

基于随机森林算法的老年肺癌根治术后复发和转移的预测模型构建与验证*

王娟 刘勇世 倪云峰 滕鑫 管利华 赵阿红 刘静 俞俊茹**

空军军医大学第二附属医院胸腔外科, 西安 710038

[摘要] **目的** 探讨随机森林算法在老年肺癌根治术后复发转移预测模型构建中的应用价值。**方法** 选取 2014 年 1 月—2016 年 1 月空军军医大学第二附属医院收治的 150 例老年肺癌根治术患者作为建模组, 将建模组相关因素的单因素分析结果中 $P < 0.05$ 的候选变量进行随机森林算法建模, 筛选术后复发和转移的相关因素, 根据重要性对影响因素排序, 另选取 2016 年 2 月—2017 年 2 月本院收治的 70 例老年肺癌根治术患者作为验证组, 对随机森林图模型进行验证。**结果** 150 例老年肺癌根治术患者随访期间 44 例 (29.33%) 复发, 51 例 (34.00%) 转移, 20 例 (13.33%) 同时出现复发和转移。术后复发变量重要性排序依次为纵隔淋巴结转移、淋巴结清扫个数、T 分期、N 分期、分化程度、病理分期、年龄、病灶直径; 术后转移变量重要性排序依次为纵隔淋巴结转移、N 分期、淋巴结转移个数、淋巴结清扫个数、T 分期、分化程度、术后放疗、病理分期、年龄。采用随机森林模型选取特征变量, 随机森林模型中纵隔淋巴结转移、淋巴结清扫个数、T 分期、N 分期指标所建模型对术后复发的预测效能为 0.904, 纵隔淋巴结转移、N 分期、淋巴结转移个数、淋巴结清扫个数、T 分期指标所建模型对术后转移的预测效能为 0.897; 外部验证随机森林模型预测老年肺癌术后复发效能为 0.905, 术后转移效能为 0.910, 与内部验证基本一致。**结论** 随机森林预测模型的构建可在一定程度上预测老年肺癌根治术后复发及转移的风险, 为手术疗效、术后辅助治疗等提供临床参考。

[关键词] 肺癌根治术; 复发; 转移; 随机森林算法; 淋巴结转移

doi: 10.3969/j.issn.1674-7593.2024.06.004

Construction and Validation of A Prediction Model for Recurrence and Metastasis of Elderly Lung Cancer Patients after Radical Surgery Based on Random Forest Algorithm

Wang Juan, Liu Yongshi, Ni Yunfeng, Teng Xin, Jian Lihua, Zhao Ahong, Liu Jing, Yun Junru**

Department of thoracic surgery, the Second Affiliated Hospital of Air Force Medical University, Xi'an 710038, China

** Corresponding author: Yun Junru, email: 459392608@qq.com

[Abstract] **Objective** To explore the application value of random forest algorithm in the construction of recurrence and metastasis prediction model for elderly lung cancer patients after radical surgery. **Methods** A total of 150 elderly lung cancer patients who underwent radical surgery at the Second Affiliated Hospital of Air Force Medical University from January 2014 to January 2016 were selected as the modeling group. Random forest algorithm modeling was conducted on the candidate variables with $P < 0.05$ in the univariate analysis results of the relevant factors in the modeling group to screen the factors related to postoperative recurrence and metastasis. The factors were ranked according to their importance. Another 70 elderly lung cancer patients who underwent radical surgery for lung cancer from February 2016 to February 2017 were selected as the validation group to validate the random forest graph model. **Results** During the follow-up period, 44 cases (29.33%) elderly patients with lung cancer undergoing radical surgery recurrence, 51 cases (34.00%) metastasis, and 20 cases (13.33%) had both recurrence and metastasis. Rank of importance of postoperative recurrence variables: mediastinal lymph node metastasis, number of lymph node dissection, T stage, N stage, degree of differentiation, pathological stage, age, lesion diameter. Rank of importance of postoperative metastasis variables: mediastinal lymph node metastasis, N stage, number of lymph node metastasis, number of lymph node dissection, T stage, degree of differentiation, postoperative chemotherapy, pathological stage, age. The random forest model was used to select the characteristic variables. In the random forest model, the model built by the mediastinal lymph node metastasis, the number of lymph nodes dissected, the T stage and the N stage indexes had an efficacy of 0.904 in predicting postoperative recurrence. The predictive efficacy of the mediastinal lymph node metastasis, N stage, number of lymph node metastasis, number of lymph node dissection and T stage was 0.897. The externally validated ran-

