An Optimization-Based Framework for Feature Selection and Parameters Determination of SVMs

Seyyid Ahmed Medjahed
University of Sciences and Technology Mohamed Boudiaf USTO-MB, Faculty of Mathematics and Computer Science,
Oran, 31000, Algeria
Email: sa.medjahed@univ-usto.dz

Mohammed Ouali
Department of Computer Science, College of Computers and Information Technology, Taif University, KSA
Email: mohammed.ouali@usherbrooke.ca

Tamazouzt Ait Saadi
University of Have, Havre, 76600, France
Email: tamazouzt.ait.saadi@univ-lehavre.fr

Abdelkader Benyettou
University of Sciences and Technology Mohamed Boudiaf USTO-MB, Faculty of Mathematics and Computer Science,
Oran, 31000, Algeria
Email: aek.benyettou@univ-usto.dz

Abstract— In this paper, feature selection and parameters determination in SVM are cast as an energy minimization procedure. The problem of feature selection and parameters determination is a very difficult problem where the number of feature is very large and where the features are highly correlated. We define the problem of feature selection and parameters determination in SVM as a combinatorial problem and we use a stochastic method that, theoretically, guarantees to reach the global optimum. Several public datasets are employed to evaluate the performance of our approach. Also, we propose to use the DNA Microarray Datasets which are characterized by the large number of features. To validate our approach, we apply it to image classification. The feature descriptors of the images were extracted by using the Pyramid Histogram of Oriented Gradients. The proposed approach was compared with twenty feature selection methods. Experimental results indicate that the classification accuracy rates of the proposed approach exceed those of other approaches.

Index Terms— Feature Selection, Parameter Determination, Learning Set Selection, Support Vector Machine, Simulated Annealing.

I. INTRODUCTION

Supervised learning is a machine learning technique that has proven its importance in many applications. One of the most popular methods and more efficient is the Support Vector Machine (SVM) that solves the problems of classification (determining the class of an observation) by finding an optimal separating hyperplan [1]. The advantage of the SVM method resides on the one hand, in the fact that it is independent of the input vector, and on the other hand, its performance in learning and its generalization ability [2], [3]. Unfortunately, the practical use of SVM is limited because the quality of SVM models and the classification accuracy rate heavily depends on a proper setting of SVM hyper-parameter C and SVM kernel parameters. Several studies have been conducted in the domain of the parameters determination: Grid Search [4], [5] is the most widely used to determine the parameters of SVM and kernel function. Other approach defined by Pai and Hong [6] that combine genetic algorithm and the SVM to generate a set of parameters values for SVM. Also, Pain and Hong in [7], [8] presented a Simulated Annealing approach to obtain parameters values of SVM and tested their approach on real data set. Ren and Bai [9] developed an approach to determine the optimal SVM parameters by using genetic algorithm and particle swarm optimization. In [10], the authors develop a novel method based on simulated annealing and SVM to determine the optimal parameter value without feature selection. These studies focused only on the determination of the parameters.

Moreover, the parameter C and the kernel function parameter are not the only factors that influence the quality of the SVM model and the classification accuracy rate, but feature selection plays an important role in the classification. The main purpose of feature selection is to find the smallest feature subset that increases the classification accuracy rate and reduces the space dimension and the computational time. This optimal features subset is obtained by eliminating the redundant, noisy and irrelevant features.

The optimal features subset is not unique; it may be possible to achieve the same accuracy rate using different subsets of feature, because if two features are correlated one can be replaced by other. Therefore, remove a relevant feature can reduce the classification accuracy rate.

Many researchers are interested in developing of a new feature selection method [11], [12], [13], [14]. The feature selection methods can be categorized as: Filter, Wrapper and Embedded methods. In [15], Chen and Hsien developed a latent semantic analysis (LSA) and web page for feature selection (WPFA), combined with the SVM to screen features. Gold et al. [16] developed a Bayesian viewpoint of SVM classifiers to adjust the parameters values in order to determine the irrelevant features. Chapelle et al. in [17] presented an automatically tuning multiple parameters and used the principal components to obtain features for the SVM technique. In [18], the authors adopted the accuracy rate of the classifier as the performance measure. Shon et al. [19], use genetic algorithm for screening the features of a dataset.

Bouguila et al. [20], [21], [22], [23], [24], [25], [26] has extensively studied the selection of relevant features in the pattern analysis field by proposing a Bayesian approach for identifying clusters of proportional data based on the selection of relevant feature. The general idea is based on the generalization of Dirichlet mixture models.

In this study, we propose to improve the quality of SVM model and the classification accuracy rate by modeling the feature selection and parameters determination problem as a combinatorial optimization problem. We use a stochastic method to solve the combinatorial problem. We also propose to select the relevant and the good instances that constitute the learning set to build a strong SVM model.

The paper is organized as follows: Section II describes the proposed approach. In Section III, experimental results are compared with those of existing approaches. In Section IV, we give the conclusion of this work and some perspectives.

