

论著·临床研究

决策树分析在提高血清维生素 D、CA153、CA125、CEA 对乳腺癌辅助诊断作用中的应用分析*

尹志辉¹,周志伟²,刘继勇²,杨保昌²,王凌凌³

(1. 河北省邢台市第三医院检验科,河北邢台 054000;2. 中国石油天然气集团公司中心医院检验科,河北廊坊 065000;3. 中国石油天然气集团公司中心医院肿瘤科,河北廊坊 065000)

摘要:目的 探讨血清指标建立决策树模型在乳腺癌患病诊断中的应用。方法 选取 2016 年 8 月至 2018 年 4 月在中国石油天然气集团公司中心医院就诊的由组织病理学确诊的新发乳腺癌患者 60 例,作为乳腺癌组;选取同期体检非肿瘤、非生殖内分泌系统疾病的女性患者 60 例,作为非乳腺癌组;再选取本研究外首次就诊乳腺科患者 60 例组成验证组,其中乳腺癌 30 例,非乳腺癌 30 例。检测两组患者肿瘤标志物糖链抗原 153 (CA153)、CA125、癌胚抗原(CEA)和 25-羟基维生素 D,首先对两组患者检测结果进行统计学分析,然后对所有检测数据建立决策树模型,分析各指标在辅助诊断乳腺癌中作用,收集验证组患者相关指标应用受试者工作曲线(ROC 曲线)对模型的临床诊断效能进行评价。**结果** (1)乳腺癌组未绝经患者维生素 D 水平低于对照组,乳腺癌组绝经患者 CEA 水平高于非乳腺癌组患者,差异具有统计学意义($P<0.05$);其他指标差异无统计学意义($P>0.05$);(2)进行自动建模分析得出,决策树 100.0%预测能力最佳模型为 CHAID 决策树,该模型显示当维生素 D 水平小于或等于 7.2 ng/mL 时,对乳腺癌阳性预测达到 100.0%;当维生素 D 水平大于 7.2 ng/mL 且小于或等于 19.0 ng/mL 时,加入是否绝经变量,对于未绝经患者阴性预测能力 76.0%,绝经患者阳性预测能力 65.7%;当维生素 D 水平大于 19.0 ng/mL 时,且 CEA 结果小于或等于 2.48 U/mL 时阴性预测能力 87.9%,且 CEA 结果大于 2.48 U/mL 时阳性预测能力 100.0%;通过 ROC 曲线分析得出决策树模型较多指标串联分析在筛查特异度方面提升明显。**结论** 决策树模型分析可通过对检验指标递进分类,以树形式组成不同判断集合,增强了检测指标在乳腺癌中的辅助诊断价值。

关键词:决策树分析; 乳腺癌; 维生素 D; 诊断价值

DOI:10.3969/j.issn.1673-4130.2019.20.012 中图法分类号:R737.9

文章编号:1673-4130(2019)20-2482-04 文献标识码:A

Application of decision tree analysis in the improvement of serum vitamin D,CA153,CA125 and CEA in the assisted diagnosis of breast cancer*

YIN Zhihui¹,ZHOU Zhiwei²,LIU Jiyong²,YANG Baochang²,WANG Lingling³

(1. Department of Clinical Laboratory, the Xingtai Third Hospital, Xingtai, Hebei 054000, China;2. Department of Clinical Laboratory;3. Department of Oncology, the Central Hospital of China National Petroleum Corporation, Langfang, Hebei 065000, China)

Abstract: Objective To explore the application of decision tree model of serum indicators in the diagnosis of breast cancer. **Methods** From August 2016 to April 2018, 60 newly diagnosed breast cancer patients diagnosed by histopathology were selected as breast cancer group, and 60 female patients with non-neoplastic and non-reproductive endocrine diseases were selected as non-breast cancer group. A total of 60 patients who first visited the department of mammary gland were included in the validation group, including 30 cases of breast cancer and 30 cases of non-breast cancer. Tumor markers (CA153, CA125, CEA) and 25-hydroxyvitamin D3 were detected in two groups. First, the results of two groups of patients were analyzed statistically. Then, a decision tree model was established for all detection data, and the role of each index in assistant diagnosis of breast cancer was analyzed. Relevant indicators of the validation group were collected and applied to the work of the subjects. The ROC curve was used to evaluate the clinical diagnostic efficacy of the model. **Results** (1) Vitamin D level of non-menopausal patients in breast cancer group was lower than that of control group, CEA