* 空军军医大学第二附属医院院级课题 (TDHLKY-2021-04)

** 通讯作者: 俞俊茹, 电子邮箱 459392608@qq.com

dom forest model predicted a postoperative recurrence efficacy of 0.905 and a postoperative metastasis efficacy of 0.910 for elderly lung cancer, which was in general agreement with the internal validation. **Conclusion** In this study, the construction of the random forest prediction model can predict the risk of recurrence and metastasis after radical resection of lung cancer in the elderly to a certain extent. It can provide clinical reference for surgical effect and postoperative adjuvant therapy.

[**Key words**] Radical surgery of lung cancer; Recurrence; Metastasis; Random forest algorithm; Lymph node metastasis

肺癌发病率、病死率均居所有恶性肿瘤前列,据统计其五年的相对存活率仅18%,根治性手术切除是临床治疗肺癌的最佳治疗选择,可彻底清除肿瘤病灶,减少癌细胞增殖、转移^[1-3]。但根治性切除后仍有30%~55%的患者出现复发、转移,导致治疗失败;且不同术后病理分期患者,五年无复发生存率存在差异^[4-6]。目前胸外科对肺癌根治术后复发转移情况十分关注,已有多项研究对术后复发转移的影响因素展开了研究。传统资料分析难以避免变量间的相互作用,随机森林算法对数据集无要求,不存在共线性、过拟合等问题,通过集成学习思想将多棵决策树进行集成,可对各特征变量在分类问题中的重要性进行评估,在多种疾病风险评估中发挥重要作用^[7-8]。本研究拟探讨随机森林算法在老年肺癌根治术后复发和转移预测模型构建中的应用价值,旨在为临床工作提供参考依据。

1 对象与方法

1.1 研究对象

选取2014年1月—2016年1月空军军医大学第二附属医院收治的150例老年肺癌根治术患者作为建模组,另选取2016年2月—2017年2月同院收治的70例行肺癌根治术的老年肺癌患者作为验证组。纳入标准:①初治原发性肺癌;②年龄 ≥ 60 岁;③经病理学检查确诊为单发病灶;④肺癌根治术成功完成,手术切缘阴性;⑤术前未接受靶向治疗、放化疗等新辅助治疗;⑥临床资料、随访结果完整可靠。排除标准:①存在其他部位恶性肿瘤;②伴有自身免疫性疾病;③伴有严重代谢性疾病;④手术死亡或围术期出现严重并发症;⑤存在肝、心、肾等严重脏器病变;⑥随访期间失访;⑦非肿瘤原因死亡;⑧存在精神疾病。本研究符合《赫尔辛基宣言》要求,经本院伦理委员会批准,患者家属自愿签署知情同意书。

1.2 方法

通过病历收集患者的性别、年龄、吸烟史、病灶部位、病灶直径、病理类型、分化程度、病理分期、T分期、N分期、纵隔淋巴结转移、淋巴结转移个数、淋巴结清扫个数、手术类型、术后并发症及术后放化疗等资料。所有肺癌患者出院后均定期进行门诊复诊或电话随访,术后1年内每3个月随访1次,第2年每6个月随访1次,之后每12个月随访1次,随访截至2021年1月,以复发或转移为随访终点,记录患者有无复发或转移出现及出现时间。复发:肿瘤局限于一侧原病灶

所在胸腔,包括同侧的肺、手术残留、胸膜与纵隔淋巴结。转移:病灶出现在远处器官或原肿瘤所在的对侧胸腔,包括对侧的肺、肝、骨、脑、锁骨上淋巴结、颈部、肾上腺等。

1.3 统计学方法

采用SPSS23.0统计学软件进行数据分析。方差齐性的正态分布计量资料以 $\bar{x} \pm s$ 表示,组间比较采用 t 检验;计数资料比较采用 χ^2 检验;采用R语言随机森林算法(Bootstrap方法)对影响术后复发转移的特征变量进行重要性排序和筛选,并采用受试者工作特征曲线(ROC)评估所构建模型的预测价值; $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 随访情况

