Models of logistic regression analysis, support vector machine, and back-propagation neural network based on serum tumor markers in colorectal cancer diagnosis

B. Zhang[1], X.L. Liang[2], H.Y. Gao[2], L.S. Ye[2] and Y.G. Wang[1]

[1]Training Department, Third Military Medical University of Chinese PLA, Shapingba, Chongqing, China
[2]Medical Department, Military General Hospital of Beijing PLA, Dongcheng, Beijing, China

Corresponding author: Y.G. Wang
E-mail: yungui.wang@aol.com

ABSTRACT. We evaluated the application of three machine learning algorithms, including logistic regression, support vector machine and back-propagation neural network, for diagnosing congenital heart disease and colorectal cancer. By inspecting related serum tumor marker levels in colorectal cancer patients and healthy subjects, early diagnosis models for colorectal cancer were built using three machine learning algorithms to assess their corresponding diagnostic values. Except for serum alpha-fetoprotein, the levels of 11 other serum markers of patients in the colorectal cancer group were higher than those in the benign colorectal cancer group ($P < 0.05$). The results of logistic regression analysis indicted that individual detection of serum carcinoembryonic antigens, CA199, CA242, CA125, and CA153 and their combined detection was effective for diagnosing colorectal cancer. Combined detection had a better diagnostic effect with a sensitivity of 94.2% and

specificity of 97.7%; combining serum carcinoembryonic antigens, CA199, CA242, CA125, and CA153, with the support vector machine diagnosis model and back-propagation, a neural network diagnosis model was built with diagnostic accuracies of 82 and 75%, sensitivities of 85 and 80%, and specificities of 80 and 70%, respectively. Colorectal cancer diagnosis models based on the three machine learning algorithms showed high diagnostic value and can help obtain evidence for the early diagnosis of colorectal cancer.

Key words: Back-propagation neural network; Colorectal cancer; Logistic regression analysis; Serum marker; Support vector machine

INTRODUCTION

Colorectal cancer is a malignant tumor with high incidence; thus, efficient screening and early diagnosis methods are important for early treatment to increase the survival rate of patients (Li, 2009). Compared to other screening methods, traditionally natural population screening of colorectal cancer shows advantages such as easy operation, specificity, and high detection ability, but exhibits low compliance and screening efficiency, resulting in misdiagnosis of colorectal patients in early stages. For example, according to a survey in 2012, 34.45% of patients were incorrectly diagnosed (Domínguez-Ayala et al., 2012). Therefore, the early diagnosis rate of colorectal cancer must be improved to enable patients to receive earlier treatment.

Herberman first proposed the concept of tumor markers at the US Tumor Immune Diagnostic Conference in 1978; with advances in medical research, markers have been increasingly used for the early diagnosis of tumors in the clinical setting (Kurman et al., 1977; Albert et al., 1988). Tumor markers show great practical value for the screening, diagnosis, prognosis, and outcome of tumor, evaluation of efficacy, and follow-up observation of high risk populations. Tumor marker examination is rapid and convenient, and thus has gained increasing attention in the screening, diagnosis, and prognosis of tumors. Recent studies have suggested that combination detection of several tumor markers can improve the sensitivity and specificity of diagnosis of tumor disease, which has great potential for clinical application (Hayes et al., 1996; Parkin et al., 2005).

As an optimal theory for small-sample leaning, the support vector machine (SVM) is currently a hot spot in studies of intelligent diagnosis. This method begins with data classification, and then the data are mapped into a higher dimensional characteristics space through nonlinear properties of the kernel function definition, showing lineal properties and the determinations of decision rules of sample classification (Vapnik and Zhang, 2000). SVM can be used to overcome classification problems such as small sample size and nonlinear with good generalization capability. In this study, during SVM modeling, the kernel function used was the radial basis function and penalty parameters.

