Breast Tumor Classification Diagnosis Based on LS-SVM

Chao Liu[1,2, 3], Bo Zhou[1], Qingzhu Li[4], Yu Chen[1],Guowei Qin[1] and Guangkuo Hu[2, 3]
[1]*Kunming University of Science and Technology;*
[2]*The First People's Hospital of Yunnan Province;*
[3]*Affiliated Hospital of Kunming University of Science and Technology;*
[4]*Yan' An Hospital of Kunming*

Keywords: Breast tumor; LS-SVM; Neural network; Support vector machine; clinical diagnostic value.

Abstract: To accurately predict breast cancer, breast cancer prediction method based on least squares support vector (LS-SVM) proposed. Patients with breast cancer through the data on the basis of 469 cases, including 400 cases of data relevance vector machine training, and the remaining 69 cases data sample tests, and finally through with neural networks, support vector machines comparison, breast cancer diagnosis model based on LS-SVM prediction accuracy is higher than the neural network and support vector machine. Has good diagnostic value of breast cancer diagnosis based on LS-SVM model, which provides a new method for breast cancer diagnosis.

1 INTRODUCTION

Cancer(LU Xin-guo,2010)is called malignant tumor in field of medicine. It is a kind of cells cancerous result at partial tissue caused by action of carious factors such as the chemical, physical, microbe and its metabolic product. It is often shown as: partial tissue abnormal cellular proliferation and form of partial lump. Cancer is a kind of disease caused by multiple causes, stages and mutations of normal cells in the body. Cancer is characterized by: ability to infinite proliferation and loss of contact inhibition phenomena at the same time, the cancer cells between viscosity reduced, easy to be lectin agglutination, this will consume a lot of nutrients of cancer patients(LIN Xiao-gang,2009);Cancer cells release a variety of toxins, causing the human body to produce a series of symptoms; cancer cells can also be spread through blood, lymphatic or direct methods such as transferred to the whole body, leading to a large amount of nutrients in the body is consumed, and viscera function is damaged. Benign tumor is easy to clean, and generally, no transfer, no recurrence, only extrusion of organs, tissues and blocking effect. Malignant tumor leads to tumor metastasis and malignant tumor destroys the normal structure and function of tissues and organs, cause necrosis bleeding merge infection, the patient eventually died due to organ failure.

Medical studies have found that the nuclear micrograph of breast tumor lesion tissue is different from that of normal tissue, but it is difficult to distinguish it by general image processing method. Therefore, it is particularly important to use scientific methods to diagnose benign or malignant breast tumors according to the nuclear microscopic images of breast tumor foci(XU Wei-yun,2003).

Based on this, this paper proposes to use LS-SVM to conduct classified diagnosis of breast tumors, and to prove its effectiveness and accuracy by comparison.

2 PRINCIPLE OF LEAST SQUARES SUPPORT VECTOR MACHINE ALGORITHM

In 1995, Corinna Cortes and Vapnik initially put forward support vector machine (SVM) which show unique advantages in solving nonlinear, high dimension, small sample, and can be applied to the function fitting other machine learning problems (WANG Yu-hong,2004).It can be expressed as follows:

For training data set, (x_i, y_i), linear function is used to fit sample, ϕ represents a mapping from the input space x to the feature space, w is the weight and b is the deviation.

$$y(x) = w^T \phi(x) + b \tag{1}$$

Equation (1) is optimized as follows:

$$\min K(w, e) = \frac{1}{2} ||w||^2 + \frac{1}{2} \gamma \sum_{i=1}^{n} e_i^2 \tag{2}$$

$$s.t. y_i - w^t x_i - b = e_i \tag{3}$$

Equation(2) γ is called regularization parameter. n is the number of training samples; e is the training error, and finally the LSSVM model can be obtained as follows:

$$\hat{y}_i = \sum_{i=1}^{n} aL(x, x_i) + b \tag{4}$$

In Equation(4), a is called as Lagrange multiplier, and L is the kernel function. The radial quantity machine (RBF) is used as the kernel function,

$$L(x, x_i) = e^{\{-\frac{||x-x_i||^2}{2\sigma^2}\}} \tag{5}$$

In Equation(5), σ is the kernel function width.