150例老年患者均进行定期随访,随访时间10~60个月,中位随访时间为22.8个月。随访期间44例复发(29.33%),其中21例为同侧肺部复发,15例为纵隔淋巴结复发,3例为气管残端复发,5例为同侧胸膜复发;51例转移(34.00%),其中13例为骨转移,12例为对侧肺部转移,6例为脑部转移,9例为肝转移,7例为肾转移,4例为对侧胸膜转移;20例(13.33%)患者术后同时出现复发和转移。

2.2 老年肺癌根治术后复发和转移的单因素分析

单因素分析显示:年龄、病灶直径、分化程度、病理分期、T分期、N分期、纵隔淋巴结转移、淋巴结清扫个数及术后放化疗均与老年肺癌根治术后复发有关($P < 0.05$),见表1;年龄、分化程度、病理分期、T分期、N分期、纵隔淋巴结转移、淋巴结转移个数、淋巴结清扫个数及术后放化疗均与老年肺癌根治术后转移有关($P < 0.05$),见表2。

2.3 随机森林模型特征变量的选取

将单因素分析结果中 $P < 0.05$ 的变量作为随机森林回归模型的构成变量(各变量赋值见表3),根据平均基尼系数对各相关因素进行重要性排序,指数越高提示该因素区分性良好。结果显示:老年肺癌根治术后复发变量重要性排序依次为纵隔淋巴结转移、淋巴结清扫个数、T分期、N分期、分化程度、病理分期、年龄、病灶直径,见图1a;老年肺癌根治术后转移的变量重要性排序依次为纵隔淋巴结转移、N分期、淋巴结转移个数、淋巴结清扫个数、T分期、分化程度、术后放化疗、病理分期、年龄,见图1b。

表 1 复发的单因素分析结果 [例 (%)]

Tab. 1 Results of univariate analysis of recurrence [n (%)]

因素	例数	复发(44 例)	未复发(106 例)	χ^2 值	P 值
性别				1.549	0.213
男	90	23(52.27)	67(63.21)		
女	60	21(47.73)	39(36.79)		
年龄				4.382	0.036
60~70 岁	64	13(29.55)	51(48.11)		
>70 岁	86	31(70.45)	55(51.89)		
吸烟史				2.875	0.090
有	56	21(47.73)	35(33.02)		
无	94	23(52.27)	71(66.98)		
病灶部位				0.056	0.814
左肺	67	19(43.18)	48(45.28)		
右肺	83	25(56.82)	58(54.72)		
病灶直径				5.041	0.025
≤3 cm	69	14(31.82)	55(51.89)		
>3 cm	81	30(68.18)	51(48.11)		
病理类型				0.318	0.957
鳞癌	35	10(22.73)	25(23.58)		
腺癌	91	28(63.64)	63(59.43)		
其他	24	6(13.64)	18(16.98)		
分化程度				26.043	<0.001
高分化	55	4(9.09)	51(48.11)		
中分化	64	22(50.00)	42(39.62)		
低分化	31	18(40.91)	13(12.26)		
病理分期				29.150	<0.001
I 期	115	21(47.73)	94(88.68)		
II~III 期	35	23(52.27)	12(11.32)		
T 分期				34.379	<0.001
T1 分期	108	17(38.64)	91(85.85)		
T2~T3 分期	42	27(61.36)	15(14.15)		
N 分期				22.691	<0.001
N0 分期	106	19(43.18)	87(82.08)		
N1~N2 分期	44	25(56.82)	19(17.92)		
纵隔淋巴结转移				22.338	<0.001
有	39	23(52.27)	16(15.09)		
无	111	21(47.73)	90(84.91)		
淋巴结转移个数				1.923	0.166
≤3 个	49	18(40.91)	31(29.25)		
>3 个	101	26(59.09)	75(70.75)		
淋巴结清扫个数				6.165	0.013
<10 个	57	10(22.73)	47(44.34)		
≥10 个	93	34(77.27)	59(55.66)		
手术类型				2.155	0.142
肺叶清除	129	35(79.55)	94(88.68)		
全肺切除	21	9(20.45)	12(11.32)		
术后并发症				0.158	0.691
有	18	6(13.64)	12(11.32)		
无	132	38(86.36)	94(88.68)		
术后放化疗				9.455	0.002
有	90	18(40.91)	72(67.92)		
无	60	26(59.09)	34(32.08)		

