Int J Clin Exp Med 2016;9(9):17153-17164
www.ijcem.com /ISSN:1940-5901/IJCEM0031047

Original Article

A study of optimized model of serum tumor markers of colorectal cancer based on intelligent algorithms

Panzhang Hou[1], Jianchao Luo[1], Jianhua Zhang[2]

[1]*Department of Tumor Radiotherapy, Henan Provincial People's Hospital, Zhengzhou, Henan Province, China;* [2]*Biomedical Engineering Technology and Data Mining Research Institution of Zhengzhou University, Zhengzhou, Henan Province, China*

Received April 22, 2016; Accepted September 1, 2016; Epub September 15, 2016; Published September 30, 2016

Abstract: Objective: To screen out serum tumor markers that has the best effects on the diagnosis, conditions monitoring and therapeutic evaluation in colorectal cancer patients and thus to provide the optimal model of simple, sensitive and noninvasive with serum markers in early auxiliary diagnosis and monitor of colorectal cancer patients. Methods: Literatures used in early supervision of colorectal cancer and published publicly during 1990-2013 were retrieved and Meta analysis was performed to screen serum markers that were of high detection value. Then 100 colorectal cancer patients and 50 patients with benign colorectal disease were recruited to test the level of serum markers which were screened out via Meta analysis with ELISA, and Logistic regression analysis, ROC curve and Bhattacharyya-SVM analysis were used to screen the optimal combination of serum markers of colorectal cancer. Results: The result of Meta analysis showed that 12 serum markers had certain correlation with colorectal cancer. At the same time, analysis of clinical data indicated that the area under the ROC curve (AUC) of CEA, CA19-9 and HSP60 when they were tested individually and jointly were 0.762, 0.752, 0.825 and 0.906. Their accuracy was 82.67%; sensitivity was 96.90%; and specificity was 90.57%. Bhattacharyya-SVM respectively adopted 12 indicators, 4 indicators (whose Bhattacharyya distance was more than 3) and 7 indicators (Bhattacharyya distance more than 2) to establish three SVM models. The classification accuracy rates of them were respectively 76.7%, 83.3% and 90.0%, sensitivity 80.0%, 85.0% and 90.0%, specificity 70.0%, 80.0% and 90.0%. The SVM prediction model established by CEA, CA50, CYFRA21-1, CA199, CA724, CA125 and UGT1A8 had the highest classification accuracy. Conclusion: The 12 serum markers such as CEA, CA242, and HSP60 are of high value for diagnosis of colorectal cancer. And the SVM models established in this study on basis of clinical validation results of these serum markers possess good predictive efficacy, which should be widely applied to clinical practice.

Keywords: Colorectal cancer, serum marker, optimized model, logistic regression, Bhattacharyya-SVM

Introduction

In China, the incidence and mortality of colorectal cancer (CRC) top the forefront in malignancies [1]. Although recent development of medical diagnosis and treatment of colorectal cancer have been greatly improved, but the morbidity and mortality remain high, which has serious impact on people's health [2]. It is the case when the vast majority of patients with colorectal cancer received treatment only in later stage. While early diagnosis of colorectal cancer is the key to improve the prognosis of patients, because patients in advanced stage not only delay the treatment effect, but also reduce the postoperative five-year survival rate. Therefore, early diagnosis and timely treatment of patients with colorectal cancer not only can improve the prognosis of patients, but also is the focus of future colorectal cancer researches [3, 4].

Tumor marker (TM) refers to substance such as gene, enzyme and protein synthesized or released by tumor cell or generated by the body as a response to the tumor. The substances only can be seen in embryonic tissues or tumor tissues in which content of the substances is higher than normal tissues and they hint the presence, activation and growth of tumors. Owing to the high sensitivity and specificity, tumor markers have become important markers for the early diagnosis of colorectal cancer as well as the evaluation of prognosis [5].

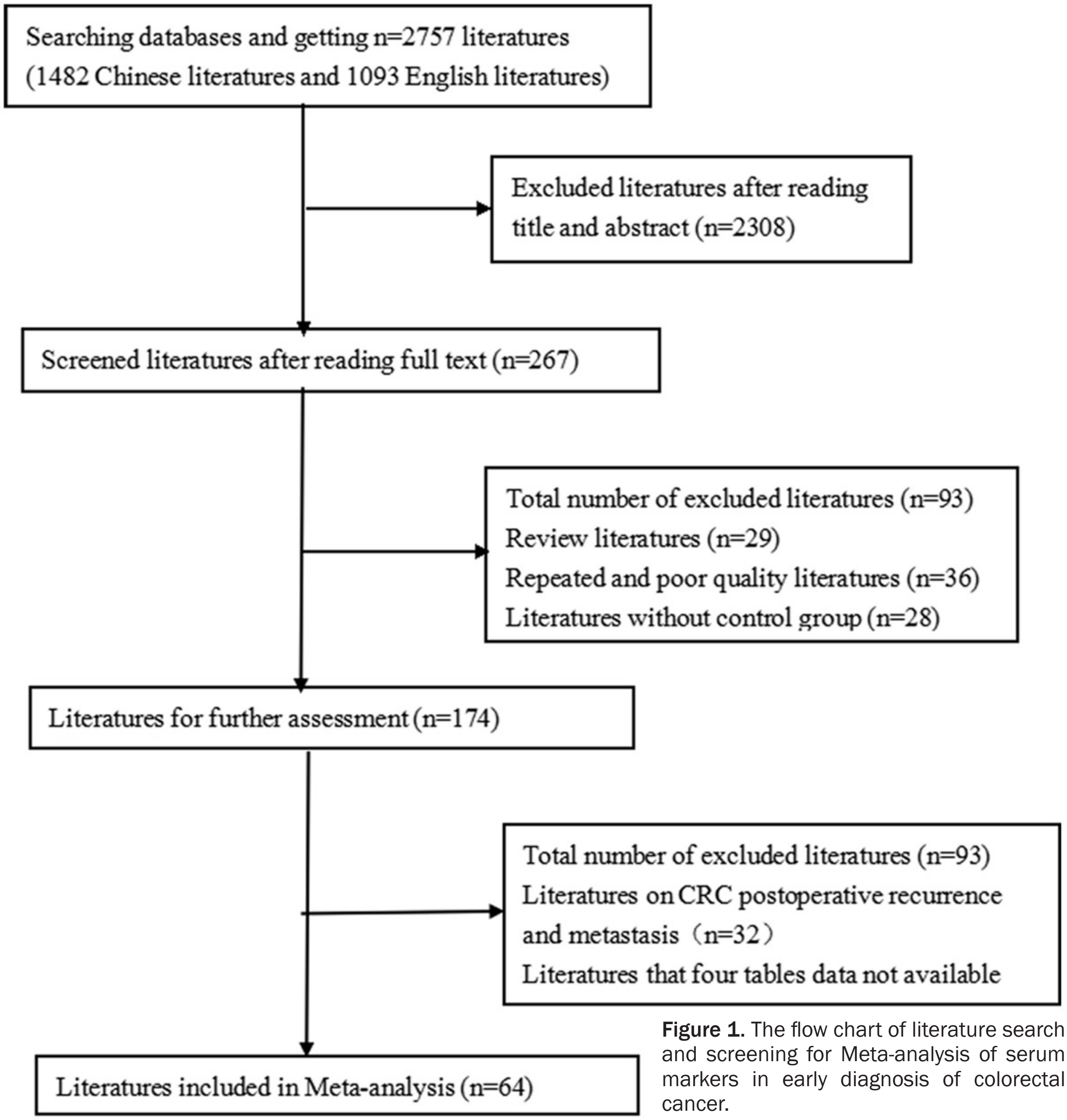
Searching databases and getting n=2757 literatures
(1482 Chinese literatures and 1093 English literatures)
Excluded literatures after reading title and abstract (n=2308)
Screened literatures after reading full text (n=267)
Total number of excluded literatures (n=93)
Review literatures (n=29)
Repeated and poor quality literatures (n=36)
Literatures without control group (n=28)
Literatures for further assessment (n=174)
Total number of excluded literatures (n=93)
Literatures on CRC postoperative recurrence and metastasis（n=32）
Literatures that four tables data not available
Literatures included in Meta-analysis (n=64)

