American Journal of Biomedical Engineering 2013, 3(6): 175-181
DOI: 10.5923/j.ajbe.20130306.06

Automated Diagnostic System for Breast Cancer Using Least Square Support Vector Machine

Hamid Fiuji[1], Behnaz N. Almasi[2], Zahra Mehdikhan[3], Bahram Bibak[4], Mohammad Pilevar[5], Omid N. Almasi[6,*]

[1]Department of Biochemistry, Faculty of Science, Payame Noor University, Mashhad, Iran
[2]Department of Medical Science, Faculty of Nursing and Midwifery, Islamic Azad University, Mashhad, Iran
[3]Department of Electrical Engineering, Islamic Azad University, Mashhad, Iran
[4]Department of Molecular Science, North Khorasan University of Medical Sciences, Bojnord, Iran
[5]Department of Animal Sciences, Faculty of Agriculture, Ferdowsi University, Mashhad, Iran
[6]Department of Electrical Engineering, Islamic Azad University, Gonabad, Iran

Abstract Breast cancer is currently going to be one of the leading causes of death among women all over the world; however, it is for sure that the early detection and accurate diagnosis of this type of cancer can assure a longer survival of the patients. Because of the effective classification and high diagnostic capability, expert systems and machine learning techniques are now gaining popularity in this field. In this study, Least square support vector machine (LS-SVM) was used for breast cancer diagnosis. The effectiveness of the LS-SVM is examined on Wisconsin Breast Cancer Dataset (WBCD) using *K*-fold cross validation method. Compared to nineteen well-known methods for the breast cancer diagnosis in the literature, the study results showed the effectiveness of the proposed method.

Keywords Breast Cancer Diagnosis, K-Fold Cross Validation, Medical Diagnosis, Least Square Support Vector Machine, Wisconsin Breast Cancer Dataset

1. Introduction

A leading cause of death among women between 40and 55 years of age, breast cancer is now the second major cause of death among women. According to the World Health Organization, every year more than 1.2 million women are diagnosed with breast cancer across the globe. Luckily, in recent years with an increased emphasis on diagnostic techniques and more effective treatments, the mortality rate from breast cancer has declined. A key factor in this approach is the early detection and accurate diagnosis of this affliction[1-3].

Undoubtedly, the evaluation of data taken from patients and decisions of experts are the most important factors in diagnosis. Therefore, the use of classifier systems in medical diagnosis has been gradually increasing. After all, expert systems and various artificial intelligence techniques for classification also help experts to a considerable extent. Classification systems can help minimize possible errors that might occur due to inexperienced experts, and also provide medical data to be examined in shorter time and more detailed[3, 4].

* Corresponding author:
o.almasi@ieee.org (Omid N. Almasi)
Published online at http://journal.sapub.org/ajbe

Proposed as effective statistical learning methods for classification[5], Support Vector Machines (SVMs) rely on support vectors (SV) to identify the decision boundaries between different classes. Nonlinearly related to the input space, SVM is based on a linear machine in a high dimensional feature space, which has allowed the development of somewhat quick training techniques, despite the large number of input variables and large training sets. SVMs have successfully been used to address many problems including handwritten digit recognition[6], object recognition[7], speaker identification[8], face detection in images[9], and text categorization[10].

The Least Square Support Vector Machine (LS-SVM) was first proposed by Suykens and et al. by modifying the formulation of standard SVM[11]. The LS-SVM was modified at two points: First, instead of inequality constraints, it takes equality constraints and changed the quadratic programming to a linear programming. Second, a squared loss function is taken from the error variable[11, 12].

In this study, LS-SVM was employed to diagnose the breast cancer. For training and testing experiments, WBCD taken from the University of California at Irvine (UCI) machine learning repository was used. It was observed that the proposed method yielded the highest classification accuracies among the nineteen other methods in the literature. In this study, the performance was evaluated by the well-known k-fold cross validation method.

