

Application of an Improved FSVM Algorithm in Breast Cancer Diagnosis

Jinzhi Zhou*, Jing Huang, Linwen Zheng

School of Information Engineering, Southwest University of Science and Technology, Mianyang, Sichuan, China.

How to cite this paper: Jinzhi Zhou, Jing Huang, Linwen Zheng. (2020) Application of an Improved FSVM Algorithm in Breast Cancer Diagnosis. *Journal of Applied Mathematics and Computation*, 4(2), 18-25. DOI:10.26855/jamc.2020.06.002

Received: April 12, 2020

Accepted: May 7, 2020

Published: May 15, 2020

***Corresponding author:** Jinzhi Zhou, School of Information Engineering, Southwest University of Science and Technology, Mianyang, Sichuan, China. Email: zhoujinzhi@swust.edu.cn

Abstract

The traditional method of constructing fuzzy support vector machine (FSVM) membership function based on the Euclidean distance between the sample and the class center treats all features equally, without considering the effect of different features on the distance between the sample and the class center. To solve this problem, a fuzzy support vector machine classification algorithm based on Relief-F feature weighting is proposed. First, the weights of each feature are calculated by the Relief-F algorithm and the features with smaller weights are removed; then the feature weights are used to calculate the weighted Euclidean distance from the sample to the center of the class; finally, the membership function is constructed based on the weighted Euclidean distance. This method considers the influence of feature importance on the classification effect, and removes the features with less weight through the weight threshold, thereby reducing the dimension of the data and improving the classification accuracy and training efficiency. And apply in the diagnosis of breast cancer. After theoretical analysis and experimental data verification, this method improves the performance of classification prediction compared to traditional support vector machines (SVM). The results are also competitive with the latest methods and have advantages in medical diagnostic applications.

Keywords

Fuzzy Support Vector Machine (FSVM), Relief-F, Feature Weighting, Membership Function, Medical Diagnostic

1. Motivation

Breast cancer is one of the most frequent malignant tumor diseases among women, which seriously threatens the life and health of women. A large number of studies have shown that early diagnosis and treatment is a key measure to improve the recovery rate of breast cancer patients and ensure their health. The computer-aided diagnosis (CAD) technology based on data mining method can help doctors detect and analyze suspicious lesions and improve the accuracy of early diagnosis of breast cancer.

In recent years, in the computer-aided diagnosis technology, due to the Support Vector Machine (SVM) method, it has largely overcome the problems of dimensional disasters and local minimization of feature attributes in traditional machine learning [1,2]. There are many advantages in nonlinear and high-dimensional pattern recognition, and it has received more and more attention in the fields of biomedicine and bioinformatics [3]. SVM has become a research hotspot in recent years due to its good performance, the main research focuses on the improvement of generalization ability and the acceleration of learning speed. One of the ways to improve its generalization ability is to introduce new parameters to traditional support vector machines (SVM), the typical method is fuzzy SVM [4,5].

Lin and Zhang Xiang introduced fuzzy mathematics to support vector machines and proposed a fuzzy support vector

machine (Fuzzy SVM, FSVM) model [6,7], which assigns corresponding membership degrees according to the different contributions of different input samples to the classification. This can reduce the impact of outliers and noise, and improve classification performance. In order to overcome the shortcomings of different categories of sample sizes, which leads to unsatisfactory classification performance, Fan Xin wei and others proposed a weighted support vector machine (Weighted SVM, WSVM) model with different weights for each type of sample [8], and compensated the impact of category differences accordingly, thereby improving the classification of small category samples, which has important practical significance for some applications that need to focus on the accuracy of small categories of samples. Zhao Hui and others integrated the ideas of WSVM and FSVM, and proposed a dual-weighted support vector machine (DWSVM) model [9] that takes into account both the difference in the number of categories and the importance of the samples.