II. THE PROPOSED SA-SVM APPROACH

The purpose of our approach is to reformulate the problem of feature selection and parameter determination (the parameter C and the Gaussian kernel parameter σ) as a combinatorial optimization problem. We define the problem as follows:

Consider the dataset $D = \{f_1,...,f_t\}$ with f_i represents the features (attributes) of the dataset.

We define the vector $V = \{V_1,...,V_t\}$, $V_i = \{0,1\}$ and the pair set $P = \{(f_1,V_1),...,(f_t,V_t)\}$ such that:

If $V_i = 1$ the feature f_i is selected and will be used for learning, else, if $V_i = 0$ the feature f_i is not selected and will not be used for learning.

On the other hand, we define a lower and upper bound for C and σ:

$$C \in [lb_c, up_c], \sigma \in [lb_\sigma, up_\sigma]$$

Finally, the decision variable has the following form:

$$X = [C\ \sigma\ V_1...V_t]$$

The energy function (cost function) $E(X)$, represents the classification error rate (classification error rate=1-

classification accuracy rate) obtained by SVM. The goal is to minimize $E(X)$ which is the objective function i.e. minimizing the error classifying in the testing set. In this type of problem, the space is not well understood and not smooth. The aim is to find in the space at which real valued energy function is minimized (finding the optimum). Nevertheless, to solve this problem, stochastic search techniques are used.

Simulated Annealing (SA) is a minimization technique that gives a good result in a large space. It has the fundamental property of finding the global minimum regardless of the initial configuration, but, this is theoretically guaranteed. While this has been shown as an asymptotic convergence towards the global minimum in infinite time, no experimental benchmarking has been achieved either to evaluate the quality of the solution or to verify that a strong minimum is reached in finite time.

Metropolis et al. [27] proposed an algorithm to simulate the behavior of the system as the Boltzmann distribution at temperature T: The simulated annealing uses this iterative procedure to achieve a state of thermal equilibrium. Kirkpatrick et al. [28] proposed to adapt this algorithm to solve optimization problems. The energy function is replaced by the objective function to minimize. The following algorithm presents the simulated annealing procedure:

1: initialize the system at an initial state X_k and calculate the function $E(X_k)$

2: choose randomly another state X_{k+1} and calculate the function $E(X_{k+1})$

3: if $E(X_{k+1}) \leq E(X_k)$ or $exp(-\Delta E/T) > rand$ then

4:   accept the transition to the new state $X_k = X_{k+1}$

5: else

6:   the transition is rejected

7: end if

8: decrease the temperature by a cooling schedule

9: repeat steps 2 to 8 until global equilibrium

The fig.1 illustrates the operation process of the proposed approach SA-SVM.

Initially, the algorithm chooses arbitrarily a solution (X_k) from the feasible region (it takes $C = 1$, $\sigma = 1$ and it uses all the features of the dataset $V_i = 1\ \forall i = 1,...,t$). In each iteration, the algorithm generates a random solution (X_{k+1}) and randomly selects a new learning set. At this step, it calculates $E(X_k)$ and $E(X_{k+1})$ the values of energy function for both solutions. ΔE denotes the difference between $E(X_k)$ and $E(X_{k+1})$. The acceptation or the rejection of a new solution is conditioned by the metropolis procedure $exp(-\Delta E/T)$ and the value of ΔE. If ΔE or $exp(-\Delta E/T) > rand$, the new solution is accepted else is rejected. In the output, the algorithm provides the optimal parameters values, the subset of relevant features and the better learning set.
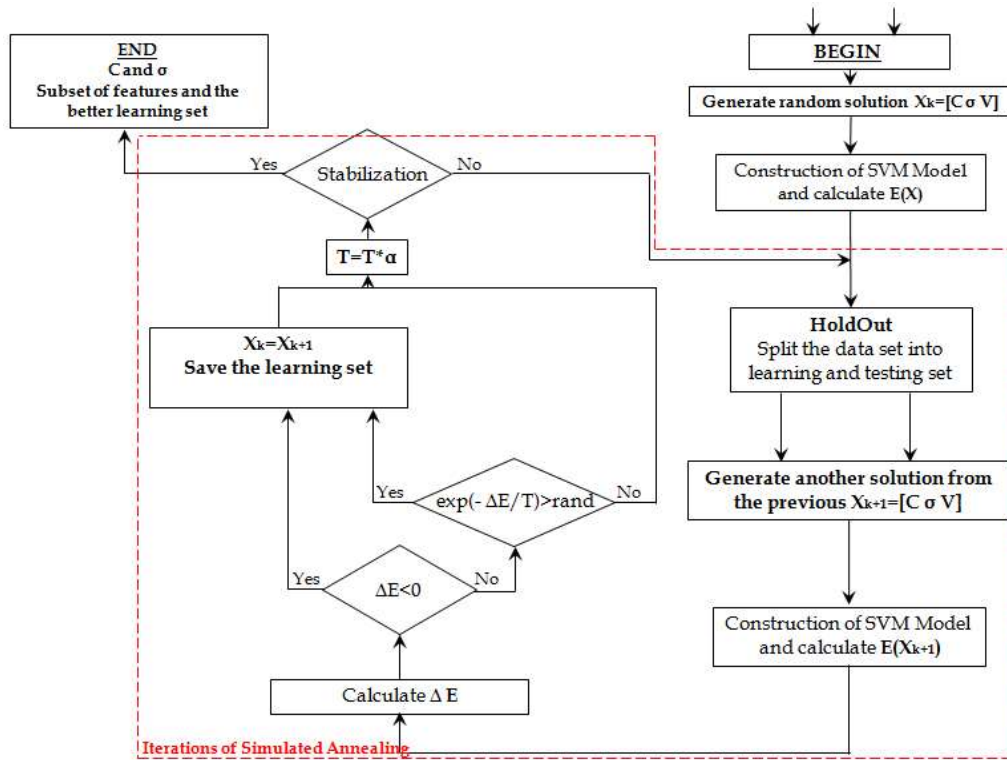
Fig. 1. The procedure of the proposed SA-SVM approach