3 LS-SVM MODEL OF BREAST TUMOR

The data of 469 cases of breast tumors (including 325 cases of benign and 144 cases of malignant) in a hospital in Yunnan province were taken as an example. Database contains 10 characteristics, the nucleus radius, texture, perimeter, area, smoothness, compactness, sag degrees, symmetry degree and fracture degree, each feature and means value, standard deviation and the worst value, namely, a total of 30 sets of data, a diagnosis of benign or malignant.

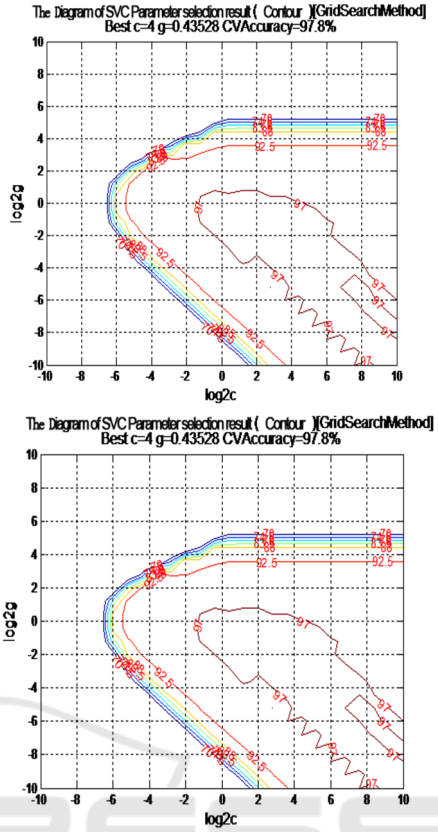

Fig 1 Parameter roughly selection.

From figure 1, c can be narrowed down to 2 ^ (-7) ~2 ^ (8), g can be narrowed down to 2 ^ (-6) ~ 2 ^ (6), it based on a rough parameter selection for parameter selection. Take c values change for: 2 ^ (-7), 2 ^ (-6.5)... 2 ^ (8).G change values for: 2 ^ (-6), 2 ^ (5.5)... 2 ^ (6).In order to see the change of accuracy more clearly, the actual change interval of the accuracy of the final parameter selection result graph was set to 0.9.

The result of fine parameter selection is shown in figure 2.As can be seen from the figure, the most parameter c is 8, g is 0.17678, and the cross validation accuracy is 97.8 %.As can be seen from figure 3, only the 69th value is different, and the rest are the same, that means the classification accuracy is 68/69=98.55%, proving that a better classification accuracy can be achieved by using the least squares support vector machine.

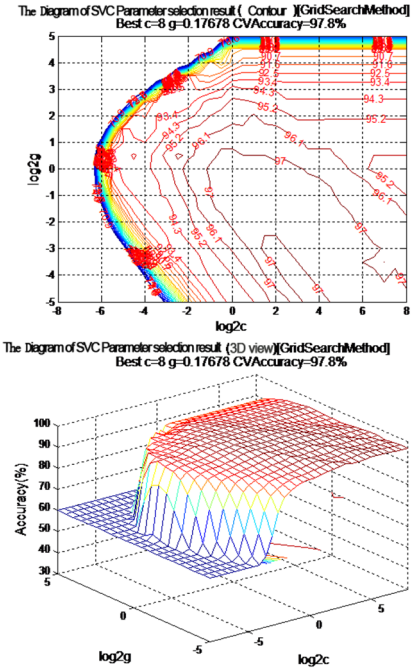

Fig 2 Parameter fine selection.