In [10], fuzzy support vector machine (FSVM) was introduced and an optimization method was proposed to improve the accuracy of sample membership and thus improve the classification accuracy of fuzzy support vector machine. Fuzzy support vector machine based on membership function of multi-hyperplane distance measurement is proposed in [11]. The ultimate goal of research and improvement of advanced technology is to apply it to people's real life in order to improve the quality of life and raise the standard of living. Therefore, in the medical and health field, the literature applied the optimized model of SVM to the identification of breast cancer, and obtained a certain improvement in diagnostic efficiency [12,13].

The above fuzzy support vector machines assume that the features of each dimension of the training sample have the same contribution to the classification, and do not consider the influence of the importance of the feature on the distance between the sample and the class center. In addition, the distance is calculated based on all the features of the sample. In reality, not all features are strongly related to classification, and some features may be weakly related, unrelated, or even negatively related to classification. In this case, the calculation of the distance may be dominated by features that are not strongly correlated, so the calculated membership often does not reflect the correlation of the sample to the classification. For example, a certain noise point is far away from the center point on some relatively important features, but it is close to the center point on some relatively unimportant features. It is possible to calculate a relatively large degree of membership. In turn, the effect of noise data on the optimal classification surface structure is not reduced but increased, which plays the opposite role, resulting in a reduction in the classification effect.

In order to solve the above problems, this paper proposes a fuzzy support vector machine classification algorithm based on Relief-F feature weighted Euclidean distance, and applies it to the diagnosis and prediction of breast cancer. Calculate the weight of each feature and calculate the weighted Euclidean distance from the sample to the center of the class through the Relief-F algorithm, and then construct a membership function based on the weighted Euclidean distance and classify, in order to obtain a higher support vector machine algorithm for disease classification correct rate. On the one hand, the algorithm considers the influence of feature importance on the classification effect, thereby generating more accurate membership; on the other hand, by setting the weight threshold to remove features with lower weights, the data is reduced in dimensionality and the training efficiency is improved. The experimental results show that the improved SVM algorithm has improved classification and recognition performance, and has a significant effect in improving the accuracy of computer-aided diagnosis of breast cancer diseases.

2. Fuzzy Support Vector Machine

Suppose the sample set T consist of n fuzzy training samples $(x_1, y_1, \varphi_1), \dots, (x_n, y_n, \varphi_n)$ with membership, where x_i in the i -th sample is the characteristic attribute of the sample, $y_i \in \{\pm 1\}$ is the category of the sample, and $\varphi_i (0 < \varphi_i \leq 1)$ is the membership of the sample, then the objective function of the fuzzy support vector machine is as follows:

$$\begin{cases} \min_{w, b, \xi} \frac{1}{2} \|w\|^2 + \sum_{i=1}^N C \varphi_i \xi_i \\ y_i (w \cdot x_i + b) \geq 1 - \xi_i \\ \xi_i \geq 0, i = 1, 2, \dots, n \end{cases} \quad (1)$$

Where w is the linear coefficient of the optimal classification surface; ξ_i is the relaxation variable of the objective function, corresponding to the allowable offset of the data point; $C \varphi_i$ is the penalty factor under the control of the membership degree φ_i , the greater the penalty factor is the constraint on the offset sample The higher the degree; b is the offset.

The solution process is similar to the standard support vector machine. The Lagrange multiplier method can be used to transform the above problem of solving the optimal hyperplane into a dual problem

$$\begin{cases} \min_a \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N a_i a_j y_i y_j K(x_i \cdot x_j) - \sum_{i=1}^N a_i \\ \text{s.t. } \sum_{i=1}^N a_i y_i = 0, \\ 0 \leq a_i \leq C \varphi_i, i = 1, 2, \dots, n \end{cases} \quad (2)$$

Where the anonymous mapping of the kernel function is $\Phi(x)$, $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$ is called the kernel function, and the kernel function should be selected as an inner product of the feature space. Obviously, the difference between fuzzy support vector machine and standard support vector machine is that the upper bound constraint of variable a_i in the dual problem changes with the fuzzy factor φ_i , this is also equivalent to using a penalty factor $C\varphi_i$ for each sample, so the determination of the fuzzy factor φ_i becomes the key to determine the performance of this fuzzy support vector machine.