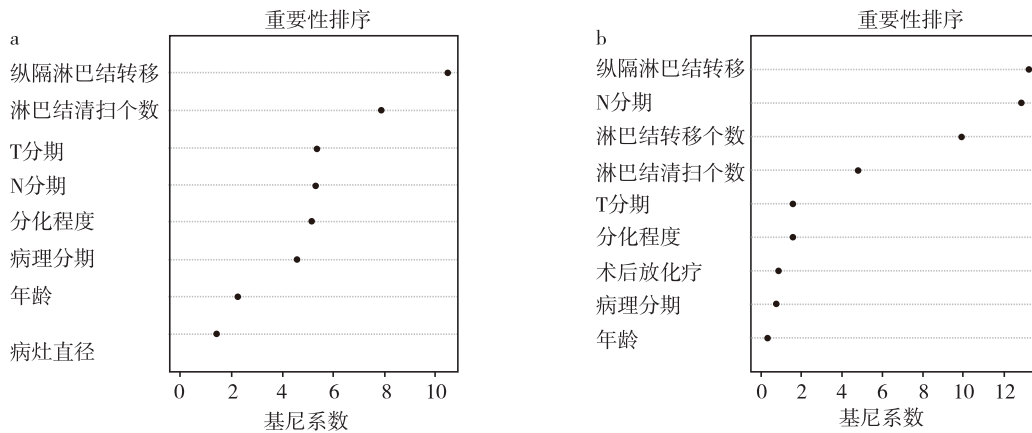
表 2 转移的单因素结果 [例 (%)]
Tab. 2 Single factor results of transfer [n (%)]

因素	例数	转移(51 例)	未转移(99 例)	χ^2 值	P 值
性别				2. 619	0. 106
男	90	26(50. 98)	64(64. 65)		
女	60	25(49. 02)	35(35. 35)		
年龄				19. 774	<0. 001
60 ~ 70 岁	64	9(17. 65)	55(55. 56)		
>70 岁	86	42(82. 35)	44(44. 44)		
吸烟史				1. 113	0. 292
有	56	22(43. 14)	34(34. 34)		
无	94	29(56. 86)	65(65. 66)		
病灶部位				0. 929	0. 335
左肺	67	20(39. 22)	47(47. 47)		
右肺	83	31(60. 78)	52(52. 53)		
病灶直径				0. 255	0. 614
≤3 cm	69	22(43. 14)	47(47. 47)		
>3 cm	81	29(56. 86)	52(52. 53)		
病理类型				0. 006	0. 997
鳞癌	35	12(23. 53)	23(23. 23)		
腺癌	91	31(60. 78)	60(60. 61)		
其他	24	8(15. 69)	16(16. 16)		
分化程度				20. 350	<0. 001
高分化	55	12(23. 53)	43(43. 43)		
中分化	64	18(35. 29)	46(46. 46)		
低分化	31	21(41. 18)	10(10. 10)		
病理分期				33. 017	<0. 001
I 期	115	25(49. 02)	90(90. 91)		
II ~ III 期	35	26(50. 98)	9(9. 09)		
T 分期				23. 844	<0. 001
T1 分期	108	24(47. 06)	84(84. 85)		
T2 ~ T3 分期	42	27(52. 94)	15(15. 15)		
N 分期				24. 371	<0. 001
N0 分期	106	23(45. 10)	83(83. 84)		
N1 ~ N2 分期	44	28(54. 90)	16(16. 16)		
纵隔淋巴结转移				25. 062	<0. 001
有	39	26(50. 98)	13(13. 13)		
无	111	25(49. 02)	86(86. 87)		
淋巴结转移个数				12. 604	<0. 001
≤3 个	49	7(13. 73)	42(42. 42)		
>3 个	101	44(86. 27)	57(57. 58)		
淋巴结清扫个数				5. 133	0. 024
<10 个	57	13(25. 49)	44(44. 44)		
≥10 个	93	38(74. 51)	55(55. 56)		
手术类型				3. 677	0. 055
肺叶清除	129	40(78. 43)	89(89. 90)		
全肺切除	21	11(21. 57)	10(10. 10)		
术后并发症				0. 994	0. 319
有	18	8(15. 69)	10(10. 10)		
无	132	43(84. 31)	89(89. 90)		
术后放化疗				38. 344	<0. 001
有	90	13(25. 49)	77(77. 78)		
无	60	38(74. 51)	22(22. 22)		