A back-propagation (BP) neural network is built as an artificial intelligence system with functions including learning, association, remembering, and pattern identification according to the discussion and learning of working principle of the human brain (Deng et al., 2012). In this study, the BP neural network model consisted of an input layer, hidden layer, and output layer, including neuron carcinoembryonic antigens (CEAs), CA199, CA242, CA125, and CA153. The primary principle of this method is that there are no connections

between neurons on the same layer, but rather between neighboring layers. Received external information can be passed successively to neurons in the intermediate layer through the input layer, and the output layer will then give the processed results. From this, the processing procedure of forward propagation is conducted. If the actual output is inconsistent with the expected result, the back propagation of deviation will begin. This cycle repeats itself until satisfactory results are obtained. During this process, the weight of every layer is continually reset to optimize the BP neural network (Hall, 2004).

This study used medical examination resources in modern big data situation, serum tumor markers were screened out through meta-analysis; next, based on SVM, logistic regression analysis, and receiver operator characteristics curve (ROC), the screened serum tumor markers were screened out again and their application values in colorectal cancer were evaluated as well. Moreover, a serum tumor marker model useful for the early diagnosis of colorectal cancer was built. The three methods (ROC, SVM, and BP neural network) were compared for their abilities to screen for an optimal diagnostic model of colorectal cancer based on serum markers. In accordance with the results, we identified one optimal combination of tumor markers through logistic regression analysis. The combination includes CEA marker, CA199, CA242, CA125, and CA153, and showed sensitivities and specificities of 94.2 and 97.7%, respectively. Based on these five markers, the SVM diagnosis model of colorectal cancer and BP neural network model of colorectal cancer were established and showed correct detection rates of 82.5 and 75%, sensitivities of 85 and 80%, and specificities of 80 and 70%, respectively. This study provides a theoretical basis for the clinical diagnosis of early colorectal cancer.

MATERIAL AND METHODS

Datasets

This study included 120 patients with colorectal cancer (68 males and 52 females) ranging from 25 to 82 years old; pathological type and differentiation degrees were in accordance with standards of the WHO. Patients included 88 cases of colorectal tubular adenocarcinoma, 19 cases of mucinous carcinoma, and 13 cases of papillary/tubular adenocarcinoma, as well as 19 low differentiated cases, 87 moderately differentiated cases, and 14 well-differentiated cases (Demetri et al., 2007). A total of 86 benign control patients (39 males and 47 females) ranging from 32 to 80 years old, including 42 clinical definite colitis, 8 cases of rectal villous and papilloma, 27 cases of colorectal polyp, and 9 cases of colorectal tubular and villous adenoma were also included. Patients who went to Military General Hospital of Beijing PLA from January 2013 to December 2013 were included as study data, and two-class taxonomy was used to divide the samples. In this study, only three colorectal cancers were collected because during the period of sample collection, these three types were most prevalent. Sample collection also depended on the collection time. The patients described above underwent surgery and were diagnosed pathologically. All patients signed on informed consent.

Serum collection and testing

For blood sample collection, 10 mL fasting venous blood was collected from each subject in the morning and the sample was centrifuged within 2 h of collection at about 1000 *g* for 10 min. Serum sample were stored at -20°C.

Blood samples from colorectal patients were examined in an enzyme-linked immunosorbent assay using the Cobas6000 Automatic biochemical analyzer (Roche, Basel, Switzerland) to determine the levels of CEA, CA50, HSP60, CYFRA21-1, TPA, AFP, CA242, CA724, CA125, CA153, and UGT1A8 in the blood. Kits and the instrument were conducted in accordance with the manufacturer instructions.

For serum tumor marker determination, when their level exceeded a critical value, the samples were considered to be positive; otherwise, they were considered to be negative. UGT1A8 was also subjected to these rules.

Logistic regression analysis and model establishment and evaluation

The levels of 12 tumor markers (CEA, CA50, HSP60, CYFRA21-1, TPA, AFP, CA199, CA242, CA724, CA125, CA153, and UGT1A8) in the rectal case group and benign rectal group were examined using two-category assignment in which the level of tumor marker examination within normal limits was assigned as 0 and higher or lower than normal limits as 1. Using these 12 indexes as covariants and using pathology diagnosis results of colorectal (colorectal patients = 1, benign ones = 0) as dependent variables, forward logistic stepwise regression analysis was used to screen serum tumor markers to verify malignant and benign colorectal cancer. After these calculations, final modeling indices were determined.

