Incremental continuous ant colony optimization For tuning support vector machine's parameters

Hiba Basim Alwan Al-Dulaimi and Ku Ruhana Ku-Mahamud

Abstract—Support Vector Machines are considered to be excellent patterns classification techniques. The process of classifying a pattern with high classification accuracy counts mainly on tuning Support Vector Machine parameters which are the generalization error parameter and the kernel function parameter. Tuning these parameters is a complex process and Ant Colony Optimization can be used to overcome the difficulty. Ant Colony Optimization originally deals with discrete optimization problems. Hence, in applying Ant Colony Optimization for optimizing Support Vector Machine parameters, which are continuous in nature, the values wil have to be discretized. The discretization process will result in loss of some information and, hence, affects the classification accuracy and seeks time. This paper presents an algorithm to optimize Support Vector Machine parameters using Incremental continuous Ant Colony Optimization without the need to discretize continuous values. Eight datasets from UCI were used to evaluate the performance of the proposed algorithm. The proposed algorithm demonstrates the credibility in terms of classification accuracy when compared to grid search techniques, GA$_{with\ feature\ chromosome}$-SVM, PSO-SVM, and GA-SVM. Experimental results of the proposed algorithm also show promising performance in terms of classification accuracy and size of features subset.

Keywords—Incremental continuous ant colony optimization, model selection, pattern classification, probability density function, support vector machine.

I. INTRODUCTION

Pattern classification is an important area in machine learning and artificial intelligence. It attaches the input samples into one of a present number of groups through an approach. The approach is found through learning the training data group [1].

Support Vector Machine (SVM) represents supervised machine learning approaches [2] and it is an excellent classifier built on statistical learning approach [3]. 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 [4]. Tuning SVM parameters and selecting optimal feature subset for SVM classifier are two problems that influence the classification accuracy. The problems affect each other [5] and [6].

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 [7] and [8]. 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 works through repetitive creation procedures where each procedure directs a dependent heuristic by intelligently mixed various ideas for exploring and exploiting the seek space [9]. The word heuristic comes from Greek and means to know, to

Hiba Basim Alwan is with the School of Computing, College of Art and Science, University Utara Malaysia, 06010 Sintok, Kedah, Malaysia, (e-mail: hiba81basim@yahoo.com).

Ku Ruhana Ku-Mahamud is a Professor with the School of Computing, College of Art and Science, University Utara Malaysia, 06010 Sintok, Kedah, Malaysia, (e-mail: ruhana@uum.edu.my).

find, to discover, or to guide an investigation [10]. 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 [9].

Ant Colony Optimization for continuous variable (ACO$_R$) uses probability density function instead of discrete probability distribution to determine the direction that an ant should follow. High running time and the application gap with the-state-of-art continuous solvers are the main disadvantages of ACO$_R$. To overcome the disadvantages, the Incremental Continuous Ant Colony Optimization (IACO$_R$) Algorithm is introduced [11].

In this study, IACO$_R$ is used to solve SVM model selection problem. The rest of the paper is organized as follows. Section II reviews several literatures on tuning SVM parameters and Section III describes the proposed hybrid algorithm. Findings are discussed in Section IV while concluding remarks and future works are presented in Section V.

II. SUPPORT VECTOR MACHINE

For binary class classification problem, given M training examples where each example is represented through a tuple (x_i, y_i) where $i = 1, ..., M$, $x_i \in R^M$ corresponds to the feature group for the ith example, and $y_i \in \{+1, -1\}$ denoted the class label, SVM need the solution of the following problem [12] and [13]:

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^{M} \xi_i \tag{1}$$

Subject to $\quad y_i(w^T \phi(x_i) + b) + \xi_i - 1, \xi_i \geq 0, i = 1, ..., M$

where C represent the penalty of misclassifying the training instances, ξ_i is positive slack parameter, and ϕ is the mapping function that maps the training instances from input space to higher dimensional feature space. Often Eq. (1) solved through solving the following dual problem:

$$\min_{\alpha} F(\alpha) = \frac{1}{2} \alpha^T Q \alpha - e^T \alpha \tag{2}$$

Subject to $\quad 0 \leq \alpha_i \leq C, i = 1, ..., M$
$\qquad\qquad y^T \alpha = 0$

where e is the vector of all ones and Q is an M by M positive semi-definite matrix. The $(i, j)^{th}$ element of Q is given through:

