Int J Clin Exp Med 2016;9(7):12612-12620
www.ijcem.com /ISSN:1940-5901/IJCEM0022612

Original Article

Forecast model establishment and screening of correlative factors of malignant and benign solitary pulmonary nodules

Chuanliang Chen[1,2], Junwu Zhao[1], Jianhua Zhang[3]

[1]The Key Laboratory of Biomedical Information Engineering of Ministry of Education, Institute of Biomedical Analytical Technology and Instrumentation, School of Life Science and Technology, Xi'an Jiaotong University, No. 28 West Xianing Rd., Xi'an, Shaanxi, China; [2]Henan Provincial People's Hospital, No. 7 Weiwu Rd., Zhengzhou, Henan, China; [3]Zhengzhou University Institute of Medical Engineering Technology and Data Mining, No. 100 Science Avenue Rd., Gaoxin Dist., Zhengzhou, Henan, China

Received December 23, 2015; Accepted May 19, 2016; Epub July 15, 2016; Published July 30, 2016

Abstract: Objectives: To screen out associated radiological indexes and serological markers based on which assistant diagnosis model was established and then to evaluated their efficacy. Methods: With 260 cases of solitary pulmonary nodules (SPN) confirmed by pathology (145 malignant cases and 115 benign cases) as study subjects, correlative factors of solitary pulmonary nodules were screened out from 9 computerized tomography (CT) radiological indexes and 8 tumor makers by logistic regression analysis. On that basis, support vector machine (SVM) model and back propagation (BP) model were established and be compared with the model including all variables. Results: SVM assistant model and BP assistant model were successfully established. When all variables were included, the accuracy, specificity and sensitivity of SVM model was 75.0%, 85.0% and 65.0%, respectively while that of BP model was 77.5%, 75.0% and 80.0%, respectively; when all screening variables were included, the accuracy, specificity and sensitivity of SVM was 80.0%, 95.0% and 65.0%, respectively while that of BP model was 67.5%, 75.0% and 60.0%, respectively. Conclusions: Compared to BP model, SVM model possessed better efficacy and through filtering feature subsets, the accuracy and specificity of SVM model could be effectively improved, which could be seen as a good assistant diagnosis approach for examination of malignant and benign SPN and offered directions for future SPN forecast model in clinical employment and development.

Keywords: Solitary pulmonary nodules, CT radiological indexes, tumor makers, logistic regression analysis, SVM assistant model, BP assistant model

Introduction

SPN is a single lesion in lung completely surrounded by lung parenchyma (functional tissue) with a diameter less than 3 cm and without associated pneumonia, atelectasis (lung collapse), pleural effusion, hilus of lung or lymphadenopathies (swollen lymph nodes) [1]. Every year, about 150,000 SPN patients were screened out throughout the world, among which malignant ones account for 5%-70%. It's reported that non small cell lung cancer patients in early stage could have a five-year survival rate of 80% after operation while that of patients getting delayed treatment for not being diagnosed the nature of SPN correctly was only 10% [2, 3]. So far lung cancer is still one of the leading causes of cancer death. Therefore, it's key to correctly diagnose the nature of SPN in decreasing the death rate of lung cancer and it's also a difficulty for medical community.

Iconography is indispensable in diagnosing malignant and benign SPN. At present, the acknowledged, the most valuable and frequently used examination method is CT test that possesses 12 indexes of CT image analysis which are the size of nodules, density, part, boundary, ground glass opacity, leaflet, vessel convergence, pleura drags, burrs, calcification, physalides and cavity thus physicians can diagnose the malignant and benign SPN through features of nodule [4]. Tumor marker is substances such as enzyme or protein that is composed of and released by tumor cells or con-