The rest of the paper is organized as follows. Section 2, briefly discusses the methods and results of previous studies on breast cancer diagnosis. Section 3 reviews basic SVM and LS-SVM concepts, respectively. Section 4 elaborates on the WBCD. Section 5 presents the experimental results achieved by applying the proposed method to diagnose breast cancer. Finally, we would make our concluding remarks in section 6.

2. Review of Literature

A great deal of approaches has been proposed to deal with automated diagnosis of breast cancer with WBCD, and most of them have managed to achieve high generalization performances. In[13], the author obtained 94.74% classification accuracy, in which10-fold cross-validation with C4.5 decision tree method was used. In[14], the researcher reached 94.99% accuracy with RIAC technique, while[15] have got to 96.8% with linear discreet analysis method. Using neuro-fuzzy techniques, the accuracy of method proposed by[16] was 95.06%. Using supervised fuzzy clustering method in[17], an accuracy of 95.57% was obtained. In[18], the fuzzy-GA method was introduced and a classification accuracy of 97.36% was achieved. In[19], three different methods, optimized learning vector quantization (LVQ), big LVQ, and artificial immune recognition system (AIRS) were applied and the obtained accuracies were 96.7%, 96.8%, and 97.2% respectively.

In[20], multilayer perceptron neural network, four different methods, combined neural network, probabilistic neural network, recurrent neural network and SVM were used respectively; and the highest classification accuracy of 97.36% was achieved by SVM. In[21], two different methods, Bayesian classifiers and artificial neural networks were applied and the obtained accuracies were 92.80% and 97.90%, respectively. In[22], the method combined with association rules and neural network were applied and accuracy of 95.60% was obtained. Moreover, in order to prepare the performance of the LS-SVM in automated diagnostics, five different variant of Artificial Neural Networks (ANNs) were employed in this study, which are frequently used in the literature. There are different kinds of ANNs, which are determined by their training algorithms and topologies. Adjusting the weights and bias of the ANN, to train an ANN, means to select a model from the set of allowed models that minimize the error of the generalization criterion. In this study, three training algorithms were used for training a three-layer ANN. The first is a well-known Levenberg-Marquardt Back Propagation (LM BP), the second is Gradient Descent Back Propagation (GD BP), the third is Gradient Descent with Momentum Back Propagation (GDM BP), and the fourth is Gradient Descent with Adaptive learning rule Back Propagation (GDA BP). The fifth comparison method is Radial Basis Function (RBF), which turned out as a famous variant of ANNs.

3. SVM for Classification

In this section, we summarize the basic SVM concepts with regard to typical two-class classification problems.

Support vector machines (SVM) originally developed by Boser et al.[23] and Vapnik[24], is based on the Vapnik–Chervonenkis (VC) theory and structural risk minimization (SRM) principle[24], by trying to find a trade-off between minimizing the training set error and maximizing the margin to achieve the best generalization ability and remain resistant to over fitting. Moreover, one major advantage of SVM is its use of convex quadratic programming, which provides only global minima; therefore, it avoids being trapped in local minima. For more details, cf.[24, 25], which give a complete description of the theory of SVM. In this section we will discuss the basic SVM concepts for typical binary-classification problems.

3.1. Linear Separable Case-Hard Margin

Let us consider a binary classification task: $\{x_i, y_i\}, i=1,\ldots,l,\ y_i \in \{-1,\ 1\},\ x_i \in R_d$, where x_i are data points and y_i are corresponding labels. They are separated with a hyper plane given by $W^T x + b = 0$, where w is an n-dimensional coefficient vector which is normal to the hyperplane and b is the offset from the origin.

There are lots of hyperplanes that can separate the two classes, whereas the decision boundary should be as far away from the data of both classes as possible, the support vector algorithm seeks an optimal separating hyper plane that maximizes the separating margin between the two classes of data. As the wider margin can acquire the better generalization ability, we can define a canonical hyper plane[24] such that $H_1 : W^T x^+ + b = +1$ for the closet points on one side and $H_2 : W^T x^- + b = -1$ for the closest on the other. Now to maximize the separating margin is equivalent to maximizing the distance between hyper plane H_1 and H_2. Hence we can get the maximal width between them: $m = (x^+ - x^-).\frac{w}{\|w\|} = \frac{2}{\|w\|}$. To maximize the margin the task is therefore:

$$\begin{aligned} & Ming(w) = \frac{1}{2}\|w\|^2 \\ & s.t. \\ & y_i(w^T x_i + b) \geq 1, \forall i \end{aligned} \quad (1)$$