表 3 赋值表

Tab. 3 Assignment table

变量	赋值	变量	赋值
性别	男 = 1, 女 = 2	T 分期	T1 分期 = 1, T2 ~ T3 分期 = 2
年龄	60 ~ 70 岁 = 1, > 70 岁 = 2	N 分期	N1 分期 = 1, N2 ~ N3 分期 = 2
吸烟史	无 = 0, 有 = 1	纵隔淋巴结转移	无 = 0, 有 = 1
病灶部位	左肺 = 1, 右肺 = 2	淋巴结转移个数	≤ 3 个 = 1, > 3 个 = 2
病灶直径	≤ 3 cm = 1, > 3 cm = 2	淋巴结清扫个数	< 10 个 = 1, ≥ 10 个 = 2
病理类型	鳞癌 = 1, 腺癌 = 2, 其他 = 3	手术类型	肺叶清除 = 1, 全肺切除 = 2
分化程度	低分化 = 1, 中分化 = 2, 高分化 = 3	术后并发症	无 = 0, 有 = 1
病理分期	I 期 = 1, II ~ III 期 = 2	术后放化疗	无 = 0, 有 = 1



注: a 为术后复发相关因素, b 为术后转移相关因素

图 1 术后复发、转移相关因素的重要性排序

Fig. 1 Ranking the importance of factors associated with postoperative recurrence and metastasis

2.4 随机森林模型的建立

建立随机森林模型, 采用对照法, 固定决策树数量 $ntree$ 不变, 特征选取数量即变量数 $mtry$ 依次取值 1 ~ 8 进行调试, 观察术后复发模型的平均袋外误差估计误差率随 $mtry$ 变化情况, 见图 2a。当袋外误差为 4 个时, 平均袋外误差估计误差率最低, 故选取特征数量为 4 个。固定 $mtry = 4$, $ntree$ 依次取 1 ~ 500, 观察平均袋外估计误差率随 $ntree$ 变化情况, 见图 2b。 $ntree$ 取值为 1 ~ 120 时平均袋外估计误差率下降迅速, 取值为 120 以后下降减缓, 取值为 300 后趋于稳定, 因此, 本研究选取决策树数量 $ntree$ 为 300, 此时获得最优模型。

建立随机森林模型, 采用对照法, 固定决策树数量 $ntree$ 不变, 特征选取数量即变量数 $mtry$ 依次取值 1 ~ 9 个进行调试, 观察术后转移模型的平均袋外误差估计误差率随 $mtry$ 变化情况, 见图 3a。当袋外误差为 5 个时, 平均袋外误差估计误差率最低, 故选取特征数量为 5 个。固定 $mtry = 5$, $ntree$ 依次取 1 ~ 500, 观察平均袋外估计误差率随 $ntree$ 变化情况, 见图 3b。 $ntree$ 取值为 1 ~ 100 时平均袋外估计误差率下降迅速, 取值为 100 以后下降减缓, 取值为 200 后趋于稳定, 因此, 本研究选取决策树数量 $ntree$ 为 200, 此时获得最优模型。