3. Design of Membership Function Based on Relief-F Feature Weighting

The design of membership function is the core of fuzzy support vector machine. The membership can reduce the impact of noisy data on classification. However, most current methods treat all features equally and do not consider the impact of different features on classification. To this end, this paper proposes a membership function based on Relief-F feature weighting to implement efficient classification. The method mainly includes feature selection based on Relief-F, calculation of weighted Euclidean distance, design of membership function and realization of feature-weighted fuzzy support vector machine. The main framework of the classification algorithm is shown in Figure 1.

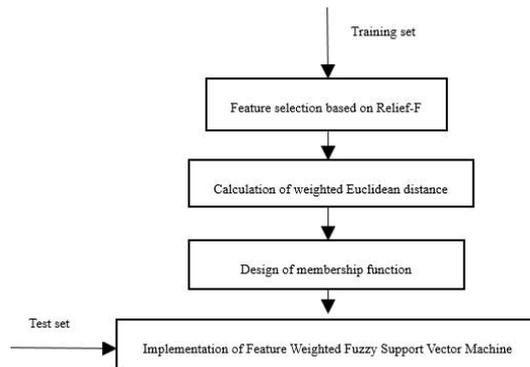


Figure 1 General framework of classification algorithm

3.1 Feature selection based on Relief-F

The basic Relief algorithm was proposed by Kira and Rendell in 1992, which is limited to solving two kinds of classification problems [14]. In 1994, Kononeill expanded the Relief algorithm and got Relief-F algorithm [15], which can handle multi-category problems, missing data problems, and regression problems.

Relief-F assigns different weights to features according to the correlation of each feature and category. Features with weights less than a certain threshold will be removed. The Relief-F algorithm can detect redundant or irrelevant attributes. The correlation between features and categories in the Relief-F algorithm is based on the ability of the features to distinguish between close-range samples. When calculating the relative distance, the features involved can affect the distance between these sample points, and then affect the selection of the nearest neighbors. In this way, the weight of the feature will be evaluated.

Suppose the training data set D , the number of samples m , the threshold δ of the feature weight, and the output is the weight β of each feature:

- 1) Set all feature weights to 0 and W to the empty set;
- 2) For $i=1$ to m do
 - a) Randomly select a sample R from D
 - b) Find out K nearest neighbor samples $H_j(j = 1, 2, 3, \dots, k)$ from the same sample set of R , and find K nearest neighbor samples $M_j(C)$ from the different sample sets of R
- 3) For $A=1$ to N all feature do

$$\beta(A) = \beta(A) - \sum_{j=1}^K \text{diff}(A, R, H_j) / (mk) + \sum_{C \in \text{class}(R)} \frac{p(C)}{1-p(\text{Class}(R))} \sum_{j=1}^k \text{diff}(A, R, M_j(C)) / mk$$

End

In this method, the feature weight of each feature (non-class label) in the sample set D can be calculated. The feature with the highest feature weight value is the feature with the highest discrimination in a given feature set, namely, the feature that contributes the most to classification. Characteristics. From this, we can use the Relief-F feature weight method to measure the correlation of each feature with respect to the classification: The larger the feature weight is, the stronger the correlation is. Assuming that each sample (except the class label) in the data set D is described by n feature (A_1, A_2, \dots, A_n) , the vector $\beta = (\beta(A_1), \beta(A_2), \dots, \beta(A_n))$ naturally describes the weight of each feature, where β_k represents the feature weight of the feature A_k .

3.2 Calculation of Weighted Euclidean Distance

In this paper, the membership function is constructed by calculating the distance from the sample to the center. Traditional algorithms mostly use Euclidean distance. However, the Euclidean distance cannot truly reflect the degree to which the sample belongs to the category. To solve this problem, based on the feature set obtained by the above feature weighting method, the weighted Euclidean distance is used to calculate the distance from the sample to the center.

1) Calculate the mean point of all the samples in the training data set D on each feature and use this as the class center O_k of the k -th class.

