少症的危险因素校准曲线图** 为了评价模型的准确度, 采用 Bootstrap 法对模型

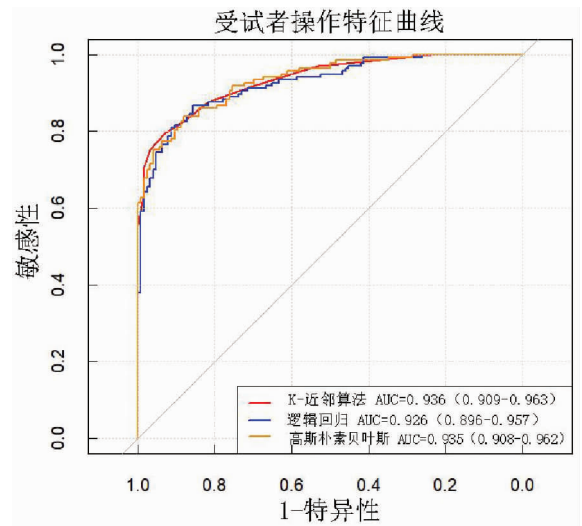


图 5 KNN、多因素 logistic 回归和 GNB 模型的 ROC 曲线  
Fig. 5 ROC curves of the KNN, multivariate logistic regression, and GNB models

进行内部验证和外部验证, 自抽样次数 B = 1 000, 并绘制校准曲线见图 6。验证结果显示, KNN、多因素 logistic 回归和 GNB 模型校正曲线均靠近理想的 45° 参考线, 表明模型预测值与实际值之间具有良好的一致性。GNB 模型预测的概率与实际概率之间的平均绝对差异约为 1.88%; KNN 模型的平均绝对误差是 2.69%; 逻辑回归模型的平均绝对误差是 2.87%。说明 KNN 模型与 GNB 模型的预测概率更接近实际概率。

