

# 基于 SMOTE 算法的声门型喉癌患者术后 下呼吸道感染预警模型构建

马灵草,董婷,戴晗青,岳红,陈海芳,臧艳姿

**摘要:**目的 构建基于 SMOTE 算法的声门型喉癌患者术后下呼吸道感染预警模型,为临床预防患者术后下呼吸道感染提供参考。**方法** 收集 596 例接受肿瘤根治手术的声门型喉癌患者临床资料,利用单因素比较和 Logistic 回归分析筛选术后下呼吸道感染的影响因素,采用 SMOTE 算法对影响因素的原始数据集进行重建,得到基于 SMOTE 算法的声门型喉癌患者术后下呼吸道感染预警模型。**结果** 术后共 38 例(6.38%)患者发生下呼吸道感染。Logistic 回归分析可得,年龄、BMI 值、吸烟史、糖尿病、肺部疾病、手术类型、术后吸痰、术中出血量、术后拔管时间和使用人工鼻是声门型喉癌患者术后下呼吸道感染的影响因素(均  $P < 0.01$ )。原始预测模型和基于 SMOTE 算法的预警模型比较,真阳性率差异无统计学意义,阳性预测值、 $F$  值、ROC 曲线下面积差异有统计学意义( $P < 0.05, P < 0.01$ )。**结论** 基于 SMOTE 算法构建的预警模型优于原始预测模型,能准确预警喉癌患者术后下呼吸道感染,可基于预测结果采取针对性措施预防喉癌患者术后下呼吸道感染。

**关键词:** SMOTE 算法; 声门型喉癌; 术后; 下呼吸道感染; 预警模型; 吸痰; 人工鼻; 出血量

**中图分类号:** R473.73 **文献标识码:** A **DOI:** 10.3870/j.issn.1001-4152.2021.08.001

**Establishment of a warning model of postoperative lower respiratory tract infection in patients with glottic carcinoma based on SMOTE algorithm** Ma Lingcao, Dong Ting, Dai Hanqing, Yue Hong, Chen Haifang, Zang Yanzi. Otolaryngology Head and Neck Surgery Department, Henan People's Hospital, Zhengzhou 450003, China

**Abstract: Objective** To construct a warning model of lower respiratory tract infection in patients with glottic carcinoma based on SMOTE algorithm, so as to provide reference for preventing postoperative lower respiratory tract infection of these patients in clinic. **Methods** The clinical data of 596 glottic carcinoma patients undergoing tumor radical operation were investigated, and single factor comparison and logistic regression analysis were utilized to screen the influencing factors of postoperative lower respiratory tract infection, then the original data set of influencing factors was reconstructed using the SMOTE algorithm to obtain a warning model of postoperative lower respiratory tract infection for glottic carcinoma patients. **Results** Totally 38 patients (6.38%) suffered from postoperative lower respiratory tract infection. Logistic regression analysis revealed that, age, BMI, smoking history, diabetes, lung disease, operation type, postoperative sputum aspiration, intraoperative bleeding volume, postoperative extubation time and use of artificial nose were the influencing factors of postoperative lower respiratory tract infection for glottic carcinoma patients ( $P < 0.01$  for all). When compared the origin warning model with the warning model based on SMOTE algorithm, there was no significant difference in TPR, while the PPV, F-score and AUC were significant ( $P < 0.05, P < 0.01$ ). **Conclusion** The warning model based on SMOTE algorithm is better than the original one, which can accurately predict postoperative lower respiratory tract infection for laryngeal cancer patients, then take targeted intervention to prevent postoperative lower respiratory tract infection.

