

• 论 著 • DOI:10.3969/j.issn.1672-9455.2020.22.005

基于血清肿瘤标志物的 Logitboost 模型用于肺腺癌和肺鳞癌鉴别的初步研究*

张桐睿¹, 申 敏², 王广丽², 李 军^{2△}, 宋歌声²

1. 山东第一医科大学研究生院, 山东泰安 271016; 2. 山东第一医科大学第一附属医院/
山东省千佛山医院放射科, 山东济南 250014

摘 要:目的 利用肺癌血清肿瘤标志物建立 Logitboost 模型,并探讨该模型用于鉴别肺鳞癌与肺腺癌的可行性。**方法** 选取 2018 年 1 月至 2019 年 6 月山东第一医科大学第一附属医院收治并经病理学证实为肺癌的患者 222 例,包括肺腺癌 125 例(肺腺癌组)、肺鳞癌 97 例(肺鳞癌组)。检测并记录两组患者血清癌胚抗原(CEA)、鳞状上皮细胞癌抗原(SCC)、神经元特异性烯醇化酶(NSE)、胃泌素释放肽前体(pro-GRP)、细胞角蛋白 19 片段(CYFRA21-1)表达水平。以肿瘤分型作为因变量,血清肿瘤标志物水平作为自变量,拟合 Logitboost 模型,采用 10 折交叉验证对模型进行评价。**结果** 肺腺癌组患者 SCC、NSE、pro-GRP、CYFRA21-1 水平低于肺鳞癌组,肺腺癌组 CEA 水平高于肺鳞癌组,差异均有统计学意义($P<0.05$)。SCC、NSE、CEA、pro-GRP、CYFRA21-1 用于鉴别肺腺癌和肺鳞癌的受试者工作特征(ROC)曲线的曲线下面积(AUC)分别为 0.642、0.577、0.647、0.579、0.627。利用 5 种肺癌血清肿瘤标志物拟合 Logitboost 模型,训练集 AUC 为 0.882,10 折交叉验证后,验证集预测准确率为 71.9%,AUC 为 0.784。**结论** 利用肺癌血清肿瘤标志物建立的 Logitboost 模型对鉴别肺鳞癌和肺腺癌有一定价值。

关键词:血清肿瘤标志物; 肺鳞癌; 肺腺癌; Logitboost 模型

中图分类号:R734.2 文献标志码:A 文章编号:1672-9455(2020)22-3244-03

A preliminary study on the differencetial diagnosis between lung adenocarcinoma and lung squamous cell carcinoma by Logitboost model based on serum tumor markers*

ZHANG Tongrui¹, SHEN Min², WANG Guangli², LI Jun^{2△}, SONG Gesheng²

1. Graduate School, Shangdong First Medical University, Taian, Shandong 271016, China ;
2. Department of Radiology, Shandong Provincial Qianfoshan Hospital/the First
Affiliated Hospital of Shandong First Medical University, Jinan, Shandong 250014, China

Abstract: Objective To establish Logitboost model with lung cancer serum tumor markers, and to explore the feasibility of the model in distinguishing squamous cell carcinoma from adenocarcinoma. **Methods** A total of 222 cases of lung cancer confirmed by pathology were collected, including 97 cases of squamous cell carcinoma (lung squamous cell carcinoma group) and 125 cases of adenocarcinoma (lung adenocarcinoma group). The expression levels of CEA, SCC, NSE, pro-GRP and CYFRA21-1 were detected and recorded for the two groups. Taking tumor classification as the dependent variable, serum tumor marker level as the independent variable, the Logitboost model was used to evaluate the effect of model fitting, and the model was evaluated by 10 fold cross validation. **Results** The expression level of SCC, NSE, pro-GRP, CYFRA21-1 in lung adenocarcinoma group was lower than that in lung squamous cell carcinoma group, and the expression level of CEA in lung adenocarcinoma group was higher than that in lung squamous cell carcinoma group, the difference was statistically significant ($P<0.05$). The AUC values of SCC, NSE, CEA, pro-GRP, CYFRA21-1 in distinguishing lung adenocarcinoma from lung squamous cell carcinoma were 0.642, 0.577, 0.647, 0.579, 0.627, respectively. Five lung cancer serum markers were used to fit the Logitboost prediction model, the prediction accuracy of validation set was 71.9%, and AUC was 0.784. **Conclusion** The Logitboost prediction model based on lung cancer serum tumor markers has certain value in distinguishing squamous cell carcinoma from lung adenocarcinoma.

