

# Incremental Continuous Ant Colony Optimization Technique for Support Vector Machine Model Selection Problem

HIBA BASIM ALWAN, KU RUHANA KU-MAHAMUD

School of Computing, College of Art and Sciences

University Utara Malaysia

06010 Sintok, Kedah

MALAYSIA

hiba81basim@yahoo.com

ruhana@uum.edu.my

*Abstract:* - Ant Colony Optimization has been used to solve Support Vector Machine model selection problem. Ant Colony Optimization originally deals with discrete optimization problem. In applying Ant Colony Optimization for optimizing Support Vector Machine parameters which are continuous variables, there is a need to discretize the continuously value into discrete value. This discretize process would result in loss of some information and hence affect the classification accuracy and seeking time. This study proposes an algorithm that can optimize Support Vector Machine parameters using Incremental Continuous Ant Colony Optimization without the need to discretize continuous value for support vector machine parameters. Seven datasets from UCI were used to evaluate the credibility of the proposed hybrid algorithm in terms of classification accuracy. Promising results were obtained when compared to grid search technique.

*Key-Words:* - Pattern Classification, Support Vector Machine, Continuous Ant Colony Optimization, Incremental Continuous Ant Colony Optimization, model selection, Probability Density Function

## 1 Introduction

Support Vector Machine (SVM) is an excellent classifier built on statistical learning approach [1]. The main concept of SVM is to obtain the Optimal Separating Hyperplane (OSH) between positive and negative samples. This can be done through maximizing the margin between two parallel hyperplanes. Once the plane is found, SVM can forecast the classification of unlabeled sample through identifying the side where the sample lies [2]. Tuning SVM parameters and selecting optimal feature subset for SVM classifier are two problems that influence the classification accuracy. These problems affect each other [3].

There is no regular methodology that accepts advance approximation of optimal values for SVM parameter. Currently, almost all SVM researches choose these variables experimentally via searching a bounded number of values and preserving those that supplies the lowest amount of errors. This approach requires a grid search through the area of variable values and requires identifying the range of executable solution and best sampling step. This is a tricky task because best sampling step changes from kernel to kernel and grid range may not be simple to identify without advance knowledge of the problem. Furthermore, when hyper parameter exceeds two of the manual prototype chosen, it may become intractable [4]. Approaches such as trial and error, grid search,

cross validation, generalization error estimation and gradient descent can be used to find optimal parameter values for SVM. Evolutionary approaches such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO) may also be utilized.

ACO algorithms have been applied to tune SVM parameters. These algorithms work through repetitive creation procedures where each procedure directs a dependent heuristic by intelligently mixed various ideas for exploring and exploiting the seek space. The learning fashions are utilized to construct information to efficiently obtain near optimal solutions. ACO algorithms deal with discrete and continuous variables. ACO that deal with continuous variables is considered as a recent research field [5].

Ant Colony Optimization for continuous variable (ACO<sub>R</sub>) uses probability density function instead of discrete probability distribution to determine the direction that an ant should follow. The main disadvantages of ACO<sub>R</sub> are the high running time and the application gap with the-state-of-art continuous solvers. To overcome these disadvantages, the Incremental Continuous Ant Colony Optimization (IACO<sub>R</sub>) Algorithm is introduced. These enhancements are as follow: the size of solution archive in IACO<sub>R</sub> begins with an initial size and increases over time until it reaches a maximum size. A variable *Growth* monitors the

percentage at which the archive grows. This algorithm also alternate from original  $ACO_R$  in mechanism for choosing the solution that directs the creation of new solutions. The new mechanism counts on a variable  $\rho \in [0, 1]$ , which monitors the probability of utilizing just the best solution in the archive as a directing solution. In order to avoid stagnation,  $IACO_R$  involve an algorithm-level diversification mechanism. The mechanism includes in restarting the algorithm and initializing the new initial archive with the best-so-far solution. The restart condition is the number of succeeding iterations, *MaxStagIter*, with a relate solution enhance lower than a certain threshold [6].

In this study,  $IACO_R$  is used to solve SVM model selection problem. The rest of the paper is organized as follows. Section 2 reviews several literatures on tuning SVM parameters and Section 3 describes the proposed hybrid algorithm. Experimental results are discussed in Section 4 while concluding remarks and future works are presented in Section 5.