2) Calculate the weighted Euclidean distance of O_k from all samples in D . The weighted Euclidean distance from the sample x_k to the class center O_k is

$$d^\beta(x_k, O_k) = \sqrt{\sum_{j=1}^n \beta_j |x_{kj} - O_{kj}|^2} \tag{3}$$

Where $d^\beta(x_k, O_k)$ represents the weighted Euclidean distance between two samples x_k and O_k , x_{kj} represents the value of the j -th feature of sample x_k , O_{kj} represents the value of the j -th feature of the class center O_k , $\beta = (\beta_1, \beta_2, \dots, \beta_j)$ is the weight vector, and $\beta_j \geq 0 (j = 1, 2, \dots, n)$ is the importance weight of each feature.

3) Calculate the corresponding weighted Euclidean distance of all categories in the training set using the above method.

3.3 Design of Membership Function

Based on the weighted Euclidean distance calculated above, a membership function based on the weighted Euclidean distance is designed. The membership function needs to first determine the hypersphere radius of each type of training set. Here, the maximum weighted Euclidean distance of each type of training set is used as the hypersphere radius of each type of training set. Suppose the radius of the hypersphere of the k -th class training set is $d_{k,max}$.

For any sample x_i in the training set D , y_i is the category to which x_i belongs, and its membership can be calculated by the following membership function:

$$\varphi_i = 1 - \frac{d^\beta_{x_i, O_{y_i}}}{d_{y_i, max} + \delta} \tag{4}$$

Where $d^\beta_{x_i, O_{y_i}}$ represents the weighted Euclidean distance from the sample x_i to the center of class y_i ; $d_{y_i, max}$ represents the radius of the hypersphere of the training set of class y_i ; δ is a small positive number set in advance to ensure that $0 < \varphi_i \leq 1$.

According to equation (4), the corresponding membership degree can be calculated for each sample of the total training set, and the fuzzy training samples can be obtained.

3.4 Implementation of Feature Weighted Fuzzy Support Vector Machine

Generally, the support vector machine constructed by kernel function with feature weighting is called feature weighting support vector machine. The feature weighted kernel function is defined as follows:

Definition1: Feature weighted kernel function: If K_p is a kernel function defined on $X * X, X \in R^n$, P is the n -order linear transformation matrix of the known set to be input, where n is the dimension of the set to be input. The Feature weighted kernel function K_p is defined as:

$$K_p(x_i, x_j) = K(Px_i, Px_j) \tag{5}$$

Where $P = \text{diag}(\beta) = \begin{pmatrix} \beta_1 & 0 & \dots & 0 \\ 0 & \beta_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \beta_n \end{pmatrix}$ is the feature weighting matrix. In general, p only considers the conditions of

the diagonal matrix. β is the feature weight vector, $\beta = (\beta_1, \beta_2, \dots, \beta_n)^T \in R^n$, $0 \ll \beta_k \ll 1$, $k = 1, 2, \dots, n$; β_k represents the weight of the k -th feature.

In order to obtain better experimental results, the feature weighted Gaussian radial basis kernel function, which performs well in nonlinear feature classification, is selected as the kernel function of the classifier

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0 \tag{6}$$

Feature-weighted Gaussian radial basis kernel function

$$K_p(x_i, x_j) = \exp(-\gamma \|Px_i - Px_j\|^2) = \exp(-\gamma ((x_i - x_j)^T P P^T (x_i - x_j))) \tag{7}$$

After determining the feature weights according to the Relief-F algorithm, using the feature weights to calculate the weighted Euclidean distance from the sample to the center of the class to design the membership function based on the weighted Euclidean distance, a weighted fuzzy support vector machine based on Relief-F feature can be constructed.

