

基于随机森林算法的甲状腺癌患者述情障碍预测模型构建*

袁红* 王惠 高璐 冯键华 徐波
广州市第一人民医院甲状腺乳腺外科(广东 广州 510180)

【摘要】目的 基于随机森林算法构建甲状腺癌患者述情障碍的预测模型。**方法** 本研究以2023年6月至2024年12月期间于本院就诊的甲状腺癌患者为调查总体, 采用便利抽样法, 共纳入234例有效样本进行问卷调查。依据述情障碍的评估结果, 将入选患者划分为发生组与未发生组。在组间比较的基础上, 深入剖析其影响因素, 并进一步构建随机森林预测模型。最后, 运用受试者工作特征(ROC)曲线对该模型的预测效能进行评价。**结果** 发生述情障碍组患者142例(60.68%)。经单因素分析筛选, 两组患者在文化程度、肿瘤大小、TNM分期、领悟社会支持、心理弹性及创伤后成长方面差异均有统计学意义(均 $P<0.05$)。多因素分析结果显示, 以上因素均为影响述情障碍发生的独立危险/保护因素(均 $P<0.05$)。构建甲状腺癌患者发生述情障碍的随机森林模型, 树的数量为100时出现的错误率最低。随机森林模型ROC曲线验证, 其曲线下面积(AUC)为0.879($P<0.05$), 敏感度和特异度分别为0.769和0.844。**结论** 基于随机森林算法构建的预测模型区分能力较优, 可作为临床评估工具用于高危人群的早期甄别。据此对目标患者实施针对性心理或行为干预, 有望改善其预后体验。

【关键词】 随机森林算法; 甲状腺癌; 述情障碍; 预测模型

【中图分类号】 R736.1

【文献标识码】 A

【基金项目】 广州市科技计划项目(2023A03J0967)

DOI:10.3969/j.issn.1009-3257.2026.7.013

Construction of a Predictive Model for Affective Disorders in Thyroid Cancer Patients Based on Random Forest Algorithm*

YUAN Hong*, WANG Hui, GAO Lu, FENG Jian-hua, XU Bo.

Department of Thyroid and Breast Surgery, Guangzhou First People's Hospital, Guangzhou 510180, Guangdong Province, China

Abstract: Objective To construct a random forest predictive model for affective disorders in thyroid cancer patients. **Methods** This study surveyed a total of 234 valid samples of thyroid cancer patients who visited our hospital between June 2023 and December 2024 using convenience sampling. According to the assessment results of affective disorders, the selected patients were divided into the occurrence group and the non occurrence group. On the basis of inter group comparison, analyze the influencing factors in depth and further construct a random forest prediction model. Finally, the receiver operating characteristic (ROC) curve was used to evaluate the predictive performance of the model. **Results** There were 142 patients (60.68%) in the group with expressive disorders. Through univariate analysis screening, significant differences were found between the two groups of patients in terms of educational level, tumor size, TNM staging, perceived social support, psychological resilience, and posttraumatic growth (all $P<0.05$). The results of multivariate analysis showed that the above factors were independent risk/protective factors affecting the occurrence of affective disorders (all $P<0.05$). Constructing a random forest model for patients with thyroid cancer who experience affective disorders, with the lowest error rate observed when the number of trees is 100. The ROC curve validation of the random forest model showed an area under the curve (AUC) of 0.879 ($P<0.05$), with sensitivity and specificity of 0.769 and 0.844, respectively. **Conclusion** The prediction model based on random forest algorithm has better discrimination ability and can be used as a clinical evaluation tool for early screening of high-risk groups. Based on this, targeted psychological or behavioral interventions for target patients are expected to improve their prognostic experience.