III. Experimental Results

A. Application in UCI Datasets

In this section, we will show the performance of the proposed approach SA-SVM through the quality of the classification accuracy rate. We used a datasets taken from UCI Machine Learning Repository (http://archive.ics.uci.edu/ml/) to evaluate the performance of the proposed SA-SVM. The table 1 represents the datasets used for our study.

Table 1. Datasets from the UCI Machine Learning Repository

Datasets	Number of classes	Number of instances	Number of features
Breast Cancer	2	699	9
Cardiotocography	2	1831	21
ILPD	2	583	9
Mammographic Mass	2	961	5
Vertebral Column	2	310	6

In the table 1, the first column represents the name of datasets used for experimentation. The second column of the table 1 contains the number of classes for each dataset. The third column represents the number of instances for each dataset. The last column is the number of features that constitute each dataset.

To calculate the classification error rate, the number of instances for the learning set and testing set must be determined. In this study, 60% of instances are considered for the learning phase and 40% of the remaining instances are used for the testing phase. A portion of the test set is considered for the validation phase.

Table 2 describes the number of instances used for learning and testing phases for each dataset.

Table 2. Number of instances used for learning and testing phases

Datasets	Missing instances	Instances for learning	Instances for test
Breast Cancer	16	411	272
Cardiotocography	0	1101	730
ILPD	0	351	232
Mammographic Mass	131	500	330
Vertebral Column	0	188	122

In the table 2, the first column represents the name of the dataset and the second contains the number of missing instances. The third and the last column contain the number of instances used in learning and testing phase respectively.

The parameters used in the proposed SA-SVM approach are set as follows: The starting temperature T_0 is set to 1000. Choosing high value of T_0 allows the algorithm to avoid falling into a local minimum. We use the geometric cooling schedule with schedule parameter $\alpha = 0.97$. Having a value around 1 allows a slow cooling.

The interval parameter of SVM model C and the kernel function parameter σ are:

$C \in [1,5000]$ and $\sigma \in [0.01,50]$

The SA-SVM stops when the value of energy function reaches 0 (classification error rate = 0) or when the energy function will stop evolving after a certain number of iterations. These parameters have proven to give good results.

The table 3 presents twenty feature selection methods which will be compared with our approach SA-SVM. Note that these methods are only used for feature selection and do not allow to determine the parameters of SVM and kernel function parameter.

Table 3. Some feature selection methods which will be used to compare with our approach SA-SVM

Methods	Full Name
mRmR	Max-Relevance Min-Redundancy [29], [30]
CMIM	Conditional Mutual Info Maximisation [31], [30]
JMI	Joint Mutual Information [32], [30]
DISR	Double Input Symmetrical Relevance [33], [30]
CIFE	Conditional Infomax Feature Extraction [34], [30]
ICAP	Interaction Capping [35], [30]
CONDRED	Conditional Redundancy [30]
BETA GAMMA	Beta Gamma [30]
MIFS	Mutual Information Feature Selection [36], [30]
CMI	Conditional Mutual Information[30]
MIM	Mutual Information Maximisation [37], [30]
RELIEF	Relief [30]
FCBF	Fast Correlation Based Filter [38], [39]
MRF	Markov Random Fields [40]
SPEC	Spectral [41], [39]
T-TEST	Student's T-test [39]
KRUSKAL-WALLIS	Kruskal-Wallis Test [42], [39]
FISHER	Fisher Score [43], [39]
GINI	Gini Index [44], [39]
GA	Genetic Algorithm

In the table 4, we show the classification accuracy rate obtained by our approach SA-SVM and the twenty other methods.

Table 4. Classification accuracy rates obtained by the proposed approach SA-SVM and the twenty feature selection methods.