Therefore, the learning task could be reduced to minimization of the primal Lagrangian:

$$MinL_p(w,b,\alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i (y_i (w^T x_i + b) - 1) \quad (2)$$

where α_i are Lagrangian multipliers, hence $\alpha_i > 0$. The minimum with respect to b and w of the Lagrangian, L_p, is given by,

$$\frac{\partial L_p}{\partial b} = 0 \rightarrow \sum_{i=1}^{n} \alpha_i y_i = 0$$
$$\frac{\partial L_p}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{n} \alpha_i y_i x_i \quad (3)$$

Now we substitute back b and w in the primal, which gives the dual Lagrangian:

$$Max \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_i x_i^T x_j$$
$$s.t. \quad (4)$$
$$\sum_{i=1}^{n} \alpha_i y_i = 0, \alpha_i \geq 0$$

Obviously, it is a quadratic optimization problem (QP) with linear constraints. From Karush Kuhn–Tucker (KKT) condition, we know that: $\alpha_i (y_i (W^T X_i + b) - 1) = 0$,Thus, only support vectors have $\alpha_i \neq 0$, which carry all the relevant information about the classification problem. Hence the solution has the form: $W = \sum_{i=1}^{n} \alpha_i y_i x_i = \sum_{i \in SV} \alpha_i y_i x_i$, where SV is the number of support vectors. And gets b from $y_i (W^T X_i + b) - 1 = 0$, where x_i is support vector. Therefore, the linear discriminant function takes the form: $g(x) = W^T x + b = \sum_{i \in SV} \alpha_i y_i x_i^T x_i$.

3.2. Linear Non-Separable Case-Soft Mmargin SVM

In practice, it is impossible to classify two classes accurately, because the data is always subject to noise or outliers, so in order to extend the support vector algorithms and solve imperfect separation, positive slack variables $\xi_i = 1, \ldots, l$ [24, 25] are introduced to allow misclassification of noisy data points, and to take into account the misclassification errors a penalty value C is introduced for the points that cross the boundaries. In fact, parameter C can be viewed as a way of controlling over-fitting. Therefore, the new optimization problem can be reformulated as follows:

$$Ming(w) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} \xi_i$$
$$s.t. \quad (5)$$
$$y_i (w^T x_i + b) \geq 1 - \xi_i, \xi_i \geq 0$$

Translate this problem into a Lagarangian dual problem

$$Max \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_i x_i^T x_j$$
$$s.t. \quad (6)$$
$$0 \leq \alpha_i \leq C, \sum_{i=1}^{n} \alpha_i y_i = 0$$

The solution to this minimizations problem is identical to the separable case except for the upper bound C on the Lagrange multipliers α_i.

3.3. Non-Linear Separable Case-Kernel Trick

In most cases, one can't linearly separate the two classes. In order to extend the linear learning machine to work well with nonlinear cases, a general idea is introduced, i.e., the original input space can be mapped into some higher-dimensional feature space where the training set is separable. With this mapping, the discriminant function is of following form:

$$g(x) = W^T \phi(x) + b = \sum_{i \in SV} \alpha_i \phi(x_i)^T \phi(x) + b \quad (7)$$

where $x_i^T x_i$ in the input space is represented as the form of $\varphi(x_i)^T \varphi(x)$ in the feature space. The functional form of the mapping $\varphi(x_i)$ does not need to be known since it is implicitly defined by the choice of kernel: $K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$. Thus, the optimization problem can be rewritten as:

$$Max \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$
$$s.t. \quad (8)$$
$$0 \leq \alpha_i \leq C, \sum_{i=1}^{n} \alpha_i y_i = 0$$

After the optimal values of α_i have been found, the decision function would be based on the sign of:

$$g(x) = \sum_{i \in SV} \alpha_i y_i K(x_i, x) + b \quad (9)$$

As a rule, any positive semi-definite functions K(x, y) that satisfy Mercer's condition could be kernel functions[26]. Kernel function is defined as a function that corresponds to a dot product of two feature vectors in some expanded feature space. There are many kernel functions that can be employed in SVM. The most commonly used kernels in SVM are listed in Table 1. In this Table σ and d are constants and those parameters must be set by a user. For MLP kernel a suitable choice for β_0 and β_1 is needed to enable the kernel function to meet Mercer's condition.

