INFORMATION TECHNOLOGY JOURNAL

Computerized Distinction of Improved Fuzzy Support Machine for Imageology Character of Benign and Malighant Pulmonary Nodules

[1]Yan Qiang, [1]Bo Pei, [2]Wei Wei, [1]Jianfeng Yang and [1]Juanjuan Zhao
[1]College of Computer Science and Technology, Taiyuan University of Technology, Taiyuan, 030024, China
[2]School of Computer Science and Engineering, Xi'an University of Technology, Xi'an, 710048, China

Abstract: In order to improve the accuracy of the solitary pulmonary nodule diagnosis with medical signs in medical imaging diagnostics, a novel computer-aided classification method is developed. In the view of the existing problems in the lung cancer diagnosis such as the large number of data and the low diagnose efficiency. In order to solve the problem, a new classification method based on the Fuzzy Support Vector Machine (FSVM) was developed to choose the lung with suspicious lesion at an early stage. In this method, the membership function was improved based on the spectral clustering theory which ensures each sample has two membership degrees that guarantees the class of the specific sample more reasonably. The proposed method was used to classify benign and malignant of the pulmonary nodules, the parameters show this method can distinguish the noise and outliers samples more effectively, compared with the traditional fuzzy support vector machine method. Thus, the results illustrated the robust to noise capability and the effective classification ability of this method.

Key words: Solitary pulmonary nodule, benign and malignant of pulmonary nodules, support vector machine, particle swarm optimization

INTRODUCTION

Lung cancer is one of the most common malignant tumors with the highest incidence at present in the world. Early detection and diagnosis of suspicious lesions allow early intervention and lead to better prognosis for the patients. With the development of the multi-slice computed tomography technology the diagnosis at the early stage of the lung cancer become possible. At this stage, the lung cancer always manifested as the Solid Pulmonary Nodule (SPN). The features of the SPNs became the main factors of benign and malignant recognition. Large number of clinical practice has proved that, in all kinds of features of SPN at the early stage, the sign of speculation will develop into cancer at a probability up to 85% in the future. So the identification of the speculation of the SPNs in the lung parenchyma has become the international hotspot.

The research on the pathological of pulmonary nodules has been an important issue in the medical imaging. Vapnik (1998) proposed the SVM method which providing a good solution to the classification and regression problems. However, when constructing the optimal hyperlane, this method think that all samples have the same effect which result in the hyperlane gotten always is not the true optimal one. In order to solve this problem, Lin and Wang (2002) proposed the FSVM method which applying the fuzzy theory to the SVM method which thought the membership function as the core in the method. The method constructs the optimal hyperlane through the different effects of the different samples which reducing the effects of the noise greatly when constructing the hyperlane. In this method, the classification according to the contribution of the samples come to the reality but the memberships of the samples in this method are only for the certain class that the samples probably belong to. So, the samples in this method are all have an only membership which neglecting the contributions to the other class.

In response to these shortcomings of the existing methods, we proposed a dual membership classification method based on the fuzzy support vector machine theory. In order to ensure a reliably classifier, we improved the membership function based on the spectral clustering theory which guaranteeing the membership degrees of the samples to each class more clearly in the two different sides.

Corresponding Author: Bo Pei, College of Computer Science and Technology, Taiyuan University of Technology, Taiyuan, 030024, China Tel: 18735128030

At present, there are so many research findings in terms of pulmonary nodules computer-aided diagnosis. Huang and Liu (2002) proposed a lung image segmentation method based on the anatomic structure to establish an effect model for lung and detect the pulmonary nodules. Brown *et al.* (2005) identified the lung nodule based on the local density maximum value method while using the simulation nodules in the experiment. Zhao *et al.* (2003) identified the lung nodule from the surrounding blood vessels and bronchial tissues based on the anatomic structure and filtered it according to the 2D and 3D information from the adaptive gray scale model. In order to detect the SPNs in the pulmonary. Lin *et al.* (2005) proposed an extended fuzzy classifier based on the neural network to train and test the samples which has gotten an excellent performance.