Figure 1. The flow chart of literature search and screening for Meta-analysis of serum markers in early diagnosis of colorectal cancer.

Therefore, screening markers in serum for early diagnosis of colorectal cancer and improving the level of early detection of patients with colorectal cancer will improve the early diagnosis of colorectal cancer patients, significantly reduce mortality and increase postoperative five-year survival rate, thus having very important clinical significance [6, 7].

A meta-analysis uses a statistical approach to systematically and quantitatively analyze the results from multiple independent studies regarding the same project. Advantages of this approach lie in that it processes many research results together, uncovers demerits of individual research, and promotes statistical test power. Furthermore, it plays important role in risk evaluation, prevention intervention, clinical diagnosis and treatment [8, 9].

Meta-analysis was used in this study for a comprehensive and quantitative analysis of the relevant literature to screen serum tumor markers in early diagnosing colorectal cancer. Logistic regression, receiver operating characteristic (ROC) curve and Bhattacharyya-SVM (support vector machine) analysis were used for analysis. Serum tumor markers screened by Meta-analysis were performed optimized screening to establish model in early auxiliary detection of colorectal cancer and evaluate its value, aiming

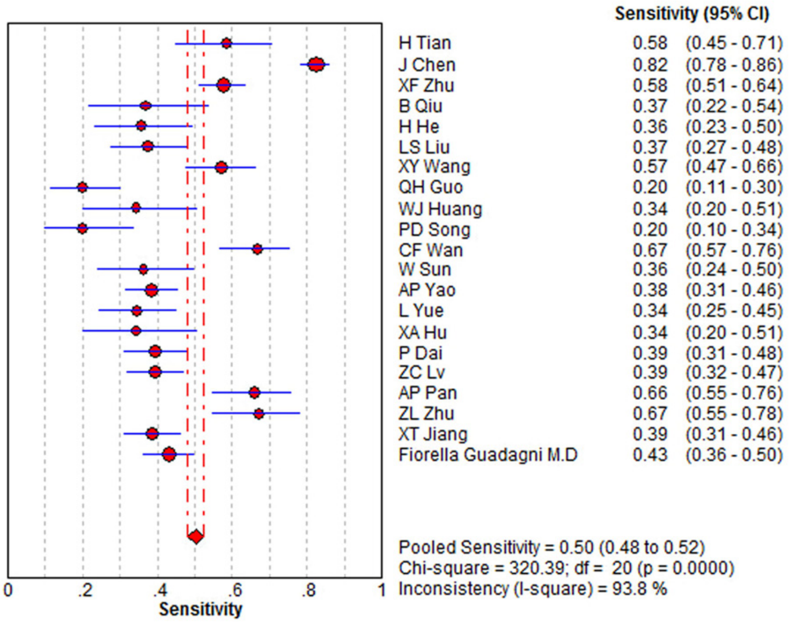

Figure 2. The forest plot of CA724 sensitivity in diagnosing colorectal cancer.

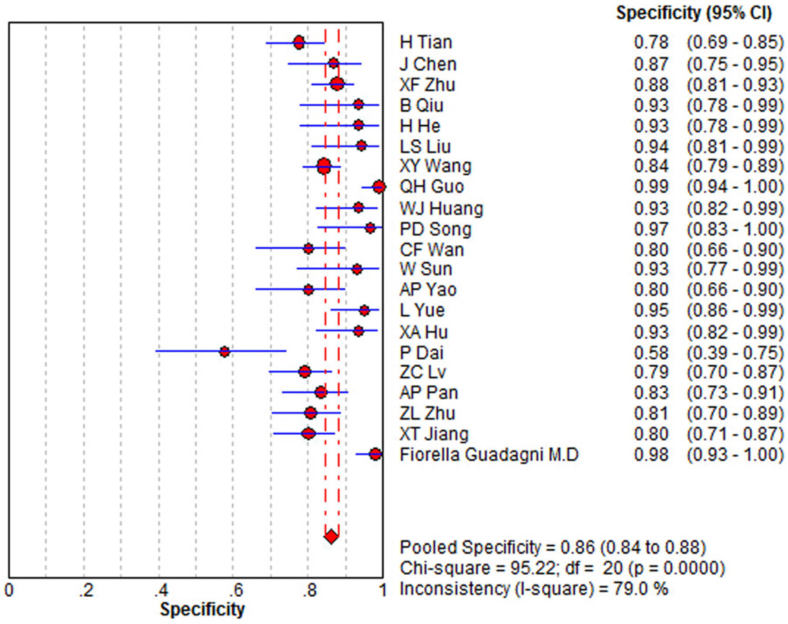

Figure 3. The forest plot of CA724 specificity in diagnosing colorectal cancer.

to provide a theoretical basis for clinical practice.

Materials and methods

Compliance with ethical standards

The studies have been approved by the ethics committee of Henan Provincial People's Hospital and have been performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. All subjects have signed the informed consent.