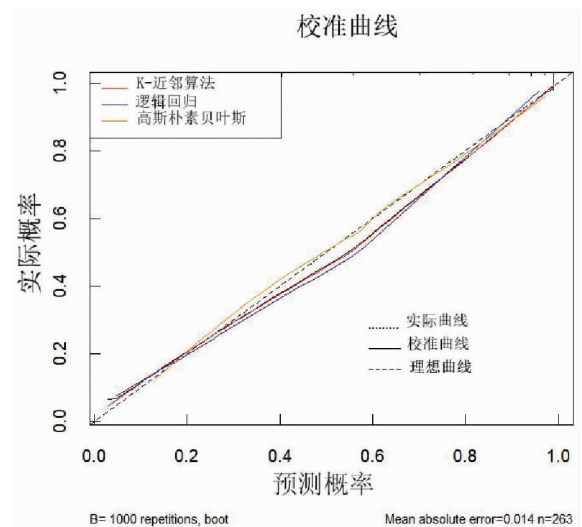


图 6 校准曲线图

Fig. 6 Calibration curve diagram

**3.6 模型的 DCA 曲线分析** 本研究通过 DCA 评估其在临床应用中该模型是否有获益。结果显示, 与多因素 logistic 回归模型相比, KNN 模型与 GNB 模型在

预测肺癌患者肌肉减少症的能力方面显示出更大的净收益(图7)。

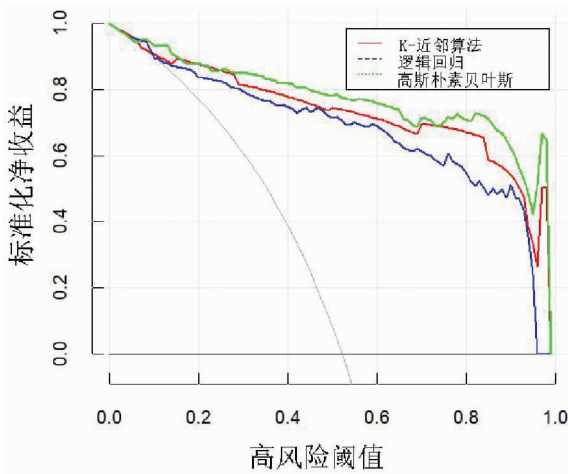


图7 KNN与多因素 logistic 回归模型 DCA 曲线分析

Fig.7 DCA curve analysis of the KNN and multivariate logistic regression models

## 4 讨论

本研究中选择了 KNN(k-近邻)机器学习模型来构建肺癌患者肌肉减少症的预测模型。根据 KNN 算法的重要性特征图,其中 IGF-1(胰岛素样生长因子-1)和 GDF-8(生长分化因子-8)是最重要的特征,分别达到 0.8718 和 0.686 的高重要性值,随后是 BMI(0.713)、IL-6(0.6754)、PSQI 评分(0.6624)、白蛋白(0.6415)、TNF- $\alpha$ (0.6203)、而吸烟(0.5854),以及低密度脂蛋白(0.5711)的重要性相对较低。这些结果表明,IGF-1(胰岛素样生长因子-1)和 GDF-8(生长分化因子-8)是影响肺癌患者合并肌肉减少症的重要因素。通过综合 IGF-1、GDF-8、IL-6 和 TNF- $\alpha$  等多个变量,KNN 模型在预测准确性方面表现优异,其 AUC 值(曲线下面积)达到 0.936。

有研究显示,肺癌患者因放化疗、营养不良等多种因素,容易发生肌肉减少症<sup>[11]</sup>,放化疗在杀死癌细胞的同时也会损伤正常的肌肉细胞,导致肌肉蛋白质的合成减少和分解增加<sup>[12]</sup>。不仅如此,化疗药物如顺铂通过抑制 mTOR 通路减少蛋白质合成,同时激活泛素-蛋白酶体系统,导致蛋白质的加速分解,最终引发肌肉萎缩与消耗<sup>[13]</sup>。同样,紫杉醇也会引起线粒体功能障碍和氧化应激,进一步加剧肌肉的损耗<sup>[14]</sup>。这些机制共同作用,导致癌症患者常见的肌肉萎缩和体重下降,严重影响患者的生活质量与治疗效果。此外,放射治疗也会通过产生大量自由基,导

致肌肉细胞的氧化应激和炎症反应,进一步加剧肌肉的损伤和衰退<sup>[15]</sup>。营养不良是另一重要因素。肺癌患者常常因恶心、呕吐和食欲不振等症状导致营养摄入不足,特别是蛋白质和必需氨基酸的摄入不足,这些营养物质是肌肉合成的基础,缺乏它们会直接影响肌肉的合成能力<sup>[16]</sup>。研究表明,癌症相关的炎症状态也会通过增加食欲抑制激素的水平和减少食欲促进激素的水平,进一步加剧营养不良的情况<sup>[17]</sup>。

肌肉减少症的发生机制复杂,涉及多种细胞因子和信号通路的相互作用。胰岛素样生长因子-1(IGF-1)和生长分化因子-8(GDF-8,亦称肌肉生长抑制素)是其中两个关键的调节因子。IGF-1 主要通过激活 PI3K/Akt 和 MAPK 信号通路,促进肌细胞的增殖、分化和存活,从而维持或增加肌肉质量<sup>[18]</sup>,IGF-1 与其受体结合后,激活下游的 Akt 信号通路,进而激活 mTOR(哺乳动物雷帕霉素靶蛋白),促进蛋白质合成并抑制蛋白质降解<sup>[19]</sup>。