Regarding each single index screened out by logistic regression and the detection level of its new variables as test variables and the pathological results of colorectal cancer as state variables, respectively, subjects' ROC was drawn. Next, according to the area under the ROC (AUC), the diagnostic value of each variable for colorectal cancer was evaluated. The AUC value ranged from 0.5 to 1.0. When AUC < 0.5, the results were not considered to be realistic; when AUC = 0.5, the results indicated that this diagnostic method was not diagnostically efficient and could not differentiate between malignant and benign cases. When the AUC value ranged from 0.5 to 0.7, the diagnostic method showed low accuracy. When the value ranged from 0.7 to 0.9, it showed some accuracy. When the value was higher than 0.9 to 1, it showed higher accuracy and better diagnostic efficacy (Yu, 1998).

Methodology

Establishment of disease diagnosis model based on SVM

Data from 206 patients were subjected to normalization processing, in which by using an algorithm, the data are limited to a specified range to ensure convenience in processing for fast convergence for program execution. The target of patients was marked as 1 and that of healthy subjects as 0. One hundred malignant tumor patients and 66 benign cases were used as the testing set sample and then entered into the SVM to be trained. The remaining 20 malignant tumor patients and 20 benign cases were entered into the trained SVM network to obtain corresponding discrimination results (1 or 0) and then compared to targets to determine the discriminate accuracy.

Establishment of diagnosis model of colorectal cancer based on BP neural network

One hundred and fifty-six cases were chosen from 206 colorectal cancer cases as a

training sample and the remaining 50 cases as a testing set. Using MATLAB R2012a, a 3-layer BP network model with a single hidden layer including an input layer, hidden layer, and output layer was built.

Statistical analysis

SPSS 17.0 statistical software (SPSS, Inc., Chicago, IL, USA) was employed to analyze significant differences in the assay data. The data are reported as means ± standard deviation. $P < 0.05$ indicated that the difference was statistically significant. Using the detection levels of serum tumor markers as independent variables and diagnostic results as dependent variables, the data were analyzed by binary logistic regression analysis. Subjects' ROCs of serological markers were calculated by combination detection, the AUC, and its sensitivity and specificity; finally, the diagnostic values of serum-related markers were evaluated. The establishment, training, and validation of the SVM model and BP neural network model were all computed using the MATLAB program.

RESULTS

Testing of serological samples

As shown in Table 1, for the 12 screening serum tumor markers, except serum AFP, the serum testing levels CEA, CA199, CA242, CA153, and CA242, etc. in the colorectal cancer group were all higher than that in the benign colorectal cancer group ($P < 0.05$).

Table 1. Twelve serum markers and their levels in colorectal cancer cases (means ± SD).

Group index	CEA (ng/mL)	CA50 (U/mL)	HSP60 (pg/mL)	CYFRA21-1 (ng/mL)	TPA (U/mL)	AFP (ng/mL)
Colorectal cancer	30.13 ± 21.42*	50.56 ± 25.25*	672.73 ± 321.72*	11.77 ± 7.34*	1.80 ± 1.39*	9.06 ± 4.41
Control	4.29 ± 1.30	12.41 ± 9.00	209.78 ± 106.58	2.13 ± 1.37	0.25 ± 0.20	6.53 ± 4.16
	CA199 (U/mL)	CA242 (U/mL)	CA724 (U/mL)	CA125 (U/mL)	CA153 (U/mL)	UGT1A8 (ng/mL)
Colorectal cancer	48.21 ± 28.72*	46.24 ± 27.72*	18.20 ± 12.00*	38.91 ± 15.35*	21.91 ± 6.29*	7.45 ± 2.25*
Colorectal cancer	25.56 ± 11.67	10.11 ± 6.60	4.36 ± 2.40	18.04 ± 11.13	15.66 ± 8.63	40.30 ± 21.52

*Significant difference compared to the control group ($P < 0.05$).

