

论著·临床研究 doi:10.3969/j.issn.1671-8348.2017.15.015

基于血清 4 项肿瘤标志物的模式识别技术对胃癌的诊断价值*

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** 探讨基于血清 CA72-4、CA242、CA19-9 和 CEA 的模式识别技术对胃癌的诊断价值。**方法** 对 212 例胃癌患者, 116 例胃良性疾病患者和 117 例健康体检者血清 4 项肿瘤标志物测定结果进行回顾性分析, 比较单项指标的诊断效能并建立主成分分析(PCA)、决策树、PCA-决策树和 Fisher 判别分析模型。**结果** 4 项指标中 CA242 对胃癌的诊断效能最佳, ROC 曲线下面积(AUC)为 0.841(95%CI:0.804~0.877)。PCA 模型表明, 胃癌组患者血清 4 项肿瘤标志物代谢明显紊乱, 与胃良性疾病患者和健康对照个体差异显著。决策树、PCA-决策树和 Fisher 判别分析模型对胃癌患者的诊断准确率分别为 58.6%、65.5% 和 58.6%, 预测准确率分别为 65.7%、77.6% 和 73.1%; 对非胃癌患者(胃良性疾病患者+健康对照)的诊断准确率分别为 94.7%、99.4% 和 97.6%, 预测准确率分别为 87.5%、96.9% 和 96.9%。**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和预测分析。

[关键词] 胃癌; 诊断; 主成分分析; 决策树; Fisher 判别分析

[中图分类号] R735.2

[文献标识码] A

[文章编号] 1671-8348(2017)15-2060-03

The diagnostic value of four serum tumor markers for gastric cancer based on pattern recognition techniques*

Gui Lin, Huang Yuanshuai[△]

(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) in patients with gastric cancer based on pattern recognition techniques. **Methods** Data of serum concentrations of CA72-4, CA242, CA19-9 and CEA of 212 patients with gastric cancer, 116 patients with benign gastric disease and 117 healthy subjects were retrospectively analyzed; and the diagnostic performance of each tumor marker, four tumor markers based principle component analysis(PCA), decision tree, PCA-decision tree and the fisher discriminant analysis models were established. **Results** CA242 had the best diagnostic effect on gastric cancer, and the area under the ROC curve (AUC) was 0.841 (95%CI:0.804—0.877). PCA model showed that the serum levels of four tumor markers in patients with gastric cancer were significantly different from those in benign and healthy patients, and obvious metabolic disorders of serum with four tumor markers were found among the patients with gastric cancer. The diagnosis accuracy of the decision tree, PCA-decision tree and the Fisher discriminant analysis models for gastric cancer patients was 58.6%, 65.5% and 58.6% respectively, and for non-gastric cancer patients (benign gastric diseases and healthy controls) was 94.7%, 99.4% and 97.6%. And the prediction accuracy of the decision tree, PCA-decision tree and the fisher discriminant analysis models for gastric cancer patients was 65.7%, 77.6% and 73.1%, and for non-gastric cancer patients was 87.5%, 96.9% and 96.9%, respectively. **Conclusion** The PCA-decision tree model of serum CA72-4, CA242, CA19-9 and CEA might be helpful for the diagnosis and prediction of patients with gastric cancer.

[Key words] stomach neoplasms; diagnosis; principle component analysis; decision tree analysis; Fisher discriminant analysis

胃癌的发病率和病死率在我国高居首位且呈逐年上升趋势^[1-2]。据统计, 2015 年我国胃癌新增病例 67.9 万, 死亡 49.8 万, 80% 以上的胃癌患者确诊时已处于进展期^[3]。因此, 早期、及时诊断胃癌刻不容缓。目前, 肿瘤标志物如癌胚抗原(CEA), 糖类抗原 CA19-9, CA72-4 和 CA242 等常用于胃癌的早期诊断, 但单项指标的灵敏度较低^[4]。多项并联或串联试验均能提高诊断的灵敏度或特异度^[5], 但并联时降低了特异度, 串联时降低了灵敏度且均不能预测分析。为此, 本研究以胃癌患者为研究对象, 采用多元统计分析中主成分分析(PCA)、决策树分析、PCA-决策树分析和 Fisher 判别分析的策略, 探讨血清 CEA、CA19-9、CA72-4 和 CA242 对胃癌的诊断和预测