**Key words:** Synthetic Minority Over-sampling Technique algorithm; glottic carcinoma; postoperative; lower respiratory tract infection; warning model; sputum aspiration; bleeding volume

声门型喉癌是喉癌最常见的分型,手术是其重要的治疗手段<sup>[1-2]</sup>,但大多数喉癌患者手术过程中需切开气管,为维持患者术中呼吸,需要建立人工气道,术后可能出现下呼吸道感染等并发症,干扰患者术后康复及肺功能恢复<sup>[3]</sup>。因此,建立感染预警模型对预防声门型喉癌患者术后下呼吸道感染具有重要意义。

喉癌患者术后下呼吸道感染率为 8%~13%<sup>[4]</sup>,感染和非感染人数不平衡。传统统计分析方法对不平衡数据处理效果欠佳<sup>[5]</sup>,对多数类预测效果高,而对少数类预测效果较差,不适合喉癌患者术后下呼吸道感染预警模型构建。为解决数量不平衡数据分类问题,有研究者引入重采样技术重建数据平衡少数类与多数类的的数据数量。少数类样本合成过抽样技术(Synthetic Minority Over-sampling Technique, SMOTE)由 Chawla 等<sup>[6]</sup>于 2011 年提出,其根据少数类属性特征随机产生相似样本,有效地平衡少数类与多数类之间数量,减少数据集过度倾斜。SMOTE 算法已成功用于建立脑损伤继发精神障碍<sup>[7]</sup>、绝经后骨质疏松性

作者单位:河南省人民医院耳鼻咽喉头颈外科,郑州大学人民医院(河南 郑州,450003)

马灵草,女,本科,主管护师

通信作者:董婷,2897280480@qq.com

科研项目:河南省医学科技攻关计划(201702196)

收稿:2020-11-08;修回:2020-12-30

骨折<sup>[8]</sup>等少数类预警模型。本研究收集接受肿瘤根治手术的声门型喉癌患者临床资料,构建基于SMOTE算法的感染预警模型,旨在为预防喉癌患者术后上呼吸道感染提供参考。

### 1 资料与方法

**1.1 一般资料** 收集2015年1月至2019年1月河南省人民医院收治的接受肿瘤根治手术声门型喉癌患者的临床资料。纳入标准:①经病理和影像诊断确诊为喉门型细胞癌;②年龄≥18岁;③临床资料完整,病历中记录有术后下呼吸道感染状况;④全喉或部分喉切除治疗;⑤手术达到根除原发癌灶标准。排除标准:①术前已合并呼吸系统疾病,先天性免疫系统缺陷和其他恶性肿瘤;②术后转入ICU昏迷、病危或病亡;③术后转院或临床资料不完整。共收集有效资料596例,男326例,女270例;未婚58例,已婚403例,离异100例,丧偶35例;小学155例,中学(含中专)265例,大专以上176例;居住地为城市296例,农村300例;使用抗生素时间1~7(3.28±0.64)d。肿瘤分期:Ⅱ期121例,Ⅲ期272例,Ⅳ期203例。有饮酒史369例,患有高血压266例。

### 1.2 方法

**1.2.1 收集临床资料** 研究者使用自制调查表收集喉癌患者信息:①人口学信息,包括性别、年龄、学历、BMI值、癌症分期等。②手术情况,包括手术类型、手术时间、术中出血量等。③术后情况,包括使用抗生素、吸痰、雾化、使用呼吸机、术后拔管时间,使用人工鼻情况和术后是否存在下呼吸道感染等。

**1.2.2 术后下呼吸道感染诊断标准** ①胸部X线检出支气管或肺部炎症改变;②静脉血检出白细胞数和(或)嗜中性粒细胞比例上升;③下呼吸道分泌物培养试验检出菌株超标。患者术后1周内出现满足①②或①③者可判断为术后下呼吸道感染<sup>[9]</sup>。

**1.2.3 实施SMOTE过抽样算法** SMOTE算法具体按照Blagus等<sup>[10]</sup>界定的7个步骤实施。本研究少数样本为感染组,增加样本倍数 $n = \text{未感染组样本量} / \text{感染组样本量} = 511 / 71 \approx 7$ ,最近邻点 $k$ 取6,SMOTE算法通过对少数类样本点与其最近邻样本点间进行随机插值,实现对少数类样本的扩充,产生的新样本不会改变原样本集的空间边界,具有较好的算法信度。

**1.2.4 统计学方法** 采用SPSS 25.0软件和R 3.5.2软件进行 $t$ 检验、 $\chi^2$ 检验、Logistic回归分析。在完成模型构建后,模型效能评估采用十折交叉法验证,效能评估指标包括真阳性率、阳性预测值、 $F$ 值、ROC曲线下面积。检验水准 $\alpha = 0.05$ 。