Diagnosis model establishment of colorectal cancer based on logistic regression analysis

Regarding the 12 indices CEA, CA50, HSP60, CYFRA21-1, TPA, AFP, CA199, CA242, CA724, CA125, CA153, and UGT1A8 as covariants and pathological diagnosis results of colorectal cancer as a dependent variable, forward logistic stepwise regression analysis was conducted and the results are shown in Table 2. As shown in Table 2, five independent variables, CA199, CA242, CA125, CA153, and CEA, were entered into the logistic regression equation. The partial regression coefficients were 0.285, 1.702, 0.161, 1.220, and -0.306, respectively, and their corresponding P values were 0.000, 0.001, 0.041, 0.013, and 0.007, respectively, which were all less than 0.05, indicating statistical significance.

Table 2. Variables in Logistic regression equation.

		B	S.E.	Wals	d.f.	Sig.	Exp (B)
Step 1[a]	CEA	-0.319	0.070	20.879	1	0.000	0.727
	Constant	4.668	0.725	41.428	1	0.000	106.478
Step 2[b]	CA199	0.272	0.071	14.889	1	0.000	1.313
	CEA	-0.306	0.113	7.327	1	0.007	0.736
	Constant	-2.015	1.310	2.366	1	0.124	0.133
Step 3[c]	CA199	0.375	0.159	5.585	1	0.018	1.456
	CA242	1.220	0.490	6.194	1	0.013	3.388
	CEA	-0.354	0.189	3.503	1	0.061	0.702
	Constant	-8.999	3.730	5.820	1	0.016	0.000
Step 4[d]	CA199	0.886	0.647	1.872	1	0.171	2.425
	CA242	3.863	2.781	1.930	1	0.165	47.611
	CA153	0.500	0.375	1.776	1	0.183	1.648
	CEA	-0.942	0.783	1.449	1	0.229	0.390
	Constant	-34.033	23.778	2.049	1	0.152	0.000
Step 5[e]	CA199	0.285	0.080	12.602	1	0.000	1.329
	CA242	1.702	0.524	10.001	1	0.001	5.256
	CA125	0.161	0.079	4.196	1	0.041	1.175
	CA153	1.220	0.490	6.194	1	0.013	3.388
	CEA	-0.306	0.113	7.327	1	0.007	0.736
	Constant	-15.857	5.787	12.537	1	0.000	0.000

[a]Variable put in step 1: CEA. [b]Variable put in step 2: CA199. [c]Variable put in step 3: CA242. [d]Variable put in step 4: CA153. [e]Variable put in step 5: CA125.

ROC curve analysis of 5 serum markers of single detection and combined detection

ROC curves are used to evaluate the quality of two-class classifiers, generally above the linear equation of y = x. As the ROC curve approaches the upper left, namely the area closer to 1, classification improved. Figure 1 shows the ROC curves of the logistic regression models of 5 serum markers detected solely and conjunctively; the sensitivity and specificity of each serum marker for single detection and combined detection and AUC are shown in Table 3. The AUC of combined detection was 0.988, which is close to 1, showing that combined detection has a better diagnosis effect than single detection, and its diagnostic sensitivity and specificity were 94.2 and 97.7%, respectively, which was higher than the values for single detection.
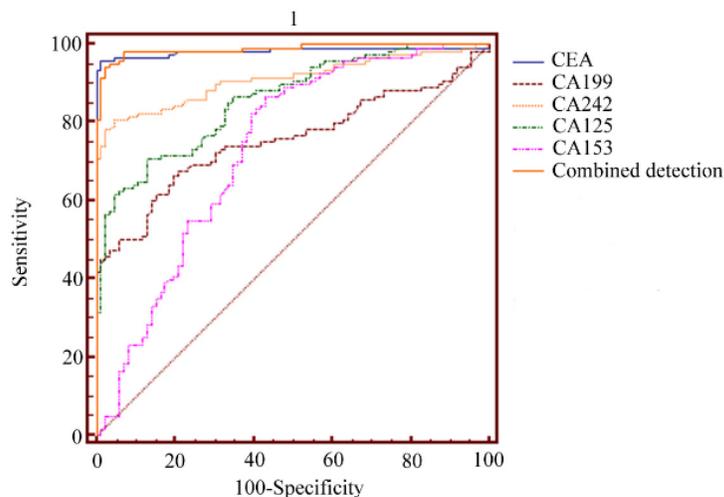