Table 1. The extraction of imaging feature

Imaging indexe	Characteristic	Assignment
Size of SPN	The length, width, height of SPN was averagedon mediastinal window	-
Location of tubercle	Superior lobe	10
	Central lobe	5
	Inferior lobe	0
Boundary characteristics	Compared pulmonary nodule edge with pulmonary parenchyma from clear to vague	0~10
Speculation	None	0
	Long spicules from long to short	1~5
	Short spicules from long to short	6~10
Lobulation	None	0
	The number of outline evaginations of every two neighboring notches which were 1, 2, 3, and 4	2, 4, 6, 8
	More than 4	9 or 10
Pleural indentation	None	0
	Pleural indentation	0
Strengthening	Homogeneous enhancement	10
	Heterogeneous enhancement	5
	None	0
Vessel convergence sign	None	0
	Vessel convergence sign	10
Cavitation sign	None	0
	Cavitation sign	10

ducted by organism's response to tumor and these substances which can reacts the existence, activity and growth of tumor, only exist in embryonic tissues or tumor tissues with a higher content than normal structures [5]. Tumor marker has the advantages over sensitivity and specificity which plays a certain indicating role in the diagnosis, typing, development and prognosis of tumor. Single tumor marker owns poor specificity and sensitivity, thus several tumor markers were used combined which clinically showed good results [6]. Physicians' subjective factors have great effects on SPN diagnosis so that the nature of SPN always was determined falsely. Therefore, it is helpful for correct diagnosis of SPN to build an effective aided diagnosis model and to decrease the interference of subjective factors. At present, the frequently-used aided diagnosis models are SVM and BP neural networks.

Imaging data and serological data of 260 patients in this study were collected including 9 basic characteristic data and 8 different concentrates of serum marker. Logistic regression was adopted to screen out imaging factors and serological factors that most likely affect the determination of the malignant and benign SPN, based on which SVM aided diagnosis model and BP neural networks were established. Moreover, this study decreased the subjunctive factors through scientific methods and made it more possible to improve the detection rate of malignant SPN and reduce the death rate of lung cancer.

Materials and methods

Research subjects

Continuous data of patients at Henan Province People's Hospital between year 2012 and 2014 and according to study of patients' clinical records, 260 SPN patients were picked out including 145 malignant ones whose pathological diagnosis results covering 88 cases of adenocarcinoma, 32 cases of squamous carcinoma, 8 cases of adenosquamous carcinoma, 2 cases of nonsamll-cell lung cancer, 4 cases of neuroendoccrine carcinoma, 2 cases of mucoepidermoid carcinoma, 4 cases of mucoepidermoid, 1 case of carcinoma mucocellulare, 1 case of metastatic renal cell carcinoma, 1 case of metastasis breast cancer, 1 case of carcinoma sareomatodes and 1 case of unclassified lung cancer and 115 benign SPN with pathological diagnosis results of 56 cases of inflammation, 13 cases of tuberculosis, 14 cases of inflammatory pseudotumor, 11 cases of mycotic infection, 5 cases of hamartoma, 4 cases of angeioma, 3 cases of pulmonary abscess, 2 cases of epithelial tumor, 1 case of bronchiectasia, 1 case of glioma peripheral of chest wall, 1 case of lymphadenoma, 1 case of secondary

osteogenic sarcoma of lung, 1 case of coccus infection, malignant dyskaryosis and 1 case of fibrocartilage. All patients have signed the informed consent.

Abstraction of imaging and serological characteristics

In this study, 9 imaging indexes (size, location, boundary characteristics, spiculation, lobulation, pleural indentation, strengthening, vessel convergence sign, and cavitation sign) were performed assignment, which is shown in **Table 1**.

What's more, 8 serological markers of blood samples of patients were detected, including carcino embryonic antigen (CEA), neuron specific enolase (NSE), Cytokerantin-19-fragment (CYFRA21-1), carbohydrate antigen 125 (CA125), cancer antigen (CA199), carbohydrate antigen 724 (CA724), squamous cell carcinoma antigen (SCC-Ag), and breast cancer antigen (CA153). Detections were performed in accordance with instructions in the kit, among which, CEA and CYFRA21-1 in serum were detected by ELISA, NSE was determined by radio immunoassay, and other markers were tested by Roche E601 automatic immune analyzer.

According to their pathological types, patients were divided into malignant SPN group and benign SPN group and then the two groups were assigned as 1 and 0, respectively.

Research methods

Logistic regression: This study employed SPSS 17.0 software to screen out all imaging indexes and serological markers with forward Binary Logistic regression analysis aiming at filtering out the factors closely related to the malignant and benign SPN. $P < 0.05$ indicates that the difference is statistically significant.