In addition, as an excellent classification method there are also so many researchers devoted in it recently. In order to determine the penalty factor C, Scholköpf *et al.* (2000) proposed v-svm method which controls the number of the support vectors and the classification errors by choosing the parameter v. In the traditional svm method the classification errors were represented by the once term resulted in the sensitivity to the noise, Song *et al.* (2002) proposed a LS-SVM method which transformed the quadratic programming problem into a problem of solving linear equation and simplified the complexity of the calculation. For the noise and outliers of samples in practice, Lin and Wang, (2002) proposed the Fuzzy-SVM method which gives different weights to each sample according to its contribution to the class.

MATERIALS AND METHODS

Dual fuzzy support vector machine model
Basic idea of the fuzzy support vector machine: Despite the good generating ability of the traditional SVM, the shortcoming of the same effects of all samples in the method when constructing the optimal hyperlane result in an unsatisfied effect as the existence of the noise points. In the fuzzy support vector machine method the membership degree to the class of each sample is determined based on the contribution to the specific class which can reduce or eliminate greatly the impacts of the noise or the isolated points to the hyperlane. To do this, the method determines the objective function which giving the different weights to the samples based on their contributions to the belonging classes.

The training set:

$$S = \{(x_1, y_1, s_1), \dots (x_i, y_i, s_i)\}$$

where:

$$x_i \in R^n, y_i \in \{-1, 1\}, \sigma \leq s_i \leq 1$$

s_i is the fuzzy membership degree of the training set (x_i, y_i, s_i). Let ε_i as the measurement of the wrong classification points, so s_i, ε_i, $i = 1, \dots, l$ is the measurement of the wrong classification points in various degrees of importance.

For the linear problems, the optimal classification problems can be eventually transformed into the quadratic programming problem as follows:

$$\begin{cases} \min\limits_{w,b,c} \dfrac{1}{2} \| w \|^2 + C \sum\limits_{i=1}^{l} s_i \varepsilon_i \\ s.t.\ y_i((w \cdot x) + b) + \varepsilon_i \geq 1, j = 1, \dots l \\ \quad \varepsilon_i \geq 0,\ j = 1, \dots, l \end{cases} \quad (1)$$

Where $C >$ is the penalty parameter, s_i is the degree of the membership to the specific class of the sample (x_i, y_i, s_i).

In order to get the dual program problem of Eq. 1 we constructed the Largrange function as follows:

$$L(w,b,\alpha,\beta) = \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{l} s_i \varepsilon_i - \\ \sum_{i=1}^{l} a_i (y_i (w \cdot x) + b) + \varepsilon_i - 1) - \sum_{i=1}^{l} \beta_i \varepsilon_i \quad (2)$$

Where $a = (a_1, \dots, a_l)^T, \beta = (\beta_1, \dots, \beta_l)^T, a_i \geq 0, \beta_i \geq 0, j = 1, \dots, l$, the final optimal solution is $a^* = (a_1^*, a_l^*)^T$ and optimal fuzzy classification function is $f(x) = sgn\{(w^* \cdot x) + b^*\}, x \in R^n$ where:

$$w^* = \sum_{j=1}^{l} a_i^* y_i x_i, b^* = y_i - \sum_{i=1}^{l} y_i a_i (x_j, x_i), \\ i \in \{i \mid 0 < a_i^* < s_i C\} \quad (3)$$

Dual fuzzy support vector machine (DFSVM) algorithm: In the traditional fuzzy support vector machine, the training samples often were separated into two disjoint classes by the optimal hyperlane based on the membership degrees. In this method, a sample can only be separated into one certain class while, there are so many outliers or noise points in actual and the classification is not always absolute (For a fuzzy set it can be 90% belongs to a collection while 10% to the other, or two collections of each half). In this study all the sample

points are not divided into two classes absolutely but the relatively two classes according to the two aspects of their membership.