Table 1. The conventional Kernel function

Name	Kernel Function expression
Linear Kernel	$k(x, x_l) = x^T x_l$
Polynomial Kernel	$k(x, x_l) = (t + x^T x_l)^d$
RBF Kernel	$k(x, x_l) = \exp(-\|x - x_l\|^2 / \sigma^2)$
MLP Kernel	$k(x, x_l) = \tanh(\beta_0 x^T x_l + \beta_1)$

3.4. Least Square Support Vector Regression

The Least Square Support Vector Regression (LS-SVR) fully described in[27], is considered as an approximation tool in this study. The formulation of SVR was modified by Suykens and *et al*. at two points: First, instead of inequality constraints, it takes equality constraints and changed the quadratic programming to a linear programming. Second, a squared loss function is taken from the error variable[27, 28]. These modifications greatly simplified the problem and can

be specifically described as follows:

$$\min J(w,e)=\frac{1}{2}w^T w+\gamma\frac{1}{2}\sum_{k=1}^{N}e_k^2$$
$$s.t. \tag{10}$$
$$y_k=w^T\varphi(x_k)+b+e_k, \qquad k=1,...,N$$

where e_k are error variables that play a similar role as the slack variables ξ_k in Vapnik SVM formulation and γ is a regularization parameter in determining the trade-off between minimizing the training errors and minimizing the model complexity.

The Lagrangian corresponding to (10) can be defined as:

$$L(w,b,e,\alpha)=J(w,e)-\sum_{K=1}^{N}\alpha_k\{w^T\varphi(x_k)+b+e_k-y_k\} \tag{11}$$

where $\alpha_k\in R$ are the Lagrange multipliers. The KKT optimality conditions for a solution can be obtained by partially differentiating with respect to w, b, e_k, and α_k

$$\begin{cases}\frac{\partial L}{\partial w}=0\rightarrow w=\sum_{k=1}^{N}\alpha_k\varphi(x_k)\\ \frac{\partial L}{\partial b}=0\rightarrow\sum_{k=1}^{N}\alpha_k=0\\ \frac{\partial L}{\partial e_k}=0\rightarrow\alpha_k=\gamma e_k, k=1,...,N\\ \frac{\partial L}{\partial \alpha_k}=0\rightarrow wT\varphi(x_k)+b+e_k-y_k=0, \quad k=1,...,N\end{cases} \tag{12}$$

After elimination of the variable w and e_k, the following linear equation can be obtained:

$$\begin{bmatrix}b\\ a\end{bmatrix}=\begin{bmatrix}0 & \vec{1}_N\\ \vec{1}_N & \Omega+\gamma^{-1}I_N\end{bmatrix}\begin{bmatrix}0\\ y\end{bmatrix} \tag{13}$$

where $y=[y_1,\ldots,y_N]$, $\vec{1}_N=[1,\ldots,1]$ and $y=[\alpha_1,\ldots,\alpha_N]$. The kernel trick is applied here as follows

$$\Omega_{kl}=\varphi(x_k)^T\varphi(x_l)=K(x_k,x_l), \qquad k,l=1,\ldots,N \tag{14}$$

where $K(.,.)$ is the kernel function meeting Mercer's condition. b and α can be obtained by the solution to the linear system

$$b=\frac{\vec{1}_N(\Omega+\gamma^{-1}I_N)^{-1}y}{\vec{1}_N^T(\Omega+\gamma^{-1}I_N)^{-1}\vec{1}_N^T} \tag{15}$$