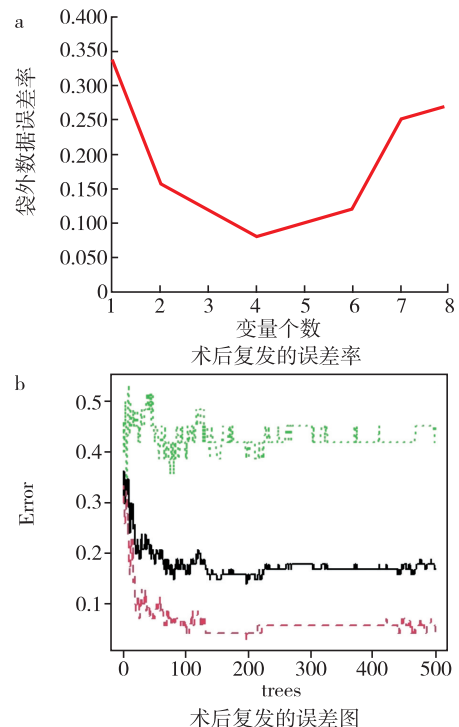


图 2 逐步随机森林袋外数据分类错误图

Fig. 2 Classification error diagram of gradual random forest bag data

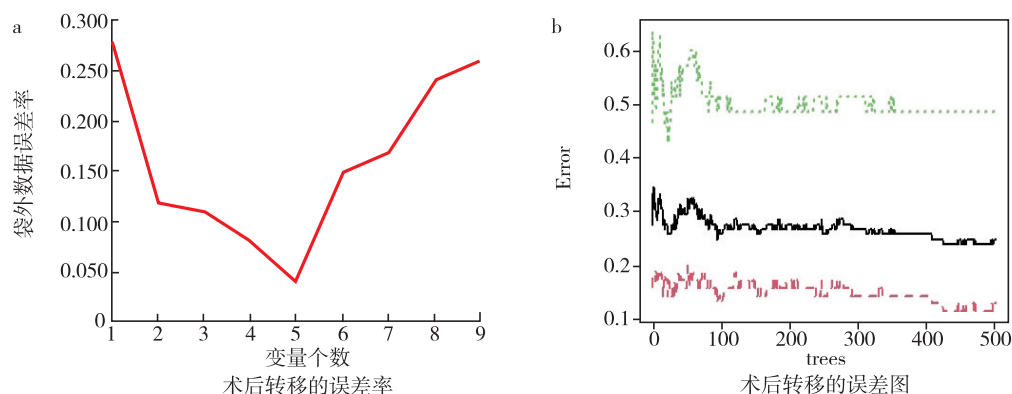


图3 随机森林误差图

Fig. 3 Random forest error diagram

2.5 随机森林图模型的验证

验证组男37例,女33例,年龄60~79岁,平均(68.76±3.85)岁;体质量指数22~25 kg/m²,平均(23.69±0.56) kg/m²;病灶部位在左肺34例,右肺36例;病灶直径1.5~4.0 cm,平均(2.79±0.43) cm;病理类型鳞癌18例,腺癌52例;术后复发21例,其中10例为同侧肺部复发,6例为纵隔淋巴结复发,1例为同侧腹膜复发,4例为气管残端复发;术后转移25例,其中8例为骨转移,7例为对侧肺部转移,3例为脑部转移,4例为肝转移,2例为肾转移,1例为对侧胸膜转移;9例患者同时出现复发转移。

内部验证:采用多维度法展示该随机森林模型对老年肺癌根治术后复发和转移的预测能力,观察到随机森林模型中纵隔淋巴结转移、淋巴结清扫个数、T分期、N分期指标所建立的预测模型对预测术后复发的效能为0.904,纵隔淋巴结转移、N分期、淋巴结转移个数、淋巴结清扫个数、T分期指标所建立的预测模型对预测术后转移的效能为0.897。

外部验证:外部验证随机森林模型预测老年肺癌术后复发效能为0.905(95% CI: 0.825~0.974, $P < 0.05$),术后转移效能为0.910(95% CI: 0.831~0.982, $P < 0.05$),与内部验证基本一致,见图4。