Methods	Average Classification Accuracy Rate (%)				
	WDBC	ILPD	Cardio	Mammo	VC
MRMR	97,11	55,15	88,68	83,l7	78,89
CMIM	97,07	58,24	95,21	82,98	81,19
JMI	97,07	58,32	95,21	82,98	83,49
DISR	96,43	58,22	97,00	82,60	83,49
CIFE	96,89	58,22	96,35	62,98	80,39
ICAP	96,89	58,24	93,21	82,98	81,19
CONDRED	95,80	58,24	96,55	82,98	78,98
BETA GAMMA	96,36	58,24	96,35	83,67	78,98
MIFS	96,89	55,15	88,68	83,27	78,89
CMI	97,07	58,26	87,60	82,98	79,49
MIM	96,76	58,24	95,21	83,07	81,19
RELIEF	95,95	60,80	97,50	79,53	79,49
FCBF	96,15	55,61	78,28	83,27	79,29
MRF	95,56	58,30	78,28	78,77	80,57
SPEC	95,79	57,53	10,33	79,33	80,32
T-TEST	93,73	59,95	18,56	79,53	80,57
KRUSKAL-WALLIS	95,99	56,73	97,85	82,98	80,37
FISHER	96,26	59,95	96,32	79,53	80,57
GINI	96,76	58,33	96,95	83,37	78,98
GA	96,32	57,05	95,96	83,07	83,69
Our Study : SA-SVM					
SVM without FS-without LSS	96,32	74,02	99,45	81,81	82,78
SVM with FS-without LSS	98,16	76,19	99,31	85,45	82,78
SVM without FS-with LSS	99,26	72,72	99,86	88,78	92,62
SVM with FS-with LSS	100,00	76,62	100,00	90,00	93,44

In our study, we have conducted experiments for 4 variants of SA-SVM:

SA-SVM without FS and without LSS: This first variant use the SA-SVM for only parameter determination by using all the features and without learning set selection. We use a single learning set and testing set over all the steps of the algorithm SA-SVM.

SA-SVM with FS and without LSS: In this variant, we use the SA-SVM for parameter determination and feature selection, and without learning set selection.

SA-SVM without FS-with LSS: In this variant, the SA-SVM is used for parameter determination without feature selection (all the features are used) and by using the learning set selection technique.

SA-SVM with FS-with LSS: This last approach, we propose to use the SA-SVM with feature selection and learning set selection. In the output, the SA-SVM provides: the optimal parameters values C and σ; the subset of relevant features and the better learning set which give good SVM model.

The analysis of the results shows that the proposed approach SA-SVM exceeded those obtained by the other approaches. We record an advantage for the SA-SVM with FS and with LSS. The classification accuracy rate improved significantly when we use the learning set technique.

The columns of table 4 contain the datasets and the rows represent the methods.

In addition, we remark that the SA-SVM with FS and with LSS provides the optimal parameter value, find the relevance features subset and select the better training set without decreasing the classification accuracy rate.

In the table 5, we illustrate the subset of relevant features selected by the SA-SVM approach and the twenty other methods.

Table 5. The selected features by the proposed SA-SVM approach and the twenty feature selection methods for each dataset

Methods	Selected Features				
	WDBC	ILPD	Cardio	Mammo	VC
MRMR	2,1,6,7	4,1	8,1	1,3	6,1
CMIM	2,6,1,8	4,6,5,1	8,17,1,12,13,18,14,19,11,20	1,2	6,1,5
JMI	16,3,1	4,6,1,5	8,13,12,14,18,17,1,19,20,11	1,2	6,4,5
DISR	2,6,9,3	4,6,1,5	8,10,17,1,13,12,14,18,19,11	1,5	6,4,5
CIFE	2,6,1	4,6,1,5	8,13,19,12,14,20,18,1,17,10	1,2	6,4,2
ICAP	2,6,1	4,6,1,5	8,17,1,12,13,18,14,19,11,20	1,2	6,1,5
CONDRED	2,3,5,8	4,6,1,5	8,12,13,14,18,19,17,1,20,10	1,2	6,3,1
BETAGAMMA	2,3,6,5	4,6,1,5	8,12,13,14,18,19,12,1,20,10	1,4	6,1,3
MIFS	2,6,1	4,1	8,1	1,3	6,1
CMI	2,4,1,8	4,6,1	8,13,19	1,2	6,4,3
MIM	2,3,6,7	4,6,5,1	8,17,1,12,13,18,14,19,11,20	1,4	6,15
RELIEF	6,4,2,3	1,5,6,8	8,4,5,18,9,7,11,19,6,2	4,3	6,3,4
FCBF	2,4,5,8	4,1	8,1	12	6,1
MRF	1,2,3,4	1,2,3,4	8,1	2,4	12,6
SPEC	9,8,4,6	6,2,3,5	6,3,7,16,2,15,10,20,2,15	43	6,2,3
T-TEST	9,5,4,8	3,2,5,4	11,12,19,17,3,21,1,14,2,12	4,3	6,1,2
KRUSKAL-WALLIS	6,2,8,4	9,2,3,1	9,10,1,2,3,4,5,6,7,8	1,2	6,1,2
FISHER	6,3,2,7	3,2,4,5	8,11,18,19,4,7,17,5,21,14	4,3	6,1,2
GINI	2,3,6,7	2,3,4,6	8,18,11,19,17,4,7,2,13	1,4	6,1,3
GA	9,6,7,1	2,4,7,6	15,21,11,3,6,4,16,2,7,8	4,1	6,42
SA-SVM with FS-without LSS	1,2,3,4	5,7,9	1,3,6,4,9,13	1,3,5	1,2
SA-SVM with FS-with LSS	2,7	1,2,5,6,7	1,5,7,10,11,12,14,17	1,3	1,2,3,4,5