$$Q_{i,j} = y_i y_j K(x_i, x_j) \tag{3}$$

$$K(x_i, x_j) \equiv \phi^T(x_i)\phi(x_j) \tag{4}$$

where $K(x_i, x_j)$ is the kernel function, Radial Basis Function (RBF) is a common one and as follow:

$$K(x_i, x) = exp^{(-\gamma\|x_i - x\|^2)} \tag{5}$$

$\{\alpha_i\}_{i=1}^{i}$ is Lagrange multipliers, and

$$w = \sum_{i=1}^{M} \alpha_i y_i \phi(x_i) \tag{6}$$

is the weight vector. The classification decision function is:

$$sgn(w^T \phi(x) + b) = sgn(\sum_{i=1}^{M} \alpha_i y_i K(x_i, x) + b) \tag{7}$$

There are two main techniques for multi-class classification problem. The techniques are One-Against-One (OAO) and One-Against-All (OAA). OAO technique, numbers of binary SVM classifier are built using the following equation:

$$C_2^v = \frac{v(v-1)}{2} \tag{8}$$

SVM binary classifiers are built and each classifier is trained utilizing two classes. A new x is classified into the majority class that is voted on through all of the decision functions. While in OAA technique, v binary SVM decision functions are built for an v-class problem. The jth $(j = 1, 2,..., v)$ decision function is trained through labeling all of the examples in the jth class with positive labels, and all of the examples that are not in the jth class with negative labels. A new x is classified into the class that has the largest decision function [12].

III. INCREMENTAL CONTINUOUS ANT COLONY OPTIMIZATION

According to [11] Incremental continuous Ant Colony Optimization (IACO$_R$) start with a small archives size, a parameter *InitArhiveSize* defines it size. IACO$_R$ starts with randomly initializing solution archive. This solution archive will be filled with solutions initially generated randomly. IACO$_R$ also characterizes a strategy alternate from the one utilized in ACO$_R$ for choosing the solution that directs the creation of new solutions. The new procedure build on parameter $p \in [0, 1]$, which monitors the probability of utilizing just the best solution in the archive as a directing solution. With a probability $1 - p$, all the solutions in the archive are utilized to create new solutions. Once a directing solution is chosen, and a new one is created exactly the same way as in ACO$_R$, they are compared according to their objective function. If the newly created solution is better than the directing solution, it replaced it in the archive. This replacement mechanism is alternate from the one utilized in ACO$_R$ in which all solutions in the archive and all the newly created ones compete. A new solution is appended to them every growth iterations till a maximum archives sizes, defines by *MaxArchiveSize*, is reached. A parameter *Growth* monitor the percentage at which the archives grows. Fast growth percentage support seeks diversification while slow growth support intensification. Each time a new solution is appended, it is initialized utilizing information from best solution in the archives. First, a new solution S_{new} is created fully in an arbitrary way, and then it is moved

in direction of the best solution in the archives S_{best} utilizing the following formula:

$$\bar{S}_{new} = S_{new} + rand(0,1)(S_{best} - S_{new}) \qquad (9)$$

where *rand(0, 1)* is an arbitrary umber in range [0, 1].

IACO$_R$ involve an algorithm-level diversification strategy for fighting stagnation. The strategy includes in restarting the algorithm and initializing the new initial archive with the best-so-far solution. The restart condition is the number of successive iterations, *MaxStagIter*, with a relative solution improvement lower than a certain threshold. An outline of IACO$_R$ is given below:

Input: *p, InitArhiveSize, Growth, MaxArchiveSize, MaxStagIter*, no. of ants, and Termination criterion
Output: Optimal Value for *C* and *γ*
k = InitArhiveSize
initialize *k* solutions and evaluate it
while Termination criterion not satisfied do
    // Generate new solutions
    if *rand (0,1) < p* then
        for *i* = 1 to no. of ants do
            Select best solution
            Sample best selected solution
            if Newly generated solution is better than S_{best} then
                Substitute newly generated solution for S_{best}
            end
        end
    else
        for *j* = 1 to *k* do
            Select *S* according to its weight
            Sample selected *S*
            Store and evaluate newly generate solutions
            if Newly generated solution is better than S_j then
                Substitute newly generated solution for S_j
            end
        end
    end
    // Archive Growth
    if current iterations are multiple of *Growth* & *k < MaxArchiveSize* then
        Initialize new solution using Eq. (4.4)
        Add new solution to the archive
        k + +
    end
    // Restart Mechanism
    if # (number) of iterations without improving S_{best} = *MaxStagIter* then
        Re-initialize *T* (solution archive) but keeping S_{best}
    end
End

IV. SOLVING MODEL SELECTION PROBLEM FOR SUPPORT VECTOR MACHINE

Kapp, Sabourin & Maupin [14] 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 [15] 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) [16] 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 [17]. 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 [18] was on the use of ACO and grid search to solve SVM model selection problem. The authors presented a novel ACO pheromone model and divided the ranges of RBF and *C* parameter into a number of grids 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 [19] 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 [20] 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.