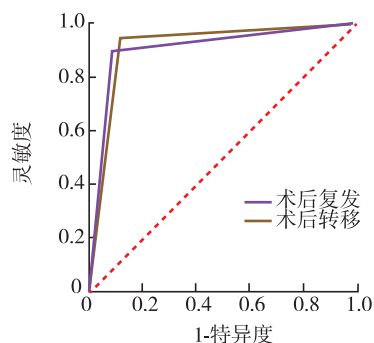


图4 ROC 曲线图

Fig. 4 ROC curve

3 讨论

肿瘤的局部复发、远处转移是肺癌根治术治疗失败、术后死亡的主要原因^[9-10]。相关数据显示,Ⅰ期非小细胞肺癌完整切除术后复发率、转移率分别为20.3%、26.6%,Ⅱ~Ⅲ期术后复发率、转移率分别为45.5%、47.7%^[11-12]。本研究随访时间为22.8个月,Ⅰ~Ⅲ期老年肺癌患者随访期间复发率、转移率分别为29.33%、34.00%,与以往报道稍有差异,可能与围术期状况、随访时间长短等因素有关。由此可见,预测老年肺癌患者术后复发和转移风险有利于指导术后干预,为临床制定诊疗措施提供理论支持。

随机森林算法基于递归分类树进行计算分析,其优势在于不受变量间共线性影响,对噪声、异常值的容忍度高,尤其适用于复杂、多维数据的分析^[13-14]。本研究尝试构建随机森林预测模型,通过基尼系数计算各个变量对分类数据中各个节点观测值的影响,分析其在老年肺癌根治术后复发转移预测模型构建中的应用价值,随机森林基尼系数减少量越多说明该变量越重要。目前高龄、吸烟史、病理分期、淋巴结转移数目等肺癌术后复发转移因素已得到多数研究认可。本研究术后复发和转移的变量重要性排序中,T分期、N分期均居于较前序位,分化程度、病理分期亦为老年肺癌患者术后复发和转移的重要因素。说明中晚期老年肺癌患者术后复发和转移率较高,主要由于T分期、N分期、病理分期与肿瘤大小、侵犯范围有关,随分期进展,完整切除可能性减小,复发风险增大。分化程度低的患者肿瘤恶性程度较高,病灶浸润、增殖、转移能力较强,易累及淋巴结,存在微血管浸润患者血行转移率增加,进而增加术后复发和转移的风险^[15-17]。有研究表明,肺癌患者淋巴结转移数量<3个时其五年生存率可达58.89%^[18]。本研究显示淋巴结转移个数与肺癌根治术后发生转移有关,转移个数>3个患者术后转移率较高,说明淋巴结转移数量是预测老年肺癌患者术后转移的价值因子,可能与大量淋巴结转移将加大根治术中淋巴结清扫难

度有关;同时大量淋巴结转移加重肿瘤细胞对内循环的侵犯,可能引起淋巴结跳跃性转移,进一步加重患者病情。

本研究采用随机森林算法,无须排除共线性问题,可直观显示各个变量对术后复发和转移的影响,纵隔淋巴结转移、淋巴结清扫个数、T 分期、N 分期对预测老年肺癌根治术后复发转移具有较高预测价值,且淋巴结转移个数有利于预测术后转移,随机森林预测模型结果显示所建模型对预测术后复发和转移的效能分别为 0.904、0.897,说明该模型对手术疗效、术后辅助治疗等具有一定的临床指导意义。但本研究为单中心小样本研究,所建模型的效果仍有待临床进一步研究证实。