B. Application in DNA Microarray Gene Expression Dataset

To validate the results obtained by the approach SA-SVM with FS and with LSS, we conduct experiments by using the DNA microarrays datasets, where there are many thousands of features, and a few tens to hundreds of samples. We propose to use the datasets describe in the following table which are widely studied in the literature and issued of microarray experiments [45-48].

These datasets were taken from the public Kent Ridge Bio-medical Data Repository with url http://sdmc.lit.org.sg/GEDatasets/ Datasets.html.

Table 6. The DNA microarrays datasets used for experimentation to validate the proposed approach SA-SVM. The number of feature represents the number of genes expression and the number of samples is the number of patients

Datasets	Number of features	Number of Samples
Leukemia	7129	72
Colon Cancer	2000	62
DLBCL	4026	47
Lung Cancer	12533	181
Ovarian Cancer	15154	253
Breast Cancer	24481	97

The table 7 shows the results obtained by the SA-SVM with FS and LSS.

Table 7. Classification accuracy rates (CAR) and the selected features obtained by the proposed approach SA-SVM with FS and LSS. In the table, "FS" represents the selected features by SA-SVM and "O.FS" represents the initial features of the dataset

Datasets	CAR (%)	FS	O.FS
Leukemia	100	2590	7129
Colon Cancer	100	991	2000
DLBCL	100	2007	4026
Lung Cancer	78.26	3587	12533
Ovarian Cancer	95.00	7535	15154
Breast Cancer	71.05	12244	24481

The first column represents the name of the datasets used to evaluate the proposed approach. The second column contains the classification accuracy rate obtained by our approach. The third column represents the number of selected features provides by the SA-SVM with FS and with LSS approach. The last column is the initial number of features in each dataset.

The analysis results clearly show that we have achieved a high classification accuracy rate with our approach. In the Leukemia, Colon Cancer and the DLBCL, we have reached 100% of classification accuracy rate.

Finally, we compare our results obtained by the SA-SVM approach with other approaches which has used the DNA Microarray datasets. The table 8 summarizes the results.

Table 8. Classification accuracy rates obtained by the proposed approach SA-SVM and compared with those obtained by other approaches.

Dataset	Average Classification Accuracy Rate (%)		
	Leukemia	Colon Cancer	DLBCL
This Study SA-SVM with FS and LSS	100	100	100
GA-EPS (Max) [49]	97.20	90.30	96.10
GA-EPS (Avg) [49]	96.40	87.30	94.80
K-TSP [50]	95.80	90.30	97.40
C4.5 [50]	73.60	80.70	80.50
Naive Bayes [49], [50]	100	58.10	80.50
K-NN [49], [50]	84.70	74.20	84.40
SVM [49], [50]	98.60	82.30	97.40
PAM [49], [50]	97.20	85.50	85.70
SVM-RFE [49]	97.20	75.80	96.10
ISVM2 [51]	82.40	83.00	-
ISM3 [51]	82.40	83.00	-
ADMM-DrSVM [51]	82.40	81.90	-
GA [52]	87.11	77.59	-
ACO [52]	86.43	76.53	-
PSO [52], [53]	86.57	75.65	-
SA [52]	85.73	78.19	-
ACO-S [52]	91.68	81.42	-

The table 8 compares the classification accuracy rate of SA-SVM with FS and with LSS with several methods defined in the literature. Clearly, we show that the proposed approach SA-SVM provides good results compared with other methods in the context of DNA Microarrays datasets.

C. Real Application-Application in Image Classification

In this section, we evaluate the performance of our SA-SVM with FS and with LSS in image classification. The objective of this experimentation is classifying images by the object categories they contain and to select the smallest relevant subset of feature descriptors that represents the relevant information in images. We use the Caltech-101 dataset (collected by Fei-Fei et al. [54]) that consists of images from 101 object categories. All the images are medium resolution about 300×300 pixels.

We use 3 binary classifications to validation the performance of the proposed approach SA-SVM:

The airplane images (800 images) VS The faces images (435 images). The dataset is split into 3 (691 images for training, 295 images for testing and 249 images for validation phase).

The cougar face images (69 images) VS The cougar body images (47 images). 57 images for training, 24 images for testing and 35 images for validation.

The water lilly images (37 images) VS The sun flower images (85 images). 61 images for training, 25 images for testing and 36 images for validation.