$$\alpha=(\Omega+\gamma^{-1}I_N)^{-1}(y-\vec{1}_N^T b) \tag{16}$$

Eventually, the resulting LS-SVR model for function estimation can be expressed as:

$$y(x)=\sum_{k=1}^{N}\alpha_k K(x_k,x_l)+b \tag{17}$$

3.5. Model Selection

LS-SVMs have two adjustable sets of parameters. One of them is called kernel parameter(s) and the other is called regularization parameter (γ). LS-SVM generalization ability depends on the proper choosing of those parameters. The best performance of SVM is realized with an optimal choice of the kernel parameter(s) and the regularization parameter. The optimal choice of those parameters is called LS-SVM's model selection problem[29-31].

Kernel parameter(s) are implicitly characterizing the geometric structure of data in high dimensional space named feature space. In the feature space the data becomes linearly separable in such a way that the maximal margin of separation between two classes is reached. The selection of kernel parameter(s) will change the shape of the separating surface in input space. Selecting improperly large or small values in kernel parameter results Over-fitting or Under-fitting in the LS-SVM model surface, so the model would be unable to accurately separate data[32, 33].

In non-separable problems, noisy training data will introduce slack variables to measure their violation of the margin. Therefore, a penalty factor γ is considered for controlling the amount of margin violation. Other words, the penalty factor γ is defined to determine the trade-off between minimizing empirical error and structural risk error and also to guarantee the accuracy of classifier outcome in the presence of noisy training data. Higher γ values cause the margin to be hard and the cost of violation to become too high, so the separating model surface over-fits the training data. In contrast, lower γ values allow the margin to be soft, which results in under-fitting separating model surface. In both cases, the generalization performance of classifier is unsatisfactory, so it makes the LS-SVM model useless [32, 34].

In this research, we employ a grid-search technique[35] using 5-fold cross-validation to find out the optimal model selection of LS-SVM.

4. The Wisconsin Breast Cancer Diagnosis Problem

In this section, we introduce the medical diagnosis problem which is the object of our study. Second to skin cancer, breast cancer is the most common cancer among women. The presence of a breast mass is an alert sign, but it is not always indicative of a malignant cancer. Fine Needle Aspiration (FNA) of breast masses is a cost-effective, non-traumatic, and mostly non-invasive diagnostic test that obtains information required to evaluate malignancy.

The Wisconsin breast cancer diagnosis (WBCD) database [36] is the result of the efforts made at the University of Wisconsin Hospital for accurately diagnosing breast masses based solely on an FNA test[37]. This dataset is generally used among researchers who use machine learning methods for breast cancer classification; therefore it allows us to

compare the performance of our method with that of others. Nine visually assessed characteristics of an FNA sample considered relevant for diagnosis were identified, and an integer value between 1 and 10 was assigned. The measured variables are as follows:

1. Clump thickness (υ_1);
2. Uniformity of cell size (υ_2);
3. Uniformity of cell shape (υ_3);
4. Marginal adhesion (υ_4);
5. Single epithelial cell size (υ_5);
6. Bare nuclei (υ_6);
7. Bland chromatin (υ_7);
8. Normal nucleoli (υ_8);
9. Mitosis (υ_9).

The diagnostics in the WBCD database were established by specialists in the field. The database itself consists of 683 cases, with each entry representing the classification for a certain group of measured values:

Case	v_1	v_2	v_3	⋯	v_9	*Diagnostic*
1	5	1	1	⋯	1	*Benign*
2	5	4	4	⋯	1	*Benign*
⋮	⋮	⋮	⋮	⋱	⋮	⋮
683	4	8	8	⋯	1	*Malignant*

Note that the diagnostics do not provide any information about the degree of benignity or malignancy. Four hundred and forty four samples of the dataset belong to benign type, and the rest are of malignant type.

5. Experimental Results and Discussion

In this section, we introduce the performance evaluation method, which is used to evaluate the proposed method. Finally, we would present the experimental results and discuss our observations of the results. The proposed automated diagnostic system for breast cancer using LS-SVM is done in MATLAB software R2008b.