* 基金项目:廊坊市科技支撑计划项目(2016013025)。

作者简介:尹志辉,男,主管技师,主要从事实验室数据管理与分析方面的研究。

本文引用格式:尹志辉,周志伟,刘继勇,等.决策树分析在提高血清维生素 D、CA153、CA125、CEA 对乳腺癌辅助诊断作用中的应用分析

[J]. 国际检验医学杂志,2019,40(20):2482-2485.

level of menopausal patients in breast cancer group was higher than that of non-breast cancer group ($P < 0.05$); there was no significant difference in other indicators ($P > 0.05$); (2) Through automatic modeling analysis, 100% prediction ability of decision tree was the best. The model is CHAID decision tree, which shows that when vitamin D ≤ 7.2 ng/mL, the positive prediction of breast cancer reaches 100%; when $7.2 \text{ ng/mL} < \text{vitamin D} \leq 19.0 \text{ ng/mL}$, whether to add menopausal variables, the negative predictive ability of premenopausal patients is 76.0%, and the positive predictive ability of menopausal patients is 65.7%; When the level of vitamin D $> 19.0 \text{ ng/mL}$, and the CEA $\leq 2.48 \text{ U/mL}$, the negative predictive ability was 87.9%, and the positive predictive ability was 100% when the CEA $> 2.48 \text{ U/mL}$. Through the analysis of ROC curve, it was concluded that the multi-index tandem analysis of decision tree model improved significantly in screening specificity. **Conclusion** Decision tree model can be used to classify the test indicators progressively and form different sets of judgements in the form of trees, which can enhance the diagnostic value of the test indicators in breast cancer.

Key words: decision tree analysis; breast cancer; vitamin D; diagnostic value

乳腺癌是女性排名第一的恶性肿瘤,严重危害女性身心健康^[1]。其发病机制尚未完全明确,目前许多研究资料显示乳腺癌是在多种因素共同作用下发生并发展的^[2-3],其中一些流行病学调查和实验研究表明:血钙水平和维生素 D 水平与乳腺癌的发生发展有密切关系^[4]。本研究通过对血清维生素 D 和糖链抗原 153(CA153)、CA125、癌胚抗原(CEA)建立决策树模型以提高上述指标在乳腺癌中的辅助诊断价值。

1 资料与方法

1.1 一般资料 选取 2016 年 8 月至 2018 年 4 月在中国石油天然气集团公司中心医院就诊由组织病理学确诊的新发乳腺癌患者 60 例,按照是否绝经分为两组,均为 30 例,非绝经组平均(46.63 ± 4.65)岁,绝经组平均(56.23 ± 2.57)岁;选取同期体检非肿瘤、非生殖内分泌系统疾病的女性患者 60 例,按照是否绝经分为两组,均为 30 例,非绝经组平均(47.26 ± 4.12)岁,绝经组平均(54.40 ± 3.81)岁;选取乳腺外科首次就诊乳腺癌患者 30 例,非乳腺癌患者 30 例,构成验证组,用于验证模型的临床诊断价值。本研究取得患者知情同意,获得医院伦理委员会批准。

1.2 仪器与试剂

1.2.1 仪器 采用贝克曼-库尔特 DXI800 化学发光分析仪,进行检测肿瘤标志物(CA153、CA125、CEA);采用罗氏 2010 化学发光分析仪,进行检测维生素 D,仪器定期保养,性能验证符合要求,室内质控在控,室间质评成绩合格。

1.2.2 试剂 所有试剂、校准品和质控品均为原装进口试剂,由美国贝克曼-库尔特公司和罗氏公司提供。

1.3 数据建模 将整理好数据以数据源节点方式接入数据挖掘软件内,以类型节点设置输入变量和目标变量,首先进行模型自动分析,C5.0 决策树、CHAID 决策树以及类神经网络模型符合要求,依次建立以上模型,进一步分析,确立最佳模型为 CHAID 决策树,并分析模型、预测变量重要性以变量结果响应图;然后将验证组患者相同变量借入已建模型对是否患有