Document retrieval

CNKI, VIP and WANFANG databases are the primary sources of Chinese documents. Pub Med and MEDLINE databases are the main sources of the English literature search. The Chinese and English key words were "colorectal cancer", "serum tumor markers" and "early diagnosis". The publication year was 1990-2013.

Literature inclusion and exclusion criteria

Inclusion criteria: (1) researches should be related English or Chinese literature about serum tumor markers influencing the early diagnosis of colorectal cancer patients before surgery; (2) retrospective studies; (3) the gold standard was the histopathological or surgery diagnosis; (4) getting the data of four tables (TP, TN, FP and FN) for individual diagnosis of serum markers.

Exclusion criteria were: (1) non-original literature studies or repeated reports; (2) literature reviews or abstracts; (3) patients without diagnosis of the gold standard; (4) only joint diagnosis results of serum marker for colorectal cancer, no single diagnosis results; (5) literature that four tables data were not available.

Data extraction and quality assessment

Data Extraction: (1) authors, publication time, journals, titles, number of cases of the experi-

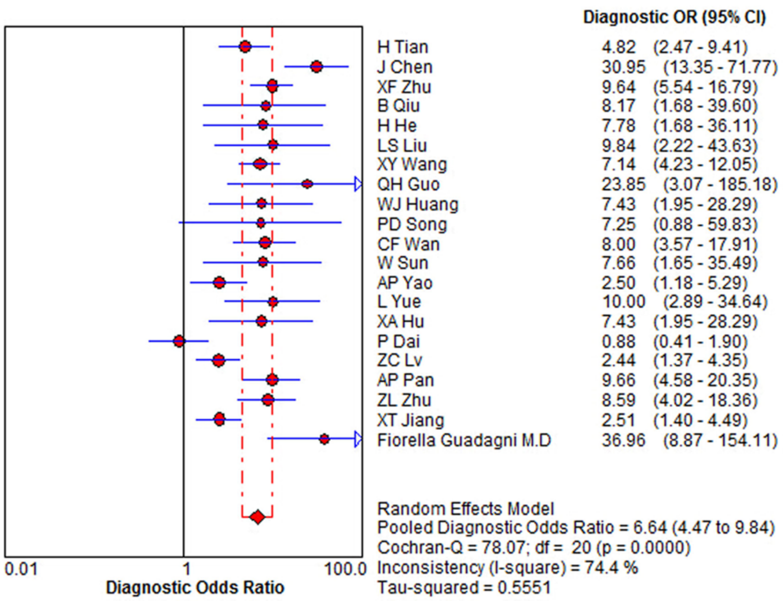

Figure 4. The diagnostic odds ratio of CA724 for colorectal cancer.

mental and control groups; (2) methodological features: cutoff value; (3) the four tabular data of diagnostic results.

Quality assessment: Each included literature was assessed by OUADAS developed by Penny [10] for quality assessment.

Clinical data

One hundred surgical treatment patients with colorectal cancer at the Henan Provincial People's Hospital in 2013-2014 were collected, in which there were 56 males and 44 females, aged 25 to 80 years, with a median age of 58.5 years. At the same period, 50 patients with benign colorectal disease were collected as the control group, including 21 males and 29 females, aged 31 to 82 years, with a median age of 53.0 years. All of the above cases were confirmed by surgery and pathology, and all subjects signed informed consent.

Clinical detection of serum markers

Cobas6000 automatic biochemical immunity analyzer (Roche, Switzerland) was used to detect the content of serum tumor markers by enzyme-linked immunosorbent assay. Specific steps followed the kit instructions.

Establishing an early prediction model of colorectal cancer based on logistic regression analysis

Serum markers were treated as covariate and the pathological diagnosis results of colorectal cancer as dependent variable, stepwise Logistic regression analysis was carried out with a forward method for establishing the early prediction model of colorectal cancer, and then the application value of Logistic regression models in the early detection of colorectal cancer was evaluated using ROC curves.

Establishing early auxiliary detection model of colorectal cancer based on Bhattacharyya-SVM

Bhattacharyya distance was used to order and screen indicators. The Bhattacharyya distance of each indicator between colorectal cancer sample and normal tissue sample is shown in formula (1) [11]. The greater the distance is, the better the effect of classification is.

$$B_i = \frac{1}{4}\frac{(\mu_{i+} - \mu_{i-})^2}{(\sigma_{i+}^2 + \sigma_{i-}^2)} + \frac{1}{2}\ln\left(\frac{\sigma_{i+}^2 + \sigma_{i-}^2}{2\sigma_{i+}\sigma_{i-}}\right) \quad (1)$$

Where, μi+ and σi+ are mean value and variance of samples in colorectal cancer group. μi- and σi- are mean value and variance of samples in benign control group. In this study, the Bhattacharyya distance was calculated by MATLAB programming.

In the next place, SVM was used to verify the specificity of indicators screened by Bhattacharyya distance. The establishment, training and verification of SVM model were realized based on MATLAB programming tool.

Statistical analysis

The literature measurement indexes included in this study were count data, and the heteroge-

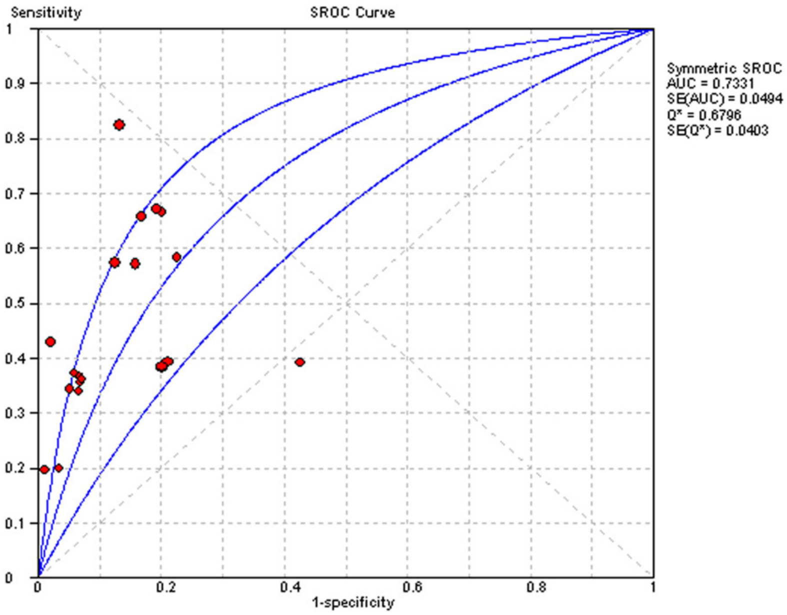

Figure 5. SROC curve of CA724 in diagnosing colorectal cancer.

neity test was carried out. If the homogeneity was better, fixed effect model was used to analyze the data; conversely, random effect model was adopted, and three effect variables with 95% CI, pooled sensitivity, pooled specificity, diagnostic odds ratio were used to analyze statistics, then summary receiver operating characteristic (SROC) curve was made. All data were performed two-tailed test of Meta-analysis, in which *P*<0.05 indicates statistically difference, while *P*<0.01 represents remarkably statistically difference. Meta-Disc software was used for Meta-analysis of diagnostic tests, and the forest plot and SROC curve were attached to explain the results. Significant difference analysis was performed by statistical software SPSS17.0 and the data were expressed by (mean value ± standard deviation). *P*<0.05 indicates significant difference. And logistic regression analysis was adopted to screen serum markers with high diagnostic value. With detection level of these serum tumor markers as test variables, and pathological diagnosis results of colorectal cancer as state variables, ROC curve was made to compare the diagnostic value.