Fig. 2. The images from the Caltech-101 dataset used for our experimentation. One per category. We used 6 categories

The feature descriptors of each image are extracted by using the Pyramid Histogram of Oriented Gradients (PHOG) [55]. We use the following parameters for the PHOG algorithm: the number of bins on the histogram is 8 with angle = 360 and 3 pyramid levels over the entire Image (region of interest is the entire image). The PHOG method provides 680 feature descriptors for each image.

The goal is to obtain a high classification accuracy rate with the smallest relevant feature descriptors subset. The results are summarized in the table 9.

Table 9. The results obtained by SA-SVM with FS and LSS, compared with SVM, LSSVM and KNN without feature selection in term of classification accuracy rate (CAR) for each binary classification.

	Airplanes *VS* Faces	Cougar Face *VS* Cougar Body	Water Lilly *VS* Sun Flower
CAR of SVM (%)	99.80	86.67	97.49
CAR of LSSVM (%)	97.56	73.33	89.36
CAR of KNN (%)	99.59	77.78	82.98
The proposed approach SA-SVM with FS and LSS			
Number of initial feature descriptors	680	680	680
Number of selection feature descriptors	358	343	332
CAR on test data (%)	100	96.80	100
CAR on validation data (%)	100	88.57	91.78

The results show that the SA-SVM provides satisfactory result. The SA-SVM approach reaches 100% classification accuracy rate on data validation and reduces the number of feature descriptors to 358 feature descriptors for the classification of airplanes VS faces images. Compared with other methods: SVM, LSSVM and KNN, SA-SVM with FS and with LSS gives a high classification accuracy rate by using the selected feature descriptors (the relevant subset feature descriptors).

IV. CONCLUSION

In this paper, we cast the problem of feature selection and parameter determination in SVM as an energy minimization problem. We modeled the problem as a combinatorial optimization problem, and we propose the simulated annealing algorithm to solve the problem. We also have proposed four variants of the proposed approach SA-SVM by integrating the learning set selection technique. The proposed SA-SVM approach optimizes the parameter of SVM model and the kernel parameter, and obtains a subset of beneficial features, also provides the good instance which can build a perfect SVM model.

Experimentations were performed on both UCI machine learning repository datasets and DNA microarrays datasets. While the results on the UCI machine learning repository datasets are satisfactory, the results obtained in DNA microarrays datasets is a very good. Also, we have conducted experimentations on the image classification to validate our proposed approach. The feature descriptors of the images were extracted by using the PHOG. The results obtained show that our SA-SVM approach gives a very good result on the one hand in term of classification, and on the other hand, it reduces the number of feature descriptors and selects the smallest relevant feature descriptors subset.

We have compared our approach with twenty feature selection method. The comparison of the obtained results with other approaches demonstrates that the proposed approach SA-SVM with FS and with LSS provides a very high classification accuracy rate. Future perspectives are the integration of the correlation term in the energy function to improve the quality of results around the quality of the selected feature and the classification accuracy rate.

REFERENCES

[1] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*," vol. 20, n. 3, pp. 273-279, 1995.

[2] M. Rychetsky, "Algorithms and architectures for machine learning based on regularized neural networks and support vector approaches," *Shaker Verlag GmBH*, Germany, 2001.

[3] J. Shawe-Taylor and N. Cristianini, "Kernel Methods for Pattern Analysis," *Cambridge University Press*, 2004.

[4] J. Wang, X. Wu, and C. Zhang., "Support vector machines based on k-means clustering for real-time business intelligence systems," *Int. J. Business Intell. Data Mining*, vol. 1, n. 1, pp. 54-65, 2005.

[5] C.-W. Hsu, C. C. Chang, and C. J. Lin, "A practical guide to support vector classification," *Technical Report, University of National Taiwan, Department of Computer Science and Information Engineering*, 2003.

[6] P. F. Pai and W. C. Hong, "Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms," *Electric Power Syst. Res.*, vol. 74, n. 3, pp. 417-425, 2005.

[7] P. F. Pai and W. C. Hong, "Support vector machines with simulated annealing algorithms in electricity load forecasting," *Energy Conversion Manage*, vol. 46, n. 17, pp. 2669-2688, 2005.

[8] P. F. Pai and W. C. Hong, "Software reliability forecasting by support vector machines with simulated annealing algorithms," *J. Syst. Softw.*, vol. 79, n. 6, pp. 747-755, 2006.

[9] Y. Ren and G. Bai, "Determination of optimal svm parameters by using ga/pso," *Journal of Computers*, vol. 5, n.8, pp. 1160-1168, 2010.

[10] J. Sartakhti, M. Zangooei, and K. Mozafari, "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, n. 2, pp. 570–579, 2012.

[11] R. Parimala and R. Nallaswamy, "Feature Selection using a Novel Particle Swarm Optimization and It's Variants," *International Journal of Information Technology and Computer Science*, vol. 4, n. 5, pp. 16-24, 2012.