## 2 Solving Model Selection Problem for Support Vector Machine

Kapp, Sabourin & Maupin [7] utilized PSO and grid search in a dynamic environment to optimize SVM parameters. The authors examined their approach on fourteen datasets and compared their work with various approaches. The results show that their approach outperforms the classical methods in terms of model complexity and calculation time. Qiu, Li, Zhang & Gu [8] proposed Bare Bones Differential Evolution (BBDE) to optimize SVM parameters. It deletes the monitor variables of PSO and substitutes the static DE monitor variables with dynamically alternating variables to generate a general parameter-free, self-adaptive, optimization algorithm. Four UCI datasets were used to test the performance of the proposed method and results were compared with grid algorithm. Results showed that BBDE-SVM and DE-SVM take shorter time to compute and produced higher classification accuracy. This implies that the parameters chosen by BBDE and DE are better than grid search. A hybrid method based on support vector machine and simulated annealing (SVM-SA) [9] has been used to diagnose hepatitis. SA was used to find the optimal value for SVM parameters. Promising results were obtained and the authors suggested employing some feature selection techniques and other learning approach to maximize the precision of their approach.

A hybridized algorithm between GA and SVM to tune its parameters which are the regularization parameter,  $C$  and RBF kernel function parameter has been proposed by Samadzadegan, Soleymani, & Abbaspour [10]. These two parameters were encoded as a real value chromosomes and the fitness value for each chromosome were computed in terms of chromosome's classification accuracy. The authors show that the hybridization between GA and SVM can improve the classification accuracy and convergence speed. The authors showed that hybridization between GA and SVM can improve the classification accuracy and convergence speed. A study by Zhang, Chen & He [11] was on the use of ACO and grid search to solve SVM model selection problem. The author presented a novel ACO pheromone model and divided the ranges of RBF and  $C$  parameter into a number of grid and let the ants select the best combination of RBF and  $C$  parameters. The proposed approach was feasible and efficient to optimize the SVM parameter and produce hopeful results in term of classification accuracy and calculation time. ACO was also used by Fang & Bai [12] to optimize both SVM parameters and kernel function parameter in continuous field. Both parameters were divided into a number of sub intervals. In each sub interval, one point is chosen unsystematically to be the location of artificial ants. Their results showed a very promising hybrid SVM model for forecasting share price in terms of accuracy and generalization ability. Grid-based ACO technique was introduced by Zhang, Chen, Zhang, & He [13] to select variables  $C$  and RBF kernel automatically for SVM. Their work provides high accuracy and less calculation time compared with other methods such as grid algorithm and cross validation approach. RBF kernel was utilized to enhance the accuracy of SVM.