SVM

SVM was first put out by Vapnik in 1995 [7]. It was based on the VC dimension theory and structure risk minimization principle. In accordance with limited sample data, to reach the optimal compromise between model's recidivity and its learning ability, it hoped to obtain the best generalization capacity. It had many distinctive advantages over solving small samples, nonlinear and high-dimensional pattern recognition and it could be generalized to the learning of function fitter and other machines. This study employed Matlab to conduct programming to establish SVM model, and the programming included data extraction, parameter selection, training set and testing set confirmation, normalization and network training and predicting of SVM; meanwhile optimizing the penalty parameter C and nuclear parameters γ to guarantee the optimum of models.

BP algorithm

BP network was a one-way conduction and multilayer feed-forward network divided into input layer, hidden layer (it could either be one layer or multilayer) and output layer. BP algorithm was a classic error correction method whose learning process consisted of forward propagation and anti-propagation and whose fundamental was putting into signal Xi through intermediate hidden node acting on output node and through nonlinear variation to get output signal Qi. If there's deviation between the network output value and the expecting output value, BP would carry out anti-propagation that deviation of output data would backtrack. Through repeated learning and training of revising connection weight and threshold, deviation decrease along gradient direction so that we can got the minimized deviation between output value and expecting output value. After the network training, trained network could be used to analyze and precast new samples [8]. The establishment of BP model used Matlab too and the programming included data extraction, parameter selection, training set and testing set confirmation, normalization and network training and predicting of BP; in the process parameters of driving function, transition function and training function were constantly optimized expecting to obtain the best results.

Evaluation of models

All cases were sorted into malignant cases and benign cases. Choosing 1-125 from the former and 146-240 from the latter as training set while picking 126-145 from the former and 241-260 from the latter as test set to build SVM model and BP neural network model. And every model would be carry out in two conditions that fitting into all of indexes or indexes screened out by Logistic regression and then to evaluate the two models from the following aspects.

Table 2. Imaging variables in Logistic regression equation

		B	S.E.	Wald	df	Sig.	Exp (B)
Step 1[a]	Spicule sign	-.340	.044	59.706	1	.000	.712
	Constant	1.055	.215	23.969	1	.000	2.872
Step 2[b]	Spicule sign	-.307	.047	43.113	1	.000	.735
	Lobulated sign	-.291	.061	22.489	1	.000	.748
	Constant	2.054	.326	39.655	1	.000	7.802

[a]Variable(s) entered on step 1: spicule sign; [b]Variable(s) entered on step 2: lobulated sign.

Table 3. Forecast classification table

Observed			Predicted		
			Type of SPN		Percentage Correct
			Malignant	Benign	
Step 1	Type of SPN	Malignant	112	33	77.2
		Benign	35	80	69.6
	Overall Percentage				73.8
Step 2	Type of SPN	Malignant	118	27	81.4
		Benign	33	82	71.3
	Overall Percentage				76.9

Table 4. Serological variables in Logistic regression equation

		B	S.E.	Wald	df	Sig.	Exp (B)
Step 1[a]	CYFRA21-1	-.686	.104	43.886	1	.000	.503
	Constant	2.029	.326	38.699	1	.000	7.603
Step 2[b]	CYFRA21-1	-.695	.109	40.886	1	.000	.499
	CA153	-.063	.016	15.468	1	.000	.939
	Constant	2.941	.426	47.686	1	.000	18.942
Step 3[c]	CYFRA21-1	-.642	.108	34.997	1	.000	.526
	SCC-Ag	-.676	.201	11.293	1	.001	.509
	CA153	-.058	.016	13.209	1	.000	.944
	Constant	3.697	.501	54.464	1	.000	40.326

[a]Variable(s) entered on step 1: CYFRA211. [b]Variable(s) entered on step 2: CA153. [c]Variable(s) entered on step 3: SCC.

Accuracy = (TP+TN)/(TP+TN+FN) × 100%

Sensitivity = TP/(TP+FN) × 100%

Specificity = TN/(TN+FP) × 100%

In those equations, TP is true positive number; TN is true negative number; FP is false positive number and FN is false negative number.

