

# The Effect of Artificial Neural Network Model Combined with Six Tumor Markers in Auxiliary Diagnosis of Lung Cancer

Feifei Feng · Yiming Wu · Yongjun Wu · Guangjin Nie · Ran Ni

Received: 21 June 2011 / Accepted: 22 August 2011 / Published online: 1 September 2011  
© Springer Science+Business Media, LLC 2011

**Abstract** To evaluate the diagnosis potential of artificial neural network (ANN) model combined with six tumor markers in auxiliary diagnosis of lung cancer, to differentiate lung cancer from lung benign disease, normal control, and gastrointestinal cancers. Serum carcino-embryonic antigen (CEA), gastrin, neurone specific enolase (NSE), sialic acid (SA), Cu/Zn, Ca were measured through different experimental procedures in 117 lung cancer patients, 93 lung benign disease patients, 111 normal control, 47 gastric cancer patients, 50 patients with colon cancer and 50 esophagus cancer patients, 19 parameters of basic information were surveyed among lung cancer, lung benign disease and normal control, then developed and evaluated ANN models to distinguish lung cancer. Using the ANN model with the six serum tumor markers and 19 parameters to distinguish lung cancer from benign lung disease and healthy people, the sensitivity was 98.3%, the specificity was 99.5% and the accuracy was 96.9%. Another three ANN models with the six serum tumor markers were employed to differentiate lung cancer from three gastrointestinal cancers, the sensitivity, specificity and accuracy of distinguishing lung cancer from gastric cancer by the ANN model of lung cancer-gastric cancer were 100%, 83.3% and 93.5%, respectively; The sensitivity, specificity and accuracy of discriminating lung cancer by lung cancer-colon cancer ANN model were 90.0%, 90.0%,

and 90.0%; And which were 86.7%, 84.6%, and 86.0%, respectively, by lung cancer-esophagus cancer ANN model. ANN model built with the six serum tumor markers could distinguish lung cancer, not only from lung benign disease and normal people, but also from three common gastrointestinal cancers. And our evidence indicates the ANN model maybe is an excellent and intelligent system to discriminate lung cancer.

**Keywords** Artificial neural network · Lung cancer · Tumor marker · Gastrointestinal cancer

## Introduction

Lung cancer is the leading cause of cancer mortality worldwide [1], with 1.2 million deaths each year. And there are 1.3 million new cases being diagnosed every year in the world. In China, 10 million lung cancer patients will be diagnosed from 2025, which is predicted by epidemiological experts. However, the five survival rate for lung cancer patients is 15% despite therapy, mainly because lung cancer patients show symptoms only when lung cancer is at an advanced and incurable stage. But when lung cancer is diagnosed earlier at asymptomatic stage, the five survival rate is high to 67% [2]. Therefore the key to improving outcomes of lung cancer at present is to diagnose it in early stage.

There are many strategies to diagnose lung cancer, such as chest radiography, sputum cytology, light-induced fluorescence endoscopy (LIFE) and serum biomarkers, and so on. But none of them can diagnose this devastating disease totally correctly. To date, more and more studies have been identifying the effects of serum tumor markers on pathology types, staging, monitoring and prognostication of lung

F. Feng (✉) · Y. Wu · Y. Wu · G. Nie  
College of Public Health, Zhengzhou University,  
Zhengzhou, People's Republic of China  
e-mail: feifeifeng9@gmail.com

R. Ni  
Department of Respiratory, The First Affiliated Hospital,  
Zhengzhou University,  
Zhengzhou, People's Republic of China

cancer [3], especially early detection [4]. In a recent study, serum squamous cell carcinoma antigen (SCC), carcinoembryonic antigen (CEA), cytokeratin 19 fragment antigen 21-1 (Cyfra21-1) and neuron specific enolase (NSE) were tested in 805 patients with lung cancer and benign pulmonary diseases and analyzed by receiver operating characteristic (ROC) curves, 37.3% of early-staged lung cancer could be diagnosed by the combination assays of the four tumor markers [5]. Furthermore it can be accepted because of its non-invasive, low cost and high sensitivity.

At present some common serum tumor markers are used for diagnosis of lung cancer. CEA is a valuable marker for diagnosing lung adenocarcinoma [6], NSE is one of the most important tumor markers that is generally acknowledged to discriminate small cell lung cancer (SCLC), not only in diagnosis, but also in staging and monitoring [7]. Gastrin is used to diagnose lung cancer, gastric cancer [8] and colon cancer. A high density of sialic acid (SA) is often expressed in tumor cell, which could help these cells get into blood system [9]. The analysis of the trace element contents (Cr, Fe, Mn, Al, Cd, Cu, Zn, Ni, Se, Pb, Ca, Mg, Sr, P) in the samples of lung cancer patients is very helpful to the early diagnosis and treatment effectiveness evaluation to the patients [10].