1) Suppose the training data set D

$$(x_1, y_1, \varphi_1), \dots, (x_n, y_n, \varphi_n), x_i \in R^n, y_i \in \{\pm 1\}, 0 < \varphi_i \leq 1$$

2) According to the Relief-F algorithm, the weights of each feature are calculated respectively, and the feature weight vector is constructed:

$$\beta = (\beta_1, \beta_2, \dots, \beta_n)^T \in R^n$$

3) Calculate the weighted Euclidean distance from each sample to the class center

$$d^\beta(x_k, O_k) = \sqrt{\sum_{j=1}^n \beta_j |x_{kj} - O_{kj}|^2}$$

4) Calculate the membership degree of each sample based on the weighted Euclidean distance, and calculate the corresponding membership degree of each sample to obtain the fuzzy training samples.

$$\varphi_i = 1 - \frac{d_{x_i, O_{y_i}}^\beta}{d_{y_i, max} + \delta}$$

5) Select the appropriate penalty parameter $C > 0$ to construct and solve the optimization problem:

$$\begin{cases} \min_a \frac{1}{2} \sum_i \sum_{j=1}^N a_i a_j y_i y_j K(x_i \cdot x_j) - \sum_{i=1}^N a_i \\ \text{s.t.} \sum_{i=1}^N a_i y_i = 0, \\ 0 \leq a_i \leq C \varphi_i, i = 1, 2, \dots, n \end{cases}$$

Where $P = \text{diag}(\beta) = \begin{pmatrix} \beta_1 & 0 & \dots & 0 \\ 0 & \beta_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \beta_n \end{pmatrix}$, get the optimal solution $a = (a_1, a_2, \dots, a_n)^T$

6) Calculate $w = \sum_{i=1}^N a_i y_i \text{diag}(\beta) x_i$, Select the sample point (x_j, y_j) corresponding to a component $0 < a_j < C$ of a , and calculation $b = y_i - \sum_{i=1}^N a_j y_j x_j^T \text{diag}(\beta) x_i$, Construct the optimal classification hyperplane $w \cdot x + b = 0$, Structure decision function $f(x) = \text{sign}[(w \cdot x) + b]$.

7) In the SVM model, an optimal support vector classifier is constructed by using the constructed feature weighted kernel function $K(Px_i, Px_j)$, and selecting an appropriate penalty parameter C and a radial basis function parameter γ .

8) Performance test of constructed RFSVM.

4. Simulation Experiments and Analysis

In order to verify the feasibility and practicability of the proposed method, the databases named ‘‘Wisconsin Breast Cancer’’ was selected from the UCI machine learning database for experiments. There are a total of 699 samples. Each sample is described by 10 features (one of which is the sample number), the value range of each feature is $[1, 10]$. Excluding 16 samples that contain unknown feature values, the remaining 683 samples are denoted as SI , and SI is a two classification data set. In order to test the stability of the application performance of the RFSVM algorithm, this paper implements the algorithm in the environment of Matlab 2016a. In the experiment, the 10-fold cross-validation method

was used to conduct the experiment. In the experiment, the breast cancer data set was divided into 10 sets of similar size, and 9 of them were taken as training sets and 1 as the test set in turn. Calculate the prediction accuracy and error.

Since each experiment will get the corresponding correct rate of the algorithm, the average value of the correct rate of these 10 results is used as the estimation of the recognition rate of the algorithm. In order to verify the improvement of the performance of the RFSVM algorithm in the diagnosis of breast cancer, the fuzzy support vector machine method based on Relief-F feature weighting (RFSVM), standard SVM method and standard Euclidean FSVM method (LFSVM) are compared. The results are shown in Table 1.

Table 1. Experimental results on breast cancer data set

Table	Average accuracy (%)	Mean absolute error (%)	Root mean square error (%)	Average training time (s)
SVM	88.7	9.6	22.3	6.2
LFSVM	93.2	7.5	19.2	5.5
RFSVM	97.7	6.5	17.5	4.5

It can be seen from Table 1 that in the diagnosis of breast cancer, the RFSVM algorithm proposed in this paper compares the other two algorithms with a relatively high average accuracy rate, less training time, and a relatively low discrimination error in error analysis. Therefore, it can be obtained that the improved RFSVM algorithm classification prediction performance has been improved, and the effect on the recognition rate of breast cancer disease has been improved.