Key words: serum tumor markers; squamous cell carcinoma; adenocarcinoma; Logitboost model

* 基金项目:山东省自然科学基金项目(ZR2015HL096)。
作者简介:张桐睿,女,硕士研究生在读,主要从事 CT、MRI 诊断的相关研究。△ 通信作者, E-mail:lijun69413@aliyun.com。

近年来,肺癌的发病率逐年升高,严重威胁人类健康,也给患者家庭带来沉重的负担^[1-2]。肺癌患者早期临床症状并无明显特征,极易被误诊、漏诊,等到发现时可能已是中、晚期,甚至发生了转移,生活质量和疗效大大降低^[3]。有研究表明,在临床诊断肿瘤时应用生物标志物检测可达到理想效果,因此,血清肿瘤标志物在肺癌的早期诊断中受到越来越多的重视,血清癌胚抗原(CEA)、鳞状上皮细胞癌抗原(SCC)、神经元特异性烯醇化酶(NSE)、胃泌素释放肽前体(pro-GRP)、细胞角蛋白 19 片段(CYFRA21-1)在肺癌早期诊断、疗效监测及预后判断等方面的作用明显^[4-5]。本文旨在利用肺癌患者血清中 CEA、SCC、NSE、pro-GRP、CYFRA21-1 的表达水平,建立二分类 Logitboost 模型,并评价该模型在鉴别肺鳞癌和肺腺癌中的准确性。

1 资料与方法

1.1 一般资料 选取 2018 年 1 月至 2019 年 6 月山东第一医科大学第一附属医院收治并经病理学证实为肺癌的患者 222 例作为研究对象,男 142 例,女 80 例,年龄 28~90 岁。纳入标准:(1)初诊患者;(2)未进行过手术、放疗和化疗;(3)未并发血液系统疾病及其他部位恶性肿瘤。入选的 222 例患者中,肺腺癌 125 例(肺腺癌组),肺鳞癌 97 例(肺鳞癌组)。两组研究对象的性别、年龄比较,差异均无统计学意义($P>0.05$)。

1.2 方法 在清晨空腹状态下采集两组患者静脉血 5 mL,静置 20 min,然后 3 000 r/min 离心 5 min 分离血清,于-20 ℃冰箱保存。血清 NSE、CYFRA21-1 水平采用电化学发光法检测,使用的仪器设备及试剂为 Roche Cobas E602 全自动电化学发光分析仪及厂家的配套试剂;CEA、SCC、pro-GRP 水平检测采用化学发光法,所用仪器设备和试剂为雅培 i2000 化学发光分析仪及厂家配套试剂。按说明书要求进行操作。

1.3 统计学处理 应用 SPSS20.0 和 Weka3.8.3 软件进行数据分析。肿瘤标志物水平以 $M(P_{25} \sim P_{75})$

表示,组间比较采用 Wilcoxon 两样本秩和检验。采用受试者工作特征(ROC)曲线分析血清肿瘤标志物对肺腺癌和肺鳞癌的鉴别诊断效能。将 5 种肺癌血清肿瘤标志物的检测数据用于拟合 Logitboost 模型,采用 10 折交叉验证对模型进行验证,并利用多种参数对模型效果进行评价。 $P<0.05$ 为差异有统计学意义。

2 结果

2.1 两组血清肿瘤标志物的比较 5 种肺癌血清肿瘤标志物中,SCC、NSE、pro-GRP、CYFRA21-1 水平:肺腺癌组低于肺鳞癌组,差异有统计学意义($P<0.05$);CEA 水平:肺腺癌组高于肺鳞癌组,差异有统计学意义($P<0.05$),见表 1。SCC、NSE、CEA、pro-GRP、CYFRA21-1 用于鉴别肺腺癌和肺鳞癌的 ROC 曲线下面积(AUC)分别为 0.642、0.577、0.647、0.579、0.627。