The sample set in the DFSVM is:

$$cz = \{z_i^A = (x_i, y_i, u_i^A), z_i^B = \{x_i, y_i, u_i^B\}, u_i^A + u_i^B = 1, i = 1, \ldots, l\}$$

where u_i^A and u_i^B are the membership degree to the two classes of the sample, respectively. So in the DSFVM, the number of the sample points is double to that in the traditional FSVM method with only one membership degree. Therefore, the proposed classification model can not only overcome the over fitting in the traditional method but also can take a full use of the limited samples in the practice. The algorithm model of the DFSVM is as follows:

$$\min_{w,b} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(u_i^A\varepsilon_i + u_i^B\eta_i)$$
$$\text{s.t. } w^T\phi(x_i) + b \geq 1 - \varepsilon_i$$
$$w^T\phi(x_i) + b \leq -1 + \eta_i u_i^A + u_i^B = 1 \tag{4}$$
$$u_i^A \geq 0, u_i^B \geq 0, \varepsilon_i \geq 0, \eta_i \geq 0, i = 1, \ldots l$$

Where ε_i and η_i are the slack variables of the classes spacing. Apparently, the Eq. 4 is a quadratic programming problem which optimal solution can be find by saddle points of the Largrange function. In order to find the optimal solution we constructed the Largrange functions as follows:

$$L(w,b,\varepsilon_k,\eta_k;\alpha_k,\beta_k,v_k,v_k) =$$
$$\frac{1}{2}w^Tw + C\sum_{i=1}^{l}u_i^A\varepsilon_i + C\sum_{i=1}^{l}(1-u_i^A)\eta_i -$$
$$\sum_{i=1}^{l}\alpha_i(w^T\phi(x_i) + b - 1 + \varepsilon_i) + \tag{5}$$
$$\sum_{i=1}^{l}\beta_i(w^T\phi(x_i) + b + 1 - \eta_i) - \sum_{i=1}^{l}v_i\varepsilon_i - \sum_{i=1}^{l}v_i\eta_i$$

Where α_k, β_k, v_k, v_k are the Non-Negative Largrange factors. The optimal solution of the origin problem equals to that of the bidirectional problem. The detailed solving process is as follows:

$$\max_{\alpha_i,\beta_i}\sum_{i=1}^{l}\alpha_i + \sum_{i=1}^{l}\beta_i -$$
$$\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i - \beta_i)(\alpha_j - \beta_j)\phi(x_i)^T\phi(x_j)$$
$$\text{s.t. } \sum_{i=1}^{l}\alpha_i = \sum_{i=1}^{l}\beta_i \tag{6}$$
$$0 \leq \alpha_i \leq Cu_i^A \quad 0 \leq \beta_i \leq C(1 - u_i^A)$$
$$i = 1, 2, \ldots, l$$

After mapping from the high dimensional space to the low, the objective function of the bidirectional

optimization problem is converted to $\varphi(x_i)^T\varphi(x_j)$. If the dimension is still high after the nonlinear transformation, the calculation of the $\varphi(x_i)^T\varphi(x_j)$ will be very difficult which is called the "dimension disaster". To solve the problem, we use the kernel function in the original input space to replace the inner product calculation. The function is as follows:

$$K(x_i, x_j) = \varphi(x_i)^T\varphi(x_j) \tag{7}$$

Finally, we can get the classifier as follows:

$$f(x) = \text{sgn}\left(\sum_{i=1}^{m}(\alpha_i - \beta_i)K(x, x_i) + b\right) \tag{8}$$

In general, the functions that content with the Mercer Condition are as follows:
The linear kernel function:

$$K(x_i, x_j) = x_i, x_j$$

The radius basis function (RBF):

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right)$$

The sigmoid function:

$$K(x_i, x_j) = \tanh[v(x_i, x_j) + \alpha]$$

Through the above model, we can find the difference between the proposed method and the traditional SVM method is mainly reflected in the way that definition the class that the sample belongs to. So, it is essential for us to find a reasonable way to construct the membership function model.