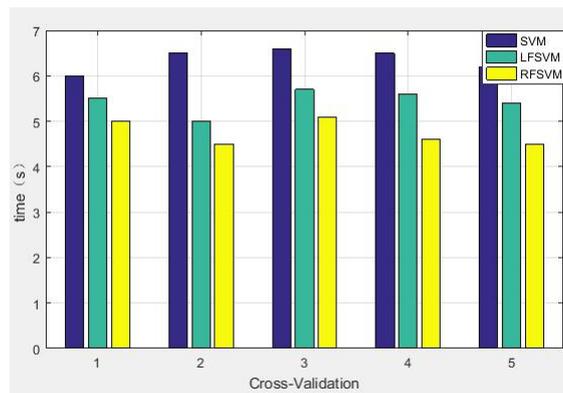


Figure 1. Training time used by the three algorithms

The time required for the first five cross-validation trainings is shown in Figure 1. It can be seen that the RFSVM takes the least training time on the breast cancer data set. This is because when SVM and LFSVM are training, the input training set is the original dimension the number of features; RFSVM undergoes the dimension reduction and weighting of the Relief-F feature selection algorithm, and the calculation of the inner product of the kernel function is less. Therefore, fuzzy support vector machines based on Relief-F feature weighting have higher training efficiency.

For the entire function class, there is no universal best algorithm. For a certain algorithm, it is only suitable for its analysis task in a specific mode, and the average performance of all algorithms on the entire function class tends to be balanced. In this paper, the above three algorithms are experimented on multiple disease data sets (Table 2) in the UCI machine learning database, and the best adaptation mode of the algorithm application is investigated.

Table 2. UCI disease data set

Data set	Quantity	Dimension	Class
Breast Cancer	683	10	2
Lung Cancer	32	56	2
Heart Disease	303	14	2
Diabetes	768	20	2
Statlog(Heart)	270	13	2

The above data sets were applied to the three algorithms mentioned previously (SVM, LFSVM and RFSVM) to perform generalization performance experiments in the Matlab 2016a environment. The results are shown in Figure 3.

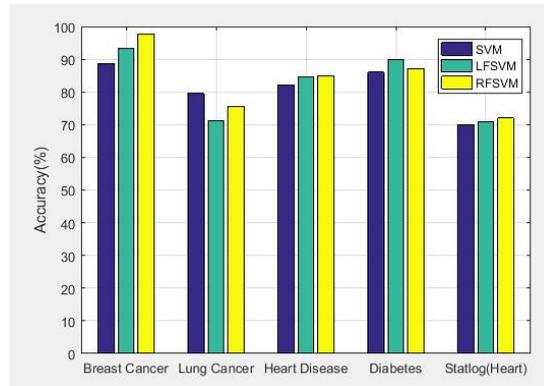


Figure 2. UCI disease data set comparison experiment

It can be seen from Figure 2 that on these five disease data sets, the average recognition performance of the RFSVM algorithm proposed in this paper is relatively highest, the recognition performance of the LSVM algorithm is second, and the recognition performance of the classic SVM algorithm is the worst. However, on the Lung cancer data set, the recognition effect of the RFSVM algorithm is not ideal. The recognition accuracy rate is reduced by about 3.5% compared with the classic SVM algorithm, and the recognition accuracy rate of the LFSVM algorithm is also reduced by about 7.2% compared to the classic SVM algorithm. This is because the sample size of the Lung cancer disease data set is too small, and the training set sample size of the RFSVM algorithm is relatively small, resulting in insufficient learning accuracy of the classification model, which reduces the recognition performance of the algorithm. This also shows that a certain algorithm is only suitable for its analysis task in a specific mode.

But overall, the recognition accuracy of the RFSVM algorithm proposed in this paper on all disease data sets is relatively higher than that of the other two algorithms, and the performance is improved. Has the highest accuracy rate and the most obvious effect. The correct rate in disease detection and diagnosis is relatively higher, and the classification performance is relatively more advantageous.