Artificial neural network (ANN) is a powerful computational tool imitating human neuronal systems, and it has shown the ability of modeling complex systems with high predictive accuracies on blind data [11]. ANN is also appreciated by its excellent fault tolerance and fast parallel processing. Now it is used in medicine widely [12]. And there have been a large number of reports on the use of ANN for diagnosis [13] and prognosis [14]. An ANN model deriving by four tumor markers, CA 125II, CA 72-4, CA 15-3 and macrophage colony stimulating factor (M-CSF), increased sensitivity about 25% over that of using CA 125II alone for detecting early stage ovarian cancer [15]. In previous studies, ANN models with tumor markers and bronchofibrosopic data [16], with tumor markers and auto-fluorescence spectrum [17] have been built to distinguish lung cancer, and both of them could diagnose lung cancer effectively.

In this paper, we developed ANN models by the six serum tumor markers of CEA, gastrin, NSE, SA, Cu/Zn, Ca combined with basic information to discriminate lung cancer, not only from lung benign disease, but also from gastrointestinal tumors.

## Materials and methods

### Research samples

The serum specimens of lung disease patients were obtained from the First Affiliated Hospital of Zhengzhou

University. The collection included 117 samples from lung cancer patients (31 lung squamous cell carcinoma, 45 lung adenocarcinoma and 41 small cell carcinoma), 44 stage I, 53 stage II, 20 stage III, average age ( $61.06 \pm 10.48$ ) year. 93 samples from lung benign disease patients, average age ( $53.06 \pm 13.85$ ) year, 111 samples from healthy people confirmed by department of physical examination, average age ( $62.63 \pm 9.72$ ) year. All the people above were surveyed 19 parameters of basic information on risk factors and symptoms of lung diseases by questionnaire, including smoking, dust and chemicals exposure, kitchen environment and related symptoms, and so on.

The serum samples of 47 gastric cancer, ( $58.57 \pm 12.36$ ) year, 50 colon cancer, ( $56.64 \pm 14.07$ ) year and 50 esophagus cancer patients, ( $60.85 \pm 11.73$ ) year, were also collected from the First Affiliated Hospital of Zhengzhou University.

The diagnosing results of all patients were carefully confirmed by experts of histopathology and/or cytopathology, and the collection of all specimens was known and the permission was got from patients or their relatives.

### Samples preparation

Five mL venous blood was collected from every fasting subject in the morning,  $37^{\circ}\text{C}$  water bath for 30 min, centrifuge at 3,000 rpm, 10 min, then the serum was separated, aliquot and stored at  $-80^{\circ}\text{C}$ , and thawed immediately before detecting.

### Serum tumor markers measurement

CEA, NSE and Gastrin were determined using radioimmunity kits (Beijing north of biological technology), according to the manufacturer's instructions.

SA and Ca (Beijing leadman biochemical technology company) were detected by spectrophotometry. Standard preparation of SA was from Sigma Company.

Cu and Zn were detected by atomic absorption spectrophotometry.

The normal critical values of CEA, Cu/Zn, sialic acid, Ca, gastrin and NSE are 20  $\mu\text{g/L}$ , 1.0, 1,040  $\text{mg/L}$ , 100  $\text{mg/L}$ , 95  $\text{ng/L}$ , 18  $\mu\text{g/L}$ , respectively.

### Statistical analysis

The lung cancer, lung benign, and control were classified by each of 19 variables, respectively. Each contingency table was tested by Chi-Square test. The groups' means of six markers were compared using ANOVA, every two groups were analyzed by post-hoc.  $P < 0.05$  was deemed as significance level. All data were analyzed by SPSS12.0 statistical package.

## Artificial neural network model establishment

### Normalization input data

The input data that did not accord with the request were normalized using linear function to range from 0 to 1. Below was the formula:

$$y = (x - \text{MinValue}) / (\text{MaxValue} - \text{MinValue}).$$

(x was the original value of input, y was transformed by above formula via x, and MaxValue, MinValue were the maximum and minimum among all data of one index, respectively.)

### Groups of training set and testing set

Based on the experiences before, the normalized data of each group were separated randomly into training set and testing set according to the ratio of 3:1. The training set was used to let ANN model learn, while testing set was to distinguish.

### Four ANN models

Four back-propagation (BP) ANN models were built by artificial neural network software, which was provided by the Center of Computer Analysis of China Pharmaceutical University (Nanjing, China). There was three-layer network structure composed of input, hidden and output layers. The number of input nodes was equal with the number of parameters into the ANN model, and the output layer had only one node, which represents the diagnostic result of the subjects.