Keywords: Random Forest Algorithm; Thyroid Cancer; Verbal Disorders; Prediction Model

甲状腺癌居内分泌系统恶性肿瘤之首, 且其发病率于近年呈现迅速上升的流行态势, 已成为备受关注的公共卫生问题^[1]。癌症本身及其治疗过程所带来的身体创伤、生理功能改变以及对未来不确定性的担忧, 给患者的心理状态带来了沉重的负担^[2]。述情障碍在癌症患者中高发, 主要表现为个体难以识别和表达自身的情绪, 缺乏幻想能力, 思维呈现出具体而局限的特点^[3]。这一障碍不仅阻碍了患者与他人之间的情感交流, 更对其自身的身心健康产生了严重影响。研究表明^[4], 患有述情障碍的甲状腺癌患者, 其负面情绪往往更为严重, 对治疗的依从性也相对较差。相较于传统回归模型, 随机森林算法在处理非线性关系和高维数据时无需预先设定变量分布, 具有更强的灵活性和预测稳定性^[5]。鉴于此, 本研究尝试构建甲状腺癌患者述情障碍

的随机森林风险预测模型, 以期辅助临床实现高风险患者的有效甄别, 并为后续干预策略的制定提供循证依据。

1 对象与方法

1.1 研究对象 根据Harrell^[6]对多因素分析样本量估算的方法, 样本量遵循变量数的10倍原则进行估算, 本研究预估可纳入17个自变量, 按20%的失访率进行扩充后, 本研究所需样本量至少为204例。采用便利抽样法选取医院2023年6月至2024年12月收治的甲状腺癌患者234例为研究对象, 并根据是否发生述情障碍分为发生述情障碍组和未发生述情障碍组, 述情障碍判定^[7]: 多伦多述情障碍量表得分 ≥ 51 分。

纳入标准: 经病理学检查确诊为甲状腺癌; 符合手术适应

【第一作者】袁红, 女, 副主任护师, 主要研究方向: 甲状腺乳腺外科。E-mail: yuanh12g@163.com

【通讯作者】袁红

证，并由同一外科团队完成手术治疗；年龄不低于18周岁；自愿参与并签署知情同意书。排除标准：伴有严重心、肺、肝、肾等重要脏器功能障碍；存在其他部位原发恶性肿瘤；合并严重精神疾患或认知功能受损。

1.2 研究工具 (1)一般资料调查表：自行设计，涵盖患者的年龄、性别、家庭人均月收入等共14项。(2)多伦多述情障碍量表^[8]：该量表涵盖外向型思维、情感描述困难及情感识别困难3个维度，总分范围20~100分，分数越高代表述情障碍程度越严重。(3)领悟社会支持量表^[9]：该量表从家庭支持、朋友支持及其他支持3方面进行评估，总分范围12~84分，得分越高反映患者感知到的社会支持越充分。(4)心理弹性量表^[10]：分为积极接受变化、坚韧、控制、忍受消极情感、精神影响5个维度，总分范围0~100分，得分越高患者心理弹性能力越强。(5)创伤后成长评定量表^[11]：该量表从与他人关系、新的可能性、个人力量、自我转变及人生感悟5方面进行评估，总分范围0~100分，得分越高反映患者的创伤后成长越充分。

1.3 收集资料的方法 为保证数据质量，研究组成员在调查启动前接受了统一的规范化培训。调查过程中，向自愿参与的患者发放纸质问卷，由其独立填写，填写完毕后当场回收。录入环节采用双人同步录入法，确保数据准确无误。本次调查共发放问卷240份，回收有效问卷234份，有效回收率为97.50%。

1.4 统计学方法 使用SPSS 26.0统计软件进行数据统计分析，正态分布的计量资料以($\bar{x} \pm s$)表示，行独立样本t检验，计数资料以[n(%)]表示，行 χ^2 检验，等级资料行秩和检验，采

因素分析采用Logistic回归；运用R4.3.3构建随机森林模型，采用受试者工作特征(ROC)曲线下面积(AUC)、敏感度、特异度评估模型的预测价值及效能。 $P<0.05$ 被视为差异有统计学意义。

2 结果

2.1 单因素分析 发生述情障碍组患者142例(60.68%)。单因素分析结果表明，两组患者在文化程度、肿瘤大小、TNM分期、领悟社会支持、心理弹性、创伤后成长方面差异有统计学意义(均 $P<0.05$)。见表1。

2.2 多因素分析 以甲状腺癌患者是否发生述情障碍(未发生=0，发生=1)为因变量，将单因素分析中差异具有统计学意义的变量作为自变量纳入多因素Logistic回归分析(文化程度：小学及以下=1、初中及高中=2，大专及以上=3，肿瘤大小<1.5cm=0、 ≥ 1.5 cm=1，TNM分期Ⅰ期=0、Ⅱ期=1，其余按原值输入)，多因素分析证实，文化程度、肿瘤大小、TNM分期、领悟社会支持、心理弹性、创伤后成长均被确认为述情障碍的独立预测因子(均 $P<0.05$)。见表2。