V. PROPOSED HYBRID ALGORIITHM

The proposed algorithm has employed the IACO$_R$ to optimize SVM classifier parameters. An ant's solution is used to represent a combination of the classifier parameters, *C* and γ, 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 [21], good performance in major cases [22], and it only needs to use one parameter, which is kernel parameter gamma (γ) [5]. The algorithm for the proposed IACO$_R$-SVM is as follow:

Input: *p*, *InitArhiveSize*, *Growth*, *MaxArchiveSize*, *MaxStagIter*, no. of ants, and Termination criterion
Output: Optimal Value for *C* and *γ*
k = InitArhiveSize
initialize *k* solutions
call SVM algorithm to evaluate *k* solutions
while Termination criterion not satisfied do
   // Generate new solutions
   if *rand (0,1) < p* then
      for *i = 1* to no. of ants do
         Select best solution
         Sample best selected solution
         if Newly generated solution is better than S_{best}
         then
            Substitute newly generated solution for S_{best}
         end
      end
   else
      for *j = 1* to *k* do
         Select *S* according to its weight
         Sample selected *S*
         Store and evaluate newly generated solutions

         if Newly generated solution is better than S_j
         then
            Substitute newly generated solution for S_j
         end
      end
   end
   // Archive Growth
   if current iterations are multiple of *Growth* & *k < MaxArchiveSize* then
      Initialize new solution using Eq. (4.4)
      Add new solution to the archive
      k + +
   end
   // Restart Mechanism
   if # (number) of iterations without improving S_{best}
   = *MaxStagIter* then
      Re-initialize *T* (solution archive) but keeping S_{best}
   end
End

The overall process for hybridize IACO$_R$ and SVM (IACO$_R$-SVM) is as depicted in Fig. 1. The main steps are (1) selecting feature subset (2) initializing solution archive and algorithm parameters, (3) solution construction for *C* and γ, (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 [12]:

$$F-Score_i = \frac{\sum_{c-1}^{v}\left(\bar{x}_i^{(c)}-\bar{x}_i\right)^2}{\sum_{c-1}^{v}\left(\frac{1}{N_i^{(c)}-1}\sum_{j=1}^{N_i^{(c)}}\left(x_{i,j}^{(c)}-\bar{x}_i^{(c)}\right)^2\right)} \quad (10)$$

where $i = 1, 2, ..., N_f$, 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 ith feature with categorical value *c*, $c \in \{1, 2, ..., v\}$, $\bar{x}_{i,j}^{(c)}$ is the jth training sample for the ith feature with categorical value *c*, $j \in \{1, 2, ..., N_i^{(c)}\}$, $\bar{x}_i$ is the ith feature, and $\bar{x}_i^{(c)}$ is the ith 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 than the threshold will be chosen and put in the feature subset and this subset will be presented to SVM. It is important to minimize the number of features by choosing those features that includes the most necessary information that would contribute to high accuracy and deleting the remaining [23].

In the initialization step, each ant established a solution path for parameter *C* and parameter *γ*. Two

solution archives are needed to design the transition probabilities for C and γ. The range for C and γ 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 γ as follow:

$$w = e^{-\frac{(l-1)^2}{2q^2k^2}} \qquad (11)$$

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 follows:

$$p_l = \frac{w_l}{\sum_{r=1}^{k} w_r} \qquad (12)$$

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 $\gamma \in [2^{-12}, 2^2]$.
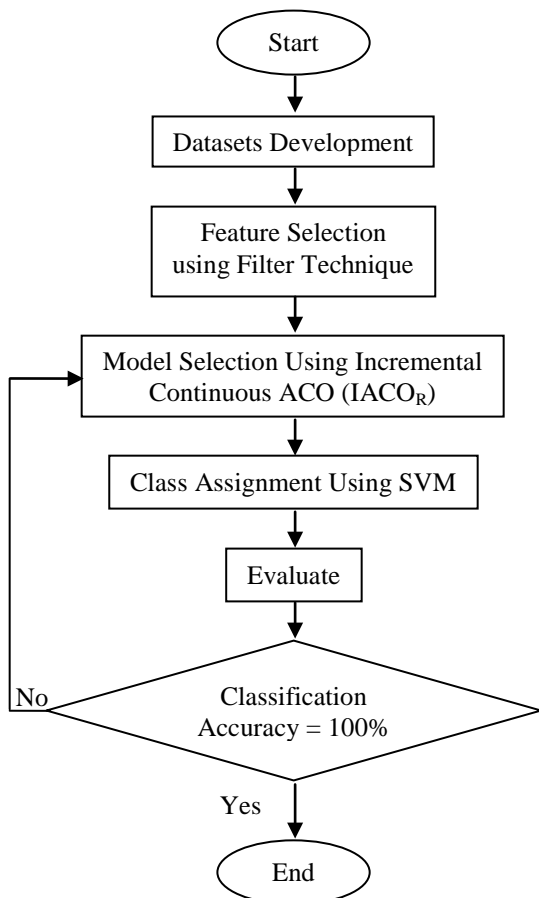

Figure 1. Hybrid IACO$_R$-SVM Algorithm

Each ant will build its own solution in the third step. This solution will be a combination of C and γ. In order to construct the solution, two transition probabilities with various solutions archives are required. The transitions will be computed as in Eq. (11) and Eq. (12).

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:

$$CV_{accuracy} = \frac{\sum_i test\_accuracy}{k}, i = 1, 2, ..., k \qquad (13)$$

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:

$$Test\ Accuracy = \frac{no.of\ correctly\ predicted\ data}{total\ testing\ data} * 100\% \qquad (14)$$

The benefits of using CV are (1) each of the test groups are independent and (2) the dependent outcomes can be enhanced [12].

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.

VI. FINDINGS

Eight datasets were used in evaluating the proposed IACO$_R$-SVM algorithm. The datasets are Australian, Pima-Indian Diabetes, Heart, Ionosphere, German, Sonar, Iris, and Vehicle datasets available from UCI Repository of Machine Learning Databases [24]. The summary of these datasets is presented in Table I. These datasets have been utilized as benchmarks to compare the execution of different classification approaches in the papers frequently.

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.

$$\bar{x} = \frac{x - min_i}{max_i - min_i} \qquad (15)$$

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 [12].

Table I Summary of UCI's Datasets Repository

Dataset	No. of Instances	No. of Features	Type of Datasets	No. of Classes
Australian	690	14	Categorical, Integer, Real	2
Diabetes	760	8	Integer, Real	2
Heart	270	13	Categorical, Real	2
Ionosphere	351	34	Integer, Real	2
German	1000	24	Categorical, Integer	2
Sonar	208	60	Real	2
Iris	150	4	Real	3
Vehicle	846	18	Integer	4

Each dataset was randomly re-arranged and divided into ten approximately equal sized subsets, one subset is a testing set and the remaining are training sets and repeated ten times to enable each fold of data to take a turn as the testing dataset. The classification accuracy of the experiment was calculated by summing the individual accuracies for each run of testing and then divided by 10. The performance of the proposed IACO$_R$-SVM was compared with the grid search approach which was considered as the basic approach to optimize SVM parameters without the ability to select features subset as well as with GA$_{with\ feature\ chromosome}$-SVM [25], ACO$_R$-SVM [8], PSO-SVM, and GA-SVM [26].

C programming language was used to implement IACO$_R$-SVM. Experiments were performed on Intel(R) Core (TM) 2 Duo CPU T5750, running at 2.00 GH$_Z$ with 4.00 GB RAM and 32-bit operating system.

Table II shows the optimal values for C and γ as well as p value which is generated randomly and use to monitor the probability of utilizing just best solution in the archive as a directing solution were produced by the proposed algorithm and these values were used to produce the classification accuracy depicted in Table III.