价值。

1 资料与方法

1.1 一般资料 选择本院 2014 年 5 月至 2016 年 6 月首诊胃癌患者 212 例作为胃癌组, 其中男 115 例, 女 97 例; 平均年龄(48.4±14.9)岁; 印戒细胞癌 95 例, 管状腺癌 67 例, 其他类型 50 例。选取消化内科胃良性疾病患者共 116 例作为胃良性疾病组, 其中男 67 例, 女 49 例; 平均年龄(47.7±18.2)岁; 急、慢性胃炎 80 例, 胃溃疡 30 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患者均经胃窥镜和病理组织学诊断确诊。健康对照组为本院体检中心门诊体检者, 共 117 例, 其中男 64 例, 女 53 例; 平均年龄(45.6±21.8)岁; 体检生化指标,

* 基金项目: 四川省卫生计生课题资助项目(120336); 西南医科大学人才基金(2014ZD-017)。 作者简介: 桂林(1986—), 技师, 硕士, 主要从事输血相关疾病方面研究。 [△] 通信作者, E-mail: 26074937@qq.com。

表 1 3 组血清 CA724,CA242,CA199 和 CEA 水平比较($\bar{x}\pm s$)

组别	CA72-4(U/L)	CA242(U/L)	CA19-9(U/L)	CEA(g/L)
健康对照组	6.25±3.36 ^{ab}	8.24±4.26 ^{ab}	31.40±12.70 ^{ab}	4.38±1.98 ^{ab}
胃良性病组	9.48±3.78 ^a	13.30±5.69 ^a	41.10±13.50 ^a	6.31±3.14 ^a
胃癌组	19.70±17.90	31.40±23.60	75.00±54.20	25.40±26.40

^a: $P<0.01$,与胃癌组比较;^b: $P<0.05$,与胃良性病比较。

心肌损伤标志物和肿瘤标志物等未见明显异常。3 组性别、年龄差异无统计学意义($P>0.05$),具有可比性。

1.2 方法 采用 MAGLUMI 2000 Plus 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/L,CA19-9 0~35.0 U/L 和 CEA 0~5.5 g/L。比较 3 组患者血清 CA72-4、CA242、CA19-9 和 CEA 水平、ROC 曲线下面积(AUC),对 3 组 PCA 进行分析。

1.3 统计学处理 采用 SPSS18.0 完成,计量资料符合正态分布、方差齐的用 $\bar{x}\pm s$ 表示,采用单因素方差分析,组间比较采用 LSD-*t* 检验,检验水准 $\alpha=0.05$,以 $P<0.05$ 为差异有统计学意义。所有数据在建模前进行归一化处理,以消除单位和量纲的影响。决策树分析以卡方自动交互检测法入选指标,采用 10 次交叉验证对模型进行质量评估。采用 Fisher 判别函数建立判别方程,以血清 CA72-4、CA242、CA19-9 和 CEA 提取主成分(PC)建立 PCA 得分图,并进一步以提取的 PC 建立 PCA-决策树模型。利用 SPSS 中随机抽样模块,抽取 70% 的个体为训练集,30% 为测试集建立决策树、PCA-决策树和 Fisher 判别分析诊断和预测模型。

2 结 果

2.1 3 组患者血清 CA72-4、CA242、CA19-9 和 CEA 水平比较 胃癌组血清 CA72-4、CA242、CA19-9 和 CEA 水平均高于健康对照组和胃良性病组,差异有统计学意义($P<0.01$)。胃良性病组与健康对照组比较,血清 CA242 水平显著升高,差异有统计学意义($P<0.05$),见表 1。

2.2 血清 CA72-4、CA242、CA19-9 和 CEA 的 AUC 比较 4 种血清肿瘤标志物对胃癌均有中等的诊断效能($AUC>0.70$)。CA242 的诊断效能优于 CEA、CA19-9 和 CA72-4,AUC[95% (CI)] 分别为:0.841(0.804~0.877)、0.816(0.775~0.857)、0.744(0.698~0.790)和 0.703(0.654~0.753)。

2.3 血清 CA72-4、CA242、CA19-9 和 CEA 的 PCA 分析 基于血清 4 种标志物的第 1 主成分(PC1)和第 2 主成分(PC2)能揭示原始数据 74.9% 的综合信息。PCA 模型中,胃癌患者体内血清 CA72-4、CA242、CA19-9 和 CEA 水平明显紊乱,个体差异大,主要分布在 PCA 模型椭圆的 1、4 象限;健康对照组个体差异较小,分布在 1、3 象限且聚类明显;良性疾病患者主要分布在椭圆中心,个体差异较小,与胃癌患者有部分重叠,3 组个体有分离的趋势,见图 1。

2.4 基于血清单项和联合肿瘤标志物的 PCA-决策树模型 单项肿瘤标志物中仅 CA242 进入决策树模型($\chi^2=142.87,P<0.01$),10 次交叉验证风险为 41.7%。以血清 CA72-4、CA242、CEA 和 CA19-9 提取 PC,PC_{方程} = 1.936_{CEA} + 0.484_{CA19-9} + 0.507_{CA242} + 0.492_{CA72-4}。4 项联合的 PCA-决策树模型的 10 次