Results

Basic information of the documents included

A total of 1482 Chinese documents and 1093 English documents were retrieved by computer. After screening, 64 documents were included in this Meta-analysis, including 56 Chinese documents and 8 English documents. The flow chart of literature search and screening is shown in **Figure 1**.

Meta-analysis of serum marker CA724

Taking Meta-analysis of serum marker CA724 for instance, 21 documents were included, 20 of which were Chinese literature, 1 of which was English literature. There were 2414 colorectal cancer patients and 1528 cases in the healthy control group. In these 21 studies, the pooled sensitivity of CA724 for colorectal cancer diagnosis was 0.50 (0.48-0.52); pooled specificity was 0.86 (0.84-0.88), the diagnostic odds ratio was 6.64 (4.47-9.84); the area under the ROC curve (AUC) was 0.7331; and a standard error was 0.0494 (**Figures 2-5**).

Meta-analysis of colorectal cancer serum tumor markers

Meta-analysis results of colorectal cancer serum tumor markers are shown in **Table 1**. As shown, the odds ratio of tumor markers CEA, CA242, HSP60, CA724, CA19-9, CA125, CA50, CYFRA21-1, TPA, and UGT1A8 for the diagnosis of colorectal cancer was all greater than 1, showing certain diagnostic value, but AFP and CA153 had no reference value in diagnosing colorectal cancer.

Determination of serum tumor markers content in colorectal cancer

Test results of tumor markers of colorectal cancer group and the benign control group are shown in **Table 2**. The table showed that serum test levels of CEA, CA50, HSP60, CYFRA21-1, TPA, CA19-9, CA242, CA724 and CA125 in the colorectal cancer group were significantly higher than those in the controls (*P*<0.05).

Table 1. Meta-analysis of twelve common serum tumor markers

Serum tumor markers	Number of documents	Number of cases	Control group	Pooled sensitivity (95% CI)	Pooled specificity (95% CI)	Diagnostic odds ratio (95% CI)
CA50	10	1024	679	0.39 (0.36-0.42)	0.82 (0.79-0.85)	3.41 (2.09-5.59)
CA19-9	39	5512	4114	0.48 (0.47-0.50)	0.79 (0.78-0.80)	4.80 (3.67-6.27)
CA125	10	2823	2346	0.45 (0.43-0.47)	0.63 (0.61-0.65)	3.65 (1.60-8.36)
CYFRA21-1	3	511	250	0.42 (0.38-0.47)	0.81 (0.76-0.86)	3.19 (2.21-4.60)
AFP	3	228	149	0.16 (0.12-0.22)	0.90 (0.84-0.94)	1.16 (0.14- 9.82)
CEA	51	6712	4996	0.56 (0.55-0.58)	0.78 (0.77-0.79)	7.37 (5.90-9.22)
CA242	22	3633	3086	0.60 (0.58-0.61)	0.79 (0.78-0.81)	7.15 (5.17-9.90)
HSP60	4	193	144	0.47 (0.40-0.54)	0.88 (0.82-0.93)	6.72 (1.93-23.45)
CA724	21	2414	1528	0.50 (0.48-0.52)	0.86 (0.84-0.88)	6.64 (4.47-9.84)
CA153	3	254	138	0.19 (0.14-0.24)	0.86 (0.79-0.92)	1.36 (0.73-2.53)
TPA	1	59	234	0.17 (0.08-0.29)	0.94 (0.90-0.97)	3.21 (1.35-7.64)
UGT1A8	1	40	30	0.88 (0.73-0.96)	0.87 (0.69-0.96)	45.5 (11.12-186.24)

Table 2. Test results of serum tumor markers of colorectal cancer patients ($\overline{x}\pm s$)

Group	CEA	CA50	HSP60	CYFRA21-1	TPA	AFP
Indicator	(ng/mL)	(U/mL)	(pg/mL)	(ng/mL)	(U/mL)	(ng/ml)
Colorectal cancer group	29.31±34.11*	50.75±33.82*	587.29±497.59*	8.75±12.15*	0.86±1.81*	16.97±4.83*
Control group	4.28±1.39	10.52±12.96	201.45±126.46	1.98±1.30	0.081±0.29	3.05±0.85
Group	CA19-9	CA242	CA724	CA125	UGT1A8	CA153
Indicator	(U/mL)	(U/mL)	(U/mL)	(U/mL)	(ng/mL)	(U/ml)
Colorectal cancer group	50.10±32.69*	43.67±31.60*	15.87±14.68*	40.14±15.35*	7.52±2.22*	20.96±8.54*
Control group	24.31±12.00	8.32±6.88	4.49±2.48	15.05±15.54	34.62±32.39	15.69±2.58
Group	CEA	CA50	HSP60	CYFRA21-1	TPA	AFP
Indicator	(ng/mL)	(U/mL)	(pg/mL)	(ng/mL)	(U/mL)	(ng/ml)
Colorectal cancer group	29.31±34.11*	50.75±33.82*	587.29±497.59*	8.75±12.15*	0.86±1.81*	16.97±4.83*
Control group	4.28±1.39	10.52±12.96	201.45±126.46	1.98±1.30	0.081±0.29	3.05±0.85
Group	CA19-9	CA242	CA724	CA125	UGT1A8	CA153
Indicator	(U/mL)	(U/mL)	(U/mL)	(U/mL)	(ng/mL)	(U/ml)
Colorectal cancer group	50.10±32.69*	43.67±31.60*	15.87±14.68*	40.14±15.35*	7.52±2.22*	20.96±8.54*
Control group	24.31±12.00	8.32±6.88	4.49±2.48	15.05±15.54	34.62±32.39	15.69±2.58

Note: *indicates that compared with the control group, the difference was significant, $P<0.05$.