Tanaka 等人的研究指出,IGF-1 水平低的患者患肌肉减少症的风险显著增加,IGF-1 在维持骨骼肌稳态中发挥核心作用,低 IGF-1 水平与 NASH 相关性肌少症密切相关,联合使用 IGF-1 和 ARB 可增强其抗肌少症效果<sup>[20]</sup>。Delagrang 等人的研究也表明,IGF-1 水平与肌肉质量和功能呈正相关,低 IGF-1 水平是肌肉减少症的一个重要预测因子<sup>[21]</sup>。GDF-8 通过与其受体结合,激活 Smad2 和 Smad3 信号通路,抑制肌细胞的增殖和分化,从而限制肌肉的生长<sup>[22]</sup>。GDF-8 的表达在癌症患者中常常升高,并与肌肉减少症的严重程度呈正相关。Kawaguchi 等人的研究表明,研究发现,肺癌组织中 GDF-8 表达升高与肌肉质量下降、预后不良密切相关<sup>[23]</sup>。GDF-8 可能通过诱导肌肉减少症和增强免疫抑制性微环境(TAM 浸润)共同促进肺癌进展<sup>[23]</sup>。Kumar 等人的研究进一步证实了这一点,发现 GDF-8 通过诱导炎症和脂肪代谢紊乱,显著加剧了肌肉减少症的发生<sup>[24]</sup>,与本研究结果一致。

此外,IL-6 和肿瘤坏死因子 TNF- $\alpha$  也是影响肌肉减少症的重要因子。IL-6 通过 JAK/STAT 信号通路促进炎症反应,增加肌肉蛋白质的分解<sup>[25]</sup>。TNF- $\alpha$  则通过激活 NF- $\kappa$ B 信号通路,诱导肌肉细胞凋亡,并抑制蛋白质合成<sup>[25]</sup>。研究表明,IL-6 表达上调可激活 JAK/STAT3 通路,促进肌肉特异性泛素化酶 MuRF1 与 MAFbx 的表达,并增强线粒体自噬活性,加剧骨骼肌萎缩,而 TNF- $\alpha$  水平升高的患者患肌肉减少症的风险是非肌肉减少症患者的 1.553 倍(95% CI:1.253~1.924)<sup>[26-27]</sup>。

除了生物学机制,肺癌患者的生活方式和健康状

况也会影响肌肉减少症的发生率。睡眠质量差、体重较低等因素都会加重患者的肌肉减少症。研究发现,睡眠质量差的患者更容易患上肌肉减少症,是因为睡眠不足会导致体内应激激素如皮质醇水平的升高,进而促进肌肉蛋白质的分解<sup>[28]</sup>。Rubio - Arias 等人研究表明,短睡眠时间(<6 小时)与可能肌少症的发生风险增加有关。睡眠不足可能通过影响激素水平和炎症反应,增加肌少症的风险<sup>[29]</sup>。同样,体重较低的患者由于脂肪和肌肉储备不足,在疾病和治疗的双重打击下,更容易发生肌肉减少症<sup>[30]</sup>。Chalerm Sri 等人的研究支持了这一发现,指出体重较低的患者肌肉减少症的发生率显著增加,OR 值为 1.95,95% CI 为 1.55 ~ 2.45<sup>[31]</sup>。吸烟不仅通过炎症和氧化应激影响肌肉健康,还可能通过影响肺功能和整体身体活动水平,间接加重肌肉减少症的风险(OR = 2.51,95% CI: 1.26 ~ 5.01)<sup>[32]</sup>。本研究中,吸烟的肺癌患者发生肌肉减少症的风险是不吸烟患者的 2.20 倍,与本研究结果一致。研究表明,LDL - C 通过其对全身炎症和代谢路径的影响,可间接加重癌症患者的肌肉减少症,非高密度脂蛋白胆固醇与高密度脂蛋白胆固醇的比值(NHHR)升高与癌症患者肌肉减少症的发生显著相关,特别是在女性患者中更为明显<sup>[33]</sup>。研究表明,低水平的白蛋白与肌肉减少症显著相关,白蛋白水平低通常反映出患者的营养状况差,可能导致肌肉蛋白的分解增加,肌肉合成减少,从而引发或加重肌肉减少症<sup>[34]</sup>。一项研究表明,低白蛋白水平与肺癌(NSCLC)患者的整体生存率显著相关,每单位增加的白蛋白 - 球蛋白比(AGR)与死亡风险降低相关(HR = 0.14,95% CI: 0.014 ~ 0.883, P = 0.038)<sup>[35]</sup>,这与本研究结果一致。