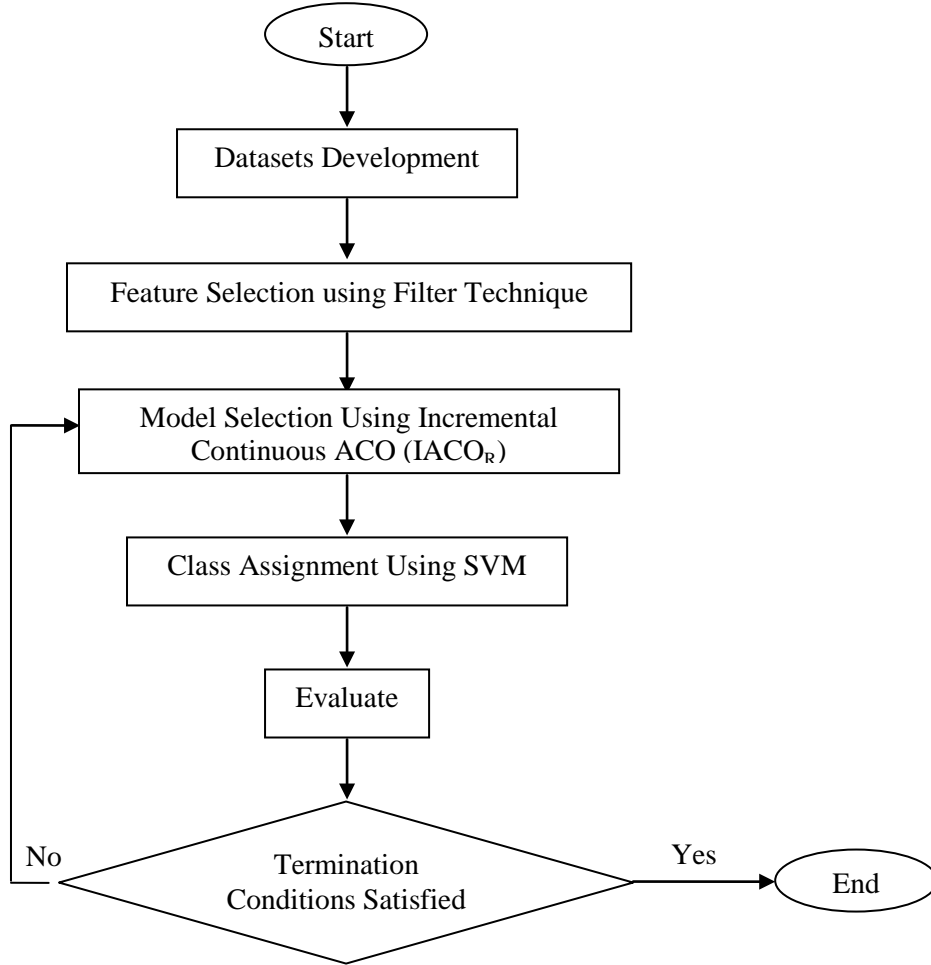
## 3 Proposed Hybrid Algorithm

The proposed algorithm has used the  $IACO_R$  to optimize SVM classifier parameters. An ant's solution is used to represent a combination of the classifier parameters,  $C$  and  $\gamma$ , based on the Radial Basis Function (RBF) kernel of the SVM classifier. The classification accuracy of the built SVM classifier is utilized to direct the updating of solution archives. Based on the solution archive, the transition probability is computed to choose a solution path for an ant. In implementing the proposed scheme, this study utilizes the RBF kernel function for SVM classifier because of its

capability to manage high dimensional data [14], good performance in major cases [15], and it only needs to use one parameter, which is kernel parameter gamma ( $\gamma$ ) [3]. The overall process for hybridize IACO<sub>R</sub> and SVM (IACO<sub>R</sub>-SVM) is as depicted in Figure 1. The main steps are (1) selecting feature subset (2) initializing solution archive and algorithm parameters, (3) solution construction for  $C$  and  $\gamma$ , (4) establishing SVM

classifier model, and (5) updating solution archives. In features subset selection step, F-score is used as a measurement to determine the feature importance. This measurement is used to judge the favouritism capability of a feature. High value of F-score indicates favourable feature. The calculation of F-score is as follow [16]:

$$F - Score_i = \frac{\sum_{c=1}^v (\bar{x}_i^{(c)} - \bar{x}_i)^2}{\sum_{c=1}^v \left( \frac{1}{N_i^{(c)} - 1} \sum_{j=1}^{N_i^{(c)}} (x_{ij}^{(c)} - \bar{x}_i^{(c)})^2 \right)} \quad (1)$$



**Figure 1: The Proposed Approaches' Flowchart**

where  $i = 1, 2, \dots, N_f$ ,  $v$  is the number of categories of target variable,  $N_f$  is the number of features,  $N_i^{(c)}$  is the number of samples of the  $i^{th}$  feature with categorical value  $c$ ,  $c \in \{1, 2, \dots, v\}$ ,  $\bar{x}_{i,j}^{(c)}$  is the  $j^{th}$  training sample for the  $i^{th}$  feature with categorical value  $c$ ,  $j \in \{1, 2, \dots, N_i^{(c)}\}$ ,  $\bar{x}_i$  is the  $i^{th}$  feature, and  $\bar{x}_i^{(c)}$  is the  $i^{th}$  feature with categorical value  $c$ . After computing the F-score for each feature in the dataset, average F-score will be computed and it will be considered as threshold for choosing feature in the feature subset. Feature with F-score equal to or greater to the threshold will be chosen and put in

the feature subset and this subset will be presented to SVM.