Results

Non-conditional Logistic regression analysis

Screening out results of imaging variables: Nine CT imaging characteristics were carried out value assignment which was regarded as covariate and the diagnosed results of SPN as dependent variable. Then stepwise logistic regression analysis was employed to filter out significant indexes in determining malignant and benign SPN and the results were as shown in **Tables 2** and **3**. After two-time variable screening, spicule sign and lobulated sign were selected into Logistic regression model. Both *P* value were less than 0.01 which had statistical significance. The forecast accuracy of prediction model was up to 76.9% of the second step from 73.8% of the first step.

Screening results of serological variables: Taking serum CEA, NSE, CYFRA21-1, CA125, CA199, CA724, SCC-Ag, and CA153 as covariates and the diagnosed results of SPN as dependent variable, stepwise logistic regression analysis was employed to filter out serological influencing factors for differentiation of benign and malignant SPN. After three-time variable screening, 5 indexes were deleted only the rest three arguments (CYFR21-1, SCC-Ag and CA153) were selected into Logistic regression equation and the results was as shown in **Table 4**. And the *P* value of the three arguments were 0.000, 0.001 and 0.000, respectively, which all had statistical significance and their correspondent partial regression coefficient all were negative showing that the concentrate of patients' serum CYFRA21-1, SCC-Ag and CA153 were related to the malignant and benign SPN, namely, the higher concentrate was, the lower rate benign SPN was. Classification table (**Table 5**) indicated the prediction of SPN classification at each step; the forecast accuracy of the first step was 75.8%, the second step was 79.6% and that of the third step increased to 80.4%.

SVM model

After optimization of penalty parameter C and nuclear parameter γ, finally the optimal parameter (C, γ) turned out to be (12, 1). Establishing SVM model in two conditions, bringing into all indexes and fitting into indexes (spicule, lobulated, CYFR21-1, SCC-Ag and CA153) screened

Table 5. Forecast classification table

	Observed		Predicted		
			Type of SPN		Percentage Correct
			Malignant	Benign	
Step 1	Type of SPN	Malignant	102	43	70.3
		Benign	20	95	82.6
	Overall Percentage				75.8
Step 2	Type of SPN	Malignant	115	30	79.3
		Benign	23	92	80.0
	Overall Percentage				79.6
Step 3	Type of SPN	Malignant	115	30	79.3
		Benign	21	94	81.7
	Overall Percentage				80.4

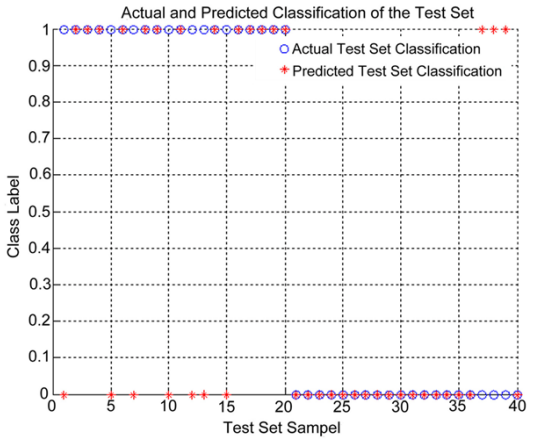

Figure 1. SVM simulation results when all variables were brought into.

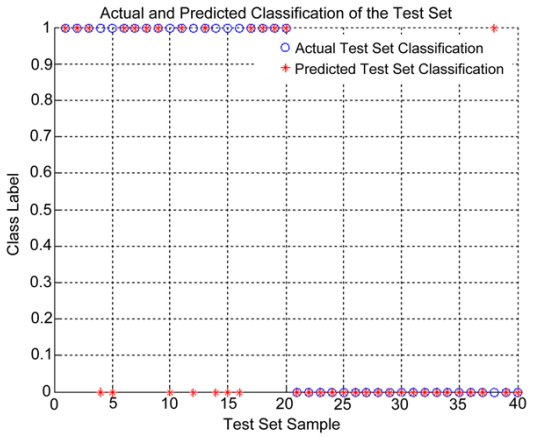

Figure 2. SVM simulation results when bring in screening variables by Logistic regression.

out by Logistic regression and 40 test samples were put into SVM to carry out simulation. The results were as shown in **Figures 1** and **2**, respectively. In the figures, ○represents target output and *suggests the actual stimulation output. When different variables were brought into, the accuracy, specificity and sensitivity of the model were shown in **Table 6**. Seen from the table, efficiency of SVM model was superior to that of bringing all variables.

