

血清 CA72-4, CA242, CA19-9 和 CEA 的 PCA-决策树模型对胃癌的诊断价值*

桂 林, 黄远帅 (西南医科大学附属医院输血科, 四川 泸州 646000)

摘要:目的 探讨血清 CA72-4, CA242, CA19-9 和 CEA 的主成分分析(PCA)-决策树模型对胃癌的诊断价值。方法 采用电化学发光法测定 193 例胃癌患者、106 例胃良性疾病和 86 例正常对照血清 CA72-4, CA242, CA19-9 和 CEA 水平并进行受试者工作特征曲线(ROC), PCA 和 PCA-决策树分析。结果 单项指标 CA72-4, CA242, CA19-9 和 CEA 的 ROC 曲线下面积(AUC)分别为 0.741(95%CI, 0.692~0.791), 0.863(95%CI, 0.827~0.898), 0.783(95%CI, 0.737~0.828) 和 0.827(95%CI, 0.785~0.869)。四指标联合的 PCA-AUC 为 0.935(95%CI, 0.912~0.958), 在临界值(PC 得分)为 44.13 时, 敏感度和特异度分别为 78.2% 和 94.8%。单项指标的决策树模型对胃癌和非胃癌(胃良性疾病+正常对照)的诊断正确率为 76.2% 和 94.8%, 预测正确率为 56.5% 和 96.5%。联合四指标的 PCA-决策树模型对胃癌和非胃癌的分类正确率为 90.3% 和 100%, 预测正确率为 72.4% 和 92.2%。结论 血清 CA72-4, CA242, CA19-9 和 CEA 的 PCA-决策树模型有助于胃癌的鉴别诊断。

关键词:胃癌; 诊断; 主成分分析; 决策树分析

中图分类号: R735.2; R730.43 文献标志码: A 文章编号: 1671-7414(2017)01-057-04

doi: 10.3969/j.issn.1671-7414.2017.01.016

Diagnostic Value of Serum CA72-4, CA242, CA19-9 and CEA for Gastric Cancer Based on PCA-Decision Tree Model

GUI Lin, HUANG Yuan-shuai (Department of Blood Transfusion, the Affiliated Hospital of Southwest Medical University, Sichuan Luzhou 646000, China)

Abstract: Objective To evaluate the diagnostic value of serum tumors CA72-4, CA242, CA19-9 and carcino-embryonic antigen (CEA) for patients with gastric cancer based on principle component analysis (PCA)-decision tree analysis. **Methods** Serum levels of CA72-4, CA242, CA19-9 and CEA in 193 patients with gastric cancer, 106 patients with benign gastric diseases and 86 normal controls were measured by electrochemiluminescence assay, and data were analyzed by the receiver operating characteristic (ROC) curve, PCA and PCA-decision tree analysis. **Results** The area under the ROC curve of CA72-4, CA242, CA19-9 and CEA was 0.741[95% confidence interval (95%CI), 0.692~0.791], 0.863 (95%CI, 0.827~0.898), 0.783 (95%CI, 0.737~0.828) and 0.827 (95%CI, 0.785~0.869), respectively. The combined four serum tumor markers in the PCA-AUC model was 0.935 (95%CI, 0.912~0.958) at the cutoff value (PC score) of 44.13 with 78.2% of sensitivity and 94.8% of specificity. The accuracy of serum CA72-4, CA242, CA19-9 and CEA for the diagnosis of gastric cancer group and nongastric cancer group (benign gastric diseases and normal controls) in the decision tree model were 76.2% and 94.8%, 56.5% and 96.5% for prediction, respectively. The combined four serum tumors for the diagnosis of gastric cancer group and nongastric cancer group in PCA-decision tree model were 90.3% and 100%, 72.4% and 92.2% for prediction, respectively. **Conclusion** The PCA-decision tree model based on serum CA72-4, CA242, CA19-9 and CEA were helpful for the diagnosis of gastric cancer.