In the initialization step, each ant established a solution path for parameter  $C$  and parameter  $\gamma$ . Two solution archives are needed to design the transition probabilities for  $C$  and  $\gamma$ . The range for  $C$  and  $\gamma$  values will be sampled according to random parameter  $k$  which is the initial archive size of solutions archives. The weight vector,  $w$  is then computed for each sample for  $C$  and  $\gamma$  as follow:

$$w_1 = \frac{1}{qk\sqrt{2\pi}} e^{-\frac{(l-1)^2}{2q^2k^2}} \quad (2)$$

where  $q$  is the algorithm's parameter to control diversification of search process. These values will be stored in solution archives. Once this step is completed, the sampling procedure will be constructed in two phases. Phase one involves choosing one of the weight vectors as following:

$$p_l = \frac{w_l}{\sum_{r=1}^k w_r} \quad (3)$$

The second phase involves sampling selecting  $w$  via a random number generator that is able to generate random numbers according to a parameterized normal distribution. This initialization will construct the transition probabilities. Like the solution archives, some important system parameters must be initialized as follows: the number of ants = 2,  $q = 0.1$ , *initial archive size* = 10, *Growth* = 5, *maximum archive size* = 15, *MaxStagIter* = 2, number of runs = 10,  $C$  range  $\in [2^{-1}, 2^{12}]$  and  $\gamma \in [2^{-12}, 2^2]$ .

The third step is related to solution construction where each ant builds its own solution. This solution will be a combination of  $C$  and  $\gamma$ . In order to construct the solution, two transition probabilities with various solutions archives are needed. These transitions will be computed according to Eq. (2) and Eq. (3).

Classifier model will be constructed in step four. Solution is generated by each ant and will be evaluated based on classification accuracy obtained by SVM model utilizing  $k$ -fold Cross Validation (CV) with the training set. In  $k$ -fold CV, training data group is partitioned into  $k$  subgroups, and the holdout approach is repeated  $k$  times. One of the  $k$  sub-groups is utilized as the test set and the remaining  $k-1$  subgroups are combined to construct the training group. The average errors along with all the  $k$  trails are calculated. CV accuracy is calculated as follows:

$$CV_{accuracy} = \frac{\sum_i test\_accuracy}{k}, i = 1, 2, \dots, k \quad (4)$$

Test accuracy is used to evaluate the percentage of samples that are classified in the right way to determine  $k$ -folds and it will be computed as follows:

$$Test\ Accuracy = \frac{\text{no.of correctly predicted data}}{\text{total testing data}} * 100\% \quad (5)$$

The benefits of using CV are (1) each of the test groups are independent and (2) the dependent outcomes can be enhanced [16].

The final step is to update the solution archives. This modification will be done by appending the newly generated group solutions that gave the best classification accuracy to solution archive and then deleting the exact number of worst solutions. This will ensure the size of solution archive does not change. This procedure guaranteed

that just good solutions are stored in the archive and it will efficiently influence the ants in the seek process.

## 4 EXPERIMENTAL RESULTS

Seven datasets were used in evaluating the proposed IACO<sub>R</sub>-SVM algorithm. The datasets are Australian, Pima-Indian Diabetes, Heart, Ionosphere, German, Sonar, Splice datasets, available from UCI Repository of Machine Learning Databases [17]. The summary of these datasets are presented in Table 1.

All input variables were scaled during data pre-processing phase to avoid features with higher numerical ranges from dominating those in lower numerical ranges and also to reduce the computational effort. The following formula was used to linearly scale each feature to [0, 1] range.

$$\bar{x} = \frac{x - \min_i}{\max_i - \min_i} \quad (6)$$

where  $x$  is the original value,  $\bar{x}$  is the scaled value, and  $\max_i$  and  $\min_i$  are the maximum and minimum values of feature  $i$ , respectively [17].

Table 1 Summarization of UCI's Datasets Repository

Dataset	No. of Instances	No. of Features	Type of Datasets
Australian Credit Approval (Statlog)	690	14	Categorical, Integer, Real
Pima-Indian Diabetes	760	8	Integer, Real
Heart (Statlog)	270	13	Categorical, Real
Ionosphere	351	34	Integer, Real
German Credit (Statlog)	1000	24	Categorical, Integer
Sonar	208	60	Real
Splice	1000	60	Categorical

Each dataset is randomly re-arranged and divided into approximately equal size subsets, one subset as testing set and the remaining as training sets. The performance of the proposed IACO<sub>R</sub>-SVM was compared with grid search approach [17] and [18] which is considered as the basic approach to optimize SVM parameters.

C programming language has been used to implement IACO<sub>R</sub>-SVM. Experiments were performed on Intel(R) Core (TM) 2 Duo CPU T5750, running at 2.00 GHz with 4.00 GB RAM and 32-bit operating system. Table 2 shows the optimal values for  $C$  and  $\gamma$  that have been produced by the proposed algorithm as well as  $p$  value which

is generated randomly and used to monitor the probability of utilizing just best solution in the archive as a directing solution.