乳腺癌进行预测分析,以临床诊断为金标准,通过 ROC 曲线评估决策树模型预测结果的临床诊断效能。

1.4 统计学处理 常规统计学分析由 SPSS22.0 完成,数据以 $\bar{x} \pm s$ 表示;采用独立样本 t 检验进行分析,以 $P < 0.05$ 为差异具有统计学意义。数据变量的决策树分析由 SPSS modeler 18.0 完成,首先采用自动建模分析,确定构建最佳模型,决策树停止条件设置为:最大层数 5 层,父分支中最小记录数为 2%,子分支中最小记录数为 1%,采用后剪枝方法进行树剪枝,以 $\alpha = 0.05$ 作为检验水准。

2 结果

2.1 一般情况 本次研究中,乳腺癌组未绝经患者平均(46.63 ± 4.65)岁和绝经患者平均(56.23 ± 2.57)岁,对照组未绝经患者平均(47.26 ± 4.12)岁和绝经患者平均(54.40 ± 3.81)岁,年龄在两组之间差异无统计学意义($t = 0.176, 0.232; P = 0.861, 0.817$)。

2.2 乳腺癌组与对照组维生素 D、CA153、CA125、CEA 比较 乳腺癌组未绝经患者维生素 D 水平低于对照组,乳腺癌绝经患者 CEA 水平高于对照组,差异具有统计学意义($P < 0.05$);其他指标在乳腺癌组与对照组的未绝经患者和绝经患者间进行比较,差异无统计学意义($P > 0.05$),见表 1。

2.2 模型建立与分析

2.2.1 纳入维生素 D、CA153、CA125、CEA 和是否绝经 5 个变量,以病理学检查结果是否乳腺癌为预测目标,通进行自动建模分析,预测精度由高到低模型依次为 CHAID 决策树、C&R 决策树、C5.0 决策树、类神经网络,总体精确度分别为 85.001%、85.000%、84.167%、84.167%。进一步对各模型结果分析,决策树 100.00%预测能力由高到低分别为:CHAID 决策树(32 例)、C5.0(27 例)和 C&R(19 例)。见图 1。

2.2.2 CHAID 决策树预测变量重要性 重要性有强到弱分别为维生素 D(0.56)、CEA(0.25)和是否绝经(0.19)。见图 2。

度与分类精度的 4 个模型,通过对 4 个模型逐一分析,最终选择分类准确度最高的 CHAID 模型。该模型结果首先对维生素 D 分成 3 层分别为 $<7.2 \text{ ng/mL}$, $>7.2 \text{ ng/mL}$ 且 $\leq 19.0 \text{ ng/mL}$, $>19.0 \text{ ng/mL}$;第一部分确定 100.000% 阳性患者,中间部分在结合是否绝经可以将阴性排除概率由 51.667% 增至 76.000%,阳性确诊率由 48.333% 增至 65.714%,第三部分结合 CEA 阈值为 2.480 U/mL ,阴性排除概率由 80.556% 增至 87.879%,阳性确诊率由 19.444% 增至 100.000%。经过 ROC 曲线分析,决策树模型在辅助诊断乳腺癌方面性能优异,灵敏度与多指标串联方法相当,但特异度、阳性预测值、阴性预测值及 AUC 均较多指标串联法有明显提升,在辅助临床医生诊断乳腺癌方面作用更大。综上所述,乳腺癌患者维生素 D 水平低于健康人群,同时结合患者是否绝经,CEA、CA125 水平进行建模分析,最终明显提高检测指标对乳腺癌的预测准确性、可靠性及一致性,在乳腺癌患者诊断与治疗中,提供有效实验室依据。

本研究不足之处,采集数据有限只进行了部分检验指标和患者基本情况建立辅助诊断模型,在小范围内进行了模型效能验证,因此所建立的预测模型有待进一步纳入更多指标和更大数据量进行完善。

4 结 论

通过乳腺癌患者和体检健康人群的肿瘤标志物水平、维生素 D 水平和患者绝经情况建立决策树数据模型;利用所建立模型对其他患者相关指标进行分析一定程度上可预测是否罹患乳腺癌以及预测准确性。

参考文献

[1] 易琳,刘兴明,林丁,等.血清 CA153、CA125、CEA 联合检测在乳腺癌诊断中的价值[J].重庆医科大学学报,2012,37(9):802-805.
[2] 陈万青,郑荣寿.中国女性乳腺癌发病死亡和生存状况[J].中国肿瘤临床,2015,42(13):668-674.
[3] 郑莹,吴春晓,张敏璐.乳腺癌在中国的流行状况和疾病特征[J].中国癌症杂志,2013,23(8):561-569.