Model of the membership function: In this study, the steps for the generation of the dual membership function are as follows.

First, separate the samples into two classes based on the spectral clustering algorithm and calculate the center of each class.

Then, choose the proper projection method to calculate the dual membership degree of the samples in each class.

Spectral clustering algorithm: The traditional clustering algorithm is constructed based on the convex spherical space which will be trapped into the "local optimum" when the samples space is not featured as convex. Recently, the emerging spectral clustering algorithm in the field of machine learning can solve the problem. Compared

with the traditional algorithm the spectral clustering is a unsupervised learning method which can get the optimal solution in any sample space.

The clustering problem is mapped into the undirected weighted graph in the spectral clustering method. First, map the original dataset into the K-dimensional space. Then, use the traditional clustering method (K-means, c-means) to cluster the samples in the K-dimensional space. The purpose of the optimal classification is that to maximize the similarity in the same sub-graph as possible as we can while minimize that in the different sub-graph. Therefore, the criterion of the classification affects the performance of the clustering. In this paper, we use the Shi-Malik method, whose normalized cut criteria for classification was proposed by Shi and Malik. The detailed steps of the algorithm are as follows.

First, construct the undirected weighted graph according to the affinity matrix converting the clustering problem into the separating problem of the graph.

The sample set $X = (x_1, x_2 \ldots x_k)$ where x_i represents the vertex of the graph, the weights between the vertexes represents the similarity degree of them. So we can construct a undirected weighted graph G (V, E, W) based on the similarities of the samples, where V is the collection of the vertexes, E is the collection of the edges and W represents the affinity matrix the elements in it are the weights between the vertexes.

The key of the step is to construct the affinity matrix W and we determine the elements in it as follows:

$$w_{ij} = \exp\left(-\frac{\|x_i - x_j\|}{2\sigma^2}\right) \quad (9)$$

Where x_i, x_j represent the samples, σ represents the core decay rate which is controlled by core radius. According to the spectral theory, a properly given σ can make the w_{ij} distributed in the K-dimensional sphere. Then, construct the Laplacian Matrix based on the affinity matrix.

After converting into an undirected weighted graph, the clustering problem can be separated by normalized cut. It can be proved that the optimal separating problem of the undirected weighted graph is a NP problem. To avoid the disadvantage, we relax the restrictions to the affinity matrix w to get the Laplacian matrix L and find out the o ptimal solution by the spectral decomposition of the L.

The key point during the procedure is to construct the Laplacian matrix. In this study, the method we adopted is as follows:

$$L = D-W \quad (10)$$

Where D is a diagonal matrix, the elements on the diagonal are the degrees of the vertexes.

Finally, calculate the eigenvalues and eigenvectors of the Laplacian matrix, decompose it and get the optimal solution.

The Laplacian matrix is a symmetrical positive semi-definite matrix; the eigenvalues of it are all non-negative as following form:

$$\lambda_n \geq \ldots \lambda_2 \geq \lambda_1 \geq 0$$

In the theory of Fiedler Vector, the optimal solution is λ_2 that is the eigenvalue of the second eigenvector. In the Fiedler Vector the classification information of the undirected weighted graph is included: Search the classification point i in x_2 based on the heuristic rule and classify the eigenvectors whose eigenvalue is larger than x_{2i} into a class, the rest are in the other class.

Generation of the membership: After the samples are clustered into two classes based on the steps above, we can definite the dual membership degree as follows:

$$\mu_i^A = \frac{d_i^B}{d_i^A + d_i^B}, \mu_i^B = \frac{d_i^A}{d_i^A + d_i^B}$$

Where $d_i^A = \|x_i - \bar{x}_A\|, d_i^B = \|x_i - \bar{x}_B\|$ and $\|\bullet\|$ represents the Euclidean space and $\bar{x}_A, \bar{x}_B$ are the average value of the two classes.

In the generation procedure of the membership for the samples, the dual membership degrees of the samples are based on the distance between the centers of the two classes which gotten by the spectral clustering method. The dual membership degrees are all between 0 and 1 which reflects the importance to both of the two classes.