## 2 结果

**2.1 声门型喉癌患者术后下呼吸道感染情况** 596例患者中,38例(感染组)术后发生下呼吸道感染,共查出菌株64株,其中革兰阴性菌44株(铜绿甲单胞

菌24株,肺炎克雷伯菌10株,大肠埃希菌4株,阴沟肠杆菌和鲍氏不动杆菌各3株),革兰阳性菌20株(金黄色葡萄球菌13株,凝固酶阴性葡萄球菌7株)。

### 2.2 术后患者下呼吸道感染影响因素的单因素分析

单因素比较显示,两组性别、婚姻状态、学历、居住地、使用抗生素时间、癌症分期、饮酒史和高血压比较差异无统计学意义(均 $P > 0.05$ ),两组有统计学差异的项目及比较见表1。

表1 术后患者下呼吸道感染影响因素的单因素分析

项目	非感染 ( $n=558$ )	感染 ( $n=38$ )	$t/\chi^2$	$P$
年龄(岁, $\bar{x} \pm s$ )	43.23±10.06	55.21±10.13	7.097	0.000
BMI( $\bar{x} \pm s$ )	15.92±2.42	14.37±2.40	-3.893	0.000
吸烟史(例)				
有	212	22	5.909	0.015
无	346	16		
糖尿病(例)				
有	144	24	24.520	0.000
无	414	14		
肺部疾病(例)				
有	131	25	32.963	0.000
无	427	13		
手术类型(例)				
部分喉切除	363	11	19.843	0.000
全喉切除	195	27		
术后雾化(例)				
是	201	20	4.207	0.040
否	357	18		
术后吸痰(例)				
是	100	25	49.185	0.000
否	458	13		
术后气道导管留置时间(例)				
<平均值	190	26	18.189	0.000
≥平均值	368	12		
术中出血量 (mL, $\bar{x} \pm s$ )	67.85±6.82	70.71±6.26	2.956	0.003
手术时间 (min, $\bar{x} \pm s$ )	42.80±4.16	44.29±4.20	2.126	0.034
使用人工鼻(例)				
是	186	25	16.388	0.000
否	372	13		

注:术后气道导管留置时间为1~5(2.45±0.58)d。

**2.3 术后下呼吸道感染影响因素的Logistic回归分析** 以是否术后下呼吸道感染为因变量 $Y$ (否=0,是=1),以单因素分析有统计学差异的变量为自变量行Logistic回归分析, $\alpha_{\text{入}} = 0.05$ , $\alpha_{\text{出}} = 0.10$ 。结果显示年龄 $X_1$ (原值)、BMI $X_2$ (原值)、吸烟史 $X_3$ (赋值:无=0,有=1)、糖尿病 $X_4$ (无=0,有=1)、肺部疾病 $X_5$ (无=0,有=1)、手术类型 $X_6$ (部分喉切除=1,全喉切除=2)、术后吸痰 $X_7$ (否=0,是=1)、术中出血量 $X_8$ (原值)、术后气道导管留置时间 $X_9$ (<平均值=0,≥平均值=1)、使用人工鼻 $X_{10}$ (否=0,是=1)为术后下呼吸道感染的主要影响因素。得到原始预警模型1:  $\text{Logit}(P) = 0.055X_1 - 0.195X_2 + 1.383X_3 +$

0.968X4+0.939X5+0.993X6+1.055X7+0.076X8+0.919X9+0.978X10-14.073,经 Hosmer-Lemeshow 检验结果显示, Logistic 回归模型拟合度良好 ( $P=0.727$ , 决定系数  $R^2=0.584$ )。

**2.4 基于 SMOTE 算法的预警模型** 选择 Logistic 回归分析纳入的影响因素, 利用 SMOTE 算法进行 14 倍 (558/38) 过抽样, 得到感染和非感染比例近似为 1 (实际抽样为非感染 = 558, 感染 = 570, 感染/非感染 = 1.02), 再对过抽样后数据进行多因素 Logistic 回归,  $\alpha_{入}=0.05$ ,  $\alpha_{出}=0.10$ 。分析结果见表 2。基于 SMOTE 算法的预警模型 2:  $\text{Logit}(P)=0.090X1-0.249X2+1.724X3+1.136X4+1.344X5+0.967X6+1.092X7+0.126X8+1.340X9+1.456X10-14.859$ , 经 Hosmer-Lemeshow 检验结果显示, 模型拟合度良好 ( $P=0.823$ , 决定系数  $R^2=0.625$ )。