2.2 5 种肺癌血清肿瘤标志物拟合 Logitboost 模型 将 5 种肺癌血清肿瘤标志物拟合 Logitboost 模型后,训练集 AUC 为 0.882;10 折交叉验证后,验证集预测准确率为 71.9%,AUC 为 0.784。见表 2。

表 1 5 种肺癌血清肿瘤标志物在两组中的结果分析

指标	组别	n	检测水平	Z	P
			$[M(P_{25} \sim P_{75}), \text{ng/mL}]$		
SCC	肺腺癌组	125	0.70(0.40~0.90)	-3.146	0.002
	肺鳞癌组	97	1.00(0.40~2.55)		
NSE	肺腺癌组	125	13.61(9.24~18.21)	-1.966	0.049
	肺鳞癌组	97	15.65(11.34~20.54)		
CEA	肺腺癌组	125	5.75(2.43~31.66)	-3.733	0.000
	肺鳞癌组	97	3.16(3.33~5.06)		
pro-GRP	肺腺癌组	125	35.21(30.41~46.57)	-2.013	0.044
	肺鳞癌组	97	40.60(31.67~52.03)		
CYFRA21-1	肺腺癌组	125	3.29(2.26~5.70)	-3.225	0.001
	肺鳞癌组	97	4.99(2.90~12.84)		

表 2 Logitboost 模型训练集、验证集结果

项目	MCC	AUC	类型	真阳性率	假阳性率	准确率	召回率	F
训练集	0.604	0.882	肺腺癌	0.864	0.268	0.806	0.864	0.834
			肺鳞癌	0.732	0.136	0.807	0.732	0.768
			平均	0.806	0.210	0.806	0.806	0.805
验证集	0.428	0.784	肺腺癌	0.784	0.361	0.737	0.784	0.760
			肺鳞癌	0.639	0.216	0.697	0.639	0.667
			平均	0.721	0.298	0.719	0.721	0.719

注:MCC 表示马修斯相关系数。

3 讨论

肿瘤标志物是反映肿瘤存在和增殖的一类生物

活性物质。机体处于正常状态下,组织中的肿瘤标志物水平很低,而肿瘤组织中肿瘤标志物水平通常远远

超过正常组织^[6]。此外,肿瘤标志物检测在肺癌的早期诊断、治疗监测及预后判断中具有重要意义^[7]。由于单一肿瘤标志物诊断特异度和灵敏度较低,因此,一般采用多种肿瘤标志物联合检测的方法来提高对恶性肿瘤诊断的灵敏度和特异度,以期在降低患者经济负担的同时,准确、有效地鉴别肿瘤的良、恶性^[8]。

CYFRA21-1 作为细胞角蛋白家族中具有重要作用的成员,当肿瘤细胞死亡或者发生溶解时,血清中 CYFRA21-1 水平明显升高,尤其在肺鳞癌中的表达较高,能作为肺癌诊断的一个首选标志物^[9]。CEA 作为肺腺癌诊断标志物之一,在肺腺癌患者中的水平较肺鳞癌和小细胞癌患者更高。NSE 作为烯醇化酶的同工酶,不仅可以有效提示神经母细胞瘤,而且能明确提示小细胞肺癌^[10]。SCC 是应用单克隆技术从相关抗原 TA4 中提纯的糖蛋白片段,存在于子宫、宫颈、肺等鳞状上皮细胞癌的细胞质中,早期肺癌患者血清中 SCC 的阳性表达率低,随着肺癌的发展而不断提高^[11-12]。有研究表明,肺癌患者病情越严重,血清 SCC 水平越高^[13]。而作为促胃泌素释放肽前体物质的 pro-GRP 是一种新型的肿瘤标志物,主要分布在神经纤维组织、脑部及肺部组织的神经内分泌细胞中。pro-GRP 在正常上皮、良性肺疾病和上皮来源的肿瘤中表达水平低,而在肺癌组织(尤其是小细胞肺癌)中表达水平极高^[14]。pro-GRP 对于小细胞肺癌诊断具有重要作用,其特异度和灵敏度甚至超过 NSE^[15]。由于肺癌组织病理的多样性和抗原表达的异质性,肺癌的早期诊断仅靠单项肿瘤标志物检测是无法实现的,而肿瘤标志物的联合检测可以提高肺癌的检出率,降低漏诊率^[16]。