Experiment: In this section, we will evaluate the performance of the proposed method by comparing it with the existing other classification methods. The existing methods include SVM and FSVM methods based on the different core functions.

The dataset in the study is from the medical image sequences repository in a hospital of Shanxi province. Considering the binary classification problem, in this paper all samples are divided into two classes, naming them, respectively as sign of speculation class A and no sign of speculation class B. In this study, we divide the samples into the training set and testing set where the testing set includes 1/3 of the samples from class A and class B, respectively while the training set includes 2/3 of the samples, respectively from the two classes.

The detailed steps of the experiments are as follows:

First, Segment the original image and get the SPN shown as Fig. 1.

After getting the suspicious lesion of the SPN, we transform the image with the Fast Fourier Transformation and get the image in frequent domain shown as Fig. 2.

According to the grayscale gotten in the procedure of the FFT, we transform the image into the binary image shown as Fig. 3.

Then, extract the features of the suspicious pulmonary nodules.

In order to quantify the feature of the speculation, we extracted the outline and marked the convex and concave points of the SPN shown as Fig. 4 and 5.

Then, count the number of the convex triangles, calculate the distance between the top vertex to the bottom and set the threshold of according to the experience to determine the speculation symptom of the SPN.
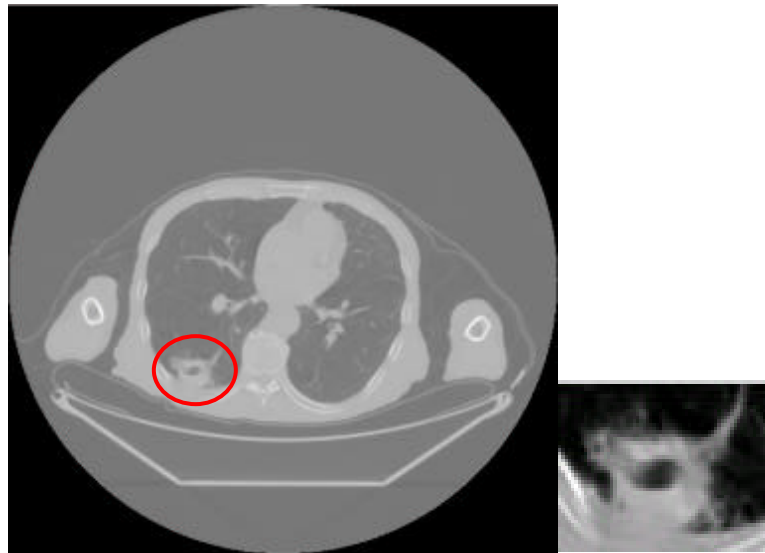
Fig. 1: Original image and the SPN
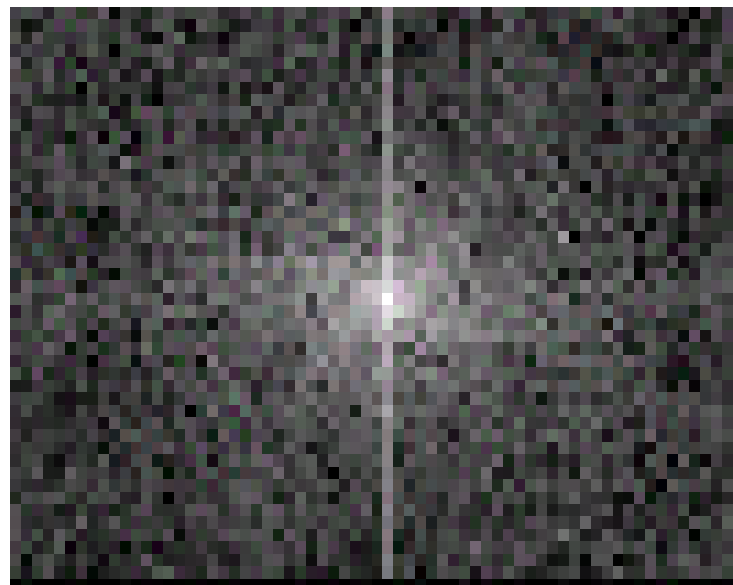
Fig. 2: SPN in frequent domain