表 2 基于 SMOTE 算法的 Logistic 回归分析结果

自变量	$\beta$	SE	Wald $\chi^2$	P	OR	95%CI
常量	-14.859	1.749	72.136	0.000	-	-
年龄	0.090	0.014	43.597	0.000	1.094	1.065~1.123
BMI	-0.249	0.046	29.354	0.000	0.780	0.713~0.853
吸烟史	1.724	0.223	59.558	0.000	5.606	3.618~8.685
糖尿病	1.136	0.229	24.563	0.000	3.115	1.988~4.883
肺部疾病	1.344	0.208	41.626	0.000	3.833	2.548~5.764
手术类型	0.967	0.218	19.602	0.000	2.630	1.714~4.035
术后吸痰	1.092	0.240	20.620	0.000	2.980	1.860~4.775
术中出血量	0.126	0.020	39.743	0.000	1.134	1.090~1.179
气管接管时间	1.340	0.208	41.691	0.000	3.820	2.543~5.737
使用人工鼻	1.456	0.217	45.083	0.000	4.290	2.804~6.563

**2.5 模型评估** 使用 ROC 曲线验证可得, 模型 2 的 ROC 曲线下面积为 0.927 (95%CI: 0.888~0.966), 显著高于模型 1 的 ROC 曲线下面积 (0.780, 95%CI: 0.734~0.846)。将模型所对应的数据分为 10 份, 每个模型轮流依次以其中 1 份为测试集, 其余 9 份为训练集进行 10 次检验, 采用 10 折交叉法对 2 个模型进行预测性能比较, 见表 3。

表 3 2 个预测模型在测试集上预测性能比较

模型	次数	真阳性率	阳性预测值	F	ROC 曲线下面积
模型 1	10	0.722±0.101	0.601±0.092	0.654±0.096	0.715±0.082
模型 2	10	0.709±0.121	0.799±0.093	0.751±0.105	0.906±0.093
t		0.854	4.117	2.756	4.208
P		0.807	0.002	0.013	0.001

### 3 讨论

#### 3.1 声门型喉癌患者术后发生下呼吸道感染率分析

本研究结果显示, 声门型喉癌患者术后发生下呼吸道感染率为 6.38%, 低于尹国华等<sup>[11]</sup>报道的 11.46%、柴丽娟等<sup>[12]</sup>报道的 12.75%, 这可能因为本研究收集病例中, 超过 50% 接受了相应的抗生素治疗, 且均为早期喉癌患者, 肿瘤细胞对其免疫系统损害程度相对较轻, 患者自身抵抗感染能力较强。本研究术后下呼吸道感染患者查出菌株主要为革兰阴性菌, 占整体检