Table 2 Optimal Value for C and  $\gamma$

Dataset	C	$\gamma$	p
Australian Credit Approval (Statlog)	2863.50	2.80	0.50
Pima-Indian Diabetes	821.83	1.15	0.33
Heart (Statlog)	3277.50	3.20	0.38
Ionosphere	2536.30	2.48	0.87
German Credit (Statlog)	205.00	0.20	0.02
Sonar	0.50	0.000244	0.91
Splice	820.74	1.15	0.76

The C and  $\gamma$  values in Table 2 have been used to produce the classification accuracy. Table 3 depicts the average number of selected features and the pattern classification accuracy of the proposed IACO<sub>R</sub>-SVM algorithm together with grid search results [17] and [18]. The proposed approach classifies patterns with higher accuracy compared to grid search for all seven datasets. The average increase in accuracy for all datasets is about 8.75%. The increased in accuracy was possible because IACO<sub>R</sub> optimized its parameters which are the regularization parameter C and gamma ( $\gamma$ ) of RBF kernel function.

Table 3 Experiment Results of the Proposed IACO<sub>R</sub>-SVM and Grid Search

Dataset	Number of Features	ACO <sub>R</sub> -SVM		Grid Search
		Classification accuracy (%)	Average number of selected features	Classification accuracy (%)
Australian Credit Approval (Statlog)	14	97.17	3.2	84.7
Pima-Indian Diabetes	8	89.06	2.3	77.3
Heart (Statlog)	13	90.30	5.7	83.7
Ionosphere	34	90.17	8.6	89.44
German Credit (Statlog)	24	94.00	5.6	76
Sonar	60	89.99	20.1	87
Splice	60	98.04	13.2	91.31

Table 4 shows the best features that were chosen by filter F-score technique to generated features subset to be introduced to SVM. All features displayed in this Table are important based on their threshold values. The reason for using filter F-score technique

to select features subset is because RBF will fail for large number of features [19].. From Table 4 it can be seen that the biggest reduction in number of features is 71% for Australian dataset while the smallest number of feature reduction is 35% for Ionosphere dataset.

Table 4 Frequencies of Selected Features

Data																							
Australian																							
Feature#	5	7	8	9																			
Frequencies	10	1	10	10																			
Pima-Indian Diabetes																							
Feature#	2	5	6	8																			
Frequencies	10	1	1	9																			
Heart																							
Feature#	3	8	9	10	12	13																	
Frequencies	7	10	10	10	10	10																	
Ionosphere																							
Feature#	3	4	5	6	7	8	9	10	11	12	13	14	15	16	21	22	23	25	27	29	31	33	
Frequencies	10	1	10	2	10	6	9	1	1	2	2	5	2	1	1	1	1	1	7	7	10	10	
German																							
Feature#	1	2	3	5	6	7	12																
Frequencies	10	10	10	9	9	2	6																
Sonar																							
Feature#	1	2	3	4	5	8	9	10	11	12	13	20	21	22	31	35	36	37	43	44	45	46	47
Frequencies	10	9	5	10	5	1	10	10	10	10	8	4	6	2	1	3	9	3	5	7	10	10	10
Feature#	48	49	50	51	52	53	54	58	59														
Frequencies	10	10	4	10	10	3	5	5	1														
Splice																							
Feature#	15	16	17	18	20	21	22	23	26	29	30	32	33	34	45	51	54	60					
Frequencies	2	5	5	5	5	1	5	1	5	5	4	5	5	5	3	5	3	1					

## 5 Conclusions

This study has investigated a hybrid IACO<sub>R</sub> and SVM technique to obtain optimal model parameters. Experimental results on seven public UCI datasets showed promising performance in terms of classification accuracy. Possible extensions can focus on the area where IACO<sub>R</sub>-SVM can simultaneously optimize both SVM parameters and features subset using mixed-variable Incremental ACO (IACO<sub>R-MV</sub>). Other kernel parameters besides RBF, application to other SVM variants and multiclass data can also be considered as possible future work in this area.

## ACKNOWLEDGEMENTS

The authors wish to thank the Ministry of Higher Education Malaysia for funding this study under Fundamental Research Grant Scheme, S/O code 12377 and RIMC, Universiti Utara Malaysia, Kedah for the administration of this study.