#### *Lung cancer-lung benign disease-healthy control ANN model*

Twenty-five input units contained six serum tumor markers and 19 parameters of basic information, hidden units were 15 and output unit was 1. The parameters of ANN program were described before [18]. We have verified the parameters are optimal. Transfer function “tansig”, train function “trainlm”, output function “purelin”, velocity for training 0.6, momentum factor 0.95, expected error 1e-5. And the models were initialized randomly. The output of the ANN represented the likelihood of different group, 1.0, 0.6 and 0.2 were set as corresponding output value of lung cancer, lung benign diseases and normal control, respectively. Therefore, when the output value was [0.8, 1.0], the sample would be determined as lung cancer, (0.4, 0.8) would be considered as lung benign diseases, and [0, 0.4] would be normal control.

#### *Lung cancer-gastric cancer (colon cancer or esophagus cancer) ANN models*

There were six input nodes (six serum tumor markers), 15 hidden nodes and one output unit. And the output value was set as 0 and 1, which represented lung cancer, and other cancer (gastric, colon, or esophagus cancer, respectively). When it was less than 0.5, the sample would be judged as lung cancer; otherwise, it would be other cancer. The other parameters were the same as lung cancer-lung benign disease-normal ANN model.

## Results

### Lung cancer-lung benign disease-normal control ANN model

#### *The results of six serum tumor markers among lung cancer, lung benign disease and normal groups*

There were all significantly statistical differences on the levels of six serum tumor markers among the three groups by ANOVA analysis ( $P < 0.05$ ) and between every two groups ( $P < 0.05$ ). The levels of CEA, Gastrin, NSE, SA and Cu/Zn in lung cancer group were higher than those of in lung benign disease and normal control groups, the level of serum Ca in lung cancer group was lower, which were described in Table 1.

#### *The results of basic information among the three groups*

There were all significantly statistical differences on classification of the three groups with 19 variables by Chi-Square Test ( $P < 0.05$ ) (Table 2). The 19 parameters include lung cancer risk factors, such as smoke, drink, environmental pollutant exposure, fume in kitchen exposure, and clinical symptoms, which are relevant with diagnosis of lung cancer.

#### *The training effect of the ANN model to distinguish lung cancer*

Two hundred forty-two samples were used to train by the ANN model of lung cancer-lung benign disease-normal control based on the parameters mentioned above, including 93 lung cancer patients (33 stage I, 44 stage II, 16 stage III.), 66 lung benign disease patients and 83 healthy people. Figure 1 showed that when the ANN model was on the 318 epochs, the expected error was as little as 1e-5, and it was the end of training. At this time, we found that all the samples of training set were determined correctly, the accuracy was 100%. The effect of fitting was perfect.

**Table 1** The results of six tumor markers among lung cancer, lung benign disease and normal control groups

Group	n	CEA ( $\mu\text{g/L}$ )	Gastrin ( $\text{ng/L}$ )	NSE ( $\mu\text{g/L}$ )	SA ( $\text{mg/L}$ )	Cu/Zn	Ca ( $\text{mg/L}$ )
Lung cancer	117	39.15 $\pm$ 22.83	154.56 $\pm$ 70.87	31.63 $\pm$ 17.86	1085.84 $\pm$ 140.09	3.12 $\pm$ 1.67	79.79 $\pm$ 16.81
Lung benign disease	93	13.99 $\pm$ 8.54*	77.59 $\pm$ 25.82*	12.23 $\pm$ 6.18*	911.11 $\pm$ 157.02*	2.10 $\pm$ 0.91*	93.74 $\pm$ 20.13*
Normal control	111	8.23 $\pm$ 2.24*#	62.05 $\pm$ 17.93*#	6.40 $\pm$ 3.15*#	720.64 $\pm$ 141.92*#	1.05 $\pm$ 0.51*#	115.48 $\pm$ 24.17*#

\* means vs lung cancer,  $P < 0.05$ ; # means vs lung benign disease,  $P < 0.05$

#### *The testing effect of the ANN model to distinguish lung cancer*

The testing set contained 24 lung cancer patients (11 stage I, 9 stage II, 4 stage III.), 27 lung benign disease patients and 28 healthy people, and tested by the ANN model which had been trained well. Table 3 showed that in the testing set, two patients with lung cancer were incorrectly distinguished as one lung benign disease patient and one normal, five patients with lung benign disease were wrongly judged as one lung cancer patient and four normal, and three normal were determined fallaciously as the lung benign disease patients. The predicted sensitivity, specificity and accuracy of the testing set were 92.0% (23/25), 98.1% (53/54), and 87.3% (69/79), respectively. While in the total samples, the predicted results were that the sensitivity was 98.3% (115/117), the specificity was 99.5% (203/204) and the accuracy was 96.9% (311/321). And all patients in stage I were diagnosed correctly.