交叉验证风险为 32.8%。4 项联合的 Fisher 判别方程(Y)为, $Y=0.657_{CEA}+0.600_{CA19-9}+0.737_{CA242}+0.558_{CA72-4}$ 。3 种模型对胃癌组和非胃癌组(胃良性病组+健康对照组)的诊断和预测准确率比较结果见表 2。

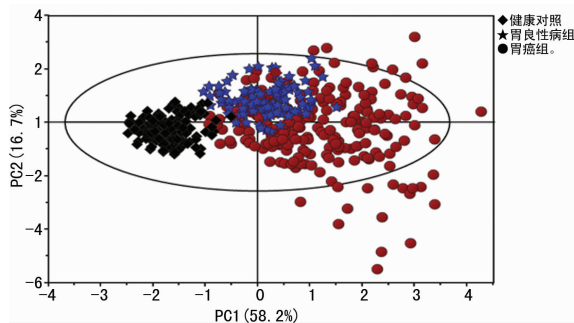


图 1 基于血清 CA72-4,CA242,CA19-9 和 CEA 联合检测的 PCA 模型

表 2 3 种模型对胃癌组和非胃癌组的诊断和预测比较 (%)

模型准确率	训练集(n=314)		预测集(n=131)	
	胃癌组 (n=145)	非胃癌组 (n=169)	胃癌组 (n=67)	非胃癌组 (n=64)
决策树分析	58.6	94.7	65.7	87.5
PCA-决策树分析	65.5	99.4	77.6	96.9
Fisher 判别分析	58.6	97.6	73.1	96.9

3 讨 论

胃癌是发病率高、治愈率低的恶性肿瘤。胃癌患者早期症状不典型,确诊时多已是晚期,故错过了手术和治疗的最佳时机^[4]。胃癌的确诊主要是通过胃镜检查 and 病理组织活检,但胃镜检查有侵入性,需专人操作,费用高,患者耐受性差。目前,由于诊断技术的局限性,胃癌的早期鉴别诊断仍较困难。血清肿瘤标志物具有灵敏、简便和易复检等优点,在临床应用广泛^[6]。CA19-9,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、病情监测和预后评估均有一定的指导意义^[7-8]。薛雯娟等^[4]证实,胃癌患者血清 CA72-4、CEA、CA242 和 CA19-9 水平显著高于胃良性疾病患者和健康人,单项指标的 AUC 均较低,与本研究结果一致。单项肿瘤标志物和多指标并联实验对胃癌的诊断仍有一定的局限性^[9-10],故探索新的数据模型对胃癌的鉴别诊断和预测分析有重要意义。

PCA 是常用的数据降维方法,通过特征根提取有代表性的 PC,用 2~3 个 PC 揭示原始数据的综合信息^[11]。王魏等^[12]采用 PCA 对胃癌患者血红蛋白的拉曼光谱进行分析,揭示了胃癌患者和健康对照个体的空间分布差异。陶海燕等^[13]利用胃溃疡患者的临床参数,成功构建了胃溃疡和溃疡型胃癌

患者的 PCA 特征谱。本研究通过血清 4 项标志物建立二维 PCA 模型,较好地揭示了胃癌患者、胃良性疾病患者和健康对照个体的空间分布特征,有助于后续建模分析。决策树分析是一类在医学诊断中常用的非线性数据挖掘方法,利用概率原理寻找最优分类节点,具有简单、分类速度快,适合处理大样本数据等优点^[14]。王辉等^[15]以血清 CEA、甲胎蛋白(AFP)、CA125、CA19-9 和 CA50 建立了胃癌-健康对照和胃癌-胃良性疾病决策树模型,该模型的诊断准确率均优于联合诊断试验和传统的 Logistic 回归分析。决策树因其输出结果易于理解和解释,而被广泛应用于各个领域,但随着样本量增加和测定指标增多,决策的准确性也会受影响^[16]。采用 PCA 对测定变量进行综合再建立 PCA-决策树模型能减少预测模型的输入量,消除数据间的干扰,提高诊断的准确率^[17]。本研究中,PCA-决策树的交叉验证风险更低,诊断和预测准确率优于决策树模型和传统的 Fisher 判别分析,故 PCA-决策树更有助于胃癌的诊断和预测分析。

由于海量医学信息不断产生,对比分析不同数据模型能更好地挖掘有价值的临床信息。本研究建立了胃癌的 PCA-决策树模型并与传统的决策树和 Fisher 判别分析模型对比分析,取得了较好的诊断和预测效果。PCA-决策树分析为胃癌的辅助诊断和预测分析提供了一种新思路。