Figure 1. ROC curve of logistic regression model of 5 serum markers of single detection and combined detection.

Table 3. Comparison of the sensitivity and specificity and the AUC of single serum marker detection and combined detection.

Items	CEA	CA125	CA153	CA199	CA242	Combining detection
Sensitivity (%)	95.8	70.8	86.7	66.7	80.8	94.2
Specificity (%)	98.8	87.2	57.0	80.2	95.3	97.7
AUC	0.984	0.860	0.732	0.756	0.912	0.988

## Diagnosis model establishment of colorectal cancer based on SVM

By combining the tumor markers CEA, CA199, CA242, CA125 and CA153, an SVM diagnostic model of colorectal cancer was built. Data from 20 cases in the case group and 20 cases in the healthy control group were entered into the model; the results are shown in Figure 2. Circle represents output and asterisk represents actual simulation output of SVM. The accuracy of the model was 82.5% and the specificity and sensitivity were 80 and 85%, respectively (Figure 3).
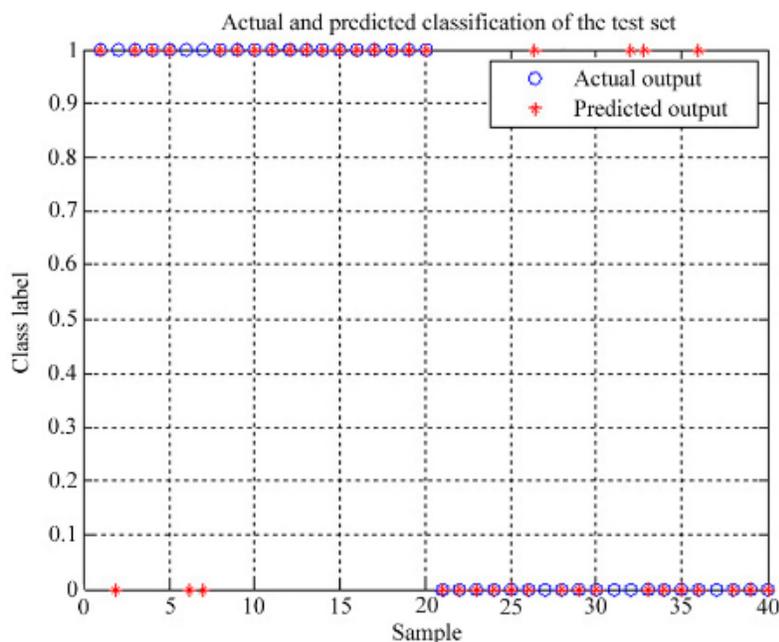
Figure 2. Accuracy testing results of SVM model established by combining 5 serum tumor markers.

Diagnosis model establishment of colorectal cancer based on BP neural value

Using 5 parameters, CEA, CA199, CA242, CA125, and CA153, a BP neural network was built. The classifier was trained until an optimal solution was obtained. The number of iterations was 37. For the test of well-trained classifiers, the training steps were as follows. First, the training samples were tested and the detection rate of colorectal cancer was 100% and the misdiagnosis rate was 0. The 40 testing set samples were further tested; 10 failed