#### Lung cancer-gastric cancer (colon cancer or esophagus cancer) ANN models

##### *The results of six serum tumor markers among lung cancer, gastric cancer, colon cancer and esophagus cancer groups*

Table 4 showed that the levels of serum CEA, Gastrin, NSE, SA and Cu/Zn were significantly different among the four cancer groups ( $P < 0.05$ ), except Ca ( $P > 0.05$ ). “\*” showed that the levels of serum tumor markers in gastrointestinal cancer groups were significantly different comparing with those of lung cancer.

##### *The training effect of the three cancer-cancer ANN models to distinguish lung cancer*

Eighty-seven lung cancer patients and 35 gastric cancer patients were used to train the lung cancer-gastric cancer ANN model, 87 lung cancer patients and 40 colon cancer patients were used to train the lung cancer-colon cancer ANN model, 87 lung cancer patients and 37 esophagus cancer patients were used to train the lung cancer-esophagus cancer ANN model. And the effect of training of the three models were perfect, the accuracies were all 100%.

#### *The testing effect of the three cancer-cancer ANN models to distinguish lung cancer*

In the testing set of lung cancer-gastric cancer ANN model, Fig. 2a illustrated that two patients (1 stage I, 1 stage II) with lung cancer were incorrectly distinguished as gastric cancer patients (two red point on the top of the dividing line on the value of 0.5), and also two gastric cancer patients were wrongly determined, so the predicted sensitivity, specificity and accuracy of the testing set were 93.3% (28/30), 83.3% (10/12) and 90.5% (38/42), respectively.

Figure 2b showed that in the testing set of lung cancer-colon cancer ANN model, three patients (1 stage I, 2 stage II) with lung cancer were incorrectly distinguished as colon cancer patients, and only one colon cancer patient was wrongly determined, so the predicted sensitivity, specificity and accuracy of the testing set were 90.0% (27/30), 90.0% (9/10), and 90.0% (9/10), respectively.

Figure 2c demonstrated in the testing set of lung cancer-esophagus cancer ANN model, four patients (2 stage I, 1 stage II, 1 stage III) with lung cancer were incorrectly distinguished as esophagus cancer patients, and also two esophagus cancer patients were wrongly determined, the predicted sensitivity, specificity and accuracy of the testing set were 86.7%(26/30), 84.6%(11/13) and 86.0%(37/43), respectively.

## Discussion

In this study, we investigated 19 parameters of basic information related with lung cancer, and the results showed that besides smoking, chemical pollutant exposure, these regular hazardous factors closely related with lung cancer, we found that kitchen environment including cooking fuel, kitchen ventilation and cooking methods, maybe had been other risk factors for women lung cancer patients. In China some poor country area, the main fuel for cooking was still coal, and there was less ventilation facilities in their kitchens, furthermore, in the middle area of China, the primary cooking way was frying. So the oil fume of cooking and smoke of fuel were stayed in kitchen and can't discharge. And there were numerous hazardous