5. Conclusions

This paper analyzes the improvement of the support vector machine algorithm and the advantages and disadvantages of solving complex problems in the medical information field, and proposes a fuzzy support vector machine based on Relief-F feature weighting. Combine the Relief-F feature selection algorithm to calculate feature weights, calculate the weighted Euclidean distance from the sample to the center of the class based on the calculated feature weights, and then realize the calculation of membership and effective classification, improve the performance of the support vector machine algorithm in classification recognition, Moreover, the feasibility and effectiveness of applying data mining algorithms to medical diagnosis to assist doctors in improving the accuracy of disease analysis and diagnosis are verified. The RFSVM algorithm has certain theoretical significance and practical value in the application of early diagnosis of cancer.

Author Contributions: All authors have equally contributed to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National natural science foundation of China (fund no. 11472297).

Acknowledgments: The authors would like to thank Tu Daoxin, Kamrul Hashmi, Lai Jianqiong and others for guiding the experimental work in the laboratory and making great contributions to the experiment and writing of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

- [1] Vapnik V. (1995). *The Nature of Statistical Learning Theory*. New York: SpringerVerlag, 1995: 91-188.
- [2] C. Burges. (1998). "A tutorial on support vector machines for pattern recognition [J]", *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121-167.
- [3] Zhang Xuegong. (2006). *On Statistical Learning Theory and Support Vector Machines [J]*. *Acta Automatica Sinica*, 2000, 26 (1): 32-41.
- [4] Alexandre R. Mello, Marcelo R. Stemmer, Alessandro L. Koerich. (2020). *Incremental and decremental fuzzy bounded twin support vector machine[J]*. Elsevier Inc., p. 526.

- [5] Tahereh Bahraini, Saeedeh Ghazi, Hadi Sadoghi Yazdi. (2020). Toward optimum fuzzy support vector machines using error distribution [J]. Elsevier Ltd, p. 90.
- [6] LIN Chun-fu, WANG Sheng-de. (2002). Fuzzy support vector machines [J]. IEEE Transactions on Neural Networks, 3(2): 464-471.
- [7] Zhang Xiang, Xiao Xiao-ling, and Xu Guang-you. (2006). Fuzzy support vector machines based on affinity among samples. Journal of Software, 17(5): 951-958.
- [8] Fan. X, Du. S, Wu. T. (2003). "Weighted Support Vector Machine Based Classification Algorithm for Uneven Class Size Problems". Chinese Journal of Image and Graphics. Vol 8, No.9, pp. 1037-1042.
- [9] Li Xiaohui, Zhu Li, Wu Jian. (2017). Target recognition of millimeter wave radiometer based on fuzzy support vector machine [J]. Chinese Journal of Microwaves, 33 (S1): 203-206.
- [10] Liu Xingchang. (2019). Research on Xixia Character Recognition Based on Improved Fuzzy Support Vector Machine [D]. Lanzhou Jiaotong University.
- [11] Zhao Hui and Rong Li-li. (2005). Classification algorithm based on combined support vector machines and its application in text categorization. Mini-Micro Systems, 26(10): 1816-1820.
- [12] Tseng Yi-Ju, Huang Chuan-En, Wen Chiao-Ni, et al. (2019). Predicting breast cancer metastasis by using serum biomarkers and clinicopathological data with machine learning technologies. [J]. International journal of medical informatics.
- [13] Kho Esther, de Boer Lisanne L, Van de Vijver Koen, et al. (2019). Hyperspectral imaging for resection margin assessment during cancer surgery. [J]. Clinical Cancer Research : An Official Journal of the American Association for Cancer Research.
- [14] Kira K, Rendell L. (1992). A practical approach to feature selection [C]//Proceedings of the 9th International Workshop on Machine Learning. San Francisco: Morgan Kaufmann, 1992: 249-256.
- [15] Kononenko I. (1994). Estimating at tributes: Analysis and extensions of RELIEF [J]. Lecture Notes in Computer Science. 784:171-182.