本研究在多个方面体现了其独特性。首先,研究综合考虑了 IGF - 1、GDF - 8、IL - 6 和 TNF -  $\alpha$  等多个生物标志物,这些因子在肌肉减少症的发生机制中具有重要作用。其次,本研究通过应用先进的机器学习算法,结合交叉验证技术,显著提高了模型的预测精度和稳定性。与 Gellhaus 等人的研究相比,本研究不仅再次验证了 IGF - 1 水平降低与肌肉减少症风险增加的相关性,还进一步将 GDF - 8、IL - 6 和 TNF -  $\alpha$  等多项生物标志物纳入分析,构建了更为全面的预测模型<sup>[36]</sup>。此外,研究还表明,结合多个生物标志物的综合模型在预测肺癌患者的肌肉减少症风险时表现最佳,这为临床干预提供了更具针对性的依据。

然而,本研究也存在一些局限性。样本量相对较小,未来需要扩大样本量以提高模型的泛化能力。虽然本研究综合了多种临床特征和生物标志物,但仍有许多潜在的影响因素未被纳入,例如患者的营养状

态、体力活动水平等。在本研究中,采用基于随机森林的方法对缺失自变量进行插补。该方法较均值或中位数插补更能保留变量间的非线性关系,提升模型预测性能。未来的研究需要在以上方面进行改进,以构建更为全面和准确的预测模型。

综上所述,基于 IGF - 1 与 GDF - 8 建立的原发性肺癌患者合并肌肉减少症的 KNN 机器学习模型对该病的区分度与一致性良好,对于评估原发性肺癌患者的健康状态和疾病预后具有重要参考价值。

**利益冲突声明** 本研究不存在任何利益冲突

#### 参考文献

- [1] 汪成洋,周飞,毕清,等. 肌肉减少症与肺癌预后的研究进展[J]. 临床肺科杂志,2024,29(6):952-956.  
Wang CY, Zhou F, Bi Q, et al. Advances in the study of prognosis of sarcopenia and lung cancer[J]. Journal of Clinical Pulmonary Medicine, 2024, 29(6): 952-956. (In Chinese)
- [2] Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022; GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA - A Cancer Journal for Clinicians, 2024, 74(3): 229-263.
- [3] Almeida NS, Rocha R, de Souza CA, et al. Energy and nutrient intake by people with and without sarcopenia diagnosed by the European working group on sarcopenia in older people: A systematic review and meta-analysis[J]. Nutrition Review, 2024, 82(12): 1666-1677.
- [4] Monaco CMF, Perry CGR, Hawke TJ. Alterations in mitochondrial functions and morphology in muscle and non-muscle tissues in type 1 diabetes: implications for metabolic health[J]. Experimental Physiology, 2020, 105(4): 565-570.
- [5] Sakai H, Asami M, Naito H, et al. Exogenous insulin-like growth factor 1 attenuates cisplatin-induced muscle atrophy in mice[J]. Journal of Cachexia, Sarcopenia and Muscle, 2021, 12(6): 1570-1581.
- [6] Smith RC, Cramer MS, Mitchell PJ, et al. Inhibition of myostatin prevents microgravity-induced loss of skeletal muscle mass and strength[J]. PLOS One, 2020, 15(4): e0230818.
- [7] Cho MR, Lee S, Song SK. A review of sarcopenia pathophysiology, diagnosis, treatment and future direction[J]. Journal of Korean Medical Science, 2022, 37(18): e146.
- [8] Widajanti N, Soelistijo S, Hadi U, et al. Association between sarcopenia and insulin-like growth factor - 1, myostatin, and insulin resistance in elderly patients undergoing hemodialysis[J]. Journal of Aging Research, 2022, 2022: 1327332.
- [9] Buysse DJ, Reynolds CF3, Monk TH, et al. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research[J]. Psychiatry Research, 1989, 28(2): 193-213.
- [10] 上海市抗癌协会肿瘤营养与支持治疗专业委员会,余震. 肌肉减少症规范化筛查、评估、诊断、治疗和监测技术标准[J]. 肿瘤代谢与营养电子杂志,2024,11(6):782-789.  
Shanghai Anti-cancer Association Tumor Nutrition and Supportive Therapy Professional Committee, Yu Z. Establishment of standardized screening, diagnosis and treatment criteria for sarcopenia and expert recommendations[J]. Electronic Journal of Metabolism and Nutrition of Cancer, 2024, 11(6): 782-789. (In