出菌类型的 68.75%, 整体检出菌株主要为铜绿假单胞菌和金黄色葡萄球菌, 与他人研究结果<sup>[13-14]</sup>相符。

#### 3.2 声门型喉癌患者术后下呼吸道感染的影响因素

##### 3.2.1 患者自身因素

本研究发现, 年龄和 BMI 是声门型喉癌患者术后下呼吸道感染的显著影响因素 (均  $P<0.01$ ), 与他人研究相符<sup>[15-16]</sup>, 表现出患者年龄越大, BMI 值越低, 其术后下呼吸道感染风险越大。这是因为人体本身的免疫功能随年龄增长而自然性下降, 造成高龄喉癌患者抵御感染菌侵袭能力较低, 增加了其术后感染风险; 而 BMI 低于正常范围的患者机体可能存在营养不良, 从而导致其自身免疫功能及抗感染效率受到限制。本研究结果提示, 吸烟史、肺部疾病均是声门型喉癌患者术后下呼吸道感染的显著影响因素 (均  $P<0.01$ ), 说明有吸烟史和合并肺部疾病患者术后下呼吸道感染风险越高, 与他人研究相符<sup>[17]</sup>。这可能因为长期吸烟能损伤患者支气管黏膜, 导致呼吸道纤毛变短、断裂和运动功能下降, 降低支气管清除感染病原体的能力, 吸烟也能增加呼吸道内诱导痰内炎性因子表达, 恶化感染程度, 而各种肺部疾病可致呼吸系统组织及器官长期受到炎性环境刺激, 造成支气管黏膜腺体增生并分泌大量黏液, 堵塞喉癌患者细支气管, 从而在下呼吸道内形成易感染菌定殖环境, 增加了患者的下呼吸道感染率。糖尿病是声门型喉癌患者术后下呼吸道感染的显著影响因素 (均  $P<0.01$ ), 这是因为手术伤害或癌症等强刺激源可致合并糖尿病喉癌患者机体内激素调节和细胞因子分泌异常, 从而出现应激性高血糖<sup>[18-19]</sup>, 高糖环境能增高喉癌患者机体内反应性氧化物表达, 诱导各种促炎因子表达上升, 产生炎性级联反应诱发和恶化其下呼吸道感染。

##### 3.2.2 手术因素

本研究结果显示, 全喉切除患者术后下呼吸道感染风险高于部分喉切除者 ( $P<0.01$ )。分析原因为全喉切除手术时切除组织范围大, 术后切口较长致下呼吸道直接暴露于外部环境, 易于细菌侵入呼吸道; 此外, 喉全切患者留置呼吸气管套管时间更久, 从而增加下呼吸道感染风险。术中出血量也是声门型喉癌患者术后下呼吸道感染的影响因素 ( $P<0.01$ ), 分析原因为术中出血量越多, 患者受到手术损伤越大, 生理状态越差, 术后身体免疫功能受到的限制越明显, 降低患者术后呼吸道及其他易感组织或器官的自抗感染能力。

##### 3.2.3 术后气道管理因素

研究显示, 术后气道导管留置时间  $>2$  h 可致气管内壁纤毛倒伏、破坏,  $>6$  h 纤毛坏死、脱落, 从而导致感染病菌易移生于下呼吸道诱发感染<sup>[20]</sup>。本研究结果也显示, 术后吸痰、术后气道导管留置时间和使用人工鼻均是声门型喉癌患者术后下呼吸道感染的影响因素 (均  $P<0.01$ ), 表现出术后接受吸痰、术后导管留置时间越久和使用人工鼻患者下呼吸道感染风险越大, 这是

因为吸痰虽能吸出患者呼吸道内分泌物,保持气道畅通,但吸痰过程中吸痰管可能把口腔内菌群带入下呼吸道和损伤呼吸道黏膜,引起呼吸道感染。而喉癌患者术后需经插管和使用人工鼻建立辅助呼吸气道,也可能造成置入通气导管损伤呼吸道黏膜,破坏呼吸道防御屏障。

**3.3 对临床护理的指导价值** 表 3 显示,2 个预警模型对训练集真阳性率差异无统计学意义,但模型 2 对训练集阳性预测值、*F* 值和 ROC 曲线下面积显著高于原始 Logistic 回归预警模型 ( $P < 0.05, P < 0.01$ ),说明使用 SMOTE 算法构建的新预警模型预测效应更优秀。按照该预警模型,护士应根据患者基本情况及合并基础疾病情况进行分级护理,积极治疗护理各种基础疾病及肺部疾病;在围术期做好患者口腔、鼻腔、手术切口及留置管切口消毒和抗感染护理,吸痰时严格遵守无菌原则,必要时采取超声雾化吸入方式湿化气道,针对有高危感染风险者密切监测其感染指标,术后首日开始采集下呼吸道分泌物进行微生物培养,基于试验结果决定选择敏感抗生素,强化呼吸道管理,同时做好环境清洁消毒;指导家属准备适宜的食谱,选择每日少量多餐的方式以提高患者肠道吸收率,提高免疫力<sup>[21]</sup>。

**4 小结**

声门型喉癌患者术后下呼吸道感染受年龄、BMI、吸烟史、糖尿病、肺部疾病、全喉切除、术后吸痰、术中出血量、术后拔管时间和使用人工鼻影响。使用 SMOTE 算法对数据进行重采样技术重建,得到的预警模型比传统 Logistic 回归建立的模型预测效应更优。本研究选择研究对象为声门型喉癌患者,未对中晚期患者进行研究,也未分析喉癌远期相关因素,需要进一步研究以完善。

**参考文献:**

[1] 杜灵彬,毛伟敏,陈万青,等. 中国 2003—2007 年喉癌发病率和死亡率分析[J]. 中华流行病学杂志,2012,33(4): 395-398.