Establishing early detection model of colorectal cancer based on logistic regression analysis

Stepwise logistic regression analysis of each indicator: With 12 indicators such as CEA, CA50, HSP60, CYFRA21-1 as the covariate, and pathological diagnosis results of colorectal cancer as the dependent variable, forward stepwise Logistic regression analysis was made, and the results are presented in **Table 3**. Moreover, as indicated in **Table 4**, 3 independent variables were eventually selected in the Logistic regression equation, namely CEA, CA19-9 and HSP60. The Logistic regression equation was Logit *P*=-0.996+1.721CEA+1.252HSP60+1.920CA19-9, thus a new set of variables were generated.

ROC curve analysis of CEA, CA19-9, HSP60 and logistic regression model

Figure 6 shows the ROC curves of the separate testing of CEA, CA19-9 as well as HSP60 and Logistic regression model joint testing of three tumor markers. **Table 5** indicated that the AUC of ROC curves of the separate detection of CEA, CA199 as wells as HSP60 and Logistic regression model joint detection of the three for the diagnosis of colorectal cancer were 0.762, 0.752, 0.825 and 0.906. *P* values were less than 0.01. The results showed CEA, CA19-9 and HSP60 all had certain diagnostic value in

Table 3. Logistic regression analysis by step forward method

			Score	df	Sig.
Step 1	Variables	CEA	17.982	1	0.000
		HSP60	10.170	1	0.001
		CYFRA21-1	2.291	1	0.130
		TPA	3.615	1	0.057
		AFP	0.104	1	0.747
		UGT1A8	3.522	1	0.061
		CA242	14.511	1	0.000
		CA724	4.334	1	0.037
		CA125	0.297	1	0.586
		CA153	0.201	1	0.654
		CA50	8.251	1	0.004
	Total statistics		32.269	11	0.001
Step 2	Variables	HSP60	7.470	1	0.006
		CYFRA211	0.370	1	0.543
		TPA	2.926	1	0.087
		AFP	2.629	1	0.105
		UGT1A8	0.883	1	0.347
		CA242	4.508	1	0.034
		CA724	0.013	1	0.909
		CA125	0.142	1	0.707
		CA153	0.241	1	0.624
		CA50	0.249	1	0.618
	Total statistics		17.147	10	0.071
Step 3	Variables	CYFRA211	0.077	1	0.781
		TPA	2.176	1	0.140
		AFP	2.942	1	0.086
		UGT1A8	0.512	1	0.474
		CA242	3.182	1	0.074
		CA724	0.007	1	0.933
		CA125	0.102	1	0.749
		CA153	0.066	1	0.797
		CA50	0.301	1	0.583
	Total statistics		10.745	9	0.294

colorectal cancer. Additionally, the diagnostic value of Logistic regression model of the combined detection of three indicators was much better than any single detection of the indicator. The detection accuracy of this model was 82.67%; the sensitivity was 96.90% and specificity was 90.57%.

Establishing early detection model of colorectal cancer based on Bhattacharyya-SVM

Bhattacharyya distance analysis of each indicator: The Bhattacharyya distance of 12 tumor markers are shown in **Table 6**. As can be seen, CA50, CEA, CA724 and CA199 had greater Bhattacharyya distance obviously. The Bhattacharyya distances of them were 4.2107, 3.4608, 3.4332 and 3.2135, respectively, which was followed by CYFRA21-1, CA125 AND UGT1A8 and the Bhattacharyya distances of them were 2.7314, 2.4567 and 2.3742.

Establishing different diagnostic models of colorectal cancer: 12 tumor markers were selected to establish the diagnostic model of colorectal cancer. Then the test data of 20 colorectal cancer patients and 10 benign controls were input into the model. The test results are shown in **Figure 7**. The hollow circles represent target output. * is the actual simulation output of SVM. As can be learned from it, the accuracy of this diagnostic model was 23/30=76.7%, sensitivity 80.0%, and specificity 70.0%.

The 4 tumor markers with Bhattacharyya distance > 3 were CEA, CA50, CA724 and CA199, which were used to establish SVM diagnostic model. The results after inputting the test data are shown in **Figure 8**. The classification accuracy of this model was 25/30=83.3%, sensitivity 17/20=85.0%, specificity 8/10=80.0%

The 7 tumor markers with Bhattacharyya distance>2 were CEA, CA50, CYFRA21-1, CA199, CA724, CA125 and UGT1A8, which were used to establish SVM diagnostic model. The results after inputting the test data are shown in **Figure 9**. The classification accuracy of this model was 27/30=90.0%, sensitivity 18/20=90.0%, and specificity 9/10=90.0%.

Comparing these three SVM models, it was found that the SVM diagnostic model based on CEA, CA50, CYFRA21-1, CA199, CA724, CA125 and UGT1A8 had the highest classification accuracy, so this SVM model was chosen.

Discussion

Currently, serum tumor markers have become one of the most commonly used method in clinic after radiological and pathological diagnosis. Carcinoembryonic antigen (CEA) is an acid glycoprotein extracted from adenocarcinoma of colon, keeping a low level in healthy human serum and being applied mostly in the malignant tumors of digestive tract. It is a good tumor

Table 4. Variables in Logistic regression equation

		B	S.E	Wals	df	Sig.	Exp (B)
Step 1[a]	CA19-9	1.839	0.420	19.158	1	0.000	6.291
	Constant	0.024	0.220	0.012	1	0.913	1.024
Step 2[b]	CEA	1.806	0.450	16.138	1	0.000	6.086
	CA19-9	1.922	0.447	18.508	1	0.000	6.834
	Constant	-0.640	0.283	5.102	1	0.024	0.527
Step 3[c]	CEA	1.721	0.462	13.911	1	0.000	5.592
	HSP60	1.252	0.472	7.044	1	0.008	3.496
	CA19-9	1.920	0.459	17.502	1	0.000	6.823
	Constant	-0.996	0.325	9.371	1	0.002	0.369

Note: [a]: the variable input in step 1 was CA199. [b]: the variable input in step 2 was CEA. [c]: the variable input in step 3 was HSP60.