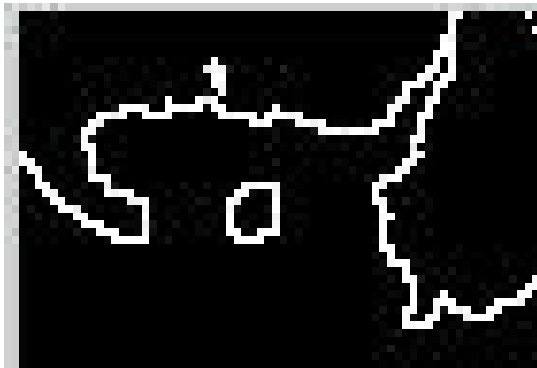Fig. 3: Binary image of the SPN

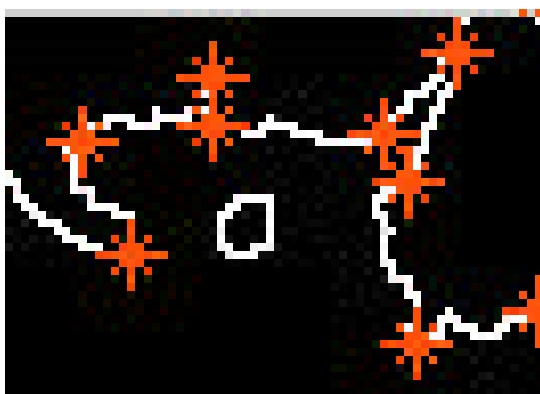Fig. 4: Outline of the SPN

Fig. 5: Marked outline of SPN

Finally, according to the extracted features construct the classifier based on the proposed method.
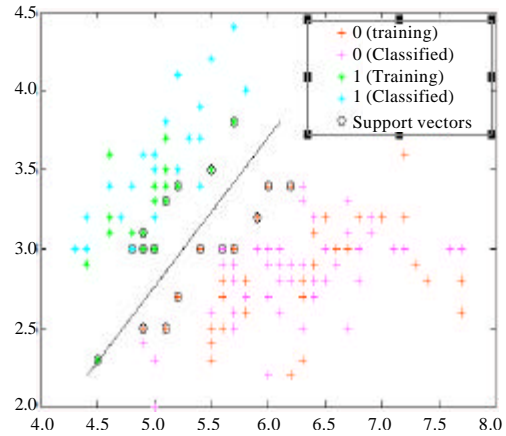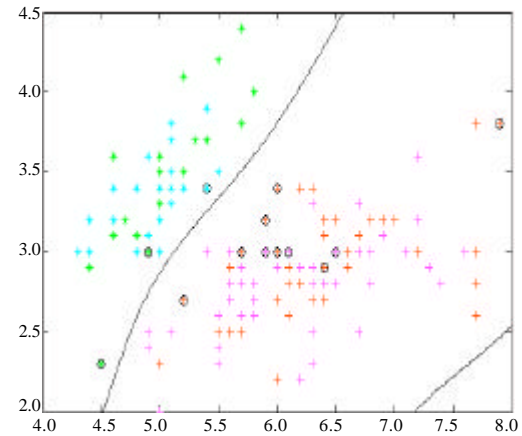

Fig. 6: Classification results based on SVM

Fig. 7: Classification result based on the FSVM

RESULTS AND DISCUSSION

According to the simulation, under the same condition, the results of the classifiers are shown as Fig. 6, 7 and 8.

As shown in the figures above, we can easily found that compared with the other two methods, the proposed method gets the fewest samples and has the best classification performance. But the DFSVM method may consume a little more time than the other two methods because of the clustering before classification while compared with the excellent classification performance the short comings are in tolerance.