(上接第 2481 页)

[12] 陈巧,王梅芳,程芳芳,等.老年原发免疫性血小板减少症患者临床特征及预后分析[J].首都医科大学学报,2018,39(2):277-281.
[13] 黄莉,姚红霞.糖皮质激素治疗初诊免疫性血小板减少症的疗效观察及预后影响因素分析[J].广西医学,2016,38(11):1578-1584.
[14] 保红云,王丽,荣冬靖,等.以血小板减少为首发症状的结缔组织病 36 例分析[J].中国医学创新,2011,8(12):140-141.
[15] 邓家栋.临床血液学[M].上海:上海科学技术出版社,

[4] OEFFINGER KC,ELIZABETH TH,FONTHAM,et al. Breast cancer screening for women at average risk 2015 guideline update from the American cancer society[J]. JAMA,2015,314(15):1599-1614.
[5] SIU AL; U. S. PREVENTIVE SERVICES TASK FORCE. Screening for Breast Cancer; U. S. Preventive Services Task Force Recommendation Statement[J]. Ann Intern Med,2016,1164(4):279-296.
[6] 卞静,赵华硕,徐岳,等.基于决策树模型的女性乳腺癌危险因素调查[J].中国妇幼保健,2015,30(8):1244-1246.
[7] 张志勇,陈从华,明炜.检测患者外周血清 CEA,CA15-3 水平在乳腺癌手术前后的价值研究[J].河北医药,2017,39(6):857-864.
[8] 罗丹,迟曰梅,王新超.CA153,CA125,CEA 联合检测在乳腺癌诊断中的临床意义[J].中国医学创新,2018,15(6):134-137.
[9] 张晓洁,张美娟,张立涛,等.CEA、CA153、CYFRA21-1、CRP 联合检测在乳腺癌辅助诊断中的临床价值探讨[J].国际检验医学杂志,2017,38(23):3352-3354.
[10] CEDOLINI C, BERTOZZI S, LONDERO A P, et al. Type of breast cancer diagnosis, screening, and survival [J]. Clin Breast Cancer, 2014, 14(4): 235-240.
[11] 刘珍,胡科.血清 CEA 和 CA15-3 水平对乳腺癌化疗疗效及预后评估的价值分析[J].国际检验医学杂志,2017,38(18):2537-2539.
[12] 钟琼.血清 CA15-3、CA125、CEA 和 PRL 检测对乳腺癌术后治疗的指导意义[J].赣南医学院学报,2015,35(6):913-917.
[13] LIBSON S, LIPPMAN M. A review of clinical aspects of breast cancer[J]. Int Rev Psychiatry, 2014, 26(1): 4-15.
[14] HU K, CALLEN DF, LI J, et al. Circulating vitamin D and overall survival in breast cancer patients: a dose-response meta-analysis of cohort studies[J]. Integr Cancer Ther, 2018, 17(2): 217-225.
[15] 武四美,文兆明,俞根.血清中维生素 D,血钙的水平与乳腺癌的关系研究[J].江西医药,2017,52(4):323-325.

(收稿日期:2019-03-16 修回日期:2019-05-28)

1985.
[16] LERTCHASITAPORN K, KASITANON N, WANG- KAEW S, et al. An evaluation of the association of leukopenia and severe infection in patients with systemic lupus erythematosus[J]. J Clin Rheumatol, 2013, 19(3): 115-120.
[17] 段晓晖,马宏杰.抗核抗体阳性免疫性血小板减少性紫癜临床观察[J].临床医药文献杂志,2017,4(18):3398.

(收稿日期:2019-03-06 修回日期:2019-05-18)