prediction. The detection rate was 75% and the specificity and sensitivity were 70 and 80%, respectively.
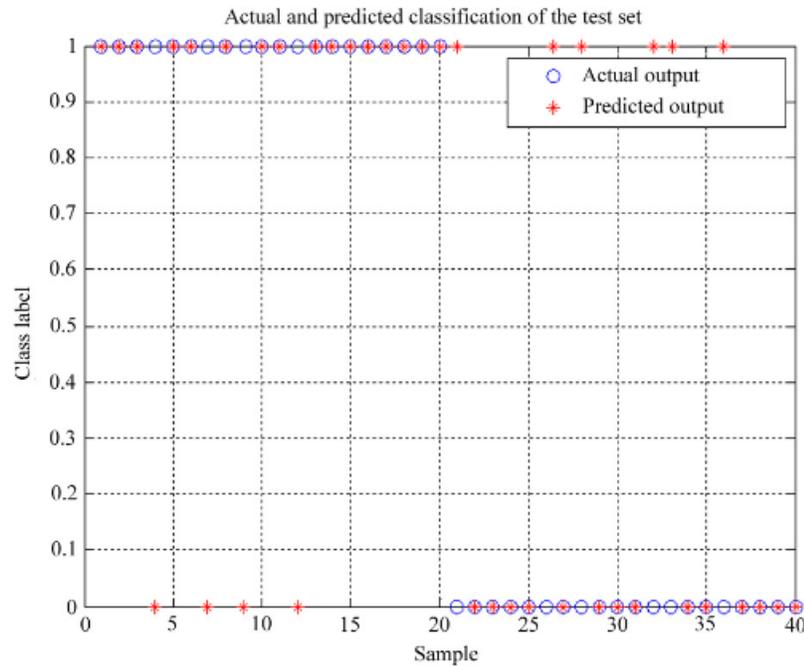
Figure 3. Accuracy of BP neural network diagnosis model built by combining 5 serum tumor makers.

DISCUSSION

Colorectal cancer is a common malignant tumor of the alimentary system and it is the fourth most prevalent in male malignant tumor and third most prevalent in females (Zhu and Wang, 1998; Li and Gu, 2004); therefore, early diagnosis of colorectal is very important (Pang et al., 2011). Herberman first posed the concept of tumor markers in 1978; with further development of medical research (Dong, 2009), markers have been used for the early diagnosis of tumors in the clinic (Kurman et al., 1977; Albert et al., 1988).

Although many tumors related markers are currently applied widely in the clinic, several tumor markers may be present in the same tumor. One tumor marker may be detected in several tumors, but single marker detection shows a low specificity and high false-positive and false-negative rates. Therefore, combining several tumor markers for combined detection improves the specificity and sensitivity of colorectal cancer diagnosis (Sheng et al., 2007). However, when all tumor markers appear positive, a tumor is thought to be present; in these cases, specificity is increased, but sensitivity is reduced and the misdiagnosis rate is increased. In contrast, if only one positive serum tumor marker is observed, the patients are considered to have a tumor, which enhances the sensitivity but also increases the false-positive rate and misdiagnosis rate.

CEA and CA199 are among the earliest tumor markers applied in the clinic and they are remarkably increased in the blood of malignant tumor patients. The high level expression of CEA in most colorectal cancer is closely related to pathological differentiation and metastasis; at later neoplasm stages, there is a greater possibility of distant metastasis or appearing positive during palindromia (Zhuang et al., 2010). CA19-9 has a low sensitivity but a high specificity, which should be noted in the detection of positive patients (Zhu et al., 2013). CA125 is an ovarian-related antigen and is clearly useful in the detection of digestive tract tumors (Guo et al., 2013; Huang et al., 2013). CA15-3 is the most significant specific marker in breast cancer; 30-50% of breast cancer patients show significantly elevated CA15-3 levels and changes in its content indicate treatment efficacy. Therefore, CA15-3 is the best marker for the diagnosis and observation of postoperative recurrence and efficacy in breast cancer patients. CA242 is a new antigen related to tumors and is observed at high levels when digestive tract tumors are present. This marker showed a higher sensitivity and specificity for pancreatic cancer and colorectal cancer, and thus may be applied for the preoperative prognosis and recurrence differentiation of colorectal cancer.