2.3 随机森林模型 构建甲状腺癌患者发生述情障碍的随机森林模型，当树的数量为100时错误率最低，见图1；影响甲状腺癌患者发生述情障碍的因素按重要性排序见图2。

2.4 随机森林模型预测性能分析 ROC分析显示，随机森林AUC为0.879(95%CI: 0.790~0.969)，敏感度和特异度分别为0.769和0.844，见图3。

表2 甲状腺癌患者发生述情障碍的多因素分析(n=234)

因素	β	S.E.	Wald	OR	95%CI	P
文化程度	-0.522	0.204	6.518	0.593	0.398~0.886	0.011
肿瘤大小	0.910	0.322	7.995	2.484	1.322~4.668	0.005
TNM分期	1.092	0.347	9.899	2.979	1.509~5.881	0.002
领悟社会支持	-0.046	0.020	5.202	0.955	0.918~0.994	0.023
心理弹性	-0.061	0.020	9.247	0.941	0.905~0.979	0.002
创伤后成长	-0.067	0.018	13.188	0.935	0.902~0.970	<0.001
常量	10.692	1.814	34.744	-	-	<0.001

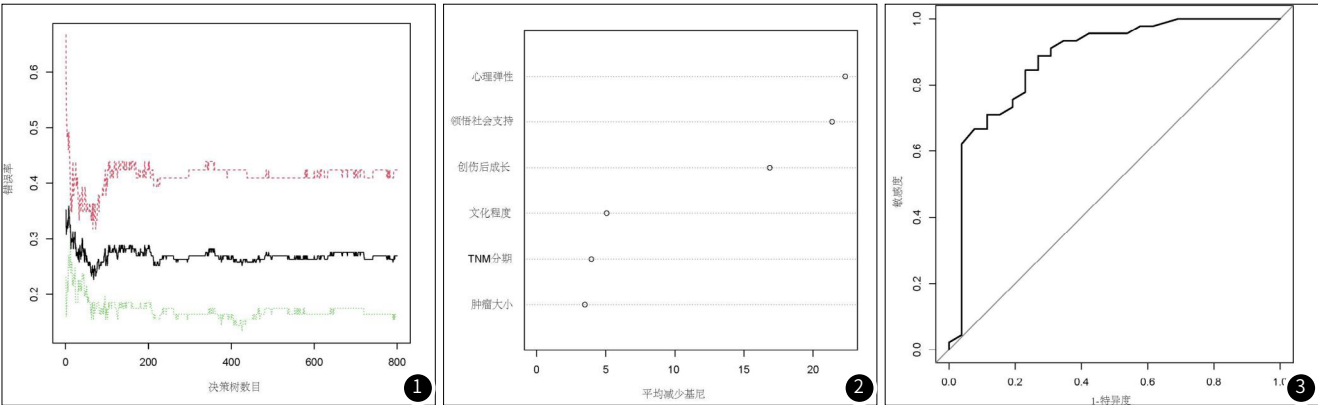


图1 决策树数量对错误率的影响。图2 随机森林模型变量重要性排序图。图3 随机森林模型ROC曲线图。

表1 甲状腺癌患者发生述情障碍的单因素分析[n=234, n(%)]