参考文献

- [1] Zong L, Abe M, Seto Y, et al. The challenge of screening for early gastric cancer in China[J]. *Lancet*, 2016, 388(10060):2606-2606.
- [2] 彭鹏,吴春晓,龚杨明,等.上海人群胃癌生存率研究[J]. *中国癌症杂志*, 2016, 26(5):414-420.
- [3] Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015[J]. *CA Cancer J Clin*, 2016, 66(2):115-132.
- [4] 薛雯娟,康艳.肿瘤标志物在胃癌早期诊断中的应用价值[J]. *实用癌症杂志*, 2016, 31(3):393-395.
- [5] 王胜,单绿虎,束新华.血清胃蛋白酶原及 CA199、CA242、CEA 联合检测在胃癌早期诊断中的价值[J]. *中华全科医学*, 2016, 14(4):646-648.
- [6] Liu H, Liu R, Zhou XY. A system for tumor heterogeneity evaluation and diagnosis based on tumor markers measured routinely in the laboratory [J]. *Clin Biochem*, 2015,

48(18):1241-1245.

- [7] 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 Pac J Cancer P*, 2015, 16(9):3867-3870.
- [8] 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, CA19-9, CA72-4 and CA50 [J]. *Acta cytologica*, 2016, 60(2):161-166.
- [9] Gwak HK, Lee JH, park SG. Preliminary evaluation of clinical utility of CYFRA21-1, CA72-4, NSE, CA19-9 and CEA in stomach cancer[J]. *Asian Pac J Cancer Prev*, 2014, 15(12):4933-4938.
- [10] Tian SB, Yu JC, Kang WM, et al. Combined detection of CEA, CA19-9, CA242 and CA50 in the diagnosis and prognosis of resectable gastric cancer [J]. *Asian Pac J Cancer P*, 2014, 15(15):6295-6300.
- [11] 王飒,盛萍,姚蓝,等.维药多伞阿魏体外抗胃癌活性部位 GC-MS 指纹图谱的研究[J]. *中草药*, 2016, 46(19):2874-2879.
- [12] 王巍,潘志峰,唐伟跃,等.胃癌患者血红蛋白的表面增强拉曼光谱分析[J]. *光谱学与光谱分析*, 2016, 36(8):1178-1184.
- [13] 陶海燕,李鹏,周萍,等.内镜下胃溃疡性病变更鉴别诊断的数学模型[J]. *中华消化内镜杂志*, 2015, 32(3):180-186.
- [14] Wang XQ, Liu Z, Lv WP, et al. Safety validation of decision trees for hepatocellular carcinoma[J]. *World J Gastroenterol*, 2015, 21(31):9394-9402.
- [15] 王辉,黄钢.肿瘤标志物检测结合支持向量机模型在胃癌诊断中的应用[J]. *中华核医学杂志*, 2010, 30(2):87-89.
- [16] 张棣,曹健.面向大数据分析的决策树算法[J]. *计算机科学*, 2016, 43(6):374-379.
- [17] 孔喜梅,木拉提·哈密提,严传波,等.基于 PCA 和 C4.5 决策树的新疆哈萨克族食管癌图像鉴别研究[J]. *科技通报*, 2016, 32(9):52-57.

(收稿日期:2016-11-20 修回日期:2017-01-10)

(上接第 2059 页)

- [27] Griffioen AW, Molema G. Angiogenesis: potentials for pharmacologic intervention in the treatment of cancer, cardiovascular diseases, and chronic inflammation[J]. *Pharmacol Rev*, 2000, 52(2):237-268.
- [28] Gasparini G, Toi M, Biganzoli E, et al. Thrombospondin-1 and -2 in node-negative breast cancer: correlation with angiogenic factors, p53, cathepsin D, hormone receptors and prognosis[J]. *Oncology*, 2001, 60(1):72-80.
- [29] Deban L, Correale C, Vetrano S, et al. Multiple pathogenic roles of microvasculature in inflammatory bowel disease: a

Jack of all trades[J]. *Am J Pathol*, 2008, 172(6):1457-1466.

- [30] Choi YS, Choi HJ, Min JK, et al. Interleukin-33 induces angiogenesis and vascular permeability through ST2/TRAF6-mediated endothelial nitric oxide production[J]. *Blood*, 2009, 114(14):3117-3126.
- [31] Sundlisæter E, Edelmann RJ, Hol J, et al. The alarmin IL-33 is a notch target in quiescent endothelial cells[J]. *Am J Pathol*, 2012, 181(3):1099-1111.

(收稿日期:2016-11-22 修回日期:2017-01-10)