Computer-assisted diagnosis models such as the linear regression model, SVM, and BP neural network model are widely employed in clinical medicine (Staff, 2011; Xie et al., 2013; Surkis et al., 2013). Serum marker levels in 120 colorectal cancer patients and 86 benign controls were evaluated in this study; the results showed that in case groups, except the level of index AFP, 11 serum markers level were all significantly higher than in the control group. Through logistic regression analysis, the optimal tumor marker combination of CEA, CA199, CA242, CA125, and CA153 was identified. However, when the serum marker CEA was detected alone, its diagnostic sensitivity and specificity were 95.8 and 98.8%, indicating that the effects of single detection of serum CEA was better than combination detection of the 5 serum markers. This may have been a false-positive result and thus the diagnosis model requires further optimization. Based on the 5 makers, a SVM diagnosis model and a BP neural network diagnosis model of colorectal cancer were built with accuracies of 82.5 and 75%, sensitivities of 85 and 80%, and specificities of 80 and 70%, respectively. Based on the predictions of the diagnosis models and compared to the three diagnosis methods involving combined detection of the 5 serum markers, the best diagnosis effect was obtained by logistic regression analysis, followed by the SVM diagnosis model, and then the BP neural network diagnosis model. Those results are in accordance with those of previous reports (Chen et al., 2009).

In medical examinations, logistic regression analysis is a widely used statistical method that functions to identify risk factors and can be used for prediction and differentiation. The binary commonly used in logistic regression analysis is easily implemented (Li et al., 1997); therefore, this method is widely applied in clinical practice. Based on our results, compared with the SVM and the BP neural network diagnosis models, logistic multifactor regression analysis showed better results for the early diagnosis of colorectal cancer, which is related to the use of multiparameters within a certain range; a larger number of parameters leads to a better diagnosis effect in the certain range. Additionally, the diagnosis effect of SVM may be linked with a selected kernel function, and for BP neural network, its diagnosis effect is related to the numbers of neurons between the input layer, hidden layer, output layer, and the specific reason needs to be analyzed further.

This study mainly used intelligent algorithms based on multiple serum tumor markers to build three diagnosis models of colorectal cancer. As auxiliary diagnostic methods in the clinic, the models have some application value, but the accuracies and sensitivities of the SVM diagnosis model and BP neural network diagnosis model for disease diagnosis require further improvement.

REFERENCES

Albert MB, Steinberg WM and Henry JP (1988). Elevated serum levels of tumor marker CA19-9 in acute cholangitis. *Dig. Dis. Sci.* 33: 1223-1225. http://dx.doi.org/10.1007/BF01536670

Chen H, Jiang C, Yu C, Zhang S, et al. (2009). Protein chips and nanomaterials for application in tumor marker immunoassays. *Biosens. Bioelectron.* 24: 3399-3411. http://dx.doi.org/10.1016/j.bios.2009.03.020

Demetri GD, Benjamin RS, Blanke CD, Blay JY, et al.; NCCN Task Force (2007). NCCN Task Force report: management of patients with gastrointestinal stromal tumor (GIST)--update of the NCCN clinical practice guidelines. *J. Natl. Compr. Canc. Netw.* 5 (Suppl 2): S1-S29, quiz S30.

Deng XT, Shi GL, Wang RX, Zhao JX, et al. (2012). [Prognostic value of admission B-type natriuretic peptide on outcome for patients with congestive heart failure]. *Zhonghua. Xin Xue Guan Bing Za Zhi* 40: 462-466.

Domínguez-Ayala M, Díezvallejo J and Comasfuentes A (2012). Missed opportunities in early diagnosis of symptomatic colorectal cancer. *Rev. Espanola De Enfermedades Digestivas Organo Oficial De La Sociedad Espanola De Patologia Digestiva* 104: 343-349. http://dx.doi.org/10.4321/S1130-01082012000700002

Dong JL (2009). Combination detection of serum CEA, CA19-9 and CA724 in colorectal cancer patients and its clinical significance. *Chin. J. Lab. Diagn* 10: 1453-1454.