Table II Optimal Values for C, γ, and p

Dataset	C	γ	p
Australian	2454.5	2.400244	0.371
Pima-Indian Diabetes	409.5	0.400244	0.428
Heart	1227.5	1.200244	0.403
Ionosphere	2045.5	2.000244	0.459
German	818.5	0.800244	0.878
Sonar	3272.5	3.200244	0.517
Iris	1636.50	1.600244	0.699
Vehicle	1227.50	1.200244	0.903

The proposed approach classifies patterns with higher accuracy compared to grid search, GA$_{with\ feature\ chromosome}$-SVM [25], ACO$_R$-SVM [8], PSO-SVM, and GA-SVM [26] for eight datasets. This is because the integration of IACO$_R$ with SVM, IACO$_R$ as an optimization approach improves SVM classification accuracy through optimizing its parameters which are the regularization parameter C and gamma (γ) of RBF kernel function. The average number of selected features to classify pattern of the proposed ACO$_R$-SVM algorithm together with GA$_{with\ feature\ chromosome}$-SVM [25] is present in Table IV.

For each iteration, IACO$_R$ generates SVM parameters' values and introduces it to SVM and SVM uses these values to classify patterns. The proposed algorithm stops if the classification accuracy or maximum number of iteration satisfies user specification, otherwise, IACO$_R$ searches for other optimal values for SVM parameters to work with. Table V shows the best features chosen by filter F-score technique to generate features subsets 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 was because RBF would fail for large numbers of features [27]. Table V shows that the biggest reduction in number of features is 77.86% for the Australian dataset while the smallest feature reduction is 50% for the Iris dataset.

Table III Classification accuracy (%)

Dataset	$IACO_R$-SVM	ACO_R-SVM	Grid Search	$GA_{with\ feature\ chromosome}$-SVM	PSO-SVM	GA-SVM
Australian	94.62	96.14	84.47	86.81	88.09	88.09
Pima-Indian Diabetes	91.25	87.79	76.58	81.97	80.19	82.98
Heart	96.28	89.99	88.15	91.11	-	-
Ionosphere	98.90	89.87	94.29	98.57	97.50	96.61
German	91.56	94.00	78.90	80.80	79.00	84.24
Sonar	98.92	90.41	90.50	95.00	88.32	95.22
Iris	99.86	-	94.09	96.00	98.00	97.56
Vehicle	93.20	-	83.94	84.74	88.71	85.87

Table IV Number of selected features

Datasets	Number of features	$IACO_R$-SVM	$GA_{with\ feature\ chromosome}$-SVM
Australian	14	3.3	6.7
Pima-indian Diabetes	8	2.6	5.1
Heart	13	6.0	7.0
Ionosphere	34	10.6	15.4
German	24	6.4	11.8
Sonar	60	21.7	28.7
Iris	4	2	1.8
Vehicle	18	9.3	10.3

Table V Frequencies of selected features

Data																
Australian																
Feature#	5	7	8	9												
Frequencies	10	3	10	10												
Pima-Indian Diabetes																
Feature#	2	4	6	8												
Frequencies	5	2	1	5												
Heart																
Feature#	3	8	9	10	11	12	13									
Frequencies	7	10	9	10	4	10	10									
Ionosphere																
Feature#	3	4	5	6	7	8	9	11	12	13						
Frequencies	10	2	10	6	10	6	10	1	1	1						
Feature#	14	15	16	18	19	21	23	25	29	31						
Frequencies	5	4	1	1	1	5	4	1	8	9						
German																
Feature#	1	2	3	5	6	7	9	12	14							
Frequencies	10	10	10	10	10	3	1	7	3							
Sonar																
Feature#	1	2	3	4	5	8	9	10	11	12	13	20	21	22	31	35
Frequencies	10	9	6	9	5	1	10	10	10	10	10	4	9	3	1	5
Feature#	36	37	43	44	45	46	47	48	49	50	51	52	54	58	59	50
Frequencies	7	2	4	8	10	9	9	10	10	2	10	10	6	3	1	2
Iris																
Feature#	3	4														
Frequencies	10	10														
Vehicle																
Feature#	1	2	3	4	5	6	7	8	10	11	13	14	15	16		
Frequencies	10	3	10	2	9	1	6	10	9	1	5	5	9	3		

VII. CONCLUSIONS

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

REFERENCES

[1] Wang, X., Liu, X., Pedrycz, W., Zhu, X., & Hu, G. (2012). Mining axiomatic fuzzy association rules for classification problems. European Journal of Operational Research, *218*(1), 202-210.

[2] Barbu T., SVM-based human cell detection technique using histograms of oriented gradients. *Paper Presented at Proceedings the 17th WSEAS International Conference on Applied Mathematics (AMATH '12) 2012*. Montreux, Switzerland.