## References

- [1] Liu, W., & Zhang, D., Feature subset selection based on improved discrete particle swarm and support vector machine algorithm. *Paper presented at the IEEE International Conference of the Information Engineering and Computer Science, (2009)*, (pp.1-4).
- [2] Vapnik, V., & Vashist, A., A new learning paradigm: learning using privileged information. *Neural networks: the official journal of the International Neural Network Society*, Vol. 22, 2009, No.5-6, pp.544-557.
- [3] Huang, C., & Wang, C., A GA-based feature selection and parameters optimization for support vector machines. *Expert system with Applications*, Vol.31, 2006, No.2, pp.231-240.
- [4] Ayat, N., Cheriet, M., & Suen, C., Automatic model selection for the optimization of SVM kernels. *Pattern Recognition*, Vol.38, 2005, No.10, pp.1733-1745.
- [5] Socha, K. & Dorigo, M., Ant colony optimization for continuous domain. *European Journal of Operational Research*, Vol.185, 2008, No.3, pp.1155-1173.
- [6] Liao, T., Dorigo, M., & Stutzle, T. (2011). Improved ant colony optimization algorithms for continuous and mixed discrete-continuous optimization problems. Retrieved from <http://http://www.swarmbots.org/~mdorigo/HomePageDorigo/thesis/dea/LiaoMAS.pdf>.
- [7] Kapp, M. N., Sabourin, R., & Maupin, P., A dynamic model selection strategy for support vector machine classifiers. *Applied Soft Computing*, Vol.12, 2012, No.8, pp.2550-2565.
- [8] Qiu, D., Li, Y., Zhang, X., & Gu, B., Support vector machine with parameter optimization by bare bones differential evolution. *Paper presented at the 7<sup>th</sup> IEEE International Conference of the Natural Computation ICNC*, (pp. 263-266), 2011.
- [9] Sartakhti, J. S., Zangoeei, M. H., & Mozafari, K., Hepatitis disease diagnosis using a novel hybrid method based on support vector machine and simulated annealing (SVM-SA). *Computer Methods and Programs in Biomedicine*, Vol. 108, 2012, No. 2, pp.570-579.
- [10] Samadzadegan, F. Soleymani, A., & Abbaspour, A., Evaluation of genetic algorithms for tuning SVM parameters in multi-class problems. *Paper presented at 11<sup>th</sup> IEEE International Symposium on Computational Intelligence and Informatic, 2010*, (pp. 323-328). Budapest, Hungary.
- [11] Zhang, X., Chen, X., & He, Z., An ACO-based algorithm for parameter optimization of support vector machines. *Expert System with Applications*, Vol.37, 2010, No.9, pp.6618-6628.
- [12] Fang, X., & Bai, T., Share Price Prediction Using Wavelet Transform and Ant Colony Algorithm for Parameters Optimization in SVM. *Paper presented at the Intelligent Systems, Global Congress on Intelligent System, 2009*, (pp.288-292).
- [13] Zhang, X. L., Chen, X. F., Zhang, Z. S., & He, Z. J., A grid-based ACO algorithm for parameters optimization in support vector machines. *Paper presented at IEEE International Conference of the Granular Computing GrC 2008*, (pp.805-808).
- [14] Moustakidis, S. & Theocharis, J., SVM-FuzCoC: A novel SVM-based feature selection method using a fuzzy complementary criterion. *Pattern Recognition*, Vol.43, 2010, No.11, pp.3712-3729.
- [15] Zhang, H., Xiang, M., Ma, C., Huang, Q., Li, W., Xie, W., Wei, Y., Yang, S., Three-class classification models of LogS and LogP derived by using GA-CG-SVM approach. *Molecular Diversity*, Vol.13, 2009, No.2, pp.261-268.
- [16] Huang C., ACO-based hybrid classification system with feature subset selection and model parameters optimization. *Neurocomputing*, Vol.73, 2009, No.1-3, pp.438-448.
- [17] *UCI Repository of machine learning databases, Department of Information and Computer Science, University of California, Irvine, CA, <http://www.ics.uci.edu/mllearn/MLRepository>, 2012.*
- [18] Ding S. & Li S. *Clonal Selection Algorithm for Feature Selection and Parameters Optimization of Support Vector Machines. Paper presented at 2<sup>nd</sup> International Symposium IEEE on knowledge acquisition and modeling 2009*, (pp. 17-20).