Guo QH, Ao HY, Cai QH, Zhao WT, et al. (2013). The diagnostic value of combination detection of CEA, CA199 and CA724 for colorectal cancer. *Intl. J. Lab. Med.* 7: 896-897.

Hall C (2004). Essential biochemistry and physiology of (NT-pro)BNP. *Eur. J. Heart Fail.* 6: 257-260. http://dx.doi.org/10.1016/j.ejheart.2003.12.015

Hayes DF, Bast RC, Desch CE, Fritsche H, Jr., et al. (1996). Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers. *J. Natl. Cancer Inst.* 88: 1456-1466. http://dx.doi.org/10.1093/jnci/88.20.1456

Huang F, Zhuang XQ and Wang BY (2013). Study on tumor factors level in peritoneal fluid drainage and its correlation with diseases of postoperative colorectal cancer patients. *Intl. J. Lab. Med.* 4: 469-470.

Kurman RJ, Scardino PT, McIntire KR, Waldmann TA, et al. (1977). Cellular localization of alpha-fetoprotein and human chorionic gonadotropin in germ cell tumors of the testis using and indirect immunoperoxidase technique. *Cancer* 40: 2136-2151. http://dx.doi.org/10.1002/1097-0142(197711)40:5<2136::AID-CNCR2820400524>3.0.CO;2-D

Li J, Yen C, Liaw D, Podsypanina K, et al. (1997). PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science* 275: 1943-1947. http://dx.doi.org/10.1126/science.275.5308.1943

Li M and Gu J (2004). Changing patterns of colorectal cancer over the recent two decades in China. *Chin. J. Gastrointest Surg* 3: 214-217.

Li P (2009). The Research of gastric cancer feature genes selection based on gene expression data. Doctoral thesis, Beijing University of Technology.

Pang WH, Liu HX, Ning XL and Chen YZ (2011). Clinical significance of combined tests with malignant tumor in digestive system. *Chin. J. Prim. Med. Pharm* 7: 913-915.

Parkin DM, Bray F, Ferlay J and Pisani P (2005). Global cancer statistics, 2002. *CA Cancer J. Clin.* 55: 74-108. http://dx.doi.org/10.3322/canjclin.55.2.74

Sheng L, An LF, Shi L, Wang JY, et al. (2007). The clinical significance of detecting the level of DcR3 and the expression of CEA mRAN in the peripheral blood of patients with colonic carcinoma. *J. Xi'an Jiaotong Univ. Med. Sci.* 2: 205-207.

Staff S; Science Staff (2011). Dealing with data. Challenges and opportunities. Introduction. *Science* 331: 692-693. http://dx.doi.org/10.1126/science.331.6018.692

Surkis A, McCrillis A, McGowan R, Williams J, et al. (2013). Informationist support for a study of the role of proteases and peptides in cancer pain. *J. eSci Libr.* 2: 9.

Vapnik DM and Zhang XG (2000). Essence of Statistical Theory. Tsinghua University Press, Beijing, 91-97.

Xie Y, Li JS, Yu XQ, Li SY, et al. (2013). An effective evaluation on quality of life of patients with stable chronic obstructive pulmonary disease by treatment of Bu-Fei Yi-Shen granule combined with acupoint sticking Therapy. *Chin. J. Integr. Med.* 4: 1-9.

Yu CH (1998). Evaluation of diagnostic butter. *Chin. J. Pracl Surg.* 11: 648-651.

Zhu HD, Peng QP and Zeng CL (2013). The detection value of serum CEA and CA19-9 in colorectal cancer patients. *Chin. J. Gen. Prac.* 5: 720-722.

Zhu ZG and Wang H (1998). Epidemiological trends of colorectal cancer. *Chin. J. Prac. Surg* 11: 648-651.

Zhuang SS, Chen JY, Fang YS and You YJ (2010). The diagnostic value of combination detection of CA199 and CEA for colorectal cancer. *J. Mod. Oncol.* 7: 1369-1371.