**Table 2** Basic information among lung cancer, lung benign disease and normal control groups

Basic information	Group			P
	Lung cancer (n=117)	Lung benign disease (n=93)	Normal control (n=111)	
Sex				
Male	96(82%)	60(64%)	65(59%)	0.00
Female	21(18%)	33(36%)	46(41%)	
Active smoking				
No-smoking	26(22%)	51(55%)	85(77%)	0.00
0–300(year sticks)	26(22%)	10(11%)	6(5%)	
≥300(year sticks)	65(56%)	32(34%)	20(18%)	
Passive smoking				
No	71(61%)	61(66%)	93(84%)	0.00
Yes	46(39%)	32(34%)	18(16%)	
Drink				
No	70(60%)	67(72%)	103(93%)	0.00
Yes	47(40%)	26(28%)	8(7%)	
Dust exposure				
No	81(69%)	69(74%)	111(100%)	0.00
Yes	36(31%)	24(26%)	0(0%)	
Chemistry exposure				
No	96(82%)	81(87%)	111(100%)	0.00
Yes	21(18%)	12(13%)	0(0%)	
Cooking fuel				
Coal/wood	95(81%)	77(83%)	2(2%)	0.00
Natural gas	22(19%)	16(17%)	109(98%)	
Ventilation in kitchen				
No	98(84%)	63(68%)	2(2%)	0.00
Yes	19(16%)	30(32%)	109(98%)	
Cooking methods				
Fry	84(72%)	73(79%)	99(89%)	0.005
Boil	33(28%)	20(21%)	12(11%)	
Cough				
No	91(78%)	23(25%)	111(100%)	0.00
Yes	26(22%)	70(75%)	0(0%)	
Expectoration				
No	102(87%)	12(13%)	111(100%)	0.00
Yes	15(13%)	81(87%)	0(0%)	
Bloody sputum				
No	81(69%)	63(68%)	111(100%)	0.00
Yes	36(31%)	30(32%)	0(0%)	
Hemoptysis				
No	88(75%)	72(77%)	111(100%)	0.00
Yes	29(25%)	21(23%)	0(0%)	
Chest distress/chest pain				
No	33(28%)	18(19%)	111(100%)	0.00
Yes	84(72%)	75(81%)	0(0%)	
Palpitation				
No	74(63%)	67(72%)	111(100%)	0.00
Yes	43(37%)	26(28%)	0(0%)	
Fatigue				
No	82(70%)	65(70%)	111(100%)	0.00
Yes	35(30%)	28(30%)	0(0%)	
Family history of cancer				
No	61(52%)	69(74%)	111(100%)	0.00
Yes	56(48%)	24(26%)	0(0%)	

**Table 2** (continued)

Basic information	Group			P
	Lung cancer (n=117)	Lung benign disease (n=93)	Normal control (n=111)	
Fever/sweating				
No	74(63%)	61(66%)	111(100%)	0.00
Yes	43(37%)	32(34%)	0(0%)	
Sore throat/hoarseness				
No	99(85%)	77(83%)	111(100%)	0.00
Yes	18(15%)	16(17%)	0(0%)	

compounds in the smoke, such as polycyclic aromatic hydrocarbons and aromatic amines [19], women often cooked in kitchen and inhaled the smoke for tens of years, eventually there were increasing occurrence rate of lung adenocarcinoma in women. Li M [20] performed a case–control study of 350 pairs nonsmoking women matched by age, and found that exposure to cooking oil fume was associated with increased risk of lung cancer in Chinese women nonsmokers, the odd ratio (OR) was 2.51, 95% confidence interval (CI) [1.80–3.51],  $P < 0.001$ .

Artificial neural network (ANN) analysis as a statistical modeling tool has demonstrated the ability to assimilate information from multiple sources and detect subtle and complex patterns. Now many researches applied tumor markers combined with ANN model to diagnose cancers [21, 22]. In this study we established an ANN model by six tumor markers and basic information to distinguish lung cancer, the effect is exciting. In the total samples, the predicted results were that the sensitivity was 98.3% (115/117), the specificity was 99.5% (203/204) and the accuracy was 96.9% (311/321). Although there were so many factors for lung cancer diagnosis and there were complex relationship among them, ANN could learn fuzzy evaluation which can't be described by mathematical methods, and deal with some complex, uncertain and nonlinear problem by imitating human intelligent behavior with excellent fault tolerance and fast parallel processing [23], especially when there were large number of samples, multi-category, multi-variable, ANN model could show more excellent capability to solve the nonlinear and unknown-data-distribution problems with better self-learning and fault tolerance, so ANN could diagnose lung cancer by these complicated factors more correctly.

Because it is difficult to obtain numerous samples of lung cancer patients, and the cost is quite expensive if many tumor markers are detected, it is a good way to add more parameters about basic information to increase input variables, and then better effect will be got with less cost.

In early stage of tumor, there are not specific symptoms to identify which kind of tumor is, or when more than two

tumors are found in the same patient, we should distinguish which is the primary tumor and which is the metastasis, because there are different treatments for different primary cancers. Pathologists are often facing the problem of tumor classification, so various selection strategies have been generally used and compared, including the ANN models. Bloom GC [24] used a protein signature to construct an ANN-based classifier for identify the tumor type from 6 similarly appearing adenocarcinomas (ovary, colon, kidney, breast, lung and stomach.), finally found that a maximum predictive accuracy of 87% and an average predictive accuracy of 82%. Another ANN model with gene microarray got accurate tumor classification and helped to extract the latent marker genes for tumor diagnosis and treatment [25].

Gastrointestinal cancers are the common cancers in clinical, encompassing esophagus, gastric and colon cancers, and the pathology of esophagus cancer often is squamous cell carcinoma, while the pathology of gastric and colon cancer are mainly adenocarcinoma, which should be distinguished with the same pathologies of lung cancer. Many serum tumor markers could be detected from various tumors, but there may be differences on quantity. Table 4 showed that except Ca, the levels of the other five tumor markers were significantly different in other three cancer groups compared with lung cancer.