因素		发生述情障碍组(n=142)	未发生述情障碍组(n=92)	t/ χ^2 /Z	P
年龄(岁)		42.34±10.26	40.65±9.45	1.269	0.206
性别	男	58(40.85)	43(46.74)	0.791	0.374
	女	84(59.15)	49(53.26)		
家庭人均月收入(元)	<3000	28(19.72)	12(13.04)	1.193	0.233
	3000~6000	74(52.11)	50(54.35)		
	>6000	40(28.17)	30(32.61)		
婚姻状况	未婚	10(7.04)	6(6.52)	0.416	0.812
	已婚	125(88.03)	83(90.22)		
	离异/丧偶	7(4.93)	3(3.26)		
工作状况	在职	122(85.92)	76(82.61)	0.469	0.493
	离退	20(14.08)	16(17.39)		
文化程度	小学及以下	38(26.76)	15(16.30)	2.708	0.007
	初中及高中	56(39.44)	30(32.61)		
	大专及以上	48(33.80)	47(51.09)		
医疗费用支付方式	医疗保险	112(78.87)	70(76.09)	0.905	0.636
	商业保险	13(9.15)	12(13.04)		
	自费	17(11.97)	10(10.87)		
居住地	城镇	98(69.01)	69(75.00)	0.979	0.322
	农村	44(30.99)	23(25.00)		
病理类型	乳头状癌	106(74.65)	72(78.26)	0.400	0.527
	滤泡癌	36(25.35)	20(21.74)		
肿瘤大小(cm)	<1.5	70(49.30)	61(66.30)	6.554	0.011
	≥1.5	72(50.70)	31(33.70)		
TNM分期[n(%)]	I 期	79(55.63)	67(72.83)	7.033	0.008
	II 期	63(44.37)	25(27.17)		
手术方式	腔镜手术	48(33.80)	37(40.22)	0.993	0.319
	开放手术	94(66.20)	55(59.78)		
切除范围	单侧腺叶	93(65.49)	66(71.74)	1.249	0.536
	单侧+对侧部分	14(9.86)	6(6.52)		
	全甲状腺	35(24.65)	20(21.74)		
诊断时间(月)	<6	74(52.11)	43(46.74)	0.645	0.422
	≥6	68(47.89)	49(53.26)		
淋巴结转移	有	65(45.77)	39(42.39)	0.259	0.611
	无	77(54.23)	53(57.61)		
领悟社会支持(分)		57.28±8.96	62.59±9.46	4.332	<0.001
心理弹性(分)		54.75±9.22	60.46±10.08	4.460	<0.001
创伤后成长(分)		50.52±8.34	55.67±9.73	4.318	<0.001

3 讨 论

本研究中,甲状腺癌患者述情障碍发生率为60.68%,同黄元琦等^[7]的研究结果相似,提示甲状腺癌患者述情障碍发生率较高。这可能与甲状腺所处的解剖位置特殊有关。目前手术切除仍是该病的主要治疗方式,但患者术后并发症发生率高、复发风险大,加之确诊及手术本身均为强烈的心理应激事件,

严重冲击患者的身心健康,诱发焦虑、抑郁等负性情绪,进而使述情障碍的发生风险显著增加。

本研究中,甲状腺癌患者发生述情障碍受多种因素影响。其中,文化程度较低的患者,可能因知识储备和认知能力的限制,在面对疾病时更难以理解和应对复杂的情绪,从而增加述情障碍的发生风险,闫晓彤等^[12]的研究与此一致。临床护理人

员应采用通俗易懂的方式开展健康宣教,以增强患者对疾病及自身情绪变化的认知。肿瘤大小较大及TNM分期较晚的患者更易发生述情障碍,这与张越等^[13]的研究结果相仿。肿瘤大小和TNM分期反映了疾病的严重程度,肿瘤较大或分期较晚往往患者往往需要承受更强烈的生理痛苦和对预后的恐惧,进而加重述情障碍。临床应加强对患者病情的观察和并发症管理,在此基础上制定个体化护理方案,同时注重患者的心理疏导与情感支持,以增强其疾病应对信心。领悟社会支持水平较低的患者,由于缺乏来自家庭、朋友和社会的有力支持,在面对疾病压力时,积累的负面情绪无法得到良好宣泄和缓解,更容易陷入述情障碍,既往研究^[14]也表达了类似的观点。临床中多引导家属和朋友关注患者情况,并积极联系社区资源,为患者提供支持,帮助其感受到外界关爱。闫晓彤等^[15]的研究认为,心理弹性水平高的患者在面对压力和逆境时具有更好的适应和恢复能力,能感受到更多积极情绪,减轻不良事件带来的负面影响,降低述情障碍的发生,本研究结果同上述研究一致。临床可通过心理干预帮助患者增强心理弹性,提高其应对疾病的能力。创伤后成长水平越高,患者发生述情障碍的可能性越低,与以往学者^[16]研究结果相似。创伤后成长是通过积极的心理调适实现自我成长的过程,有利于患者从创伤中自我恢复,进而不易发生述情障碍。临床护理工作中应及时回应患者的合理诉求,耐心解答其疑虑,使患者切实感受到人文关怀,从而增强其疾病适应能力,促进创伤后成长水平的提升。