BP model

After the optimization of all parameters, this study finally determined the optimal parameter combination, namely, driving function was tansig, transition function was purelin, training function was trainlm, hidden layer number was 1, hidden layer node was 10, output layer node was 1, threshold of training time was 10,000 and accuracy was 0.001. The training process of BP model was as following. When all variables were fitted into, we can see from the error fitting curve in **Figure 3** that, as the number of training times ran up to 54, the accuracy of the model achieved 0.001 so that the training stopped. And the training results were shown in **Figure 4** in which vertical ordinate represented the actual output of BP network and horizontal ordinate was the target output. When the value of actual output equated that of the target output, the sample point would on diagonal line. Seen from the figure, most sample points located near the diagonal line showing that the training results were good.

Forty test samples were brought into the model to carry out simulation and the results were shown in **Figure 5** in which ○represented target output and Δindicated actual simulation output. When variables filtered out by Logistic regression were executed BP model training, the accuracy could achieve requirement and the training results were good. The simulation results were shown in **Figure 6**. The two BP models were evaluated from accuracy, specificity and sensitivity and the results were as shown in **Table 7**. We can see that when screening out variables were brought into, the diagnosing efficiency was not as good as that of when all variables were fitted into.

Discussion

Lung cancer is a common malignant tumor all over the world and has been one of the main

Table 6. Evaluation of SVM models bring into different variables

Group	Accuracy	Specificity	Sensitivity
When bringing into all variables	75.0%	85.0%	65.0%
When bringing into screening out variables	80.0%	95.0%	65.0%

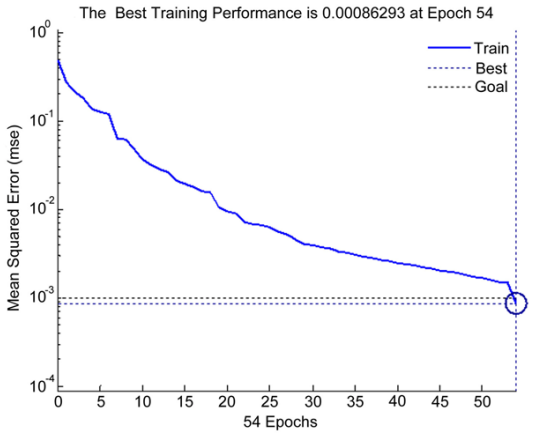

Figure 3. Error fitting curve.

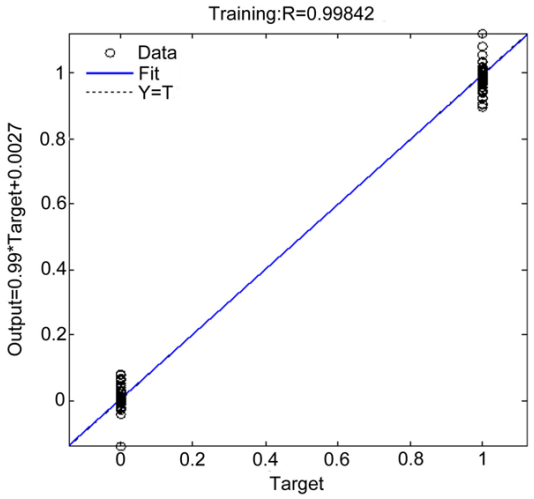

Figure 4. Training results.

reasons of cancer death. At the early stage of lung cancer, most patients have no symptom or minor ones, which are easily been ignored. When found, it often had metastasized, thus leading a decrease in patients' survival rate. The early expression of lung cancer usually is tubercle, and then the tubercle will develop into different lumps; therefore, it's the key to correctly determine the malignant and benign SPN in preventing lung cancer and reducing its death rate [9].