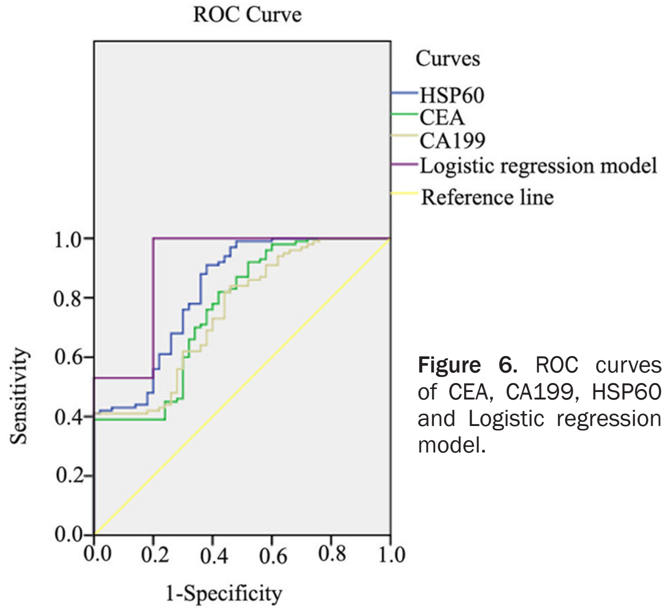

Figure 6. ROC curves of CEA, CA199, HSP60 and Logistic regression model.

marker in the diagnosis, efficacy judgment, disease progression and prognosis detection of colorectal cancer, breast cancer and lung cancer [12]. CA242 is a relatively new tumor marker in recent clinical application, which has a sensitive testing effect on pancreatic and colon cancer [13]. Liu [14] found that the positive rate of pancreatic cancer serum markers CEA and CA242 were 36.4% and 42.7%, and those in the benign controls were 14.8% and 23.6%. This study, taking CNKI, VIP, WANFANG, Pub Med and MEDLINE databases as the main source of the literature search, collected relevant literature about serum markers for the early diagnosis of colorectal cancer, and then Meta-analysis of diagnostic tests was used for comprehensive and quantitative analysis of related literature to study the efficacy for the early diagnosis of colorectal cancer. It was found when tested individually, common serum markers CEA, CA242, HSP60, CA724, CA19-9, CA125, CA50 and CYFRA21-1 all had certain detection value in colorectal cancer. The pooled sensitivity of CEA was 0.56 (0.55-0.58), pooled specificity 0.78 (0.77-0.79) and diagnostic odds ratio 7.37 (5.90-9.22). It has higher diagnostic efficiency than other serum markers and the best determination effect of diagnostic test.

Heat shock protein 60 (HSP60) is an important member of the family of HSPs. In 2010, Cappello F [15] found HSP was closely related to the transformation and canceration of cells. HSP were over expressed in colorectal cancer cell lines. In this paper, HSP60 was conducted diagnostic test Meta-analysis, indicating that its pooled sensitivity was 0.47 (0.40-0.54), pooled specificity 0.88 (0.82-0.93) and diagnostic odds ratio 6.72 (1.93-23.45). It has higher diagnostic efficacy than CA724, CA19-9 and CA125 and can be used as a new type of serum markers for early diagnosis of colorectal cancer. Alpha-fetoprotein (AFP) is mainly synthesized in the fetal liver. Since higher level of AFP was found in about 80% of hepatic carcinoma patients, it has been considered as the specific tumor markers of primary liver cancer [16, 17]. The study found that, the diagnostic odds ratio of AFP in colorectal cancer detection was 1.16 (0.14-9.82), showing no diagnostic value in colorectal cancer.

Bagaria [18] used ROC curve to analyze the sensitivity and specificity of CEA and CA19-9 in detecting esophagus, stomach and colorectal cancer, and found that CEA had a high sensitivity in colorectal cancer; CA19-9 had a high sensitivity in gastric cancer; and different serum tumor markers had different manifestations in various cancers. Wang [19] also analyzed the application value of CEA, AFP and CA19-9 in the diagnosis of colorectal cancer. The sensitivity and accuracy of CEA were all the highest. Its AUC of ROC curve reached 0.88. By using the similar analytical method, three indicators CEA, CA19-9 and HSP60 in this study were screened

Table 5. The area under the ROC curve of CEA, CA19-9, HSP60 and Logistic regression model

Variable	AUC	Standard error[a]	Sig.[b]	Approximation 95% CI	
				Lower limit	Upper limit
HSP60	0.825	0.036	0.000	0.754	0.897
CEA	0.762	0.042	0.000	0.680	0.844
CA19-9	0.752	0.041	0.000	0.671	0.833
Logistic regression model	0.906	0.029	0.000	0.850	0.962

Note: [a]: under condition of non parametric hypothesis. [b]: Null hypothesis: true area = 0.5.

Table 6. The Bhattacharyya distance between two sets of samples of tumor markers

Indicator	CEA	CA50	HSP60	CYFRA21-1	TPA	CA199
Bhattacharyya distance	3.4608	4.2107	1.2176	2.7314	0.9357	3.2135
Indicator	AFP	CA242	CA724	CA125	CA153	UGT1A8
Bhattacharyya distance	1.0877	1.7578	3.4332	2.4567	1.0739	2.3742

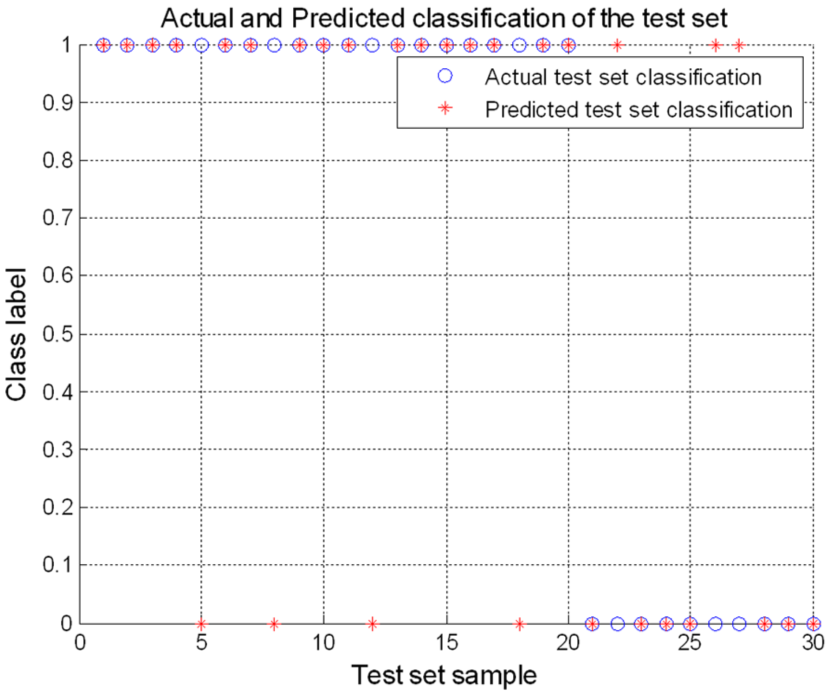

Figure 7. Test results of SVM model based on 12 tumor markers.

from 12 tumor markers CEA, HSP60, CYFRA21-1, TPA, CA19-9, CA242, CA50, CA724, CA125 and UGT1A8, and then included in the Logistic regression model. The AUC of the model was 0.906, significantly higher than that of any of the three tumor markers. The results showed that Logistic regression analysis, as a statistical method, can improve the diagnostic specificity and sensitivity by combining Meta-analysis, and can be used as an important method for the identification of benign or malignant colorectal cancer clinically.