The ANN model with six tumor markers can distinguish lung cancer from lung benign disease and normal control, while whether it could identify lung cancer from other cancers or not, so we established three ANN models to distinguish lung cancer from esophagus, gastric and colon cancer, respectively. We found that the sensitivity and specificity of identifying lung cancer from gastric cancer were 93.3% and 83.3%, 90.0% and 90.0% from colon cancer, and 86.7% and 84.6% from esophagus cancer. These findings demonstrated the ANN model constructed by the six serum tumor markers had a highly accurate to identify lung cancer from gastrointestinal tumors.

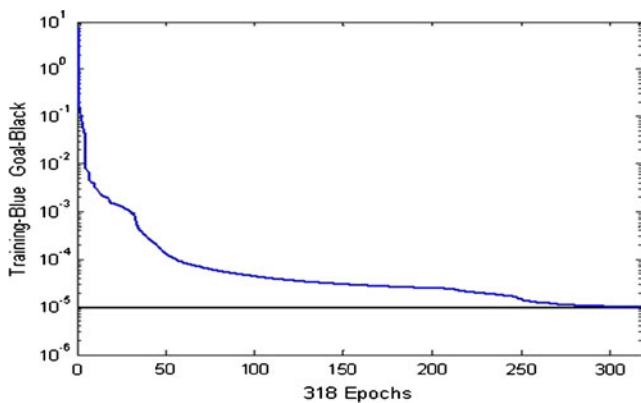


Fig. 1 Training with trainlm

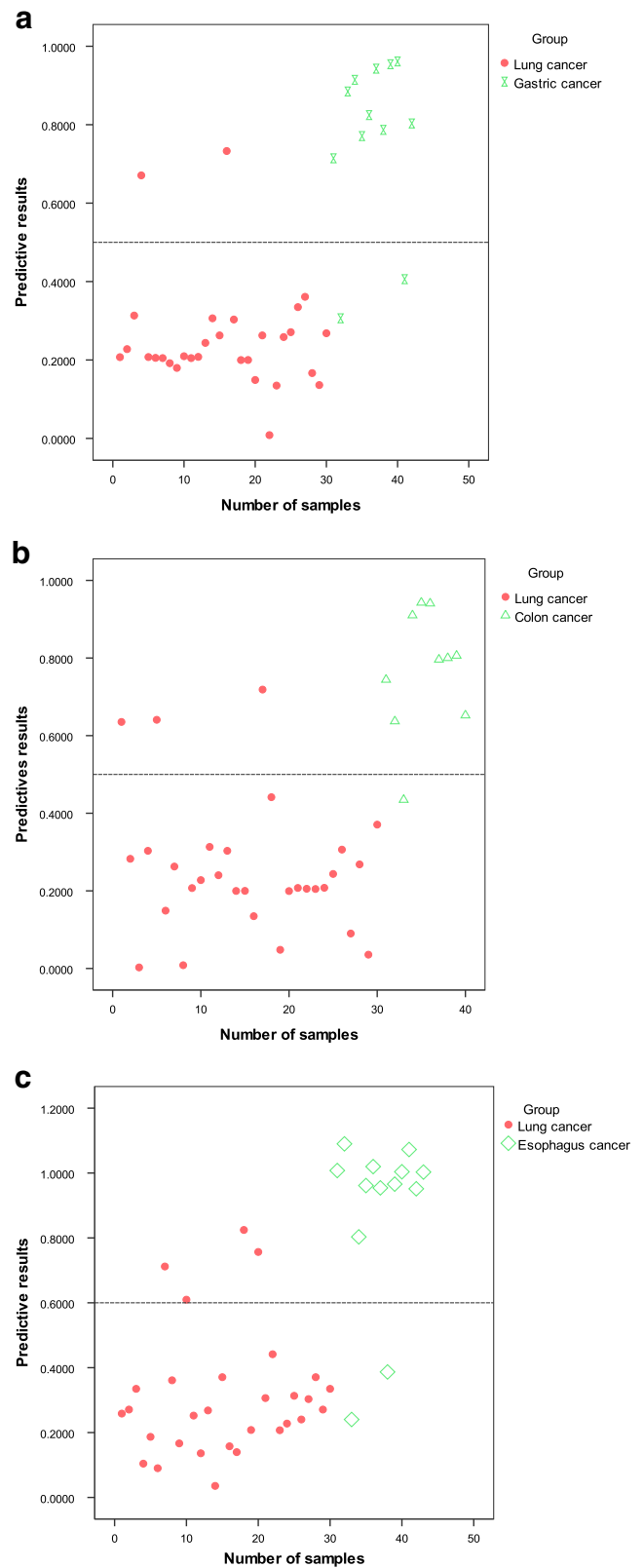
Table 3 The predictive results of ANN model to distinguish lung cancer in lung cancer-lung benign disease-normal control groups