Logitboost 算法是一种基于机器学习的判别分类方法,在不能满足传统判别方法应用条件的数据中优势明显。本研究建立了二分类 Logitboost 模型,采用 10 折交叉验证对模型效果进行评价。结果显示,将 5 种肿瘤标志物的检测数据拟合 Logitboost 模型后,训练集 AUC 为 0.882,验证集 AUC 为 0.784,二者 AUC 接近,说明 Logitboost 模型用于鉴别肺腺癌与肺鳞癌具有可行性。

综上所述,通过对肺癌患者血清 CEA、SCC、NSE、pro-GRP、CYFRA21-1 5 种肿瘤标志物的联合检测,建立 Logitboost 模型,有助于肺腺癌和肺鳞癌的鉴别。

参考文献

[1] CHEN W Q,ZHENG R S,ZENG H M,et al. The incidence and mortality of major cancers in China,2012[J]. Chin J Cancer,2016,35(1):73-78.
[2] 杜汉芳,许海军,林丽. 多种血清肿瘤标志物联合检测在

肺癌诊断中的应用研究[J]. 国际检验医学杂志,2019,40(1):89-91.
[3] WANG L,WANG D,ZHENG G,et al. Clinical evaluation and therapeutic monitoring value of serum tumor markers in lung cancer[J]. Int J Biol Markers,2016,31(1):e80-e87.
[4] 陈维. 血清 ProGRP、TSGF、CYFRA21-1、NSE、CEA、CA125 的优化组合在早期肺癌诊断中的价值[J]. 数理医药学杂志,2016,26(4):495-497.
[5] MIE M,KAI T,LE T,et al. Selection of DNA aptamers with affinity for pro-gastrin-releasing peptide (proGRP), a tumor marker for small cell lung cancer[J]. Appl Biochem Biotechnol,2013,169(1):250-255.
[6] 赖其廷,李敏菁,罗志扬. 多种肿瘤标志物联合检测在肺癌诊断中的应用价值分析[J]. 疾病监测与控制,2018,12(1):19-20.
[7] HUANG C,MA R,XU Y,et al. Wnt2 promotes non-small cell lung cancer progression by activating WNT/ β -catenin pathway[J]. Am J Cancer Res,2015,5(3):1032-1046.
[8] WU Y L,SUN Y,ZHOU CC,et al. Survival without common toxicity criteria grade 3/4 toxicity following second-line treatment with pemetrexed for nonsquamous non-small cell lung cancer in Chinese patients[J]. Chin Med J (Engl),2013,126(24):4624-4628.
[9] 薛雯娟,康艳. 血清肿瘤标志物在肺癌诊断中的应用价值[J]. 中国肿瘤临床与康复,2019,26(10):1212-1215.
[10] 罗文娟,刘广国,苏州,等. 五种血清肿瘤标志物在肺癌诊断中的临床应用[J]. 临床和实验医学杂志,2018,17(18):1967-1971.
[11] 王俊杰. 联合检测 ProGRP 和 NSE 对小细胞肺癌的诊断意义[J]. 实验与检验医学,2017,35(3):378-379.
[12] 廖莎莎. 肿瘤标志物检验在肺癌检验与诊断中的应用效果观察[J]. 当代医学,2019,25(34):119-121.
[13] MA L,XIE X W,WANG H Y,et al. Clinical evaluation of tumor markers for diagnosis in patients with non-small cell lung cancer in China[J]. Asian Pac J Cancer Prev,2015,16(12):4891-4894.
[14] MOLINA R,MARRADES R M,MARIA AUGÉ J A,et al. Assessment of a combined panel of six serum tumor markers for lung cancer[J]. Am J Respir Crit Care Med,2016,193(4):427-437.
[15] 崔连鸾,高海燕,朱华,等. 多肿瘤标志物联合检测在肺癌诊断中的应用价值[J]. 中国实验诊断学,2013,33(12):2221-2222.
[16] 倪军,郭子健,张力. 单独与联合检测四项肺癌血清肿瘤标志物在肺癌诊断中的价值[J]. 中华内科杂志,2016,55(1):25-30.