[2] 黄冠江,罗梦思,张靖萱,等. T<sub>1</sub>(1a)期声门型喉癌的外科治疗进展[J]. 临床耳鼻咽喉头颈外科杂志,2017,31(9):76-79.

[3] 吴莉,邱云,姜妍,等. 喉癌全切手术后气管感染对肺功能影响的临床观察[J]. 临床肺科杂志,2017,22(3):496-498.

[4] 童巧玲,沈志森,应焱燕,等. 喉癌患者手术治疗的远期疗效及影响因素分析[J]. 浙江医学,2018,40(23):17-20.

[5] 孙涛,吴海丰,梁志刚,等. SMOTE 算法在不平衡数据中

的应用[J]. 北京生物医学工程,2012,31(5):528-530.

[6] Chawla N V, Bowyer K W, Hall L O, et al. SMOTE: Synthetic Minority Over-sampling Technique[J]. J Art Intelli Res,2011,16(1):321-357.

[7] 孙秀彬,辛涛,薛付忠,等. 基于 SMOTE 算法的颅脑损伤患者继发精神障碍预警模型[J]. 中国卫生统计,2013,30(6):790-793.

[8] 章轶立,魏戎,聂佩芸,等. 基于 SMOTE 算法和决策树的绝经后骨质疏松性骨折分类模型建构[J]. 中国骨质疏松杂志,2019,25(1):7-11.

[9] 胡必杰. 下呼吸道感染实验诊断规范[M]. 上海:上海科学技术出版社,2006:45-46.

[10] Blagus R, Lusa L. SMOTE for high-dimensional class-imbalanced data[J]. BMC Bioinformatics,2013,14(1):1-16.

[11] 尹国华,岳文江,付高洁,等. 喉癌术后下呼吸道感染危险因素与护理管理[J]. 临床耳鼻咽喉头颈外科杂志,2017,16(2):149-151.

[12] 柴丽娟,姚爱勤,陈伟,等. 喉癌术后气管切开患者下呼吸道感染的易感因素与预防[J]. 中华医院感染学杂志,2018,28(14):140-143.

[13] 孙秀霞,丁元平,季民,等. 喉癌患者喉部分切除术后感染调查研究[J]. 中华医院感染学杂志,2015,25(21):4962-4964.

[14] 毛振宇,江涛. 喉癌术后并发下呼吸道感染的调查分析[J]. 中华全科医学,2015,13(8):178-180.

[15] 郇振伟,曹晓林,李勇,等. 喉癌患者术后医院感染相关因素及病原学分析[J]. 中华医院感染学杂志,2017,27(13):3044-3047.

[16] 韩瑞玲,周燕菱,吕梦颖. 喉癌手术后气管切开后下呼吸道感染的护理干预[J]. 中华医院感染学杂志,2013,23(23):5735-5736.

[17] 朱文伟,丁滢燕,陈志华. 老年患者气管插管全麻后下呼吸道感染相关因素分析[J]. 中华医院感染学杂志,2013,23(7):1586-1587.

[18] 谭智文,杨丽君,白玛扎西,等. 不同 BMI 分级对藏族恶性肿瘤住院患者生活质量的影响[J]. 现代肿瘤医学,2017,25(13):91-94.

[19] 祖晓玲. 随机森林算法预测医院患者院内感染的应用研究[D]. 唐山:燕山大学,2016.

[20] Gong H L, Zhou L, Wu H T, et al. Long-term clinical outcomes of supracricoid partial laryngectomy with cricohyoidoepiglottopexy for glottic carcinoma [J]. Acta Otolaryngol,2019,139(9):803-809.

[21] 杜馨雯,冯先琼,胡紫宜,等. 喉癌患者手术前后症状变化及其影响因素分析[J]. 护理学杂志,2019,34(20):20-23.

(本文编辑 韩燕红)