In statistics, many methods are used to measure the specificity of certain feature. Currently the most widely used is the distance metric. Distance measurement can also be referred to as “dispersion criterion” or “classification separability criterion”. Distance as an important concept in statistical pattern recognition, includes Bhattacharyya distance, Euclidean distance and Mahalanobis distance [20], while Bhattacharyya distance is suitable for data with high and low dimensionality, so its application is wide. Statistical pattern recognition found that the greater the distance between different categories is, the greater the separability of category is, and the lower the corresponding classification error rate is. In the actual application process, subsets of these features are usually selected.

In this paper, tumor markers with high specificity in colorectal cancer screened by Bhattacharyya distance were CA50, CEA, CA724 and CA199. Studies reported that CEA and CA724 have a high value for the diagnosis of colorectal cancer [21]. Jolanda et al [22] has found that 7.3% of patients with colorectal cancer had elevated CA19-9, while 55.4% of the patients had elevated CA19-9 and CEA. Eskelinen et al [23] examined serum CEA, CA50 and CA242 levels of 138 patients with colorectal cancer and 104 healthy controls.

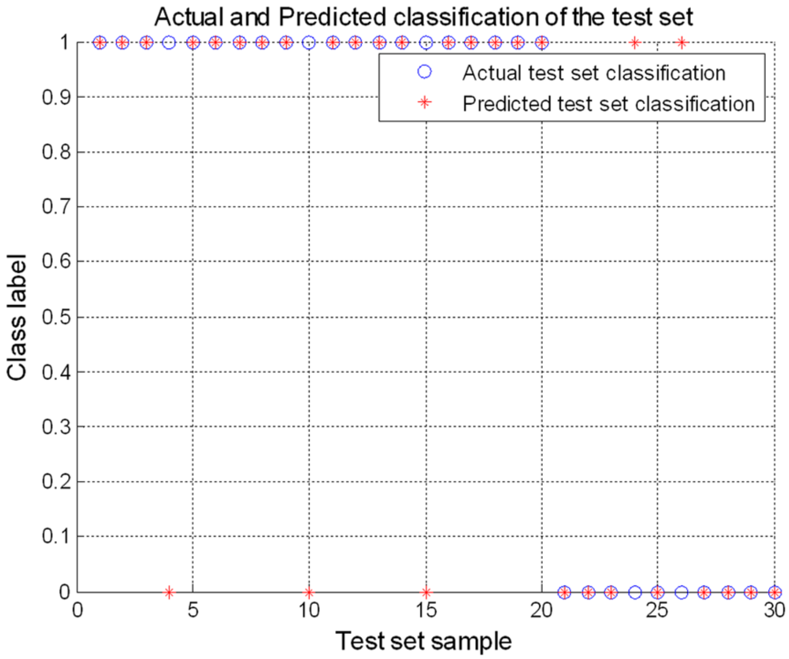

Figure 8. Test results of SVM model based on 4 tumor markers.

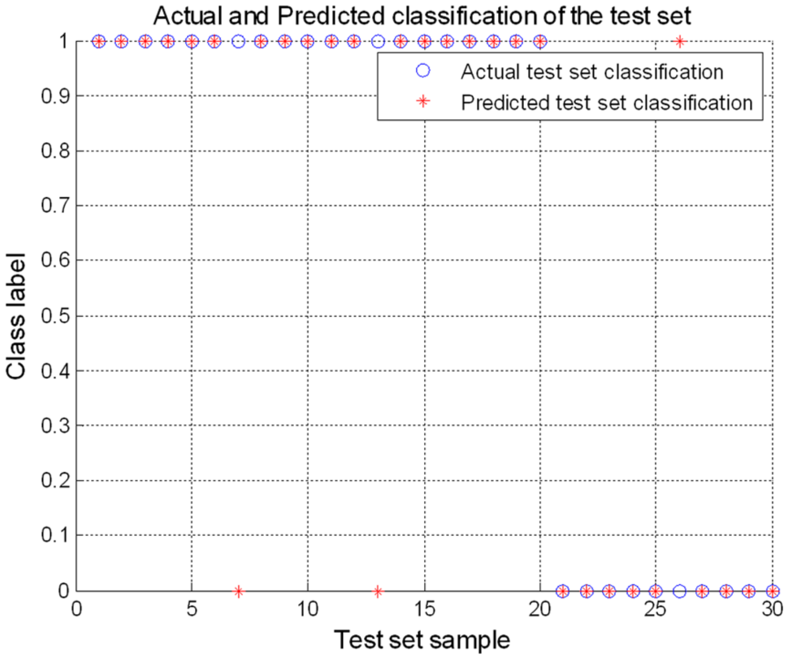

Figure 9. Test results of SVM model based on 7 tumor markers.

Then, clinical data of these tumor markers were analyzed. With 90% as the specificity for division, CEA content was 2.5 ng/mL, CA50 was 17 U/mL and CA242 was 17 U/mL. The results of sensitivity showed that CEA was 0.63 and CA50 was 0.30. CEA was the most important predictive indicator in colorectal cancer, followed by CA242. To calculate the contribution rate of tumor marker tests, the diagnostic value (DS) was used. The sensitivity of DS in the detection of colorectal cancer was 0.47; specificity was 0.88; and the effectiveness was 0.67. Based on this study, serum CEA and CA242 can be the discrimination criteria in preoperative evaluation of patients with colorectal cancer.

The results of this study showed that when a joint SVM model was established using 12 indicators, the classification accuracy of the model was 76.7%, sensitivity 70.0%, and specificity 70.0%, which was less ideal. When establishing the SVM model using 7 markers (CEA, NSE, CYFRA21-l, AFP, and CA724), the discrimination accuracy, sensitivity and specificity were all 90.0%. It suggested that the choice of too many markers in the identification of benign or malignant colorectal tumors may interfere with the discrimination rate of useful indicators due to redundancy indexes, resulting in the low discrimination accuracy. In the results of this study, when establishing SVM model using CEA, NSE, CA724 and AFP, which had the highest Bhattacharyya distance, the discrimination accuracy of the model was 83.3%, sensitivity 85.0%, and specificity 80.0%, which were lower than the model established by the 7 indicators. Liu [14] has also found that the joint of CA19-9, NSE, CEA, CA242 and CA125 was very helpful for diagnosis of pancreatic cancer. In short, the less the better is not the law of indicator selection, for too few indexes may lead to the instability and occasionality of result.