Group by ANN model	Group by pathology in testing set			Total	Group by pathology in total samples			Total
	Lung cancer	Lung benign disease	Normal control		Lung cancer	Lung benign disease	Normal control	
Lung cancer	23	1	0	24	115	1	0	116
Lung benign disease	1	23	3	27	1	88	3	92
Normal control	1	4	23	28	1	4	108	113
Total	25	28	26	79	117	93	111	321
Sensitivity	92.0% (23/25)			98.3% (115/117)				
Specificity	98.1% (53/54)			99.5% (203/204)				
Accuracy	87.3% (69/79)			96.9% (311/321)				

**Table 4** The results of 6 tumor markers among four cancer groups

Group	n	CEA (μg/L)	Gastrin (ng/L)	NSE (μg/L)	SA (mg/L)	Cu/Zn	Ca (mg/L)
Lung cancer	117	39.15±22.83	154.56±70.87	31.63±17.86	1085.84±140.09	3.12±1.67	79.79±16.81
Gastric cancer	47	19.07±5.64*	169.51±34.43	23.37±7.28*	1064.91±134.53	1.98±0.64*	73.29±12.00
Colon cancer	50	29.98±11.02*	37.20±17.31*	6.92±2.92*	1018.50±202.13*	2.13±0.74*	78.27±13.19
Esophagus cancer	50	9.907±2.21*	149.98±29.48	20.09±5.01*	931.59±137.41*	3.368±0.376	79.393±14.49
P		0.000	0.000	0.000	0.000	0.000	0.06

\* means vs lung cancer,  $P < 0.05$



**Fig. 2** The predictive results of testing set by lung cancer-other cancers ANN models (a) lung cancer-gastric cancer (b) lung cancer-colon cancer (c) lung cancer-esophagus cancer

## Conclusion

The ANN model built with the six serum tumor markers could distinguish lung cancer, not only from lung benign disease and normal control, but also from other common gastrointestinal cancers, and these are strong evidences to prove the ANN model is an excellent tool of auxiliary diagnosing lung cancer. For the further plan, we will continue to collect samples of other kinds of cancers and prove the model again, then use it to screen high risk people of lung cancer for primary prevention, based on its good accuracy and low cost, and do cohort study to prove whether the person, who was identified as high risk person by the ANN model, will be diagnosed lung cancer for following years. And perfect the ANN model to be the truly intelligent tool of distinguishing lung cancer in early stage, for secondary prevention.

**Acknowledgments** This study was financially supported by grants from the Nature Science Foundation of China (NO. 30972457 and 81001240).