- Chinese)
- [11] 魏芯芯,薛伟,李利廷.肺癌患者合并肌肉减少症对预后的影响[J].肿瘤代谢与营养电子杂志,2023,10(5):679-683.  
Wei XX, Xue W, Li LT. The influence of sarcopenia on the prognosis of lung cancer patients [J]. Electronic Journal of Metabolism and Nutrition of Cancer, 2023, 10(5): 679-683. (In Chinese)
- [12] Collao N, D'souza D, Messeiller L, et al. Radiation induces long-term muscle fibrosis and promotes a fibrotic phenotype in fibro-adipogenic progenitors [J]. Journal of Cachexia, Sarcopenia and Muscle, 2023, 14(5): 2335-2349.
- [13] Conte E, Bresciani E, Rizzi L, et al. Cisplatin-Induced skeletal muscle dysfunction: mechanisms and counteracting therapeutic strategies [J]. International Journal of Molecular Sciences, 2020, 21(4): 1242.
- [14] Hain BA, Waning DL. Bone-Muscle crosstalk: musculoskeletal complications of chemotherapy [J]. Current Osteoporosis Reports, 2022, 20(6): 433-441.
- [15] D'souza D, Roubos S, Larkin J, et al. The late effects of radiation therapy on skeletal muscle morphology and progenitor cell content are influenced by diet-induced obesity and exercise training in male mice [J]. Scientific Reports, 2019, 9(1): 6691.
- [16] Ahmad SS, Ahmad K, Lee EJ, et al. Implications of Insulin-Like growth factor-1 in skeletal muscle and various diseases [J]. Cells, 2020, 9(8): 1773.
- [17] 王艳芳,江波.炎症指数/营养评分及肌肉减少症与肺癌免疫治疗预后相关性的研究进展 [J]. 临床肿瘤学杂志,2024,29(10): 1028-1032.  
Wang YF, Jiang B. Research progress of correlation between inflammation index/nutrition score, sarcopenia and prognosis of lung cancer by immunotherapy [J]. Chinese Clinical Oncology, 2024, 29(10): 1028-1032. (In Chinese)
- [18] Yoshida TDH, Delafontaine P. Mechanisms of IGF-1-Mediated regulation of skeletal muscle hypertrophy and atrophy [J]. Cells, 2020, 9(9): 1970.
- [19] Schiaffino S, Reggiani C, Akimoto T, et al. Molecular mechanisms of skeletal muscle hypertrophy [J]. Journal of Neuromuscular Diseases, 2021, 8(2): 169-183.
- [20] Tanaka M, Kaji K, Nishimura N, et al. Blockade of angiotensin II modulates insulin-like growth factor 1-mediated skeletal muscle homeostasis in experimental steatohepatitis [J]. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 2024, 1871(2): 119649.
- [21] Delagrangé M, Rousseau V, Cessans C, et al. Low bone mass in Noonan syndrome children correlates with decreased muscle mass and low IGF-1 levels [J]. Bone, 2021, 153: 116170.
- [22] Cawthon PM, Patel S, Newman AB, et al. Evaluation of associations of growth differentiation factor-11, growth differentiation factor-8, and their binding proteins, follistatin and follistatin-like protein-3, with measures of skeletal muscle mass, muscle strength, and physical function in older adults [J]. The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, 2023, 78(11): 2051-2059.
- [23] Kawaguchi Y, Watanabe A, Shiratori T, et al. Myostatin expression in lung cancer induces sarcopenia and promotes cancer progression [J]. General Thoracic and Cardiovascular Surgery, 2024, 72(4): 232-239.