All the experiments reported here are implemented using RBF kernels for the following reasons:

When the relation between desired output and input attributes is nonlinear, the RBF kernel non-linearly maps datasets into the feature space so that it can handle the datasets. The number of hyper-parameters is the second reason which influences the complexity of model selection. The RBF kernel has less hyper-parameter than the polynomial kernel. Eventually, the RBF kernel is numerically less difficult[38-41].

5.1. Performance Evaluation Methods

In this study, *k*-fold cross validation method was used for performance evaluation of breast cancer diagnosis using LS-SVM. *k*-fold cross validation is a way to improve over the holdout method. The data set is divided into *k* subsets, and the holdout method is repeated *k* times. Every time, one of the *k* subsets is used as the test set, and the other *k-1*subsets are gathered to form a training set. Then the average error across all *k* trials is calculated. The advantage of this method is that it is less significant for this method how the data gets divided. Every data point gets to be in a test set only once, and gets to be in a training set *k-1* times. As *k* increases, the variance of the resulting estimate reduces. The downside of this method is that the training algorithm must rerun *k* times from scratch, in other words, it takes *k* times computation to make an evaluation. To randomly divide the data into a test and training set *k* different times is a variant of this method. The advantage of this method is that you can independently choose how large you wish each test set to be and after how many trials you average should be over[42].

A confusion matrix[43] contains information about actual and predicted classifications done by a classifier. Performance of such a system is commonly evaluated using the data in the matrix. Table 2 shows the confusion matrix for a two-class classifier. In Table 2, TP is the number of true positives (benign breast tumor); FN, the number of false negatives (malignant breast tumor); TN, the number of true negatives; and FP, the number of false positives.

Table 2. Confusion matrix

	Predecited negative	Predicted positive
Actual negative	TN	FP
Actual positive	FN	TP

Table 3. The best parameter pair (γ, σ)

Partition	γ	σ
80-20%training-test	1.6784	2.4449

Table 4. Classification accuracies obtained with LS-SVM and other classifiers from literature

Method	Classification accuracy (%)
BP-GD	87.25
BP-GDM	88.62
BP-GDA	91.38
Bayesian classifiers[21]	92.80
BP-LM	94.52
C4.5[13]	94.74
RIAC[14]	94.99
ANFIS[16]	95.06
RBF	95.14
Supervised-FCM[17]	95.57
ANN-association rules[22]	95.60
Optimized-LVQ[19]	96.70
Nu-SVM[20]	96.79
Big LVQ[19]	96.80
LDA[15]	96.80
AIRS[19]	97.20
Fuzzy-GA[18]	97.36
SVM[20]	97.36
ANN[21]	97.40
LS-SVM	**97.81**

The optimal model selection of LS-SVM model (γ , σ) is presented in Table 3.

5.2. Results and Discussion

We conducted some experiments on the WBCD dataset mentioned in section 4, so that we can evaluate the effectiveness of LS-SVM. We compared our results with those of earlier methods. Table 4 shows the classification accuracies of our method and nineteen previous methods. As the results show, our method using 10-fold cross validation has obtained the highest classification accuracy, 97.81% reported up to now. Table 5 presents the confusion matrix for a LS-SVM classifier.

Given the research findings, the SVM-based model that we have developed yielded very promising results in classifying the breast cancer. We believe that the proposed system could be very helpful for physicians in their final decisions about their patients. Using such a tool, they can make reasonably accurate decisions.

Table 5. Confusion matrix

	Benign	Malignant
Benign	88	1
Malignant	2	46

6. Conclusions

Classification systems used in medical decision making, provide medical data to be examined in a shorter time and more detail. Based on statistical data for breast cancer in the world, this affliction is among the most prevalent types of cancer. In this study, a medical decision making system based on LS-SVM was applied in diagnosing breast cancer and the most accurate learning methods were evaluated. To diagnose breast cancer in a fully automatic manner using LS-SVM , experiments were conducted on the WBCD dataset. The experiment results strongly suggest that LS-SVM could be helpful in diagnosis of breast cancer. Compared to nineteen well-known methods in the literature, the experiment results demonstrated that the proposed method was more effective than other 19 methods in the breast cancer diagnosis.