Keywords: gastric cancer; diagnosis; principle component analysis; decision tree analysis

胃癌是发生率高、临床治愈率低的恶性肿瘤, 早期诊断、及时治疗对提高胃癌患者的生存率极为重要。近年来, 多种标志物已广泛用于胃癌的筛查, 但单指标诊断胃癌的灵敏度不尽人意^[1]。联合检测多种标志物如癌胚抗原(CEA)、糖类抗原 CA19-9, CA72-4, CA125 和 C 反应蛋白(CRP)等在提高胃癌诊断灵敏度的同时降低了特异度^[2]。

因此, 寻找合适的诊断和预测模型有望更灵敏、特异地实现胃癌的鉴别诊断。主成分分析法(PCA)是常用的多元统计分析方法, 可用于多指标数据信息的挖掘, 而决策树常用于诊断模型的构建和预测分析^[3]。因此, 本研究借鉴代谢靶标分析的策略, 以胃癌患者为研究对象, 探讨血清 CA72-4, CA242, CA19-9 和 CEA 的 PCA-决策树模型对胃

* 基金项目: 四川省卫生计生厅课题资助项目(120336); 西南医科大学人才基金(2014ZD-017)。

作者简介: 桂 林(1986-), 女, 技师, 研究方向: 输血相关疾病, E-mail: 459531301@qq.com。

通讯作者: 黄远帅, Tel: 0830-3165740, E-mail: riys@live.cn。

癌的诊断价值。

1 材料与方法

1.1 研究对象 193例胃癌患者为我院2014年5月~2016年2月首诊病人,其中男性108例,女性85例,平均年龄 49.1 ± 14.5 岁,其中印戒细胞癌90例,管状腺癌60例,其它类型43例。胃良性疾病对照共106例,其中急、慢性胃炎共80例,胃溃疡20例,胃平滑肌瘤4例,胃食管反流2例。胃癌组和胃部良性疾病组均经胃窥镜和病理组织学诊断确诊。正常对照为体检人员,共86例,体检肝功能、心肌损伤标志物和肿瘤标志物全套未见明显异常。

1.2 试剂和仪器 采用 Elecsy 电化学发光仪及配套试剂测定血清 CA72-4, CA242, CA19-9 和 CEA,其参考区间分别为 CA72-4(0~6.0)U/L,

CA242(0~12.8)U/ml, CA19-9(0~35.0)U/ml 和 CEA(0~5.5)ng/ml。

1.3 统计学分析 所有统计学分析均采用 SPSS 18.0 完成,数据以均数±标准差($\bar{x} \pm s$)表示,采用单因素方差分析,组间比较采用最小显著差法(Least-significant difference, LSD)检验, $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 三组患者血清 CA72-4, CA242, CA19-9 和 CEA 水平比较 见表1。胃癌组血清 CA72-4, CA242, CA19-9 和 CEA 水平均高于正常组和胃良性疾病组,差异均有统计学意义($P < 0.001$)。胃良性疾病组与正常对照组比较,血清 CA19-9, CA242 和 CA72-4 水平均升高,差异有统计学意义(P 值分别为 0.019, 0.020 和 0.022)。

表1 胃癌组、良性疾病组和正常组血清 CA72-4, CA242, CA19-9 和 CEA 水平比较($\bar{x} \pm s$)

项目	胃癌组	胃良性疾病组	健康对照组	F值	P值
CA72-4(U/L)	19.9 ± 18.3^{ab}	9.91 ± 3.54^c	4.01 ± 2.60	49.078	0.000
CA242(U/ml)	32.1 ± 24.1^{ab}	14.0 ± 4.77^c	6.87 ± 2.13	79.612	0.000
CA19-9(U/ml)	79.6 ± 54.8^{ab}	35.4 ± 14.7^c	27.5 ± 9.22	61.406	0.000
CEA(ng/ml)	25.9 ± 27.1^{ab}	5.46 ± 2.98	4.60 ± 3.99	54.720	0.000

注: a 胃癌组与对照组比较 $P < 0.001$; b 胃癌组与胃良性疾病组比较 $P < 0.001$; c 胃良性疾病组与对照组比较 $P < 0.05$ 。

2.2 三组患者血清 CA72-4, CA242, CA19-9 和 CEA 单项和联合指标的 PCA-ROC 曲线下面积(AUC)比较 见表2。四项指标物的 AUC 均较高(0.741~0.863),其中标志物 CA242 对胃癌的 AUC 最佳[0.863, 95% 置信区间(CI) 0.827~0.898],在临界值为 19.89 U/ml 时,灵敏度和特