Traditional CT imaging examination and serum markers check can no longer meet people's needs of improving SPN detection rate so that diagnosis-aided model emerged as the development of science and technology. In 1997, Swensen, et al [10] established the first model for determining the malignant and benign SPN using Logistic regression. And he analyzed retrospectively 629 SPN patients, regarding their age, gender, smoking history, history of malignant tumor of chest, history of asbestos exposure and chronic interstitial or obstructive pulmonary disease, and imaging data including the size, location, boundary characters (such as spicule sign and lobulated sign) of tubercle and other features (such as inanition sign) as study subjects. And then through screening, 3 clinical features (age, smoking history and more than 5 years cancer history) and 3 imaging characters (size of tubercle, spicule sign and located on superior lobe) were decided as independent risk factors for malignant tubercle to establish model for the diagnosis of malignant and benign SPN which had a good accuracy and its area under ROC reached 0.8328±0.0226. Later, Gould, et al [11] built VA model after analyzed 375 SPN patients retrospectively. His screening results of independent risk factors were ages, smoking history, size of tubercle and cease-smoking time, based on which a model was built showing a good accuracy (area under ROC was 0.79). Similarly, computer aided diagnosis system (CAD) was widely employed in the diagnosis of malignant and benign SPN among which the greatest use was SVM model and BP model. The area under ROC of SVM model established, with plenty of collected data, by Sun, et al [12] indicating that it has high accuracy in the diagnosis of malignant and benign SPN and has clinical application value. Kuruvilla [13] built BP neural network lung cancer diagnosis model based on CT imaging features and the accuracy rate of the model is 91.1% suggesting that both SVM model and BP model have good developing foreground.

Currently, one of key problems that hasn't been completely solved among malignant and benign SPN diagnosis models is that how to select appropriate feature subsets. Excessive features will increase the computation complexity and the possibility of over-fitting while too fewer features will bring out an unreliable categorizer. Thus, this study first screen out patients' 17

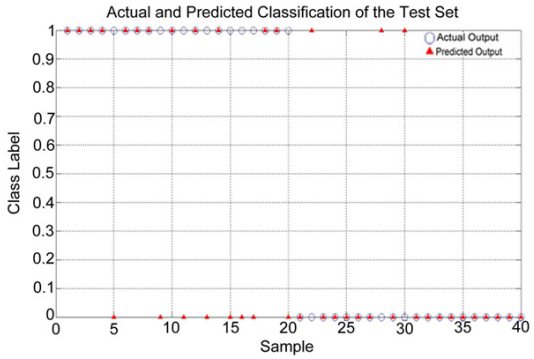

Figure 5. Simulation results of BP model when bring into all variables.

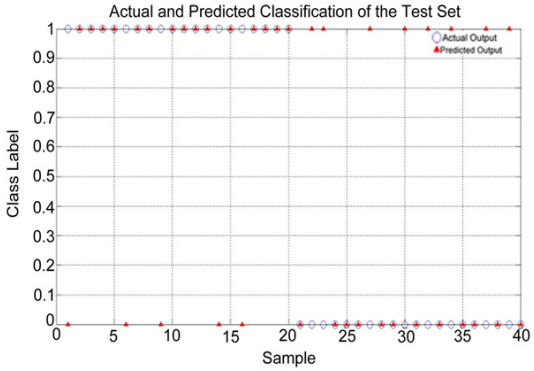

Figure 6. Simulation results of BP model when bring into the variables screened out by Logistic regression.