[12] B. Izadi, B. Ranjbarian, S. Ketabi and F. N. Mofakham, "Performance Analysis of Classification Methods and Alternative Linear Programming Integrated with Fuzzy Delphi Feature Selection," *International Journal of Information Technology and Computer Science*, vol. 5, n. 10, pp. 9-20, 2013.

[13] S. Goswami and A. Chakrabarti, "Feature Selection: A Practitioner View," *International Journal of Information Technology and Computer Science*, vol. 6, n. 11, pp. 66-77, 2014.

[14] I. Lavy and A. Yadin, "Support Vector Machine as Feature Selection Method in Classifier Ensembles," *International Journal of Modern Education and Computer Science*, vol. 6, n. 3, pp. 1-10, 2014.

[15] R.-C. Chen and C.-H. Hsieh, "Web page classification based on a support vector machine using a weighed vote schema," *Expert Syst. Appl.*, vol.31, n. 2, pp.427-435, 2006.

[16] C. Gold and P. Sollich, "Bayesian approach to feature selection and parameter tuning for support vector machine classifiers," *Neural Netw*, vol. 18, n, 5-6, pp. 693-701, 2005.

[17] O. Chapelle, V. Vapnik, O. Bousquet, and S. Mukherjee, "Choosing multiple parameters for support vector machines," *Mach. Learn.*, vol. 46, n. 1-3, pp.131-159, 2002.

[18] R. Kohavi and G. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, n. 1-2, pp. 273-324, 1997.

[19] T. Shon, Y. Kim, C. Lee, and J. Moon, "A machine learning framework for network anomaly detection using svm and ga," *Proceedings of the IEEEWorkshop on Information Assurance and Security*, vol. 2, 2005.

[20] O. Amayri and N. Bouguila, "On online high-dimensional spherical data clustering and feature selection," *Eng. Appl. of AI*, vol. 26, n. 4, pp. 1386-1398, 2013.

[21] N. Bouguila and D. Ziou, "A countably infinite mixture model for clustering and feature selection," *Knowl. Inf. Syst.*, vol. 33, n. 2, pp. 351-370, 2012.

[22] W. Fan and N. Bouguila, "Variational learning of a dirichlet process of generalized dirichlet distributions for simultaneous clustering and feature selection," *Pattern Recognition*, vol. 46, n. 10, pp. 2754-2769, 2013.

[23] T. Elguebaly and N. Bouguila, "Simultaneous bayesian clustering and feature selection using rjmcmc-based learning of finite generalized dirichlet mixture models," *Signal Processing*, vol. 93, n. 6, pp.1531-1546, 2013.

[24] W. Fan, N. Bouguila, and H. Sallay, "Anomaly intrusion detection using incremental learning of an infinite mixture model with feature selection," *RSKT 2013*, 2013.

[25] W. Fan and N. Bouguila, "Online learning of a dirichlet process mixture of generalized dirichlet distributions for simultaneous clustering and localized feature selection," *Journal of Machine Learning Research*, vol.25, 2012.

[26] T. Bdiri and N. Bouguila, "Bayesian learning of inverted dirichlet mixtures for svm kernels generation," *Neural Computing and Applications*, vol. 23, n. 5, pp. 1443-1458, 2013.

[27] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *Journal of Chemical Physics*, vol. 21, n. 6, 1953.

[28] S. Kirkpatrick, J. C. D. Gelatt, and M. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, 1983.

[29] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and minr edundancy," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 27, n.8, pp. 1226–1238, 2005.

[30] G. Brown, A. Pocock, M.-J. Zhao, and M. Luj án, "Conditional likelihood maximisation: A unifying framework for information theoretic feature selection," *The Journal of Machine Learning Research*, vol. 13, pp. 27-66, 2012.

[31] F. Fleuret. Fast binary feature selection with conditional mutual information. Journal of Machine Learning Research, 2004, 5:1531–1555.

[32] H. Yang and J. Moody, "Feature selection based on joint mutual information," *In Proceedings of International ICSC Symposium on Advances in Intelligent Data Analysis*, pp. 22–25, 1999.

[33] P. Meyer and G. Bontempi, "On the use of variable complementarity for feature selection in cancer classification," *Applications of Evolutionary Computing*, vol. 3907, pp. 91–102, 2006.

[34] L. Dahua and T. Xiaoou, "Conditional infomax learning: An integrated framework for feature extraction and fusion," *Computer Vision ECCV*, vol. 3951, pp. 68–82, 2006.

[35] J. A. "Machine learning based on attribute interactions. Fakulteta za racunalniˇstvo in informatiko," *Univerza v Ljubljani*, 2005.

[36] R. Battiti, :Using mutual information for selecting features in supervised neural net learning," *IEEE Transactions on Neural Networks,* vol. 5, n. 4, pp. 537 –550, 1994.

[37] D. D. Lewis, "Feature selection and feature extraction for text categorization," *In Proceedings of Speech and Natural Language Workshop. Morgan Kaufmann*, pp. 212–217, 1992.

[38] L. Yu and H. Liu, "Efficient feature selection via analysis of relevance and redundancy" *Journal of Machine Learning Research*, vol. 5, pp.1205–1224, 2004.