- [24] Kumar A, Prajapati P, Singh G, et al. Salbutamol attenuates diabetic skeletal muscle atrophy by reducing oxidative stress, myostatin/gdf-8, and pro-inflammatory cytokines in rats [J]. Pharmaceutics, 2023, 15(8): 2101.
- [25] Hirata Y, Nomura K, Kato D, et al. A piezo1/klf15/il-6 axis mediates immobilization-induced muscle atrophy [J]. Journal of Clinical Investigation, 2022, 132(10): 1-13.
- [26] Demir L, Oflazoğlu U. The relationship between sarcopenia and serum irisin and TNF- $\alpha$  levels in newly diagnosed cancer patients [J]. Supportive Care in Cancer, 2023, 31(10): 586.
- [27] Huang ZW, Zhong L, Zhu JW, et al. Inhibition of IL-6/JAK/STAT3 pathway rescues denervation-induced skeletal muscle atrophy [J]. Annals of Translational Medicine, 2020, 8(24): 1681.
- [28] Chen LF, Li QY, Huang XY, et al. Association between sleep duration and possible sarcopenia in middle-aged and elderly Chinese individuals: evidence from the China health and retirement longitudinal study [J]. BMC Geriatrics, 2024, 24(1): 594.
- [29] Rubio-Arias JÁ, Rodríguez-Fernández R, Andreu L, et al. Effect of sleep quality on the prevalence of sarcopenia in older adults: a systematic review with meta-analysis [J]. Journal of Clinical Medicine, 2019, 8(12): 2156.
- [30] Guven DC, Aksun MS, Cakir IY, et al. The association of BMI and sarcopenia with survival in patients with glioblastoma multiforme [J]. Future Oncology, 2021, 17(32): 4405-4413.
- [31] Chalerm Sri C, Aekplakom W, Srinonprasert V. Body mass index combined with possible sarcopenia status is better than BMI or possible sarcopenia status alone for predicting All-Cause mortality among Asian Community-Dwelling older adults [J]. Frontiers in Nutrition, 2022, 9: 881121.
- [32] Sha TT, Zhang YQ, Joshi AD, et al. Smoking, alcohol and risk of sarcopenia: a Mendelian randomisation study [J]. BMJ Open, 2025, 15(4): e091656.
- [33] He R, Ye YJ, Zhu QL, et al. Association between non-high-density lipoprotein cholesterol to high-density lipoprotein cholesterol ratio and sarcopenia in individuals with cancer: a cross-sectional study [J]. Lipids in Health and Disease, 2024, 23(1): 217.
- [34] Zhang C, Zhang L, Zeng L, et al. Associations of serum albumin and dietary protein intake with all-cause mortality in community-dwelling older adults at risk of sarcopenia [J]. Heliyon, 2024, 10(8): e29734.
- [35] Guo XL, Shao JJ, Zhai BQ, et al. Relationship and prognostic significance between preoperative serum albumin to globulin ratio and CT features of non-small cell lung cancer [J]. European Journal of Radiology, 2020, 128: 109039.
- [36] Gellhaus B, Böker KO, Schilling AF, et al. Therapeutic Consequences of targeting the IGF-1/PI3K/AKT/FOXO3 axis in sarcopenia: a narrative review [J]. Cells (Basel, Switzerland), 2023, 12(24): 2787.