indexes aiming at simplifying and increasing the accuracy of the model. The screening results demonstrated that imaging indexes (spicular sign and lobulated sign) and serum indexes (CYFRA21-1, SCC-Ag, CA153) have relation to the malignant and benign SPN. The research showed that the appearance of spicular signs was believed that it was resulted from malignant tumor proliferating along lung interstitium and it had a positive predictive value of 90% for malignant tubercle [14]. The expression of lobulated sign is that several notches around the border of focus and there're outline evagination between every two neighboring notches. A number studies have indicated that spicular sign and lobulated sign are the significant sign of malignant SPN [15, 16]. CYFRA21-1 is an acidic protein mainly existed in endochylema of epithelial origins tumor cells caused by lung cancer or esophageal cancer, ect. When tumor cells dissolve or necrose, CYFRA21-1 would be released into blood thus it was one of the tumor markers of lung cancer. Cui, et al [17] carried out Meta on documents about CYFRA21-1 in diagnosing non small lung cancer. In his analysis the sensitivity and specificity of CYFRA21-1 in diagnosing non small lung cancer was 0.72 and 0.94, respectively, and the area under the ROC was 0.95, which indicated CYFRA21-1 was a very helpful tumor marker for CYFRA21-1 in diagnosing non small lung cancer. SCC-AG was a tissue antigen purified from cervical squamous cell carcinoma by Kato and Torigoe in 1977 and Liu, et al [18] confirmed that SCC-AG had high value in diagnosing lung squamous cell carcinoma. CA153 was mainly in adenocarcinoma, mostly in lung adenocarcinoma and breast cancer. In the study of Wang, et al [19], AUC value of CA153 in diagnosing lung adenocarcinoma was 0.838 and the specificity and sensitivity was 85.2%, 73.2%, respectively; that in diagnosing lung squamous cell carcinoma was 0.716, 91.2% and 57.6%, respectively; that in diagnosing non small lung cancer was 0.812, 94.1% and 61.5%, respectively, which demonstrated that CA153 was more suitable in diagnosis of lung adenocarcinoma.

In this study, two conditions, all variables and screening variables were brought in to establish SVM model and BP model, and the two models efficiency was evaluated. The results suggested that when bring into all variables, the accuracy, specificity and sensitivity of SVM model was 75.0%, 85.0% and 65.0%, respectively while that of BP model was 77.5%, 75.0% and 80.0%, respectively; when filtering into screening variables, they were 80.0%, 95.0% and 65.0% while that of BP model were 67.5%, 75.0% and 60.0%, respectively. We can see from the results that after screening the variables, though the efficiency of SVM model could be improved effectively, that of BP model didn't. Generally speaking, SVM model is superior to BP model in accuracy and specificity but not in sensitivity. At present, there are few documents study the influences of feature subsets on the efficiency of prediction models, except this study. And results of the study show that for the established SVM model, screening out appropriate feature subsets could not only simplify model but also improve the diagnostic efficiency. Seventeen imaging indexes and serological markers were fitted in the research of Zhao, et al [20] to build SVM model whose

Table 7. Evaluation of BP model when filtering different variables

Group	Accuracy	Specificity	Sensitivity
When bring into all variables	77.5%	75.0%	80.0%
When filtering into screening out variables	67.5%	75.0%	60.0%

accuracy, sensitivity and specificity was 80.0%, 85.7% and 66.7%, respectively, which is equal to the efficiency of the model in this study. But the difference is that only 5 indexes were brought in the study which simplified the model. The efficiency of their BP model is similarly to ours.

To sum up, this study lays foundation for clinically build efficient SPN assistant diagnosis model and provides direction for the future application and development of SPN prediction model, which is significant in decreasing patients' medical expenses and diagnosis and treatment cost of hospital.

Acknowledgements

I thank all authors who have contributed to this paper for advice and comments, especially my tutor, Junwu Zhao. Moreover, the author would like to thank People's Hospital of Zhengzhou for providing research materials and experimental base. Project supported by the Natural Science Foundation of Henan Province, China (Grant No. 152300410151).

Disclosure of conflict of interest

None.

Address correspondence to: Junwu Zhao, The Key Laboratory of Biomedical Information Engineering of Ministry of Education, Institute of Biomedical Analytical Technology and Instrumentation, School of Life Science and Technology, Xi'an Jiaotong University, No. 28 West Xianing Rd., Xi'an, Shaanxi, China. Tel: 0371-65580012; Fax: 860371565-77092; E-mail: junwuzhaoxj@126.com

References

[1] Ost D, Fein AM and Feinsilver SH. The solitary pulmonary nodule. N Engl J Med 2003; 25: 2535-2542.

[2] Yankelevitz DF and Henschke CI. Small solitary pulmonary nodules. Radio Clin North Am 2000; 38: 471-478.

[3] Pastorino U. Current status of lung cancer screening. Thorac Surg Clin 2013; 23: 129-140.