## References

- Jemal, A., Siegel, R., Ward, E., Hao, Y., Xu, J., and Thun, M. J., Cancer statistics. *CA Cancer J. Clin.* 59:225–249, 2009.
- Ghosal, R., Kloer, P., and Lewis, K. E., A review of novel biological tools used in screening for the early detection of lung cancer. *Postgrad. Med. J.* 85:358–363, 2009.
- Greenberg, A. K., and Lee, M. S., Biomarkers for lung cancer: clinical uses. *Curr. Opin. Pulm. Med.* 134:249–255, 2007.
- Ma, P. C., Blaszkowsky, L., Bharti, A., Ladanyi, A., Kraeft, S. K., Bruno, A., Skarin, A. T., Chen, L. B., and Salgia, R., Circulating tumor cells and serum tumor biomarkers in small cell lung cancer. *Anticancer. Res.* 23:49–62, 2003.
- Chu, X. Y., Hou, X. B., Song, W. A., Xue, Z. Q., Wang, B., and Zhang, L. B., Diagnostic values of SCC, CEA, Cyfra21-1 and NSE for lung cancer in patients with suspicious pulmonary masses: a single center analysis. *Canc. Biol. Ther.* 11, 2011.
- Niho, S., and Shinkai, T., Tumor markers in lung cancer. *Gan To Kagaku Ryoho/Canc. Chemother.* 28:2089–2093, 2001.
- Giovanella, L., Ceriani, L., Bandera, M., and Garancini, S., Immunoradiometric assay of chromogranin A in the diagnosis of small cell lung cancer: comparative evaluation with neuron-specific enolase. *Int. J. Biol. Markers.* 16:50–55, 2001.
- Chao, C., and Hellmich, M. R., Gastrin, inflammation, and carcinogenesis. *Curr. Opin. Endocrinol. Diabetes. Obes.* 17:33–39, 2010.
- de Castro, J., Rodríguez, M. C., Martínez-Zorzano, V. S., Hernández-Hernández, A., Llanillo, M., and Sánchez-Yagüe, J., Erythrocyte and platelet phospholipid fatty acids as markers of advanced non-small cell lung cancer: comparison with serum levels of sialic acid, TPS and Cyfra 21-1. *Canc. Investig.* 26:407–418, 2008.
- Dou, L., and Xu, W., Determination of 14 elements in the body fluid and hair of lung cancer patients by microwave digestion with ICP-MS. *Chin. J. Lung Canc.* 13:817–820, 2010.
- Caron, J., Mangé, A., Guillot, B., and Solassol, J., Highly sensitive detection of melanoma based on serum proteomic profiling. *J. Canc. Res. Clin. Oncol.* 135:1257–1264, 2009.
- Cartwright, H. M., Artificial neural networks in biology and chemistry: the evolution of a new analytical tool. *Meth. Mol. Biol.* 458:1–13, 2008.
- Mat-Isa, N. A., Mashor, M. Y., and Othman, N. H., An automated cervical pre-cancerous diagnostic system. *Artif. Intell. Med.* 42: 1–11, 2008.
- Bassi, P., Sacco, E., De Marco, V., Aragona, M., and Volpe, A., Prognostic accuracy of an artificial neural network in patients undergoing radical cystectomy for bladder cancer: a comparison with logistic regression analysis. *Brit. J. Urol. Int.* 99:1007–1012, 2007.
- Zhang, Z., Yu, Y., Xu, F., Berchuck, A., van Haaften-Day, C., Havrilesky, L. J., de Bruijn, H. W., van der Zee, A. G., Woolas, R. P., Jacobs, I. J., Skates, S., Chan, D. W., and Bast, R. C., Jr., Combining multiple serum tumor markers improves detection of stage I epithelial ovarian cancer. *Gynecol. Oncol.* 107:526–531, 2007.
- Feng, F. F., Wu, Y. J., Zhang, C., and Wu, Y. M., Application of artificial neural network model established by tumor markers and bronchofibroscopic data in auxiliary diagnosis of lung cancer. *Nat. Comput.* 2:118–125, 2009.
- Wu, Y. J., Hao, Y. H., Wu, W. C., and Wu, Y. M., Value of auto-fluorescence spectrum combined with tumor markers in diagnosis of lung cancer. *Spectrosc. Spect. Anal.* 29:2787–2791, 2009.
- Wu, Y. J., Wu, Y. M., Wang, J., Yan, Z., Qu, L. B., Xiang, B. R., and Zhang, Y. G., An optimal marker group-coupled artificial neural network for diagnosis of lung cancer. *Expert Syst. Appl.* 38:11329–11334, 2011.
- Wu, M. T., Lee, L. H., Ho, C. K., Wu, S. C., Lin, L. Y., Cheng, B. H., Liu, C. L., Yang, C. Y., Tsai, H. T., and Wu, T. N., Environmental exposure to cooking oil fumes and cervical intraepithelial neoplasm. *Environ. Res.* 4:25–32, 2004.
- Li, M., Yin, Z., Guan, P., Li, X., Cui, Z., Zhang, J., Bai, W., He, Q., and Zhou, B., XRCC1 polymorphisms, cooking oil fume and lung cancer in Chinese women nonsmokers. *Lung Canc.* 62:145–151, 2008.
- Yu, J. K., Yang, M. Q., Jiang, T. J., and Zheng, S., The optimal combination of serum tumor markers with bioinformatics in diagnosis of colorectal carcinoma. *J. Zhejiang. Univ. Med. Sci.* 33:407–410, 2004.
- Donacn, M., Yu, Y., Artioli, G., Banna, G., Feng, W., Bast, R. C., Jr., Zhang, Z., and Nicoletto, M. O., Combined use of biomarkers for detection of ovarian cancer in high-risk women. *Tumor Biol.* 31:209–215, 2010.
- Patel, J. L., and Goyal, R. K., Applications of artificial neural networks in medical science. *Curr. Clin. Pharmacol.* 2:217–226, 2007.
- Bloom, G. C., Eschrich, S., Zhou, J. X., Coppola, D., and Yeatman, T. J., Elucidation of a protein signature discriminating six common types of adenocarcinoma. *Int. J. Canc.* 120:769–775, 2007.
- Liu, B., Cui, Q., Jiang, T., and Ma, S., A combinational feature selection and ensemble neural network method for classification of gene expression data. *BMC Bioinforma.* 5:136, 2004.