本研究随机森林模型AUC为0.879($P<0.05$),提示随机森林模型预测性能较好,应用于临床可有效识别高风险发生述情障碍患者。且敏感度和特异度分别为0.769和0.844,其中灵敏度理想,表明模型对述情障碍高风险患者具有较高的检出能力,漏诊风险较低;特异度尚可,提示模型在排除非述情障碍患者方面具有一定鉴别力,误诊率可控。此外,随机森林算法输出的变量重要性排序反映了各因素在预测述情障碍中的相对贡献,临床可据此合理分配有限的护理资源,优先关注高风险患者,提高护理效率。

本研究所构建的随机森林模型展现出良好的预测效能,有助于临床早期甄别述情障碍高危个体并实施分层干预。然本研究属单中心小样本调查,样本代表性有限,可能存在一定的选择偏倚及抽样误差,后续需通过大样本、多中心研究进一步验证模型的泛化能力。

参考文献

- [1] Han B, Zheng R, Zeng H, et al. Cancer incidence and mortality in China, 2022[J]. J Natl Cancer Cent, 2024, 4(1): 47-53.
- [2] 梁昌翠, 杨惠. 社会支持对甲状腺癌术后病人病耻感影响: 心理一致感的中介效应[J]. 蚌埠医学院学报, 2024, 49(9): 1244-1248.
- [3] 苏芮, 牛巧红, 张颖, 等. 结直肠癌患者述情障碍的研究进展[J]. 心理月刊, 2024, 19(17): 210-212.
- [4] 苏珊珊, 蒋文晖, 王书婷, 等. 焦虑障碍患者的焦虑水平、述情障碍程度对其生活质量的影响[J]. 上海交通大学学报(医学版), 2024, 44(5): 584-590.
- [5] 谢文亮, 王淑芳, 李旭光, 等. ICU患者肠内营养相关性腹泻风险预测模型的构建及验证[J]. 中华护理杂志, 2022, 57(19): 2324-2332.
- [6] Harrell F E. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis[M]. Berlin: Springer, 2010.
- [7] 黄元琦, 梁波, 赵小乐, 等. 甲状腺癌患者述情障碍现状及其影响因素分析[J]. 中国癌症防治杂志, 2022, 14(4): 440-444.
- [8] 夏朝云, 王东波, 潘小娅. 多伦多述情障碍自陈量表在大学生中的适用性研究[J]. 浙江医学, 2001, (12): 52-54.
- [9] 黄丽, 姜乾金, 任蔚红. 应对方式、社会支持与癌症病人心身症状的相关性研究[J]. 中国心理卫生杂志, 1996, 10(4): 160-161.
- [10] Yu X, Zhang J. Factor analysis and psychometric evaluation of the Connor-Davidson Resilience Scale (CD-RISC) with Chinese people[J]. Soc Behav Pers, 2007, 35(1): 19-30.
- [11] 汪际, 陈瑶, 王艳波, 等. 创伤后成长评定量表的修订及信效度分析[J]. 护理学杂志, 2011, 26(14): 26-28.
- [12] 闫晓彤, 郝春艳. 老年慢性阻塞性肺疾病病人述情障碍现状及影响因素分析[J]. 全科护理, 2024, 22(17): 3193-3197.
- [13] 张越, 周凤芹. 首发分化型甲状腺癌术后患者述情障碍现状及其影响因素[J]. 中国校医, 2024, 38(1): 41-44.
- [14] 岳芸, 邓丽云, 孙晓静, 等. 述情障碍在慢性精神分裂症患者社会支持和生活质量间的中介效应[J]. 心理月刊, 2024, 19(2): 109-111.
- [15] 闫晓彤, 梁嘉卿, 刘巧丽, 等. COPD病人心理弹性在领悟社会支持与述情障碍间的中介效应[J]. 循证护理, 2024, 10(4): 755-760.
- [16] 沈婕. 述情障碍对遭受工作场所暴力护士创伤后成长的影响[J]. 当代护士(下旬刊), 2019, 26(7): 27-30.

(收稿日期: 2025-06-21)

(校对编辑: 李清芸)

(排版编辑: 刘潍嘉)