REFERENCES

[1] D. West, P. Mangiameli, R. Rampal, and V. West, "Ensemble strategies for a medical diagnosis decision support system: A breast cancer diagnosis application," European Journal of Operational Research, Vol. 162, No. 2, 2005, pp. 532–551.

[2] K. Polat and S. Güneş, "Breast cancer diagnosis using least square support vector machine," Digital Signal Processing, Vol. 17, No.4, 2007, pp. 694–701.

[3] H.-L. Chen, B. Yang, J. Liu, and D.-Y. Liu, "A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis," Expert Systems with Applications, Vol. 38, No.7, 2011, pp. 9014–9022.

[4] M. F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis," Expert systems with applications, Vol. 36, No.2, 2009, pp. 3240–3247.

[5] V. N. Vapnik, "Statistical Learning Theory," New York: Wiley, 1998.

[6] B. Scholkopf, S. Kah-Kay, C. J. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. N. Vapnik, "Comparing support vector machines with Gaussian kernels to radial basis function classifiers," IEEE Transactions on Signal Processing, Vol. 45, No.11, 1997, pp. 2758–2765.

[7] M. Pontil and A. Verri, "Support vector machines for 3-D object recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence , Vol. 20, No. 6, 1998, pp. 637–646.

[8] V. Wan and W. M. Campbell, "Support vector machines for speaker verification and identification," Proceedings of IEEE Workshop Neural Networks for Signal Processing, 2000, pp. 775–784.

[9] E. Osuna, R. Freund, and F. Girosi, "Training support vector machines: Application to face detection," In Proceedings of computer vision and pattern recognition, 1997, pp. 130–136.

[10] T. Joachims, "Transductive inference for text classification using support vector machines," In Proceedings of international conference machine learning, vol. 99, 1999, pp. 200–209.

[11] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, "Least squares support vector machines," World Scientific Publishing, 2002, Singapore.

[12] J. A. K. Suykens, J. Vandewalle, and B. De Moor, "Optimal control by least squares support vector machines," Neural Networks, Vol.14, No.1, 2001, pp. 23–35.

[13] J. R. Quinlan, "Improved use of continuous attributes in C4.5," Journal of Artificial Intelligence Research, Vol.4, 1996, pp. 77–90.

[14] H. J. Hamiton, N. Shan, and N. Cercone, "RIAC: A rule induction algorithm based on approximate classification," Computer Science Department, University of Regina, 1996.

[15] B. Ster and A. Dobnikar, "Neural networks in medical diagnosis: Comparison with other methods," In Proceedings of the international conference on engineering applications of neural networks, 1996, pp. 427–430.

[16] D. Nauck and R. Kruse, "Obtaining interpretable fuzzy classification rules from medical data," Artificial Intelligence in Medicine, Vol. 16, No. 2, 1999, pp. 149–169.

[17] J. Abonyi and F. Szeifert, "Supervised fuzzy clustering for the identification of fuzzy classifiers," Pattern Recognition Letters, Vol. 24, No. 14, 2003, pp. 2195–2207.

[18] C. A. Pena-Reyes and M. Sipper, "A fuzzy-genetic approach to breast cancer diagnosis," Artificial intelligence in medicine, Vol. 17, No.2, 1999, pp. 131–155.

[19] D. E. Goodman, L. Boggess, and A. Watkins, "Artificial immune system classification of multiple-class problems," In Proceedings of the Artificial Neural Networks in Engineering ANNIE 2, 2002, pp. 179–183.

[20] E. D. Ubeyli, "Implementing automated diagnostic systems

for breast cancer detection," Expert Systems with Applications, Vol. 33, No.4, 2007, pp. 1054–1062.

[21] I. Maglogiannis, E. Zafiropoulos, and I. Anagnostopoulos, "An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers," Applied Intelligence, Vol. 30, No.1, 2009, pp. 24–36.

[22] Murat Karabatak and M. Cevdet Ince, "An expert system for detection of breast cancer based on association rules and neural network", Expert Systems with Applications, Vol. 36, No.2, 2009, pp. 3465–3469.