异度分别为 59.6% 和 95.8%。基于血清 CA72-4, CA242, CA19-9 和 CEA 四指标的 PCA-AUC 最高(0.935, 95% CI: 0.912~0.958),在临界值(即第一主成分 PC1 得分)为 44.13 时,灵敏度和特异度分别为 78.2% 和 94.8%。

表2 三组患者血清单项和联合指标的 PCA-ROC 曲线下面积(AUC)比较

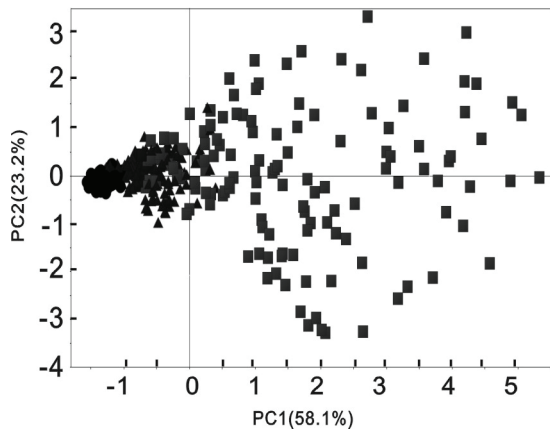
项目	AUC(95% CI)	灵敏度(%)	特异度(%)	临界值
CEA(ng/ml)	0.827(0.785~0.869)	56.5	99.5	10.58
CA19-9(U/ml)	0.783(0.737~0.828)	48.2	100	63.14
CA242(U/ml)	0.863(0.827~0.898)	59.6	95.8	19.89
CA72-4(U/L)	0.741(0.692~0.791)	47.2	98.4	14.58
PC1	0.935(0.912~0.958)	78.2	94.8	44.13

2.3 血清 CA72-4, CA242, CA19-9 和 CEA 的 PCA 分析 见图1。本研究中,为消除不同肿瘤标志物单位和数值的影响,将数据标准化处理[(测定值-均值)/标准差]后建立 PCA 模型,该模型中前两个主成分能揭示数据 81.3% 的综合信息,提示 PCA 模型构建成功。由图1可知,正常对照组血清 CA72-4, CA242, CA19-9 和 CEA 代谢未见异常,个体分布在最左侧,聚类良好。胃良性疾病组和胃癌组血清 CA72-4, CA242, CA19-9 和 CEA 代

谢明显紊乱,其中胃癌组最显著,表现为个体分布离散、主要分布在 PCA 图最左侧且与胃部良性疾病组有部分重叠。

2.4 基于血清 CA72-4, CA242, CA19-9 和 CEA 单项和四项联合的 PCA-决策树模型 单项指标的决策树模型在节点 1(CA242 \leq 9.00 U/ml)和 3(CA242 \geq 21.6 U/ml)时对胃良性疾病和胃癌的诊断正确率分别为 76.2% 和 94.8%,对胃癌组和非胃癌组(胃良性疾病组+健康对照组)的预测正

准确率分别为 43.5%(84/193)和 96.9%(186/192)。四项指标联合的 PCA-决策树模型中 $PC1=0.5118 \times CEA + 0.487 \times CA199 + 0.499 \times CA242 + 0.501 \times CA724$ 。该模型在节点 1 和 3 时对胃良性疾病和胃癌的诊断正确率分别为 90.3%, 0 和 9.7%, 100%(图 2), 对胃癌组和非胃癌组(胃良性疾病组+健康对照组)的预测正确率分别为 72.4%(139/192)和 92.2%(178/193)。