[4] Truong MT, Sabloff BS and Ko JP. Multidetector CT of solitary pulmonary nodules. Radiol Clin North Am 2010; 48: 141-155.

[5] Molina R, Augé JM, Bosch X, Escudero JM, Viñolas N, Marrades R, Ramírez J, Carcereny E and Filella X. Usefulness of serum tumor markers, including progastrin-releasing peptide, in patients with lung cancer: correlation with histology. Tumour Biol 2009; 3: 121-129.

[6] Ma L, Xie XW, Wang HY, Ma LY and Wen ZG. Clinical evaluation of tumor markers for diagnosis in patients with non-small cell lung cancer in China. Asian Pac J Cancer Prev 2015; 12: 4891-4894.

[7] Vapnik N. The nature of statistical learning theory. Berlin: Springer-Verlag; 1995. pp. 1-5.

[8] Chen YC, Ke WC and Chiu HW. Risk classification of cancer survival using ANN with gene expression data from multiple laboratories. Compu Biol Med 2014; 48: 1-7.

[9] Yi CA, Lee KS, Kim EA, Han J, Kim H, Kwon OJ, Jeong YJ and Kim S. Solitary pulmonary nodules: dynamic enhanced multi-detector row CT study and comparison with vascular endothelial growth factor and microvessel density. Radiology 2014; 233: 191-199.

[10] Swensen SJ, Silverstein MD, Ilstrup DM, Schleck CD and Edell ES. The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules. Arch Intern Med 1997; 8: 849-855.

[11] Gould MK, Ananth L and Barnett PG. A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules. Chest 2007; 2: 383-388.

[12] Sun T, Wang J, Li X, Lv P, Liu F, Luo Y, Gao Q, Zhu H and Guo X. Comparative evaluation of support vector machines for computer aided diagnosis of lung cancer in CT based on a multi-dimensional data set. Comput Methods Programs Biomed 2013; 111: 519-524.

[13] Kuruvilla J and Gunavathi K. Lung cancer classification using neural networks for CT images. Comput Methods Programs Biomed 2014; 113: 202-209.

[14] Gurney JW, Lyddon DM and Mckay JA. Determining the likehood of malignancy in solitary pulmonary nodules with Bayesian analysis. Part II. Application. Radiology 1993; 186: 415-422.

[15] McWilliams A, Tammemaqi MC, Mayo JR, Roberts H, Liu G, Soghrati K, Yasufuku K, Martel S, Laberge F, Gingras M, Atkar-Khattra S, Berg CD, Evans K, Finley R, Yee J, English J, Nasute P, Goffin J, Puksa S, Stewart L, Tsai S, Johnston MR, Manos D, Nicholas G, Goss GD, Seely JM, Amjadi K, Tremblay A, Burrowes P,

MacEachern P, Bhatia R, Tsao MS and Lam S. Probability of cancer in pulmonary nodules detect on first screening CT. New Engl J Med 2013; 10: 910-919.

[16] Xiao F, Liu D, Guo Y, Shi B, Song Z, Tian Y and Liang C. Novel and convenient method to evaluate the character of solitary pulmonary nodule-comparison of three mathematical prediction models and further stratification of risk factors. PLos One 2013; 10: e78271.

[17] Cui C, Sun X, Zhang J, Han D and Gu J. The value of serum Cyfra21-1 as a biomarker in the diagnosis of patients with non-small cell lung cancer: a meta-analysis. J Cancer Res Ther 2014; 10: C131-134.

[18] Liu L, Liu B, Zhu LL, Zhang W and Li Y. Clinical significance of CYFRA21-1, Scc-Ag and telomerase activity in serum and pleural effusion of patients with squamous-cell lung cancer. Bioanalysis 2012; 4: 2367-2374.

[19] Wang XF, Wu YH, Wang MS and Wang YS. CEA, AFP, CA125, CA153 and CA199 in malignant pleural effusions predict the cause. Asian Pac J Cancer Prev 2014; 15: 363-368.

[20] Zhao Z, Chen J, Yin X, Song H, Wang X and Wang J. Establishing assistant diagnosis models of solitary pulmonary nodules based on intelligent algorithms. Cell Physiol Biochem 2015; 35: 2463-2471.