[39] Z. Zhao, F. Morstatter, S. Sharma, S. Alelyani, A. Anaud, and H. Liu, "Advancing feature selection research-asu feature selection repository," *technical report*, 2010.

[40] Z. H. Cheng Q. and C. J. "The fisher-markov selector: fast selecting maximally separable feature subset for multiclass classification with applications to high-dimensional data," *IEEE Tran Pattern Anal Mch Intell*, vol. 33, n. 6, pp. 1217–1233, 2011.

[41] H. Liu and Z. Zhao, "Spectral feature selection for supervised and unsupervised learning," *Proceedings of the 24th International Conference on Machine Learning*, 2007.

[42] L. J. Wei, "Asymptotic conservativeness and efficiency of kruskal-wallis test for k dependent samples," *Journal of the American Statistical Association*, vol. 76, n. 376, pp. 1006–1009, 1981.

[43] R. Duda, P. Hart, and D. Stork, "Pattern Classification," *2nd ed. John Wiley & Sons, New York*, 2001.

[44] T. M. Cover and J. A. Thomas, "Elements of Information Theory," *Wiley*, 1991.

[45] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol.46, n. 1-3, pp. 389-422, 2002.

[46] S. Shah and A. Kusiak, "Cancer gene search with data-mining and genetic algorithms," *Computers in Biology and Medicine*, 2002.

[47] R. Mallika and V. Saravanan, "An svm based classification method for cancer data using minimum microarray gene expressions," *World Academy of Science, Engineering and Technology*, 2010.

[48] A. E., Garcia-Ni, eto J., L. Jourdan, and E. Talbi, "Gene selection in cancer classification using pso/svm and ga/svm hybrid algorithms," *Evolutionary Computation, 2007. CEC 2007. IEEE Congress on*, 2007.

[49] T. Muchenxuan, L. Kun-Hong, X. Chungaui, and J. Wenbin, "An ensemble of svm classifiers based on gene pairs," *Computers in Biology and Medecine*, 2013.

[50] A. Tan, D. Naiman, L. Xu, R. Winslow, and D. Geman, "Simple decision rules for classifying human cancers from gene expression profiles," *Bioinformatics*, vol. 21, 2005.

[51] L. Dehua, Q. Hui, D. Guang, and Z. Zhihua, "An iterative svm approach to feature selection and classification in high-dimensional datasets," *Pattern Recogn.*, vol. 46, n. 9, pp. 2531–2537, 2013.

[52] Y. Lia, G. Wanga, H. Chend, L. Shia, and L. Qina, "An ant colony optimization based dimension reduction method for high-dimensional datasets," *Journal of Bionic Engineering*, vol. 10, n. 2, pp. 231–241, 2013.

[53] L. Y, W. G, and C. H, "An improved particle swarm optimization for feature selection," *Journal of Bionic Engineering*, vol. 8, pp.191–200, 2011.

[54] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories," *In IEEE CVPR Workshop of Generative Model Based Vision*, 2004.

[55] B. Anna, Z. Andrew, and M. Xavier, "Representing shape with a spatial pyramid kernel," *CIVR '07 Proceedings of the 6th ACM international conference on Image and video retrieval*, 2007.

Authors' Profiles

Seyyid Ahmed Medjahed graduated with M.Sc., of Engineering of Data and Knowledge at Oran University, Algeria. He is a PhD student at University of Sciences and Technology Mohamed Boudief USTO-MB, of Mathematics and Computer Science, Oran, Algeria. He started teaching as Assistant Professor in 2012 at Relizane University Center, Algeria. His area of research interests includes: Machine Learning and Image Processing.

Mohammed Ouali received the Doctor degree in Real Time Informatics, Robotics, and Automatic Control from Mines ParisTech, France, and the Ph.D. in Mathematics and Computer Science from Sherbrooke University, Canada. Dr. Ouali has more than 15 years in both industry and

academia. His research interests include signal processing, pattern recognition, big data, and computer vision.

Tamazouzt Ait Saadi teacher and researcher at Mostaganem University, Algeria since September 2003. She obtained her certificate of engineer in computer science in 1989. In 2003, she had her M.Sc. and continues his doctoral at Havre University, France.

Abdelkader Benyettou received his BSc of engineering in 1982 from the Institute of Telecommunications of Oran, and the MSc degree in 1986 from the University of Sciences and Technology of Oran, Algeria. In 1987, he joined the Computer Sciences Research Center of Nancy, France, where he worked until 1991 on Arabic speech recognition by expert systems and received his PhD in electrical engineering in 1993 from the University of Sciences and Technology of Oran. Since 2003, he has been a professor at the University of Sciences and Technology of Oran. His interests are in the area of speech and image processing, Arabic speech recognition, neural networks, and machine learning. He has been the director of the Signal-Speech-Image— SIMPA Laboratory, Department of Computer Sciences, Faculty of Sciences, University of Sciences and Technology of Oran, since 2002.