Conclusions

This study used Meta-analysis for comprehensive and quantitative analysis of the relevant literature on serum markers in the diagnosis of colorectal cancer and screened serum markers which can be used for early diagnosis of colorectal cancer. Logistic regression analysis and Bhattacharyya-SVM analysis were for further screening on the 12 serum markers screened by Meta-analysis, assessed its value in the diagnosis of colorectal cancer, and established two serum marker models in favor of the early diagnosis of colorectal cancer. The SVM model in the judgment of benign or malignant colorectal cancer based on CEA, NSE, CYFRA21-I, AFP, CA724 and other seven indicators was better than the Logistic regression model based on CEA, CA199 and HSP60, so this study provides a reference for the early diagnosis of colorectal cancer.

Disclosure of conflict of interest

None.

Address correspondence to: Jianchao Luo, Department of Tumor Radiotherapy, Henan Provincial People's Hospital, No.7 Weiwu Road, Zhenghzhou 450003, Henan Province, China. Tel: +86136-07649987; E-mail: luojc2015@126.com

References

[1] Jemal A, Siegel R, Ward E, Ward Y, Hao J and Xu T. Cancer statistics, 2008. CA Cancer J Clin 2008; 58: 71-96.

[2] Song M, Garrett WS and Chan AT. Nutrients, food, and colorectal cancer prevention. Gastroenterology 2015; 148: 1244-1260.

[3] Toiyama Y, Tanaka K, Kitajima T, Shimura T, Kawamura M, Kawamoto A, Okugawa Y, Saigusa S, Hiro J, Inoue Y, Mohri Y, Goel A and Kusunoki M. Elevated serum angiopoietin-like protein 2 correlates with the metastatic properties of colorectal cancer: a serum biomarker for early diagnosis and recurrence. Clin Cancer Res 2014; 20: 6175-6186.

[4] Fernández-Esparrach G, Alberghina N, Subtil JC, Vázquez-Sequeiros E, Florio V, Zozaya F, Araujo I and Ginès A. Endoscopic ultrasound-guided fine needle aspiration is highly accurate for the diagnosis of perirectal recurrence of colorectal cancer. Dis Colon Rectum 2015; 58: 469-473.

[5] Li S, Wang J, Lu Y and Fan D. Screening and early diagnosis of colorectal cancer in China: a 12 year retrospect (1994-2006). J Cancer Res Clin Oncol 2007; 133: 679-686.

[6] Herberman RB, Ortaldo JR. Natural killer cells: their roles in defenses against disease. Science 1981; 214: 24-30.

[7] Albert MB, Steinberg WM, Henry JP. Elevated serum levels of tumor marker CA19-9 in acute cholangitis. Dig Dis Sci 1988; 33: 1223-1225.

[8] Chaiyakunapruk N, Saokaew S, Sruamsiri R, Dilokthornsakul P. Systematic review and network meta-analysis in health technology assessment. J Med Assoc Thai 2014; 97: 33-42.

[9] Vahey NA, Nicholson E, Barnes-Holmes D. A meta-analysis of criterion effects for the Implicit Relational Assessment Procedure (IRAP) in the clinical domain. J Behav Ther Exp Psychiatry 2015; 48: 59-65.

[10] Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003; 3: 25.

[11] Al-Shuneigat JM, Mahgoub SS, Huq F. Colorectal carcinoma: nucleosomes, carcinoembryonic antigen and ca 19-9 as apoptotic markers; a comparative study. J Biomed Sci 2011; 18: 50.

[12] Gold P, Freedman SO. Demonstration of tumor-specific antigens in human colonic carcinoma by immunological tolerance and absorption techniques. J Exp Med 1965; 121: 439-462.

[13] Yang XQ, Chen C, Peng CW, Liu SP, Li Y. Carbohydrate antigen 242 highly consists with carbohydrate antigen 19-9 in diagnosis and prognosis of colorectal cancer: study on 185 cases. Med Oncol 2012; 29: 1030-1036.

[14] Liu F, Du F, Chen X. Multiple tumor marker protein chip detection system in diagnosis of pancreatic cancer. World J Surg Oncol 2014; 12: 333.

[15] Cappello F, David S, Peri G, Farina F, Conway de Macario E, Macario AJ, Zummo G. Hsp60: molecular anatomy and role in colorectal cancer diagnosis and treatment. Front Biosci (Schol Ed) 2010; 3: 341-351.

[16] Chen QW, Cheng CS, Chen H, Ning ZY, Tang SF, Zhang X, Zhu XY, Vargulick S, Shen YH, Hua YQ, Xie J, Shi WD, Gao HF, Xu LT, Feng LY, Lin JH, Chen Z, Liu LM, Ping B, Meng ZQ. Effectiveness and complications of ultrasound guided fine needle aspiration for primary liver cancer in a Chinese population with serum α-fetoprotein levels ≤200 ng/ml--a study based on 4,312 patients. PLoS One 2014; 9: e101536.

[17] Zheng L, Gong W, Liang P, Huang X, You N, Han KQ, Li YM, Li J. Effects of AFP-activated PI3K/Akt signaling pathway on cell proliferation of liver cancer. Tumour Biol 2014; 35: 4095-4099.

[18] Bagaria B, Sood S, Sharma R, Lalwani S. Comparative study of CEA and CA19-9 in esophageal, gastric and colon cancers individually and in combination (ROC curve analysis). Cancer Biol Med 2013; 10: 148-157.

[19] Wang YR, Yan JX, Wang LN. The diagnostic value of serum carcino-embryonic antigen, alpha fetoprotein and carbohydrate antigen 19-9 for colorectal cancer. J Cancer Res Ther 2014; 10: 307-309.

[20] FuKunaga K. Introduction to statistical pattern recognition. 2nd edition. Boston: Academic Press; 1990.

[21] Gebauer G, Muller RW. Tumor marker concentrations in normal and malignant tissues of colorectal cancer patients and their prognostic relevance. Anticancer Res 1997; 17: 2939-2942.

[22] Stiksma J, Grootendorst DC, van der linden PW. CA19-9 as a marker in addition to CEA to monitor colorectal cancer. Clin Colorectal Cancer 2014; 13: 239-244.

[23] Eskelinen M, Pasanen P, Kulju A, Janatuinen E, Miettinen P, Poikolainen E, Tarvainen R, Nuutinen P, Pääkkönen M, Alhava E. Clinical evaluation of serum tumour markers CEA, CA 50 and CA 242 in colorectal cancer. Anticancer Res 1994; 14: 1427-1432.