[3] 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).

[4] 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.

[5] 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.

[6] A. Hiba & M. Ku, Incremental continuous ant colony optimization technique for support vector machine model selection problem. *Paper Presented at Proceedings the 17th WSEAS International Conference on Applied Mathematics (AMATH '12) 2012*. Montreux, Switzerland.

[7] 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.

[8] A. Hiba & M. Ku, Optimizing support vector machine parameters using continuous ant colony optimization. *Paper Presented at Proceedings the 7th ICCCT International Conference on Computing and Convergence Technology (ICCIT, ICEI and ICACT) 2012*. Seoul, South Korea.

[9] Socha, K. & Dorigo, M., Ant colony optimization for continuous domain. *European Journal of Operational Research,* Vol.185, 2008, No.3, pp.1155-1173.

[10] M. Ku & A. Mustafa, New heuristic function in ant colony system for job scheduling in grid computing. *Paper Presented at Proceedings the 17th WSEAS International Conference on Applied Mathematics (AMATH '12) 2012*. Montreux, Switzerland.

[11] 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.

[12] Huang C., ACO-based hybrid classification system with feature subset selection and model parameters optimization. *Neuro computing,*Vol. 73, 2009, No.1-3, pp.438-448.

[13] R. Fan, P. Chen, C. Lin, "Working Set Selection Using Second Order Information for Training Support Vector Machines". Machine Learning Research, vol. 6 (2005), pp. 1889-1918, 2005.

[14] 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.

[15] Qiu, D., Li, Y., Zhang, X., & Gu, B., Support vector machine with parameter optimization by bare bones differential evolution. *Paper presented at the 7th IEEE International Conference of the Natural Computation ICNC,* (pp. 263-266), 2011.

[16] Sartakhti, J. S., Zangooei, 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.

[17] Samadzadegan, F. Soleymani, A., &Abbaspour, A., Evaluation of genetic algorithms for tuning SVM parameters in multi-class problems. *Paper presented at 11th IEEE International Symposium on Computational Intelligence and Informatic, 2010,* (pp. 323-328). Budapest, Hungary.

[18] 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.

[19] 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).

[20] 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).

[21] 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.

[22] 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.

[23] A. Hamada & S. Brahim. A comparison of wavelet and curvelet for lung cancer diagnosis with a new cluster k-nearest neighbor classifier. *Paper Presented at Proceedings the 17th WSEAS International Conference on Applied Mathematics (AMATH '12) 2012*. Montreux, Switzerland.

[24] *UCI Repository of machine learning databases, Department of Information and Computer Science, University of California, Irvine, CA, <http://www.ics.uci.edu/mlearn/MLRepository>, 2012.*

[25] M. Zhao, C. Fu, L. Ji, K. Tang, M. Zhou, "Feature Selection and Parameter Optimization for Support Vector Machines: A New Approach Based on Genetic Algorithm with Feature Chromosomes". Expert System with Applications, vol. 38(2011), pp. 5197-5204, 2011.

[26] S. Lin, K. Ying, S. Chen, Z. Lee, "Particle Swarm Optimization for Parameter Determination and Feature Selection of Support Vector Machines". Expert System with Applications, vol. 35(2008), pp. 1817-1824, 2008.

[27] C. Huang, J. Dun, "A Distributed PSO-SVM Hybrid System with Feature Selection and Parameter Optimization". Applied Soft Computing, ELSEVER, vol. 8, no. 4, pp. 1381-1391, 2008.

Hiba Basim Alwan holds a Bachelor in Computer Science from Al-Mansour University College, Baghdad-Iraq in 2004 and a Masters degree in Computer Science from University of Technology, Baghdad-Iraq, in 2006. As an academic, her research interests include Ant Colony Optimization, Genetic Algorithm, Machine Learning, Support Vector Machine, and Classification. She is currently a PhD. student in the School of Computing at the College of Arts and Science, UUM.

Prof. Dr. Ku Ruhana Ku-Mahamud holds a Bachelor in Mathematical Science and a Masters degree in Computing, both from Bradford University, United Kingdom in 1983 and 1986 respectively. Her PhD in Computer Science was obtained from University Pertanian Malaysia in 1994. As an academic, her research interests include computer system performance modeling and swarm intelligence. She is currently a professor in the School of Computing at the College of Arts and Science, UUM.