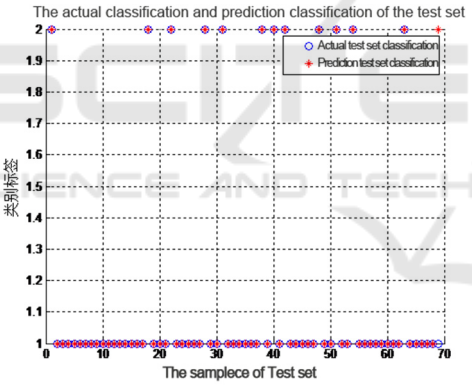

Fig 3 Actual classification & Prediction classification.

4 DIFFERENTIAL DIAGNOSIS AND COMPARISON OF BREAST TUMORS

According to the established LS-SVM model, the diagnostic accuracy and running time of diagnosis were obtained by MATLAB programming, and then the sensitivity and specificity of diagnosis were obtained by the method of ten-fold cross validation (HE Ya-peng, 2012). .From table 1 LS-SVM running time, accuracy, sensitivity and specificity degree were higher than that of SVM and BP neural network. It illustrates the breast tumor classification diagnosis model based on the LS-SVM for breast tumor.

Table 1 Comparison of diagnostic performance.

Method	Time (t)	Accuracy (%)	Sensitivity (%)	Specificity (%)
LS-SVM	3.242	97.80	94.45	95.58
SVM	3.363	94.35	93.63	94.32
BP neural network	3.745	91.77	91.23	91.77

5 CONCLUSIONS

The benign tumor by treatment and reasonable diet will gradually recover, but malignant tumors grow quickly, growth often extends into the surrounding tissue, few coated surface, often have systemic symptom such as transfer, so difficult to cure. At present, the highest incidence of breast cancer is women's diseases; cancer has become the world's highest incidence of female breast cancer, so accurate and timely diagnosis for breast neoplasms is particularly important.

This article used LS-SVM in breast cancer diagnosis, by solving the linear equation to obtain the optimal solution. Compared with the traditional SVM (WANG Bo, 2012)and BP neural network (LI Xiao-feng, 2008), it has huge advantages in terms of calculation speed, cost of computing and high accuracy. Therefore, the classification diagnosis model based on LS-SVM is very suitable for breast diagnosis, and the simulation results prove its accuracy as well.

REFERENCES

1. Lu, X., Lin, Y., &Luo, J. 2010. Classification Algorithm Combined GCM with CCM in Cancer Recognition. *Journal of SoftwarestSoftw*, *21*(11), 2838-2851.
2. Lin X. G. Pan, Y. J., Guo, Y. C. 2009. The study of cervical cancer cells model based on UV absorption spectrum. *Spectroscopy and Spectral Analysis*, *29*(9), 2547-2550.
3. Xu, W. Y., Chen, L., He, S., Li, Y. C., Yang, Y. H., Wang, A. Q., &Xie, G. 2003. Quantitative pathologic technique in prognostic identification of breast carcinoma with negative lymph node. *Chinese journal of oncology*, *25*(5), 461-463.
4. Wang, Y. H., Huang, D. X., Gao, D. J., & Jin, Y. H. 2004. Nonlinear predictive control based on LS-SVM. *Control and Decision*, *19*, 383-387.

5. He, Y. P., Zhuang, S. N., Zhang, Y. H., & Zhu, X. H. 2012. Cross validation based robust-SL 0 algorithm for target parameter extraction. *Systems Engineering and Electronics*, *34*(1), 64-68.
6. WANG, B., DU, X. X., & JIN, M. 2012. Application of Breast Tumor Diagnosis Based on Learning Vector Quantization Neural Network [J]. *Computer Simulation*, *8*, 042.
7. Li, X. F.,&Shen, Y. 2008. Support vector machines based computer-aided diagnosis system of breast tumor with ultrasound images. *JOURNAL OF OPTOELECTRONICS LASER*, *19*(1), 115-119.