For the binary classification problem, we use the confusion matrix as Table 1. TP and TN represent the number of the positive and negative samples that are classified correctly while FP and FN represent the number of that are incorrectly classified.
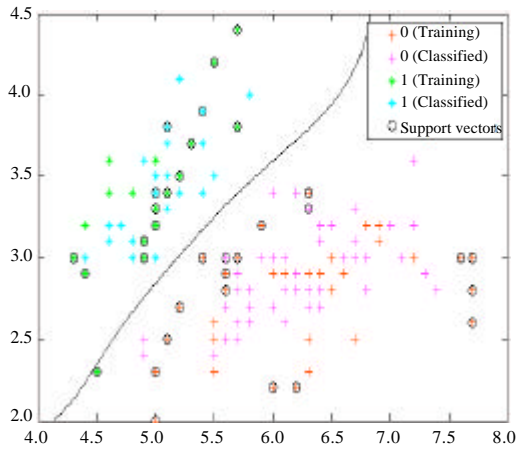
Fig. 8: Classification results based on the DFSVM

Table 1: Confusion matrix for binary classification

	Predicted	

Actual	Positive	Negative
Positive	True positive (TP)	False negative (FN)
Negative	False positive (FP)	True negative (TN)

Table 2: Simulation results of different models

Model	Kernel function	TPR (%)	TNR (%)	Accuracy (%)
SVM	RBF	88.98	91.21	90.96
FSVM	RBF	89.39	92.52	90.96
DFSVM	RBF	92.25	94.33	93.29

Therefore, we can calculate the true positive rate and the true negative rate according to the table.

The results of the different classifiers are shown in Table 2. In this study, we choose Matlab R2010a as the simulation tool and 2.66 GHz CPU and 1 GB of RAM as the hardware environment, Windows 7 as the operating system.

According the results of the simulation, the proposed method has a higher classification performance compared with the SVM and FSVM method in the aspect of TPR, TNR and the total accuracy rate.

CONCLUSION

In this study, we proposed a Dual Fuzzy Support Vector Machine (DFSVM) method based on the traditional SVM which ensures all the samples have two membership degrees using the spectral clustering based membership function. The method was used to classify the benign and malignant of the pulmonary nodules. The simulation results show that the proposed method outperforms especially when dealing with small samples in terms of accuracy and efficiency. Thus, the experiment results illustrated the validity of the method. What's more, the proposed method has also suggested an efficiency method to the small sample classification problem.

REFERENCES

Brown, M.S., J.G. Goldin, S. Rogers, H.J. Kim and R.D. Suh *et al.*, 2005. Computer-aided lung nodule detection in CT: Results of large-scale observer test. Acad. Radiol., 12: 681-686.

Huang, H.P. and Y.H. Liu, 2002. Fuzzy support vector machines for pattern recognition and data mining. J. Fuzzy Syst., 4: 826-835.

Lin, C.F. and S.D. Wang, 2002. Fuzzy support vector machines. IEEE Trans. Neural Networks, 13: 464-471.

Lin, D.T., C.R. Yan and W.T. Chen, 2005. Automatics detection of pulmonary nodules on CT images with a neural network-based fuzzy system. Computerized Med. Imag. Graphics, 29: 447-458.

Scholköpf, B., A. Smola, R.C. Williamson and P.L. Bartlett, 2000. New support vector algorithms. Neural Comput., 12: 1207-1245.

Song, Q., W. Hu and W. Xie, 2002. Robust support vector machine with bullet hole image classification. IEEE Trans. Syst. Man Cybernetic-Part C, 32: 440-448.

Vapnik, V.N., 1998. Statistical Learning Theory. 1st Edn., John Wiley and Sons, New York.

Zhao, B., G. Gamsu, M.S. Ginsbers, L. Jiang and L.H. Schwartz, 2003. Automatic detection of small lung nodules on CT utlizing a local density maximum algorithem. J. Applied Clin. Med. Phys., 4: 248-260.