参考文献

- [1] Al Zreibi C, Gibault L, Fabre E, et al. Surgery for small-cell lung cancer [J]. *Rev Mal Respir*, 2021, 38 (8): 840–847.
- [2] Deng H, Liu J, Cai X, et al. Radical minimally invasive surgery after immuno-chemotherapy in initially-unresectable stage IIIB non-small cell lung cancer [J]. *Ann Surg*, 2022, 275 (3): e600–e602.
- [3] Hui Z, Men Y, Hu C, et al. Effect of postoperative radiotherapy for patients with pIIIA-N2 non-small cell lung cancer after complete resection and adjuvant chemotherapy: the phase 3 PORT-C randomized clinical trial [J]. *JAMA Oncol*, 2021, 7 (8): 1178–1185.
- [4] Chapman B V, Ning M S, Farnia B, et al. Postoperative radiotherapy for locally advanced NSCLC: implications for shifting to conformal, high-risk fields [J]. *Clin Lung Cancer*, 2021, 22 (3): 225–233. e7.
- [5] Fu F, Zhang Y, Gao Z, et al. Development and validation of a five-gene model to predict postoperative brain metastasis in operable lung adenocarcinoma [J]. *Int J Cancer*, 2020, 147 (2): 584–592.
- [6] 刘显平, 李晓, 杨帆. 非小细胞肺癌根治性切除术后复发转移模式研究进展 [J]. *中国肺癌杂志*, 2022, 25 (1): 26–33.
Liu X P, Li X, Yang F. Pattern of recurrence and metastasis after radical resection of non-small cell lung cancer [J]. *Chin J Lung Cancer*, 2022, 25 (1): 26–33.
- [7] Kinoshita T, Goto T. Links between inflammation and postoperative cancer recurrence [J]. *J Clin Med*, 2021, 10 (2): 228.
- [8] Ellis D E, Hubbard R A, Willis A W, et al. Comparing LASSO and random forest models for predicting neurological dysfunction among fluoroquinolone users [J]. *Pharmacoepidemiol Drug Saf*, 2022, 31 (4): 393–403.
- [9] Zou Y, Zhao L, Zhang J, et al. Development and internal validation of machine learning algorithms for end-stage renal disease risk prediction model of people with type 2 diabetes mellitus and diabetic kidney disease [J]. *Ren Fail*, 2022, 44 (1): 562–570.
- [10] 饶德新, 潘丽红, 刘恒, 等. 血清 miR-134-5p 和趋化因子 20 对非小细胞肺癌患者肺癌根治术后复发转移的预测价值 [J]. *中华实用诊断与治疗杂志*, 2022, 36 (2): 130–135.
Rao D X, Pan L H, Liu H, et al. Values of serum miR-134-5p and chemokine 20 to the prediction of recurrence/metastasis after radical resection of non-small cell lung cancer [J]. *J Chin Pract Diagn Ther*, 2022, 36 (2): 130–135.
- [11] Zhong Y, She Y, Deng J, et al. Deep Learning for Prediction of N2 Metastasis and Survival for Clinical Stage I Non-Small Cell Lung Cancer [J]. *Radiology*, 2022, 302 (1): 200–211.
- [12] 赵方超, 王伟健, 刘建明. 非小细胞肺癌术后复发转移的风险模型构建及预测能力的验证 [J]. *肿瘤防治研究*, 2020, 47 (4): 273–277.
Zhao F C, Wang W J, Liu J M. Construction of risk model for postoperative recurrence and metastasis of non-small cell lung cancer and verification of predictive ability [J]. *Cancer Res Prev Treat*, 2020, 47 (4): 273–277.
- [13] Austin A M, Ramkumar N, Gladders B, et al. Using a cohort study of diabetes and peripheral artery disease to compare logistic regression and machine learning via random forest modeling [J]. *BMC Med Res Methodol*, 2022, 22 (1): 300.
- [14] Chong-Wen W, Sha-Sha L, Xu E. Predictors of rapid eye movement sleep behavior disorder in patients with Parkinson's disease based on random forest and decision tree [J]. *PLoS One*, 2022, 17 (6): e0269392.
- [15] Terada Y, Isaka M, Harada H, et al. Radiotherapy for local recurrence of non-small-cell lung cancer after lobectomy and lymph node dissection—can local recurrence be radically cured by radiation? [J]. *Jpn J Clin Oncol*, 2020, 50 (4): 425–433.
- [16] Jin J, Xu Y, Hu X, et al. Postoperative radiotherapy option based on mediastinal lymph node reclassification for patients with pN2 non-small-cell lung cancer [J]. *Curr Oncol*, 2020, 27 (3): e283–e293.
- [17] 邱培, 郑希, 谭小波, 等. 非小细胞肺癌患者微创切除术后复发与转移的相关因素分析 [J]. *实用癌症杂志*, 2020, 35 (4): 598–602.
Qiu P, Zheng X, Tan X B, et al. Related factors influencing the recurrence and metastasis of non-small cell lung cancer after minimally invasive resection [J]. *Pract J Cancer*, 2020, 35 (4): 598–602.
- [18] Kukhon F R, Lan X, Helgeson S A, et al. Occult lymph node metastasis in radiologic stage I non-small cell lung cancer: The role of endobronchial ultrasound [J]. *Clin Respir J*, 2021, 15 (6): 676–682.

(2023-10-25 收稿)