[23] B. E. Boser, I. M Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," In Fifth annual workshop on computational learning theory, 1992, pp. 144–152.

[24] Corinna Cortes and Vladimir Vapnik, "Support-vector networks", Machine Learning, Vol. 20, No.3, 1995, pp. 273–297.

[25] N. Cristianini and J. Shawe-Taylor, "An introduction to support vector machines: And other kernel-based learning methods," Cambridge, UK: Cambridge University Press, 2000.

[26] A. J. Smola, "Learning with kernels: Support vector machines, regularization, optimization, and beyond," The MIT Press, 2002.

[27] J. A. K. Suykens, "Support vector machines: a nonlinear modeling and control perspective," European J. of Control, Vol. 7, No. 2–3, 2001, pp. 311–327.

[28] C.-C. Chuang, "Fuzzy weighted support vector regression with a fuzzy partition," IEEE Transaction on System, Man, Cybern. B, Cybern., Vol. 37, No. 3, 2007, pp. 630–640.

[29] X. Peng and Y. Wang, "A geometric method for model selection in support vector machine," Expert Systems with Applications, Vol. 36, No.3, 2009, pp. 5745–5749.

[30] S. Wang, B. Meng, "Parameter selection algorithm for support vector machine," Procedia Environmental Sciences, Vol. 11, 2011, pp. 538–544.

[31] O. Chapelle, V N. Vapnik, O. Bousquet, and S. Mukherjee, "Choosing multiple parameters for support vector machines," Machine Learning, Vol. 46, No. 1, 2002, pp. 131–159.

[32] S. S. Keerthi, "Efficient Tuning of SVM Hyperparameters Using Radius/Margin Bound and Iterative Algorithms," IEEE Transaction on Neural Networks, Vol. 13, No. 5, pp.1225–1229.

[33] P. Williams, S. Li, J. Feng, and S. Wu, "A geometrical method to improve performance of the support vector machine," IEEE Transaction on Neural Networks, Vol. 18, No. 3, 2007,pp. 942–947.

[34] S. Ding and X. Liu, "Evolutionary computing optimization for parameter determination and feature selection of support vector machines," IEEE Conference on Computational Intelligence and Software Engineering, 2009, pp. 1-5.

[35] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," Technical report, Department of Computer Science and Information Engineering, National Taiwan University, Taipei, 2003. Available at http://www.csie.ntu.edu.tw/cjlin/libsvm/.

[36] C. J. Merz and P. M. Murphy, "UCI repository of machine learning databases," http:www.ics.uci.edu:~mlearn: MLRepository.html, 1996.

[37] O. L. Mangasarian, R. Setiono, and W. H. Wolberg, "Pattern recognition via linear programming: Theory and application to medical diagnosis," In: Coleman TF, Li Y, editors. Large-Scale Numerical Optimization. SIAM, 1990, pp. 22–31.

[38] S. S. keerthi and C. –J. Lin, "Asymptotic behavior of support vector machines with Gaussion kenerl," Neural Computation, Vol. 15, No. 7, 2003, pp.1667–1689.

[39] H.-T. Lin and C.-J. Lin, "A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO –type methods," Technical report, Department of Computer Science , National Taiwan University, 2003, pp. 1–32.

[40] A. Bordes, S. Ertekin , J. Weston, and L. Bottou ,"Fast Kernel Classifiers with Online and Active Learning," The Journal of Machine Learning Research, Vol. 6 , 2005, pp. 1579–1619.

[41] J. Sun, C. Zheng, X. Li, and Y. Zhou, "Analysis of the Distance Between Two Classes for Tuning SVM Hyperparameters," IEEE Transaction on Neural Networks, Vol. 21, NO. 2, 2010, pp. 305–318.

[42] Jeff Schneider's home page, http://www.cs.cmu.edu/~schneide/tut5/node42. html, last accessed August 2006.

[43] R. Kohavi, and F. Provost, "Editorial for the Special Issue on Applications of Machine Learning and the Knowledge Discovery Process," Vol. 30, No. 2–3, 1998.