(●)正常对照, (▲)胃良性疾病组, (■)胃癌组

图 1 基于血清 CA72-4, CA242, CA19-9 和 CEA 联合检测的 PCA 模型

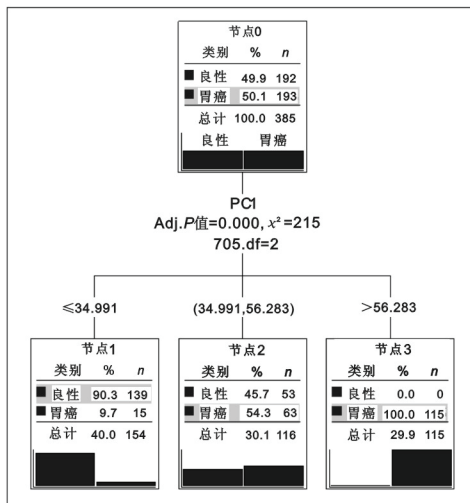


图 2 基于血清 CA724, CA242, CA199 和 CEA 的 PCA-决策树胃癌诊断模型

3 讨论 目前,由于诊断技术的局限性,胃癌的早期诊断和鉴别诊断仍较困难。CA199, CA72-4, CA242 和 CEA 均是临床上常用的肿瘤标志物,对肿瘤的疗效判断、病情监测和预后估计均有一定的临床价值^[4,5]。CEA 是大肠癌组织产生的一种糖蛋白,广泛存在于内胚叶起源的消化道肿瘤,故 CEA 诊断胃癌的特异度较高灵敏度较低^[6],这与本研究结果一致。最新研究表明,CEA 可能是预测胃癌患者生存率的独立危险因素,监测血清

CEA 水平异常的胃癌患者并积极采取有效的治疗措施有助于提高胃癌患者的生存率^[7]。CA19-9 是一种黏蛋白型糖蛋白肿瘤标志物,为细胞膜上的糖脂质。CA19-9 与胃癌肿瘤组织的大小、淋巴结转移和肿瘤浸润的深度等密切相关,是评估胃癌是否浸润周边组织和胃癌患者预后的指标^[8]。CA72-4 是一种高分子糖蛋白,也是诊断胃癌的标志物之一。邢瑞青等^[9]研究发现,胃癌组血清 CA72-4 水平高于萎缩性胃炎组,CA72-4 对胃癌和胃部良性疾病的鉴别诊断有一定的参考价值。CA242 作为一种唾液酸化的糖类抗原,常用于胰腺癌和结肠癌的辅助诊断,故 CA242 用于胃癌的报道较少见。Tian 等^[10]发现,CA242 用于辅助诊断胃癌的灵敏度优于 CA199 和 CEA,与本研究结果一致。此外,本研究发现胃癌组血清 CA72-4, CA242, CA19-9 和 CEA 单项指标诊断胃癌的灵敏度均较低,诊断效能中等,故单项指标对胃癌的诊断仍有一定的局限性。

PCA 是一种基于投影技术的数据分析方法^[11],在保证数据信息损失最少的原则下对高维空间变量进行降维处理,将多指标转化为少数几个综合指标(即 PC)。以 PC1 对 PC2 作图,从而直观地在二维平面图上观察样本的构成和分布特点。李建更等^[12]将 PCA 联合偏最小二乘(PLS)两种降维对胃癌进行亚型分类,诊断的正确率最高达到 100%。本研究中,基于血清 CA199, CA72-4, CA242 和 CEA 四指标的 PCA 模型的 AUC 和灵敏度均优于单一指标,效果较为理想。决策树分析是一种常用的数据挖掘技术,在食品和计算机领域应用广泛,近年来也用于疾病的辅助诊断^[13]。决策树利用概率分析的原理寻找最有分辨力的节点,分层递归,直到所有子集包含同一类型的数据。决策树具有简单、分类速度快、适合处理大样本数据等优点。目前,鲜有 PCA-决策树用于胃癌诊断的相关报道。本研究中,基于 PCA-决策树模型的诊断和预测准确率均优于单一指标的决策树模型。因此,基于血清 CA199, CA72-4, CA242 和 CEA 的 PCA-决策树模型更有助于胃癌的辅助诊断。

在胃癌的辅助诊断中,PCA-决策树结合血清 CA199, CA72-4, CA242 和 CEA 诊断胃癌效果较理想。尽管如此,由于患者本身的差异和诊断数据的误差等因素,准确诊断胃癌仍较困难。本研究标本量比较有限,且尚未全面考察不同组合肿瘤标志物对胃癌的鉴别诊断和分期诊断价值,还需后续增大样本量深入研究。

参考文献:

[1] 黄道桂,任传路,虞红艳,等. 血清 PG I, PG II, TK1,

- TSGF, CEA 与 CA724 联合检测在胃癌诊断中的应用价值[J]. 现代检验医学杂志, 2015, 30(1): 101-104.
- Huang DG, Ren CL, Yu HY, et al. Application value of combined detection of PG I, PG II, TK1, TSGF, CEA and CA724 in diagnosis of Gastric cancer[J]. Journal of Modern Laboratory Medicine, 2015, 30(1): 101-104.
- [2] 林益振, 蒋相霞, 陈翔, 等. D-D, CRP 联合肿瘤标志物检测对胃癌早期诊断的价值[J]. 中国卫生检验杂志, 2014, 24(5): 695-698.
- Lin YZ, Jiang XX, Chen X, et al. Joint detection value of D-D, CRP and tumor markers in early diagnosis of gastric cancer[J]. Chinese Journal of Health Laboratory Technology, 2014, 24(5): 695-698.
- [3] Mohammadzadeh F, Noorkojuri H, Pourhoseingholi MA, et al. Predicting the probability of mortality of gastric cancer patients using decision tree[J]. Irish Journal of Medical Science, 2015, 184(2): 277-284.
- [4] Yin LK, Sun XQ, Mou DZ. Value of combined detection of serum CEA, CA72-4, CA19-9 and TSGF in the diagnosis of Gastric Cancer[J]. Asian Pacific Journal of Cancer Prevention, 2015, 16(9): 3867-3870.
- [5] Virgilio E, Giarnieri E, Montagnini M, et al. Analyzing gastric lavage of gastric cancer patients; a prospective observational study on cytopathology and determination of intragastric CEA, CA 19.9, CA 72.4 and CA 50[J]. Acta Cytologica, 2016, 60(2): 161-166.
- [6] Gwak HK, Lee JH, Park SG. Preliminary evaluation of clinical utility of CYFRA 21-1, CA 72-4, NSE, CA19-9 and CEA in stomach cancer[J]. Asian Pacific Journal of Cancer Prevention, 2014, 15(12): 4933-4938.
- [7] Deng K, Yang L, Hu B, et al. The prognostic significance of pretreatment serum CEA levels in gastric cancer: a meta-analysis including 14 651 patients[J]. PLoS One, 2015, 10(4): e0124151.
- [8] Han ES, Lee HH, Lee JS, et al. At which stage of gastric cancer progression do levels of carcinoembryonic antigen and carbohydrate antigen 19-9 increase? Application in advanced gastric cancer treatment[J]. J Gastric Cancer, 2014, 14(2): 123-128.
- [9] 邢瑞青, 吴永昌, 邹海, 等. 不同血清肿瘤标志物对胃癌诊断价值的临床研究[J]. 现代检验医学杂志, 2014, 29(6): 102-106.
- Xing RQ, Wu YC, Zou H, et al. Diagnostic value of different serum tumor markers in gastric cancer[J]. Journal of Modern Laboratory Medicine, 2014, 29(6): 102-106.
- [10] Tian SB, Yu JC, Kang WM, et al. Combined detection of CEA, CA 19-9, CA 242 and CA 50 in the diagnosis and prognosis of resectable gastric cancer[J]. Asian Pacific Journal of Cancer Prevention, 2014, 15(15): 6295-6300.
- [11] Trygg J, Holmes E, Lundstedt T. Chemometrics in metabonomics[J]. Journal of Proteome Research, 2007, 6(2): 469-479.
- [12] 李建更, 李萍, 李君, 等. PCA 和 PLS 应用于胃癌亚型分类研究[J]. 生物物理学报, 2009, 25(2): 141-147.
- Li JG, Li P, Li J, et al. PCA and PLS for gastric cancer subtype classification[J]. Acta Biophysica Sinica, 2009, 25(2): 141-147.
- [13] 王旭晨, 陈小惠. 统计关联规则决策树在医疗数据中的应用[J]. 微型机与应用, 2016, 35(15): 78-81.
- Wang XC, Chen XH. An application of statistical association rules decision tree in medical treatment data[J]. Microcomputer & Its Application, 2016, 35(15): 78-81.